Stony Brook University

A Study of Structure-Function Relationships of DNA Polymerase Activity on DNA Interstrand Crosslinks

A Dissertation Presented

by

Upasana Roy

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Chemistry

Stony Brook University

May 2017

Stony Brook University

The Graduate School

Upasana Roy

We, the dissertation committee for the above candidate for the

Doctor of Philosophy degree, hereby recommend

acceptance of this dissertation.

Orlando D. Schärer, Ph.D. – Dissertation Advisor
Professor, School of Life Sciences, UNIST, South Korea

Nicole Sampson, Ph.D. - Chairperson of Defense
Professor, Department of Chemistry

Carlos Simmerling, Ph.D. – Co-Advisor
Professor, Department of Chemistry

Miguel Garcia-Diaz, Ph.D.
Professor, Department of Pharmacological Sciences

This dissertation is accepted by the Graduate School

Charles Taber
Dean of the Graduate School

Abstract of the Dissertation

A Study of Structure-Function Relationships of DNA Polymerase Activity on DNA Interstrand Crosslinks

by

Upasana Roy

Doctor of Philosophy

in

Chemistry

Stony Brook University

2017

DNA interstrand crosslinks (ICLs) are DNA lesions that covalently link two strands of a DNA duplex, thereby acting as a block to DNA replication and transcription. As a result ICLs are particularly toxic to dividing cells, and ICL-forming agents such as cisplatin, nitrogen mustards and MMC are widely used in cancer chemotherapy. However, there are complex cellular pathways that repair ICLs, contributing to the problem of resistance to these anti-tumor agents. There appear to be multiple ICL repair pathways operating in cells but the majority of ICLs in vertebrates are repaired in a replication-coupled manner. A key step in this repair pathway is the unhooking of the ICL from one of the strands by nucleases, followed by translesion synthesis across the unhooked ICL by DNA polymerases. The structure of the unhooked ICL substrate determines key outcomes such as efficiency and fidelity of the repair process. However the

position of incisions by nucleases and the structure of unhooked ICL substrates are not known. Using a strategy developed in our laboratory, I synthesized structurally diverse model unhooked ICLs to investigate the effect of ICL structure on the approach, insertion and extension abilities of replicative and TLS polymerases. I demonstrated that the distortion induced in the duplex by the crosslink, as well as the duplex context of the ICL are important factors that influence the nature of translesion synthesis by DNA polymerases. Surprisingly, I found that the most processed form of an ICL – a single crosslinked base that is believed to be a putative intermediate during ICL repair – can be bypassed by replicative polymerases and that ICLs may be repaired without the help of TLS polymerases in some cases. This work provides insight into how structures of unhooked ICLs influence biological outcomes such as occurrence of resistance and secondary tumor formation during the translesion synthesis step in ICL repair.

TABLE OF CONTENTS

List of Figures ..... **ix**
List of Tables ..... **xi**
List of Abbreviations ..... **xii**
Acknowledgements ..... **xiv**

CHAPTER 1

Overview **1**
DNA crosslinking agents **2**
Nitrogen mustards 3
DNA interstrand crosslink repair pathways **6**
Replication-dependent ICL repair 6
Dual-fork convergence pathway 6
Glycosylase-dependent pathway 7
Replication fork ICL traverse pathway 8
Unhooking in replication-coupled ICL repair 10
Replication-independent ICL repair 11
Overview 11
Unhooking in replication-independent ICL repair 13
TLS polymerases in DNA interstrand crosslink repair **14**
Results from genetic studies 15
Results from functional assays 18
Results from biochemical assays 20
Recruitment of TLS polymerases to ICL repair pathways **24**
TLS polymerases in chemotherapy **26**
References **28**

CHAPTER 2

Preface **37**
Introduction **38**
Materials **42**
Reagents and Buffers 42
Equipment and Consumables 44
Methods **46**
Preparation of NM ICLs 46
Native and 2'FdG-containing NM ICL analogs 46
7CdG-containing NM ICL analogs 47
Resection of NM ICLs by USER 48
Purification and Characterization of NM ICLs 49
Primer Extension Assays with NM ICL analogs **54**
Notes **55**
Summary and Conclusions **57**
References **59**

CHAPTER 3

Preface **61**
Introduction **62**
Materials and Methods **66**
Synthesis of 2-aminoethyl 7-deazaguanosine 66
Oligonucleotides and Primers 67
Preparation of 20bp and 6bp ICLs 67
Preparation of Single Nucleotide (1 nt) ICLs 69
Enzymes 70
Polymerase assays 70

Single nucleotide insertion assays 71
Analysis of single nucleotide insertion assays 71
Results 72
Design of NM ICL substrates mimicking unhooked repair intermediates 72
Synthesis of the Single Nucleotide NM ICL 74
Reducing duplex around ICL facilitates bypass 76
NM ICL-induced distortion facilitates approach and insertion, but inhibits extension by pol η 79
Pol η is more accurate on NM ICLs than on undamaged DNA 84
Discussion 86
Influence of helix distortion and ICL resection on polymerase activity 86
Bypass of a single nucleotide NM ICL 87
What could the role of pol η in ICL repair be? 89
Conclusions 90
Acknowledgements 91
References 92

CHAPTER 4

Introduction 98
Materials and Methods 101
Oligonucleotides used for synthesis of biotinylated substrates 101
Synthesis of biotinylated substrates 102
Enzymes 104
Polymerase assays 104
Results 105
Duplex resection and distortion facilitates approach

by pol ε and pol δ 105
Pol ε and pol δ can fully bypass the 1nt ICL 109
Synthesis of 93mer ICL substrates with biotinylated ends for studies with PCNA 110
Discussion 113
Effect of duplex resection and distortion on polymerase activity 113
Bypass of the single nucleotide NM ICL 115
Reconstituting translesion synthesis at unhooked ICLs 116
References 117

Chapter 5

Conclusions and Perspectives 120

LIST OF FIGURES

Chapter 1

1.1 DNA crosslinking agents 3

1.2 G-G nitrogen mustard ICL 4

1.3 Depurination of nitrogen mustard (NM) ICLs 5

1.4 Models for Replication-Dependent ICL Repair Pathways. 9

1.5 Model of Replication-Independent ICL repair 13

1.6 Biochemical activity of TLS polymerases on ICL substrates 22

Chapter 2

2.1 Native and stable NM ICLs 40

2.2 Scheme for the formation of NM ICLs 41

2.3 Purification of stable NM ICL analogs 48

2.4 Analysis of purified NM ICLs by 15% PAGE 53

2.5 NM ICL analogs block the polymerase reaction by Klenow (exo-) 55

Chapter 3

3.1 Scheme for generation of NM-like ICLs with various structures 66

3.2 Strategy for synthesis of the 5a NM ICL mimic 68

3.3 MALDI-TOF spectrum of the single nucleotide ICL 69

3.4 ICL substrates used in this study 73

3.5 Synthesis of the single nucleotide ICL 75

3.6 Shortening of duplex around the ICL facilitates bypass 77

3.7 Duplex distortion facilitates approach and insertion by pol η at the ICL 80

3.8 Duplex distortion by NM ICLs facilitates approach by Klenow 82

3.9 Reaction of Klenow and pol η with two distinct 6a ICLs 83

3.10 pol η is more accurate across NM ICLs than undamaged DNA 85

Chapter 4

4.1 Duplex shortening and distortion facilitates approach by pol ε (WT) 106

4.2 Duplex shortening and distortion facilitates approach by pol ε (exo-) 108

4.3 Duplex shortening and distortion facilitates approach by pol δ (exo-) 109

4.4 Scheme for synthesis of biotinylated ICL substrates 111

4.5 Synthesis and purification of longer substrates with biotinylated ends 112

LIST OF TABLES

Table 1: Known roles of TLS polymerases in ICL repair 16

Table 2: MALDI-ToF MS of synthesized NM ICLs 52

LIST OF ABBREVIATIONS

ICL : interstrand crosslink

TLS : translesion synthesis

NM : nitrogen mustard

MMC : mitomycin C

FA : Fanconi Anemia

DSB : double stranded break

HR : homologous recombination

NER : nucleotide excision repair

RIR : replication-independent repair

GG-NER : global genome nucleotide excision repair

TC-NER : transcription coupled nucleotide excision repair

MMR : mismatch repair

ssDNA : single stranded DNA

dsDNA : double stranded DNA

HY : hydrazine

DMEDA : dimethylethylenediamine

CPD : cyclopyrimidine dimer

XP-V : xeroderma pigmentosum – variant

ACKNOWLEDGEMENTS

First and foremost, I'd like to thank Orlando, my advisor, for being an excellent mentor these past five years. He gave me the freedom to explore and follow my own scientific path, and had the patience to let me learn at my own pace. He has always been extremely approachable and supportive, and I deeply value all his advice and encouragement over the years. Whenever I needed to discuss anything, Orlando made the time for it, and I'm thankful for his thoughtful feedback and guidance throughout. But what I appreciate the most, is that he maintained the lab as a "judgement-free zone" where it's safe to admit you made mistakes, and be honest about priorities outside research. I can't stress enough how important that has been for my ability to grow as a scientist, so thank you Orlando!

I would also like to thank my committee members, Dr. Nicole Sampson and Dr. Carlos Simmerling for all their help and guidance over the years. Their input and suggestions have helped me think critically about my project and learn a lot in the process. Many thanks also to the Chemistry Department staff- Mike, Sandhya, Charmine and especially Katherine, who was my go-to person for all kinds of questions, and helped me navigate through graduate school.

A big thank you to all the Schärer lab folks for all the good times we had– Alejandra, Julie, Shivam, Burak, Yuxin and Alexandra. A special thanks to Julie for being a great mentor during my rotation, and for taking on disproportionately many responsibilities to keep the lab running smoothly. To Alejandra, who has been much more than just a colleague, I am deeply thankful for her sincere

friendship, love and support. She was the pulse of the lab, making sure it never got boring, difficult or lonely. I always looked forward to our daily chats and fun conversations, and relied on her to cheer me up when things weren't going well. She has been my closest confidant, an extremely caring friend and an unconditional cheerleader, and I am so grateful to her for always being there for me, and rooting for my success.

I have also been extremely lucky in having a great circle of friends outside the lab- Tiffany, Luisa, Coray, Koushik and Pratik. I'm grateful for their friendship and for all the fun times we've had. A special thanks to Tiffany, for being a wonderful friend, and for trusting me with responsibilities in GWISE (even though I had my doubts). It turned out to be a very rewarding and enriching experience, and I'm thankful to have been a part of it.

I'm immensely grateful to Aatish, for all his love and support over the years. I'm inspired by his sense of wonder and curiosity, and am glad for our shared love for science. Thanks to him, I've also had a rich and adventurous life outside the lab, which is inextricably tied to my happy Ph.D. experience. Also, thanks to him for being an annoyingly good editor.

None of this would have been possible without the support of my family, and I'm especially grateful to my parents for all their efforts and sacrifices, which afforded me many privileges they didn't have. They were the first to introduce me to the world of research- and although both my parents earned their Ph.D. in a time with no internet, they are still impressed by mine today. I'm deeply thankful for their unconditional love, and it is to them, this thesis is dedicated.

Dedicated to my parents, Arati and Kunal B. Roy

ICLs are toxic to a cell,

Unless the bypass goes well,

There is no question,

Approach, insertion and extension,

Are affected by distortion, in a nutshell.

CHAPTER 1: INTRODUCTION

This chapter has been adapted from the manuscript "*Involvement of Translesion Synthesis DNA Polymerases in DNA Interstrand Crosslink Repair*" by Upasana Roy and Orlando D. Schärer published in DNA Repair, Volume 44, pages 33-41 in 2016.

OVERVIEW

DNA interstrand crosslinks (ICLs) are lesions that covalently link the two strands of a DNA helix together, preventing their separation and blocking essential processes such as transcription and replication. As a result, ICLs are particularly toxic to rapidly dividing cells such as cancer cells, and ICL-forming agents have been widely used as anti-cancer drugs for decades (Chabner et al., 2005). However, there are several cellular pathways that repair and remove ICLs, leading to resistance of tumors against these drugs. Depending on the type of ICL and the pathway, this repair can be mutagenic and contribute to secondary tumor formation (Deans and West, 2011). The study of ICL repair mechanisms can therefore provide insight into the development of secondary tumor formation and chemoresistance against ICL-inducing anti-cancer therapeutics.

There appear to be multiple pathways by which ICLs are repaired, depending on the cell-cycle phase and type of ICL, although the mechanistic understanding for most of them is still limited (Clauson et al., 2013). However a common, key step in all of these pathways is believed to involve a partially

processed intermediate that serves as a substrate for a specialized group of translesion synthesis (TLS) DNA polymerases to restore the integrity of one of the two affected strands. Due to their large active site, TLS polymerases are able to synthesize DNA using the crosslinked intermediate as a template, in a lesion "bypass" reaction. However, they are error-prone and can introduce mutations during the bypass step, potentially contributing to secondary tumors in a clinical setting. There are multiple TLS polymerases in mammals, and the role of each polymerase in ICL repair has not been clearly defined (Roy and Scharer, 2016). The choice of polymerase, as well as efficiency and accuracy of translesion DNA synthesis is influenced by the structure of the ICL substrate.

ICLs are formed by DNA crosslinking agents, which are a group of structurally diverse, bifunctional electrophilic compounds that can react with two bases. Nitrogen mustards are one such class of DNA crosslinking agents that preferentially react with N7 of guanine bases in a 5'-GNC-3' sequence to form ICLs (Millard et al., 1990; Millard et al., 1991; Rink et al., 1993). Using a synthetic, stable nitrogen mustard mimic and its structural variants, this work investigates how ICL structure affects the bypass activity of DNA polymerases.

DNA CROSSLINKING AGENTS

There are several classes of bifunctional compounds that act as DNA crosslinking agents, including those used in anti cancer therapy such as mitomycin C (MMC), cisplatin and nitrogen mustards (NM) (**Fig. 1.1**). Treating DNA with crosslinking agents produces a mixture of products:

a) Mono-adducts – products of a reaction with only one base

b) *Intra* strand crosslinks – products of a reaction with two bases on the *same* strand of DNA

c) *Inter* strand crosslinks (ICLs) – products of a reaction with two bases on *opposite* strands of the DNA.

Although ICLs are formed less frequently than the other adducts, they are considered the most toxic type of lesion for most crosslinking agents (Noll et al., 2006; Scharer, 2005).

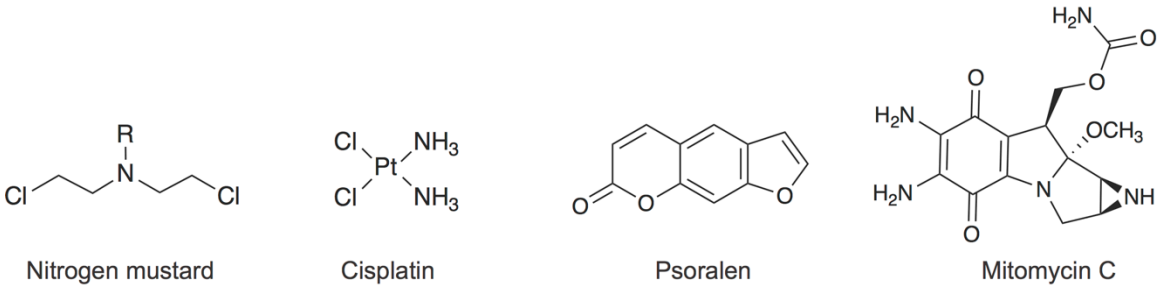


Figure 1.1: DNA crosslinking agents

Nitrogen mustards

Nitrogen mustards (NMs) were the first drugs to be used in anti-cancer therapy by Goodman and Gilman in the 1940s (Goodman et al., 1946), giving rise to the field of cancer chemotherapy. NMs were first used to treat Hodgkin's lymphoma and leukemia, but are now also used against non-Hodgkin's lymphoma, ovarian cancer, breast cancer, testicular cancer and small cell lung cancer among others (Chabner and Roberts, 2005). Nitrogen mustards are bis-(beta-chloroethyl)- amines, that react preferentially with N7 of guanine bases in a 5'-GNC-3' sequence to produce ICLs that lie in the major groove of the

crosslinked duplex (Millard et al., 1990; Millard et al., 1991; Povirk and Shuker, 1994; Rink et al., 1993) (**Fig. 1.2A**). The mechanism of ICL formation is a two-step process: the nitrogen mustard forms an aziridinium ion which reacts with the N7 of a guanine base to first form a monoadduct, followed by reaction with a second guanine base on the opposite strand to form the ICL (Rink et al., 1993) (**Fig. 1.2B**). Formation of the ICL distorts the B-form DNA around the ICL and introduces a bend of about 20° in the helical axis (Guainazzi et al., 2010; Rink and Hopkins, 1995).

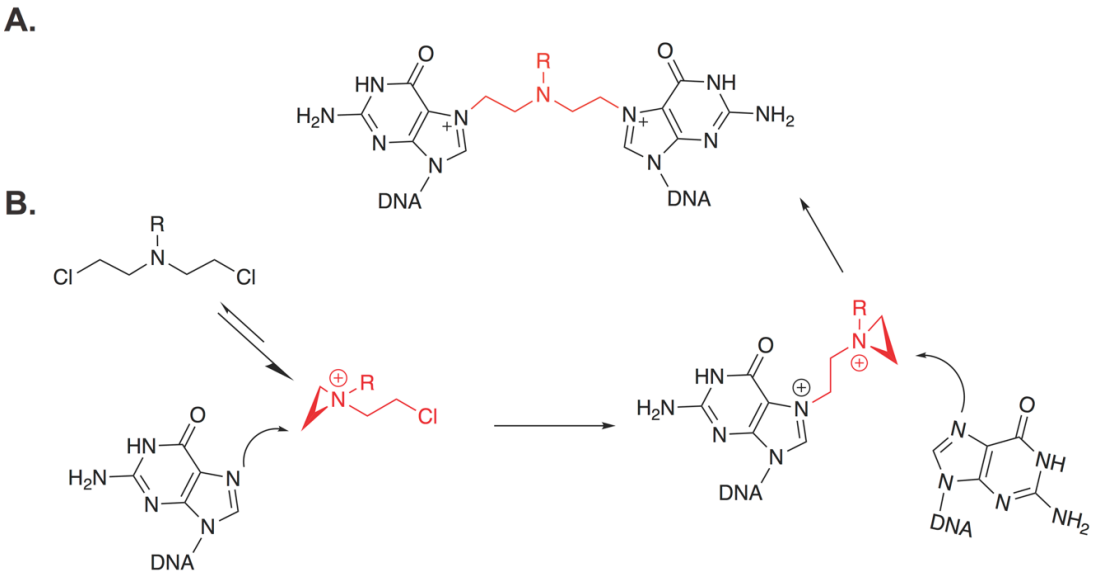


Figure 1.2: G-G nitrogen mustard ICL. A. Structure of a nitrogen mustard ICL crosslinking N7 of two guanine bases. **B.** Mechanism of nitrogen mustard ICL formation. The crosslink and azidirinium intermediates are shown in red.

Only 1-5% of the total DNA adducts produced by nitrogen mustards are ICLs, with the rest being monoadducts and intrastrand crosslinks (Povirk and Shuker, 1994). The low yield of ICLs has been a limiting factor in using nitrogen

mustard ICLs for cellular and biochemical studies. In addition, the crosslinked N7 atoms in the ICL carry a positive charge, making it hydrolytically unstable and prone to depurination (**Fig. 1.3**) (Masta et al., 1994). As a result using the native nitrogen mustard ICL in biochemical studies has been problematic.

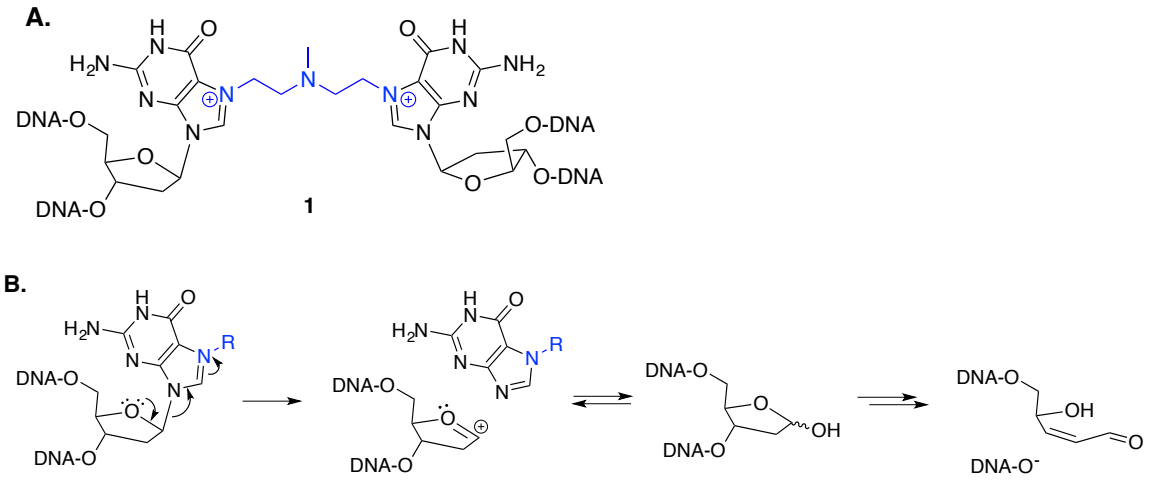


Figure 1.3: **Depurination of nitrogen mustard (NM) ICLs**. **A.** Structure of an NM ICL between two guanines. The crosslinked N7 atoms carry a positive charge in an NM ICL. **B.** NM ICL depurination reaction. *Adapted from Castaño, A[1], Roy, U., & Schärer, O. D. (2017) "Preparation of nitrogen mustard DNA interstrand crosslinks and their stable analogs for biochemical and cell biological studies". Methods in Enzymology (in press)*

Previous work in our lab has established a synthetic strategy to generate stable, site-specific nitrogen mustard ICL mimics in a high yield (Angelov et al., 2009; Guainazzi et al., 2010; Mukherjee et al., 2014). Using this strategy (discussed in detail in Chapter 2), I synthesized structurally diverse ICL substrates for the biochemical studies with DNA polymerases discussed in this thesis.

DNA INTERSTRAND CROSSLINK REPAIR PATHWAYS

Replication-dependent ICL repair

The majority of ICLs are repaired in a replication-coupled manner in vertebrates. There appear to be multiple replication-associated ICL repair pathways, and several are still not well understood. We know most about the "dual-fork convergence pathway" that requires FANCD2-I and is initiated when two replication forks converge on an ICL, and have some first glimpses at a glycosylase-dependent pathway that does not require FANCD2-I. Another replication-associated pathway has been observed, termed "replication traverse pathway" in which the replication fork continues downstream of the ICL without removing it, however the mechanistic details of this process are not known. These pathways are discussed here with a focus on the possible structures of the ICL substrates that are encountered by DNA polymerases in these pathways.

Dual-fork convergence pathway

The most well defined mechanism has been elucidated by the Walter group using plasmids containing site-specific ICLs with replication-competent *Xenopus* egg extracts (Räschle et al., 2008). In this system, two replication forks converge on an ICL and their leading strands pause 20 to 40 nucleotides from the crosslink due to presence of the replicative helicase (Zhang et al., 2015) (**Fig. 1.4A, i**). The stalled helicase is removed from the vicinity of the ICL by BRCA1/BARD1 (Long et al., 2014), allowing one of the leading strands to proceed to one nucleotide before the ICL. At this point, the Fanconi anemia pathway is activated, resulting in the ubiquitylation of FANCD2-I (Knipscheer et

al., 2009), followed by unhooking of the ICL by incisions involving ERCC1-XPF-SLX4 on the strand opposite the approached replication fork (**Fig. 1.4A, ii**) (Hodskinson et al., 2014; Klein Douwel et al., 2014). The unhooking incisions are made on the same parental DNA strand, one on either side of the crosslink. This leads to formation of a double-stranded break (DSB) in that daughter duplex, while the unhooked ICL remains attached to the other parental strand (**Fig. 1.4A iii-iv**). Translesion synthesis by DNA polymerases across the unhooked ICL allows the stalled leading strand to be extended past the ICL and eventually be ligated to the downstream lagging strand, generating a template for repair of the DSB by homologous recombination (HR) (**Fig. 1.4A, v-vii**). The unhooked ICL remnant is no longer very toxic and is eventually likely removed by NER.

Glycosylase-dependent pathway

Recently, the Walter group identified a new replication-associated pathway of ICL repair that also requires fork convergence (Semlow et al., 2016). They found that ICLs formed by psoralen and abasic sites are resolved in a FANCD2-I-independent manner. In this pathway, after two replication forks converge on the ICL, the glycosylase NEIL3 cleaves the N-glycosidic bond between the crosslinked base and its deoxyribose sugar, thereby breaking the covalent linkage between the two DNA strands. This yields an abasic site on the cleaved strand, and a monoadduct on the other strand, both of which can act as substrates for translesion synthesis, thereby allowing the replication forks to continue. Blocking this pathway resulted in the ICLs being unhooked by dual incisions in the conventional FANCD2-I dependent manner. Interestingly, they

found that the glycosylase-dependent pathway resolves psoralen and abasic site ICLs, but not cisplatin ICLs, indicating that the choice of ICL repair pathway depends on the structure and type of ICL.

Replication fork ICL traverse pathway

A study by Seidman and colleagues revealed another replication coupled pathway (**Fig. 1.4B**). They used an ICL repair-specific DNA fiber assay with fluorescently labeled psoralen ICLs and dual labeling to map the progression of a replication fork around the ICL in mammalian cells (Huang et al., 2013). Surprisingly, although dual fork convergence (~20% of the events) was observed, the main pathway observed (~50% of the events) was different. The replication fork was found to 'traverse' the ICL without any unhooking incisions, leaving an intact ICL behind (**Fig. 1.4B, i-ii**). These observations suggest that ICLs may also be repaired in a post-replicative manner, once the replication fork has traversed past the ICL. Notably, both ICL traverse and the dual fork convergence pathways ultimately lead to an X-shaped structure around the ICL (**Fig. 1.4A ii, 4B ii**), which could undergo similar unhooking and ICL bypass steps to complete repair.

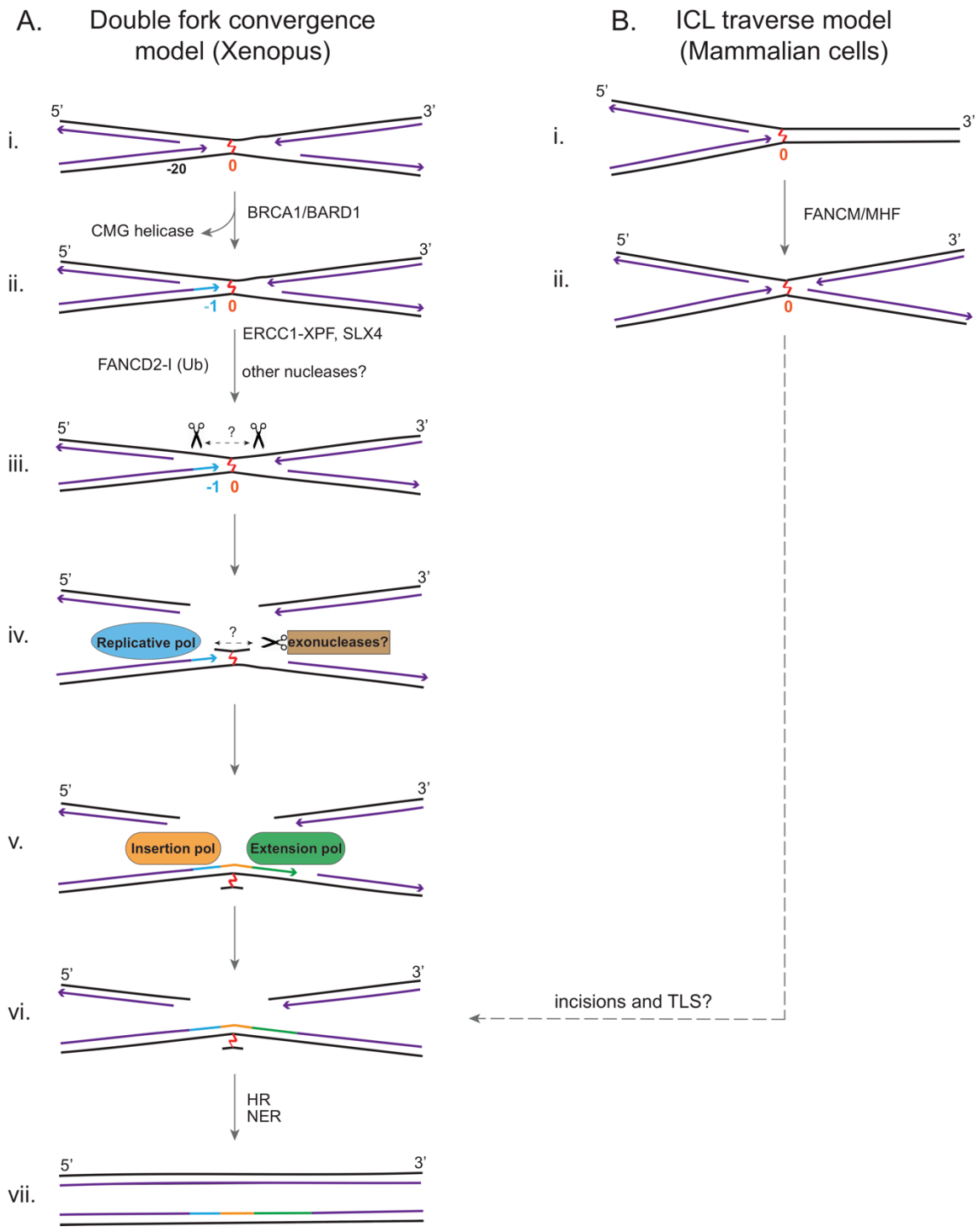


Figure 1.4: Models for Replication-Dependent ICL Repair Pathways. **A: Double fork convergence model** (i) Two replication forks converge on an ICL, stalling 20-40 nt away. (ii) Removal of the CMG helicase by BRCA1 allows one of the leading strands to approach within 1nt of the ICL (-1). (iii) Activation of the FA pathway leads to ubiquitination of FANCD2-I, which is required for unhooking of the ICL by SLX4/ERCC1-XPF, and possibly other nucleases. The position of these incisions have not been determined, and the amount of duplex surrounding the ICL is unknown. (iv) The unhooked ICL could then be processed further by exonucleases to trim the duplex around the ICL, making it more amenable to bypass by DNA polymerases.

(v,vi) An insertion polymerase inserts nucleotide(s) opposite the ICL and an extension polymerase extends the insertion product further. (vi) Ligation to downstream Okazaki fragments restores one daughter duplex, and (vii) is used to restore the other daughter duplex by HR. The ICL remnant on one strand is likely removed by NER to complete the repair of both daughter duplexes. **B: ICL traverse model**. (i) A single fork collides with the ICL, and (ii) in a FANCM/MHF dependent manner 'traverses' the ICL to continue replication on the other side of the crosslink without unhooking it. The later steps of the pathway are not known, but could involve incisions and TLS for post-replicative repair. *Adapted from Roy, U., & Schärer, O. D. (2016). "Involvement of Translesion Synthesis DNA Polymerases in DNA Interstrand Crosslink Repair". DNA Repair, (44): 33-41*

Unhooking in replication-coupled ICL repair

Apart from the glycosylase-dependent pathway, one of the central steps in ICL repair pathways is unhooking, in which incisions around the ICL release it from one of the two crosslinked strands (**Fig. 1.4A, iii-iv**). From the point of view of translesion synthesis across the ICL, a critical concern is where the incisions take place during unhooking, and whether an unhooked ICL is processed further by exonucleases. These factors determine the structure of intermediates that DNA polymerases encounter and, consequently, which polymerase(s) may act on an ICL.

Determining the identity of the nucleases involved in ICL unhooking has been challenging. Genetic deletion of several nucleases, including XPF, MUS81, FAN1, SNM1A and SLX1 renders cells sensitive to ICL forming agents to various degrees (Zhang and Walter, 2014). Experiments in *Xenopus* extracts showed that XPF-ERCC1 is essential for unhooking and that this activity requires SLX4 and ubiquitylated FANCD2-I (Klein Douwel et al., 2014; Knipscheer et al., 2009). This suggests a model in which XPF makes the incision on the 3' side of the ICL (**Fig. 1.4, iii**). The identity of the endonuclease making the cut on the other side has not been established. Possible candidates are SLX1 or again XPF, which could be making both cuts (Zhang and Walter, 2014).

Currently, there is no experimental system available to determine where on the parental strand the unhooking incisions occur and what the length of the duplex surrounding an unhooked ICL is. Additionally, the unhooked ICL may be further processed by an exonuclease such as SNM1A, which can digest DNA past an ICL (Allerston et al., 2015; Sengerová et al., 2012; Wang et al., 2011). SNM1A may also act on intermediates nicked only on one side of the ICL, digesting DNA past the ICL leaving an unhooked lesion behind. Although nothing is known about how ICLs are eventually unhooked in the traverse pathway, it is possible that the X-shaped intermediates are processed similarly by the same group of nucleases (**Fig. 1.4 A&B, ii**).

Multiple *in vitro* studies have shown that resection of the duplex around an ICL is crucial for translesion synthesis by TLS polymerases, with effective bypass only occurring with ICLs embedded in short (< 6 base pairs) duplexes (see Section 1.4.3) (Ho et al., 2011; Klug et al., 2012; Minko et al., 2008a; Yamanaka et al., 2010; Zietlow et al., 2009). Therefore, the degree of processing of unhooked ICLs is crucial for how DNA polymerases interact with them.

Replication-independent ICL repair

Overview

The replication-independent repair (RIR) of ICLs is believed to be a minor pathway in vertebrates, in particular because ICLs are much more deleterious during replication. However, it is likely to play an important role in post-mitotic cells where endogenous ICLs could block transcription of essential genes. ICL

repair in G1 has furthermore been shown to diminish the burden of ICLs before a cell enters S-phase (Enoiu et al., 2012). Much of what we know about RIR ICL repair is based on studies with reporter plasmids harboring a site-specific ICL (Enoiu et al., 2012; Hlavin et al., 2010; Shen et al., 2006; Wang et al., 2001; Williams et al., 2012; Zheng et al., 2003). Nucleotide excision repair (NER) proteins have been shown to be involved in this ICL repair pathway (**Fig. 1.5**). Depending on the ICL and assay conditions, both branches of NER, global genome NER (GG-NER) and transcription-coupled NER (TC-NER), (which differ in the damage recognition step, but otherwise share the same set of common factors) (Marteijn et al., 2014) were shown to be involved. The recruitment of NER proteins to ICLs has been directly demonstrated in G1 cells, lending support to the studies conducted with reporter plasmids (Muniandy et al., 2009). In addition to NER factors, mismatch repair (MMR) and HMG proteins have been shown to interact with NER proteins at ICLs and may therefore also contribute to replication-independent repair of ICLs (Mukherjee and Vasquez, 2016; Zhao et al., 2009).

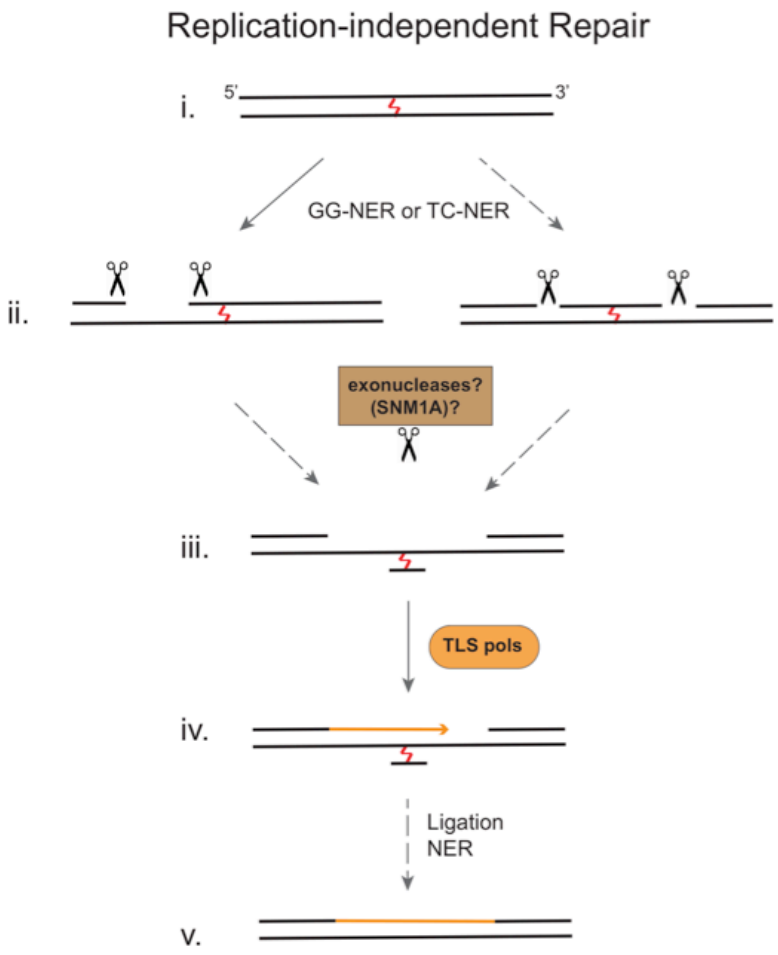


Figure 1.5: Model of replication-independent ICL repair. (i) Global genome NER (GG-NER) as well as transcription coupled NER (TC-NER) proteins are involved in replication independent repair of ICLs. (ii) Although dual incisions 5' to the ICL have been observed, unhooking incisions on either side of the ICL may also occur. (iii) These intermediates may be further processed by exonucleases like SNM1A to facilitate translesion synthesis. (iv) TLS polymerases carry out gap filling and (v) the ICL remnant is likely removed by NER. Solid arrows indicate details obtained from experimental observations and dashed arrows indicate indirect evidence or speculation. *Adapted from Roy, U., & Schärer, O. D. (2016). "Involvement of Translesion Synthesis DNA Polymerases in DNA Interstrand Crosslink Repair". DNA Repair, (44): 33-41*

Unhooking in replication-independent ICL repair

Although the involvement of NER in replication-independent repair has been clearly demonstrated, how incisions around the ICL occur is not immediately obvious. NER incisions require the opening of the DNA duplex around the lesion – a step that would be blocked by ICLs (Scharer, 2013).

Consistent with this idea, it has been shown that NER-proficient extracts incise psoralen and alkyl ICLs with both incisions occurring 5' to the ICL (Bessho et al., 1997; Mu et al., 2000; Smeaton et al., 2008) (**Fig. 1.5, ii**). As incisions on one side of the ICL do not enable a polymerase to bypass the ICL, another incision on the 3' side of the ICL would be required to generate an unhooked substrate. Although it is not yet known how this happens, it has been shown that NER nuclease ERCC1-XPF together with RPA (Mu et al., 2000), and the exonucleases SNM1A and FAN1 can degrade one strand of an ICL-containing duplex (Pizzolato et al., 2015; Wang et al., 2011; Wang et al., 2014). In support of such a scenario, a recent study found CSB and SNM1A to interact at trioxsalen ICLs (Iyama et al., 2015) raising the possibility that SNM1A may be directly recruited to TC-NER complexes for processing of these lesions. Interestingly, NER-independent incisions around an ICL have also been observed, but the factors responsible have not yet been identified (Smeaton et al., 2008).

TLS POLYMERASES IN DNA INTERSTRAND CROSSLINK REPAIR

The preceding section outlined how the unhooking step is key in determining the nature of substrate encountered by DNA polymerases. Translesion synthesis often involves multiple polymerases, with one carrying out the insertion of a nucleotide opposite the lesion, and another carrying out further extension (Shachar et al., 2009). There are 15 DNA polymerases in mammals

and many of them have been implicated in the repair of ICLs based on genetic or biochemical studies. Most of the TLS polymerases belong to the Y-family of polymerases - pol η, pol ι, pol κ and Rev1. However, pol ζ (composed of catalytic subunit Rev3 and regulatory subunit Rev7) belongs to the B-family of DNA polymerases.

The existence of multiple ICL repair pathways, the vastly different structures of (unhooked) ICLs, potential redundancy among DNA polymerases and limited options in studying these pathways at a mechanistic level have made it difficult to definitely assign the roles of various DNA polymerases in ICL repair. Discussed below is an overview of our current understanding of the role of TLS polymerases in ICL repair, based on the types of assays used in various studies.

Results from genetic studies

Genetic studies involving treatment of cells with crosslinking agents have implicated multiple polymerases in ICL repair (Table 1). A limitation of such experiments is that they do not provide any information about which ICL pathway a polymerase is involved in. Results from such studies are further complicated by the fact that crosslinking agents can also form *intra*strand crosslinks, which can have overlapping toxic effects with the ICLs, but are typically addressed by NER, not ICL repair.

Nonetheless, such assays have clearly revealed that deletion of the REV1 and REV3 genes leads to a clear hypersensitivity to cisplatin and mitomycin C exposure (Gan et al., 2008; Nojima et al., 2005; Okada et al., 2005; Sharma and Canman, 2012; Simpson and Sale, 2003; Sonoda et al., 2003). REV3 together

with REV7 and the accessory subunits POLD2 and POLD3 constitutes Pol ζ, a B-family polymerase (Lee et al., 2014; Sale et al., 2012), which is believed to frequently work together with REV1, a Y-family polymerase with dCMP transferase activity. The role of REV1 and Pol ζ seems to be important for both replication-dependent and -independent ICL repair, explaining the exquisite sensitivity of REV1 and REV3 deficient cell lines to crosslinking agents (Enoiu et al., 2012; Nojima et al., 2005; Räschle et al., 2008; Sarkar et al., 2006; Shen et al., 2006; Sonoda et al., 2003).

TABLE 1 : Known roles of TLS polymerases in ICL repair

Pols	Sensitivity	Replication dependent	RIR	In vitro assays
Pol ζ	Cisplatin MMC NM	Cisplatin	Cisplatin Psoralen MMC	n/d*
REV1	Cisplatin	Cisplatin	Cisplatin Psoralen MMC	n/d*
Pol η	Cisplatin Psoralen	n/a	Psoralen MMC	Cisplatin NM- like Acrolein-like** AP ICL***
Pol κ	MMC	n/a	MMC Acrolein-like**	Cisplatin NM-like Acrolein-like**
Pol ν	MMC Cisplatin	n/a	n/a	N6-N6A (major groove)

* n/d : Current biochemical data is not consistent with known role in ICL repair
** Acrolein-like: N2-N2 dG ICL
*** AP ICL: Oxidized abasic site ICL
Adapted from Roy, U., & Schärer, O. D. (2016). "Involvement of Translesion Synthesis DNA Polymerases in DNA Interstrand Crosslink Repair". DNA Repair, (44): 33-41

The involvement of other TLS polymerases is less unequivocal. The main role of Pol η, deficient in XP-V (*Xeroderma pigmentosum-variant*) cells, is to

accurately bypass UV-lesions during replication (Sale et al., 2012). Its open yet rigid active site also allows for the insertion of dNTPs opposite cisplatin *intra*strand crosslinks (Biertumpfel et al., 2010; Zhao et al., 2012), and this active site architecture may be suitable for the bypass of certain ICLs. Pol η deficient cells are indeed sensitive to crosslinking agents such as cisplatin or psoralen (Albertella et al., 2005; Chen et al., 2006; Misra and Vos, 1993; Mogi et al., 2008), but this sensitivity is less pronounced than that for REV1 or REV3 (Nojima et al., 2005), suggesting a less central role of Pol η in ICL repair.

Another Y-family TLS polymerase, Pol κ, is believed to be especially important for the bypass of minor groove DNA adducts (Sale et al., 2012). Consistent with this property, Pol κ-/- cells are sensitive to MMC, which forms ICLs in the minor groove (Minko et al., 2008a; Ogi et al., 2002; Takeiri et al., 2014; Williams et al., 2012). Interestingly, Pol κ -/- cells were found to be more sensitive to both MMC and cisplatin in the G0/G1 phase of the cell cycle, suggesting a more important role for Pol κ in replication-independent repair processes (Williams et al., 2012).

The last polymerase that appears to be involved in the repair of ICLs based on cellular sensitivity is Pol ν. Knock down of Pol ν renders cells hypersensitive to MMC and cisplatin (Moldovan et al., 2010; Zietlow et al., 2009), although the main activity of the enzyme *in vitro* seems to be on major groove ICLs (Yamanaka et al., 2010) (see below). Apart from a role in translesion synthesis, Pol ν could also be functioning at later stages of ICL repair, such as homologous recombination, as depletion of Pol ν sensitized cells to DSB forming

agents and also leads to reduced rates of homologous recombination (Moldovan et al., 2010).

Results from functional assays

The information on the role of DNA polymerases in ICL repair based on functional ICL repair assays is rather limited, relying mainly on studies in *Xenopus* extracts for replication-dependent repair and on plasmid reporter assays for replication-independent repair (Table 1). The results from these studies have provided some initial valuable information about the involvement of TLS polymerases in ICL repair.

Experiments in replication-competent *Xenopus* egg extracts have shown that REV1 and Pol ζ are required for extension (but not insertion) of the leading strand past a cisplatin ICL, while they are not essential for repair of non-distorting nitrogen mustard-like ICLs (Budzowska et al., 2015; Räschle et al., 2008). Pol ζ has a well-known role as an extension polymerase, with the ability to efficiently extend mismatched primer termini of insertion products of a variety of lesions (Gan et al., 2008; Haracska et al., 2003; Lee et al., 2014), while REV1 is known to serve as a hub protein to coordinate the activities of multiple TLS polymerases (Sharma et al., 2012; Wojtaszek et al., 2012). The unique roles of these two TLS enzymes are likely also critical for ICL repair. The role of other TLS polymerases in replication-dependent ICL repair, particularly in insertion opposite the ICL remains to be elucidated. Due to the possible redundancy of polymerases in some of the steps, dissecting the role of each polymerase is not likely to be straightforward. Interestingly, the repair of ICLs has been found to exhibit a

mutagenicity rate of a few percent, with various mutations clustered around the site of the ICL (Budzowska et al., 2015). Depletion of REV1 reduced this mutation rate, consistent with a role for TLS in lesion-induced mutagenesis.

The requirement of polymerases for replication-independent repair has been mainly determined by reporter plasmid systems and has revealed that the involvement of polymerases, at least in part, is dependent on the structure of the ICL. REV1 and Pol ζ were found to be required for replication-independent repair of cisplatin, psoralen and MMC ICLs in mammalian cells and nitrogen mustard ICLs in yeast (Enoiu et al., 2012; Sarkar et al., 2006; Shen et al., 2006), while Pol η was involved in the replication-independent repair of psoralen and MMC, but not cisplatin ICLs (Enoiu et al., 2012; Wang et al., 2001; Zheng et al., 2003). A definitive role for other TLS polymerases has not yet been established in this system. In a replication-independent *in vitro* system in *Xenopus* extracts, the repair of a minor groove acrolein-like N2-N2 trimethylene crosslink was specifically dependent on Pol κ (Williams et al., 2012). Immunodepletion of Pol κ greatly diminished repair efficiency of an ICL-containing plasmid, whereas depletion of Pol ζ did not have any effect. While these studies have shed some light on how different TLS polymerases are required for ICLs with different structures, it remains to be seen what influence the different assay systems and organisms used have on these results.

Results from biochemical assays

Biochemical assays using defined ICL substrates and purified DNA polymerases by contrast have provided more defined answers (Table 1). Although translesion synthesis can be studied at a single nucleotide resolution in this way, one limitation has been to design ICL substrates that are physiologically relevant. Given that the exact structure of unhooked ICLs that TLS polymerases encounter in cells are not known, a variety of different types of ICLs have been used. While some degree of selectivity for certain ICL structures was to be expected for various polymerases based on their substrate specificities for lesions on one strand of DNA, a key observation first made by Lloyd and coworkers was that the amount of duplex surrounding an ICL also dramatically affects the efficiency of bypass reaction (Minko et al., 2008a). While the bypass reactions are almost always completely inhibited by an ICL embedded in a long stable duplex, a certain amount of bypass is often possible if duplex is shortened to a few nucleotides (Ho et al., 2011; Klug et al., 2012; Minko et al., 2008a; Yamanaka et al., 2010). These findings highlight the importance of the unhooking step in ICL repair for polymerase bypass, as the position of the endonucleolytic incisions, and possible further processing by exonucleases, will determine what the unhooked ICL looks like when the polymerase encounters it. Biochemical assays have used a variety of unhooked ICL structures (**Fig. 1.6**), and while the diversity of ICL structures, polymerases and assay conditions used makes it challenging to compare the different studies, a number of general conclusions can be reached:

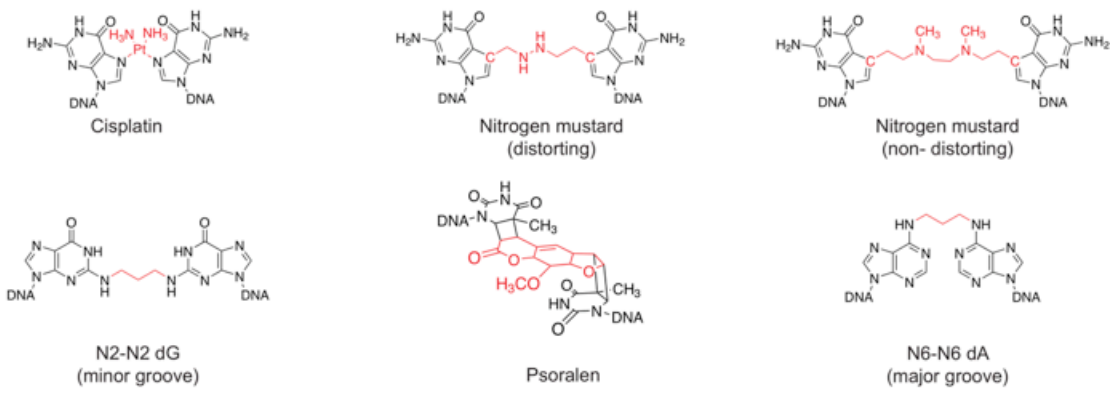
1) Multiple TLS polymerases have the ability to insert dNTPs opposite ICLs and extend them to full length products, even if their involvement in ICL repair is not clear from genetic or functional studies. This could explain why deficiency in individual TLS polymerases does not always lead to hypersensitivity to exposure of cells to crosslinking agents.

2) The amount of duplex surrounding an ICL is a key parameter for the efficiency of insertion and extension. While some enzymes can insert a dNTP opposite an ICL in a long (12-20bp) duplex, none can extend these substrates to a full length product. By contrast, ICLs in a short duplex (2-6 bp) can be extended to full length product by a number of TLS polymerases.

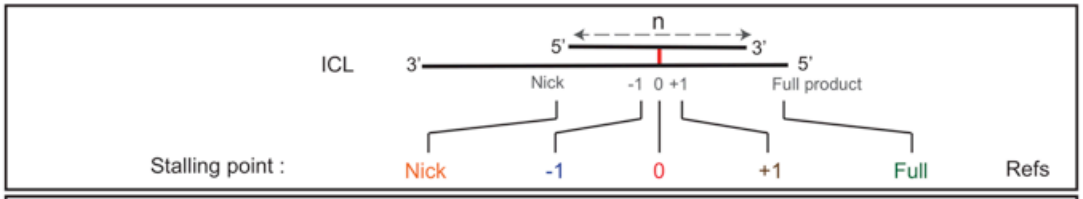
3) The structure of an ICL greatly influences the activity of TLS polymerases. In cases where insertion and extension is possible, polymerases approach helix-distorting ICLs more easily than non-distorting ones (likely because of strand displacement), but hinder extension as multiple polymerases stall at or within a few bases past the ICLs. By contrast, for non-distorting ICLs, the approach is more challenging, while insertion and bypass occur more readily.

An overview of results obtained from polymerase studies with ICL-containing substrates is shown in Figure 1.6, and the most important findings for each polymerase are summarized below.

A.



B.



Cisplatin	n= 20nt	η,κ,ι		η,κ,ι			
	n= 6nt	ι	κ,ι	η,κ,ι	η	η,κ	
NM (distorting)	n= 20nt	η,ι	ι	η	η,κ		
	n= 6nt		ι	κ,ι	η,κ,ι	η,κ	
	n= 1nt					η	
NM (non distorting)	n= 20nt	η,ι	η,κ				
	n= 6nt	η	ι			η,κ,ι	
N2-N2 dG (min. groove)	n= 10-14nt	η	η,κ,ν	η,κ			
	n= 2nt		η,ν	η,ι	η	η,κ	
Psoralen	n= 12	η,κ,ι	η,κ,ν	ι			
N6-N6 dA (maj. groove)	n= 9nt (5')	ν	ν	ν		ν	
	n= 1nt		ν	ν		ν	

Figure 1.6: Biochemical activity of TLS polymerases on ICL substrates. A. Chemical structures of ICLs used in biochemical studies. The crosslinks between bases are highlighted in red. **B.** Primer extension activity of TLS polymerase η,κ,ν and ι across diverse ICL substrates with varying amount of duplex (n) surrounding the crosslink. For polymerases that were tested with the substrates shown, the main stalling points at the start of the duplex (nick, orange), 1nt before the ICL (-1, blue), at the ICL (0, red), 1nt after the ICL (+1, brown), as well as complete extension to full product (Full, green) are indicated for each ICL substrate. Although conditions used to generate the data listed in this figure varied greatly, ICLs embedded in longer duplexes were not bypassed efficiently by TLS polymerases. *Adapted from Roy, U., & Schärer, O. D. (2016). "Involvement of Translesion Synthesis DNA Polymerases in DNA Interstrand Crosslink Repair". DNA Repair, (44): 33-41*

REV1/Pol ζ: Surprisingly, while experiments with purified REV1 or Pol ζ did reveal some insertion of dNTPs opposite ICLs, no extension or cooperation of the two enzymes was observed (Ho et al., 2011; Minko et al., 2008b). Given the clear importance of REV1 and Pol ζ in ICL repair from genetic and functional assays, it is possible that proper *in vitro* activity might require the two additional subunits PolD2 and PolD3. The new four subunit Pol ζ complex was found to be more active in translesion synthesis across cisplatin *intra*strand lesions (Lee et al., 2014) and testing the activity of this complex will likely be required to reveal the ability of Pol ζ to bypass ICLs.

Pol η: Pol η is the enzyme that has been most extensively studied with ICLs and is able to carry out insertion and in many cases extension across a diverse set of major and minor groove ICLs - including cisplatin, nitrogen mustard, acrolein mimics and ICLs formed at abasic sites (**Fig. 1.6**) (Ho et al., 2011; Klug et al., 2012; Xu et al., 2015). Interestingly, the pattern of bypass by Pol η is similar for various ICLs, stalling predominantly at 0, +1 and +2 positions, and the structure-function relationships mentioned above apply in particular to studies with Pol η. These observations are consistent with the structure of Pol η, which has the largest active site among Y-family polymerases and can therefore accommodate a variety of lesions (Biertumpfel et al., 2010). The rigid molecular splint guiding the primer-template strands in the active site of

Pol η may also help the polymerase reaction in the presence of a bulky unhooked ICL.

Pol κ: Pol κ can bypass diverse ICL structures *in vitro* including cisplatin, nitrogen mustard and acrolein ICLs and, consistent with genetic findings, is particularly effective in bypassing the minor groove acrolein lesions. (Ho et al., 2011; Minko et al., 2008a).

Pol v: Pol v is proficient in bypassing a variety of major groove ICLs, but not minor groove lesions or psoralen ICLs (Yamanaka et al., 2010; Zietlow et al., 2009). While the biological importance of Pol v in ICL repair is still elusive it might be a complement to the preference of Pol κ for minor groove adducts.

Pol ι: Pol ι was able to insert a base across psoralen, cisplatin and N2-dG ICLs derived from acrolein, but could only carry out an extension reaction on the non-distorting nitrogen mustard like ICL. (Ho et al., 2011; Klug et al., 2012; Smith et al., 2012). Therefore, Pol ι could act as an insertion polymerase for some ICLs, although its *in vivo* role in ICL repair remains to be demonstrated.

RECRUITMENT OF TLS POLYMERASES TO ICL REPAIR PATHWAYS

The activity of TLS polymerases needs to be a highly regulated process, as these enzymes are much more error-prone than replicative polymerases. The mechanisms by which the activity of TLS polymerases is regulated at lesions on one strand of DNA has been studied in some detail. Uncoupling of the helicase

and replicative polymerase at single stranded (ss) DNA lesions leads to long stretches of ssDNA, which are covered by RPA and lead to the activation of the E2-E3 ubiquitin ligase RAD6-RAD18 and PCNA ubiquitination (Davies et al., 2008). This mono-ubiquitination of PCNA at Lys164 leads to recruitment of TLS polymerases which interact with ubiquitinated PCNA via their PCNA and ubiquitin binding motifs (Sale et al., 2012). An important difference in replication-coupled ICL repair is that both the helicase and polymerases stall at ICLs, preventing their uncoupling and precluding the formation of long ssDNA stretches. Consistent with this difference, it has been shown that damaging agents like UV (which mostly form intrastrand lesions) elicited strong PCNA monoubiquitination, whereas MMC treatment (which mostly forms ICLs) did not (Hicks et al., 2010). Furthermore, REV1 was recruited to cellular foci induced by MMC independently of PCNA ubiquitination.

So how are TLS polymerases recruited to sites of ICL repair? An obvious candidate would be ubiquitinated FANCD2/FANCI, which is required for the incision and TLS steps in *Xenopus* egg extracts. However, it has been shown that while mutation rates are reduced in cells with deficiencies in the FA core complex, they are increased in FANCD2-deficient cells, suggesting that the core complex, but not FANCD2 promotes TLS (Niedzwiedz et al., 2004). More recent evidence suggests that the FA core complex directly interacts with REV1 (Kim et al., 2012; Mirchandani et al., 2008). These findings are also supported by functional studies in *Xenopus* egg extracts, where the depletion of FANCA led to a reduction in REV1 binding to ICLs, while FANCD2-I depletion did not

(Budzowska et al., 2015). Given that REV1 interacts with the other Y-family polymerases (Guo et al., 2003; Wojtaszek et al., 2012) this recruitment could then facilitate the formation of a TLS complex involving multiple polymerases at the site of the lesion.

The situation is likely to be more complex however. For example, unlike REV1, Pol η does not appear to be regulated by FA in response to ICLs (Hicks et al., 2010; Kim et al., 2012). It has recently been shown that Pol η interacts with the Pol δ subunit POLD2 (Baldeck et al., 2015). Furthermore, Pol ζ has also been shown to share subunits with Pol δ and this interaction enhances the TLS activity of Pol ζ (Baranovskiy et al., 2012; Lee et al., 2014). It is therefore possible that these connections with a replicative polymerase may also facilitate TLS activity in ICL repair. The regulation of the activity of TLS polymerases during ICL repair therefore appears to be complex and the elucidation of the underlying mechanisms will require many additional studies.

TLS POLYMERASES IN CHEMOTHERAPY

ICL-inducing agents are widely used in anti-cancer therapy, however clinical efficacy is limited by the development of resistance and well as secondary tumors (Deans and West, 2011). Error-prone translesion synthesis during ICL repair is a key step contributing to resistance as well as increased therapy-induced mutation load and has been implicated in the emergence of secondary tumors. Studies have shown that knock-down of REV3 and REV1 not only lead to cisplatin sensitivity in a variety of experimental tumor models, but also led to a

significant reduction in drug-induced mutagenesis (Doles et al., 2010; Xie et al., 2010).

Consistent with this, increased expression of TLS polymerases in multiple cancers has been correlated with a poor prognosis and response to chemotherapy. Elevated levels of Pol η and Pol ζ were found to mediate resistance to cisplatin treatment in ovarian cancer stem cells and cervical cancer cells, respectively (Srivastava et al., 2015; Yang et al., 2015). Pol η levels were also found to be elevated in head and neck squamous cell carcinomas and lower Pol η level was significantly correlated with better response to cisplatin and gemcitabine therapy in patients (Zhou et al., 2013). Altogether, these studies suggest TLS polymerase levels could be a useful predictor for therapeutic outcomes and that inhibition of TLS polymerase in cancer therapy could lead to a dual benefit – reduced occurrence of resistance and reduced secondary tumor formation. A more detailed understanding of the roles of individual TLS polymerases in the contribution to ICL repair to specific agents will be an important guide to determine which polymerases may be specifically targeted for treatment modalities involving a variety of crosslinking agents.

REFERENCES

Albertella, M.R., Green, C.M., Lehmann, A.R., and O'Connor, M.J. (2005). A role for polymerase η in the cellular tolerance to cisplatin-induced damage. *Cancer Res. 65*, 9799-9806.

Allerston, C.K., Lee, S.Y., Newman, J.A., Schofield, C.J., McHugh, P.J., and Gileadi, O. (2015). The structures of the SNM1A and SNM1B/Apollo nuclease domains reveal a potential basis for their distinct DNA processing activities. *Nucleic Acids Res. 43*, 11047-11060.

Angelov, T., Guainazzi, A., and Scharer, O.D. (2009). Generation of DNA interstrand cross-links by post-synthetic reductive amination. *Org Lett 11*, 661-664.

Baldeck, N., Janel-Bintz, R., Wagner, J., Tissier, A., Fuchs, R.P., Burkovics, P., Haracska, L., Despras, E., Bichara, M., Chatton, B.*, et al.* (2015). FF483-484 motif of human Poleta mediates its interaction with the POLD2 subunit of Poldelta and contributes to DNA damage tolerance. *Nucleic Acids Res. 43*, 2116-2125.

Baranovskiy, A.G., Lada, A.G., Siebler, H.M., Zhang, Y., Pavlov, Y.I., and Tahirov, T.H. (2012). DNA polymerase delta and zeta switch by sharing accessory subunits of DNA polymerase delta. *J. Biol. Chem. 287*, 17281-17287.

Bessho, T., Mu, D., and Sancar, A. (1997). Initiation of DNA interstrand cross-link repair in humans: the nucleotide excision repair system makes dual incisions 5' to the cross-linked base and removes a 22- to 28-nucleotide-long damage-free strand. *Mol Cell Biol. 17*, 6822-6830.

Biertumpfel, C., Zhao, Y., Kondo, Y., Ramon-Maiques, S., Gregory, M., Lee, J.Y., Masutani, C., Lehmann, A.R., Hanaoka, F., and Yang, W. (2010). Structure and mechanism of human DNA polymerase eta. *Nature 465*, 1044-1048.

Budzowska, M., Graham, T.G., Sobeck, A., Waga, S., and Walter, J.C. (2015). Regulation of the Rev1-pol zeta complex during bypass of a DNA interstrand cross-link. *EMBO J 34*, 1971-1985.

Chabner, B.A., Amrein, P.C., Druker, B.J., Michaelson, M.D., Mistsiades, C.S., Goss, P.E., Ryan, D.P., Ramachandra, S., Richardson, P.G., Supko, J.G.*, et al.* (2005). Antineoplastic Agents. In Goodman & Gilman's The Pharmacological Basis of Therapeutics, L.L. Brunton, J.S. Lazo, and K.L. Parker, eds. (New York: McGraw-Hill), pp. 1315-1403.

Chabner, B.A., and Roberts, T.G., Jr. (2005). Timeline: Chemotherapy and the war on cancer. *Nat Rev Cancer 5*, 65-72.

Chen, Y.-w., Cleaver, J.E., Hanaoka, F., Chang, C.-f., and Chou, K.-m. (2006). A novel role of DNA polymerase η in modulating cellular sensitivity to chemotherapeutic agents. *Mol Cancer Res. 4*, 257-265.

Clauson, C., Scharer, O.D., and Niedernhofer, L. (2013). Advances in understanding the complex mechanisms of DNA interstrand cross-link repair. *Cold Spring Harb Perspect Med 3,* a012732.

Davies, A.A., Huttner, D., Daigaku, Y., Chen, S., and Ulrich, H.D. (2008). Activation of ubiquitin-dependent DNA damage bypass is mediated by replication protein a. *Mol Cell 29*, 625-636.

Deans, A.J., and West, S.C. (2011). DNA interstrand crosslink repair and cancer. *Nat Rev Cancer 11*, 467-480.

Doles, J., Oliver, T.G., Cameron, E.R., Hsu, G., Jacks, T., Walker, G.C., and Hemann, M.T. (2010). Suppression of Rev3, the catalytic subunit of Pol{zeta}, sensitizes drug-resistant lung tumors to chemotherapy. *Proc Natl Acad Sci U S A 107*, 20786-20791.

Enoiu, M., Jiricny, J., and Schärer, O.D. (2012). Repair of cisplatin-induced DNA interstrand crosslinks by a replication-independent pathway involving transcription-coupled repair and translesion synthesis. *Nucleic Acids Res. 40*, 8953-8964.

Gan, G.N., Wittschieben, J.P., Wittschieben, B.O., and Wood, R.D. (2008). DNA polymerase zeta (pol [zeta]) in higher eukaryotes. *Cell Res. 18*, 174-183.

Goodman, L.S., Wintrobe, M.M., and et al. (1946). Nitrogen mustard therapy; use of methyl-bis (beta-chloroethyl) amine hydrochloride and tris (beta-chloroethyl) amine hydrochloride for Hodgkin's disease, lymphosarcoma, leukemia and certain allied and miscellaneous disorders. *J Am Med Assoc 132*, 126-132.

Guainazzi, A., Campbell, A.J., Angelov, T., Simmerling, C., and Scharer, O.D. (2010). Synthesis and molecular modeling of a nitrogen mustard DNA interstrand crosslink. *Chemistry 16*, 12100-12103.

Guo, C., Fischhaber, P.L., Luk-Paszyc, M.J., Masuda, Y., Zhou, J., Kamiya, K., Kisker, C., and Friedberg, E.C. (2003). Mouse Rev1 protein interacts with multiple DNA polymerases involved in translesion DNA synthesis. *EMBO J 22*, 6621-6630.

Haracska, L., Prakash, S., and Prakash, L. (2003). Yeast DNA polymerase zeta is an efficient extender of primer ends opposite from 7,8-dihydro-8-Oxoguanine and O6-methylguanine. *Mol Cell Biol 23*, 1453-1459.

Hicks, J.K., Chute, C.L., Paulsen, M.T., Ragland, R.L., Howlett, N.G., Gueranger, Q., Glover, T.W., and Canman, C.E. (2010). Differential roles for DNA

polymerases eta, zeta, and REV1 in lesion bypass of intrastrand versus interstrand DNA cross-links. *Mol Cell Biol 30*, 1217-1230.

Hlavin, E.M., Smeaton, M.B., Noronha, A.M., Wilds, C.J., and Miller, P.S. (2010). Cross-link structure affects replication-independent DNA interstrand cross-link repair in mammalian cells. *Biochemistry 49*, 3977-3988.

Ho, T.V., Guainazzi, A., Derkunt, S.B., Enoiu, M., and Schärer, O.D. (2011). Structure-dependent bypass of DNA interstrand crosslinks by translesion synthesis polymerases. *Nucleic Acids Res. 39*, 7455-7464.

Hodskinson, M.R., Silhan, J., Crossan, G.P., Garaycoechea, J.I., Mukherjee, S., Johnson, C.M., Scharer, O.D., and Patel, K.J. (2014). Mouse SLX4 is a tumor suppressor that stimulates the activity of the nuclease XPF-ERCC1 in DNA crosslink repair. *Mol Cell 54*, 472-484.

Huang, J., Liu, S., Bellani, M.A., Thazhathveetil, A.K., Ling, C., de Winter, J.P., Wang, Y., Wang, W., and Seidman, M.M. (2013). The DNA translocase FANCM/MHF promotes replication traverse of DNA interstrand crosslinks. *Mol Cell 52*, 434-446.

Iyama, T., Lee, S.Y., Berquist, B.R., Gileadi, O., Bohr, V.A., Seidman, M.M., McHugh, P.J., and Wilson, D.M., 3rd (2015). CSB interacts with SNM1A and promotes DNA interstrand crosslink processing. *Nucleic Acids Res. 43*, 247-258.

Kim, H., Yang, K., Dejsuphong, D., and D'Andrea, A.D. (2012). Regulation of Rev1 by the Fanconi anemia core complex. *Nat Struct Mol Biol. 19*, 164-170.

Klein Douwel, D., Boonen, R.A., Long, D.T., Szypowska, A.A., Raschle, M., Walter, J.C., and Knipscheer, P. (2014). XPF-ERCC1 acts in Unhooking DNA interstrand crosslinks in cooperation with FANCD2 and FANCP/SLX4. *Mol Cell 54*, 460-471.

Klug, A.R., Harbut, M.B., Lloyd, R.S., and Minko, I.G. (2012). Replication bypass of N^2-deoxyguanosine interstrand cross-links by human DNA polymerases η and ι. *Chem Res Toxicol 25*, 755-762.

Knipscheer, P., Räschle, M., Smogorzewska, A., Enoiu, M., Ho, T.V., Schärer, O.D., Elledge, S.J., and Walter, J.C. (2009). The Fanconi anemia pathway promotes replication-dependent DNA interstrand cross-link repair. *Science 326*, 1698-1701.

Lee, Y.S., Gregory, M.T., and Yang, W. (2014). Human Pol zeta purified with accessory subunits is active in translesion DNA synthesis and complements Pol eta in cisplatin bypass. *Proc Natl Acad Sci U S A 111*, 2954-2959.

Long, D.T., Joukov, V., Budzowska, M., and Walter, J.C. (2014). BRCA1 promotes unloading of the CMG helicase from a stalled DNA replication fork. *Mol Cell 56*, 174-185.

Marteijn, J.A., Lans, H., Vermeulen, W., and Hoeijmakers, J.H. (2014). Understanding nucleotide excision repair and its roles in cancer and ageing. *Nat Rev Mol Cell Biol 15*, 465-481.

Masta, A., Gray, P.J., and Phillips, D.R. (1994). Molecular basis of nitrogen mustard effects on transcription processes: role of depurination. *Nucleic Acids Res. 22*, 3880-3886.

Millard, J.T., Raucher, S., and Hopkins, P.B. (1990). Mechlorethamine cross-links deoxyguanosine residues at 5'-GNC sequences in duplex DNA fragments. J *Am Chem Soc 112*, 2459-2460.

Millard, J.T., Weidner, M.F., Kirchner, J.J., Ribeiro, S., and Hopkins, P.B. (1991). Sequence preferences of DNA interstrand crosslinking agents: quantitation of interstrand crosslink locations in DNA duplex fragments containing multiple crosslinkable sites. *Nucleic Acids Res 19*, 1885-1891.

Minko, I.G., Harbut, M.B., Kozekov, I.D., Kozekova, A., Jakobs, P.M., Olson, S.B., Moses, R.E., Harris, T.M., Rizzo, C.J., and Lloyd, R.S. (2008a). Role for DNA polymerase κ in the processing of N^2-N^2-guanine interstrand cross-links. *J Biol Chem 283*, 17075-17082.

Minko, I.G., Yamanaka, K., Kozekov, I.D., Kozekova, A., Indiani, C., O'Donnell, M.E., Jiang, Q., Goodman, M.F., Rizzo, C.J., and Lloyd, R.S. (2008b). Replication bypass of the acrolein-mediated deoxyguanine DNA-peptide cross-links by DNA polymerases of the DinB family. *Chem Res Toxicol 21*, 1983-1990.

Mirchandani, K.D., McCaffrey, R.M., and D'Andrea, A.D. (2008). The Fanconi anemia core complex is required for efficient point mutagenesis and Rev1 foci assembly. *DNA Repair (Amst) 7*, 902-911.

Misra, R.R., and Vos, J.M. (1993). Defective replication of psoralen adducts detected at the gene-specific level in xeroderma pigmentosum variant cells. *Mol Cell Biol 13*, 1002-1012.

Mogi, S., Butcher, C.E., and Oh, D.H. (2008). DNA polymerase η reduces the γ-H2AX response to psoralen interstrand crosslinks in human cells. *Exp Cell Res 314*, 887-895.

Moldovan, G.-L., Madhavan, M.V., Mirchandani, K.D., McCaffrey, R.M., Vinciguerra, P., and D'Andrea, A.D. (2010). DNA polymerase POLN participates in cross-link repair and homologous recombination. *Mol Cell Biol 30*, 1088-1096.

Mu, D., Bessho, T., Nechev, L.V., Chen, D.J., Harris, T.M., Hearst, J.E., and Sancar, A. (2000). DNA interstrand cross-links induce futile repair synthesis in mammalian cell extracts. *Mol Cell Biol 20*, 2446-2454.

Mukherjee, A., and Vasquez, K.M. (2016). HMGB1 interacts with XPA to facilitate the processing of DNA interstrand crosslinks in human cells. *Nucleic Acids Res 44*, 1151-1160.

Mukherjee, S., Guainazzi, A., and Scharer, O.D. (2014). Synthesis of structurally diverse major groove DNA interstrand crosslinks using three different aldehyde precursors. Nucleic Acids Res. *42*, 7429-35

Muniandy, P.A., Thapa, D., Thazhathveetil, A.K., Liu, S.-t., and Seidman, M.M. (2009). Repair of laser-localized DNA interstrand cross-links in G1 phase mammalian cells. *J Biol Chem 284*, 27908-27917.

Niedzwiedz, W., Mosedale, G., Johnson, M., Ong, C.Y., Pace, P., and Patel, K.J. (2004). The Fanconi anaemia gene *FANCC* promotes homologous recombination and error-prone DNA repair. *Mol Cell 15*, 607-620.

Nojima, K., Hochegger, H., Saberi, A., Fukushima, T., Kikuchi, K., Yoshimura, M., Orelli, B.J., Bishop, D.K., Hirano, S., Ohzeki, M.*, et al.* (2005). Multiple repair pathways mediate tolerance to chemotherapeutic cross-linking agents in vertebrate cells. *Cancer Res 65*, 11704-11711.

Noll, D.M., Mason, T.M., and Miller, P.S. (2006). Formation and repair of interstrand cross-links in DNA. *Chem Rev 106*, 277-301.

Ogi, T., Shinkai, Y., Tanaka, K., and Ohmori, H. (2002). Polkappa protects mammalian cells against the lethal and mutagenic effects of benzo[a]pyrene. *Proc Natl Acad Sci U S A 99*, 15548-15553.

Okada, T., Sonoda, E., Yoshimura, M., Kawano, Y., Saya, H., Kohzaki, M., and Takeda, S. (2005). Multiple roles of vertebrate REV genes in DNA repair and recombination. *Mol Cell Biol 25*, 6103-6111.

Pizzolato, J., Mukherjee, S., Scharer, O.D., and Jiricny, J. (2015). FANCD2-associated nuclease 1, but not exonuclease 1 or flap endonuclease 1, is able to unhook DNA interstrand cross-links in vitro. *J Biol Chem 290*, 22602-22611.

Povirk, L.F., and Shuker, D.E. (1994). DNA damage and mutagenesis induced by nitrogen mustards. *Mutat Res 318*, 205-226.

Räschle, M., Knipscheer, P., Enoiu, M., Angelov, T., Sun, J., Griffith, J.D., Ellenberger, T.E., Schärer, O.D., and Walter, J.C. (2008). Mechanism of replication-coupled DNA interstrand crosslink repair. *Cell 134*, 969-980.

Rink, S.M., and Hopkins, P.B. (1995). A mechlorethamine-induced DNA interstrand cross-link bends duplex DNA. *Biochemistry 34*, 1439-1445.

Rink, S.M., Solomon, M.S., Taylor, M.J., Rajur, S.B., McLaughlin, L.W., and Hopkins, P.B. (1993). Covalent structure of a nitrogen mustard-induced DNA interstrand cross-link: an N7-to-N7 linkage of deoxyguanosine residues at the duplex sequence 5'-d(GNC). *J Am Chem Soc 115*, 2551-2557.

Roy, U., and Scharer, O.D. (2016). Involvement of translesion synthesis DNA polymerases in DNA interstrand crosslink repair. *DNA Repair (Amst) 44*, 33-41.

Sale, J.E., Lehmann, A.R., and Woodgate, R. (2012). Y-family DNA polymerases and their role in tolerance of cellular DNA damage. *Nat Rev Mol Cell Biol 13*, 141-152.

Sarkar, S., Davies, A.A., Ulrich, H.D., and McHugh, P.J. (2006). DNA interstrand crosslink repair during G1 involves nucleotide excision repair and DNA polymerase zeta. *EMBO J 25*, 1285-1294.

Scharer, O.D. (2005). DNA interstrand crosslinks: natural and drug-induced DNA adducts that induce unique cellular responses. *Chembiochem 6*, 27-32.

Scharer, O.D. (2013). Nucleotide excision repair in eukaryotes. *Cold Spring Harbor perspectives in biology 5*, a012609.

Semlow, D.R., Zhang, J., Budzowska, M., Drohat, A.C., and Walter, J.C. (2016). Replication-Dependent Unhooking of DNA Interstrand Cross-Links by the NEIL3 Glycosylase. *Cell 167*, 498-511 e414.

Sengerová, B., Allerston, C.K., Abu, M., Lee, S.Y., Hartley, J., Kiakos, K., Schofield, C.J., Hartley, J.A., Gileadi, O., and McHugh, P.J. (2012). Characterization of the human SNM1A and SNM1B/Apollo DNA repair exonucleases. *J Biol Chem 287*, 26254-26267.

Shachar, S., Ziv, O., Avkin, S., Adar, S., Wittschieben, J., Reissner, T., Chaney, S., Friedberg, E.C., Wang, Z., Carell, T.*, et al.* (2009). Two-polymerase mechanisms dictate error-free and error-prone translesion DNA synthesis in mammals. *EMBO J 28*, 383-393.

Sharma, S., and Canman, C.E. (2012). REV1 and DNA polymerase zeta in DNA interstrand crosslink repair. *Environ Mol Mutagen 53*, 725-740.

Sharma, S., Helchowski, C.M., and Canman, C.E. (2012). The roles of DNA polymerase ζ and the Y family DNA polymerases in promoting or preventing genome instability. *Mutation Research.*

Shen, X., Jun, S., O'Neal, L.E., Sonoda, E., Bemark, M., Sale, J.E., and Li, L. (2006). REV3 and REV1 play major roles in recombination-independent repair of

DNA interstrand cross-links mediated by monoubiquitinated proliferating cell nuclear antigen (PCNA). *J Biol Chem 281*, 13869-13872.

Simpson, L.J., and Sale, J.E. (2003). Rev1 is essential for DNA damage tolerance and non-templated immunoglobulin gene mutation in a vertebrate cell line. *EMBO J 22*, 1654-1664.

Smeaton, M.B., Hlavin, E.M., McGregor Mason, T., Noronha, A.M., Wilds, C.J., and Miller, P.S. (2008). Distortion-dependent unhooking of interstrand cross-links in mammalian cell extracts. *Biochemistry 47*, 9920-9930.

Smith, L.A., Makarova, A.V., Samson, L., Thiesen, K.E., Dhar, A., and Bessho, T. (2012). Bypass of a psoralen DNA interstrand cross-link by DNA polymerases beta, iota, and kappa in vitro. *Biochemistry 51*, 8931-8938.

Sonoda, E., Okada, T., Zhao, G.Y., Tateishi, S., Araki, K., Yamaizumi, M., Yagi, T., Verkaik, N.S., van Gent, D.C., Takata, M.*, et al.* (2003). Multiple roles of Rev3, the catalytic subunit of pol[zeta] in maintaining genome stability in vertebrates. *EMBO J 22*, 3188-3197.

Srivastava, A.K., Han, C., Zhao, R., Cui, T., Dai, Y., Mao, C., Zhao, W., Zhang, X., Yu, J., and Wang, Q.E. (2015). Enhanced expression of DNA polymerase eta contributes to cisplatin resistance of ovarian cancer stem cells. *Proc Natl Acad Sci U S A 112*, 4411-4416.

Takeiri, A., Wada, N.A., Motoyama, S., Matsuzaki, K., Tateishi, H., Matsumoto, K., Niimi, N., Sassa, A., Gruz, P., Masumura, K.*, et al.* (2014). In vivo evidence that DNA polymerase kappa is responsible for error-free bypass across DNA cross-links induced by mitomycin C. *DNA Repair (Amst) 24*, 113-121.

Wang, A.T., Sengerová, B., Cattell, E., Inagawa, T., Hartley, J.M., Kiakos, K., Burgess-Brown, N.A., Swift, L.P., Enzlin, J.H., Schofield, C.J.*, et al.* (2011). Human SNM1A and XPF–ERCC1 collaborate to initiate DNA interstrand cross-link repair. *Genes & Dev 25*, 1859-1870.

Wang, R., Persky, N.S., Yoo, B., Ouerfelli, O., Smogorzewska, A., Elledge, S.J., and Pavletich, N.P. (2014). DNA repair. Mechanism of DNA interstrand cross-link processing by repair nuclease FAN1. *Science 346*, 1127-1130.

Wang, X., Peterson, C.A., Zheng, H., Nairn, R.S., Legerski, R.J., and Li, L. (2001). Involvement of nucleotide excision repair in a recombination-independent and error-prone pathway of DNA interstrand cross-link repair. *Mol Cell Biol 21*, 713-720.

Williams, Hannah L., Gottesman, Max E., and Gautier, J. (2012). Replication-independent repair of DNA interstrand crosslinks. *Mol Cell 47*, 140-147.

Wojtaszek, J., Lee, C.J., D'Souza, S., Minesinger, B., Kim, H., D'Andrea, A.D., Walker, G.C., and Zhou, P. (2012). Structural basis of Rev1-mediated assembly of a quaternary vertebrate translesion polymerase complex consisting of Rev1, heterodimeric polymerase (Pol) zeta, and Pol kappa. *J Biol Chem 287*, 33836-33846.

Xie, K., Doles, J., Hemann, M.T., and Walker, G.C. (2010). Error-prone translesion synthesis mediates acquired chemoresistance. *Proc Natl Acad Sci U S A 107*, 20792-20797.

Xu, W., Ouellette, A., Ghosh, S., O'Neill, T.C., Greenberg, M.M., and Zhao, L. (2015). Mutagenic Bypass of an Oxidized Abasic Lesion-Induced DNA Interstrand Cross-Link Analogue by Human Translesion Synthesis DNA Polymerases. *Biochemistry 54*, 7409-7422.

Yamanaka, K., Minko, I.G., Takata, K.-i., Kolbanovskiy, A., Kozekov, I.D., Wood, R.D., Rizzo, C.J., and Lloyd, R.S. (2010). Novel enzymatic function of DNA polymerase ν in translesion DNA synthesis past major groove DNA−peptide and DNA−DNA cross-links. *Chem Res Toxicol 23*, 689-695.

Yang, L., Shi, T., Liu, F., Ren, C., Wang, Z., Li, Y., Tu, X., Yang, G., and Cheng, X. (2015). REV3L, a promising target in regulating the chemosensitivity of cervical cancer cells. *PLoS One 10*, e0120334.

Zhang, J., Dewar, J.M., Budzowska, M., Motnenko, A., Cohn, M.A., and Walter, J.C. (2015). DNA interstrand cross-link repair requires replication-fork convergence. *Nat Struct & Mol Biol 22*, 242-247.

Zhang, J., and Walter, J.C. (2014). Mechanism and regulation of incisions during DNA interstrand cross-link repair. *DNA Repair (Amst) 19*, 135-142.

Zhao, J., Jain, A., Iyer, R.R., Modrich, P.L., and Vasquez, K.M. (2009). Mismatch repair and nucleotide excision repair proteins cooperate in the recognition of DNA interstrand crosslinks. *Nucleic Acids Res 37*, 4420-4429.

Zhao, Y., Biertumpfel, C., Gregory, M.T., Hua, Y.J., Hanaoka, F., and Yang, W. (2012). Structural basis of human DNA polymerase eta-mediated chemoresistance to cisplatin. *Proc Natl Acad Sci U S A 109*, 7269-7274.

Zheng, H., Wang, X., Warren, A.J., Legerski, R.J., Nairn, R.S., Hamilton, J.W., and Li, L. (2003). Nucleotide excision repair- and polymerase η-mediated error-prone removal of mitomycin C interstrand cross-links. *Mol Cell Biol 23*, 754-761.

Zhou, W., Chen, Y.W., Liu, X., Chu, P., Loria, S., Wang, Y., Yen, Y., and Chou, K.M. (2013). Expression of DNA translesion synthesis polymerase eta in head and neck squamous cell cancer predicts resistance to gemcitabine and cisplatin-based chemotherapy. *PLoS One 8*, e83978.

Zietlow, L., Smith, L.A., Bessho, M., and Bessho, T. (2009). Evidence for the involvement of human DNA polymerase N in the repair of DNA interstrand cross-links. *Biochemistry 48*, 11817-11824.

CHAPTER 2

PREFACE

This chapter has been adapted from the methods paper "*Preparation of nitrogen mustard DNA interstrand crosslinks and their stable analogs for biochemical and cell biological studies*" by Alejandra Castaño, Upasana Roy and Orlando D. Schärer published online (in press) in *Methods in Enzymology*.

This chapter discusses the limitations of using native NM ICLs in ICL repair studies, and outlines two strategies developed in our lab to synthesize stable NM ICL analogs that are better suited for biochemical and cell biological studies. The first strategy involves incorporation of modified 7-deaza-dG (7CdG) bases carrying alkyl ICL precursors (Angelov et al., 2009; Guainazzi et al., 2010; Mukherjee et al., 2014). The second strategy involves incorporation of 2'-deoxy-2'-β-fluoroarabino dG (2'FdG) to stabilize the glycosidic bond in an NM ICL. Experiments with the 2'FdG NM ICL analog were performed by my colleague Alejandra Castaño. The modified 7CdG base was synthesized by Todor Angelov, Angelo Guainazzi and Shivam Mukherjee, and all experiments using the 7CdG analog were performed my me.

In the context of this thesis, the following chapter describes the synthesis of 7CdG NM ICL substrates used for all experiments discussed in Chapter 3 and Chapter 4. It outlines the preparation of ICL substrates containing structurally diverse NM ICL analogs, embedded in a duplex of 20bp or 6bp, for primer extension assays with DNA polymerases. The use of such substrates in ICL repair studies, their advantages and limitations are discussed.

INTRODUCTION

Bis-(2-chloroethyl)-amine derivatives or nitrogen mustards (NM) are bifunctional alkylating agents widely used in the clinic to treat a variety of cancers (Chabner et al., 2005). The bifunctional nature of NMs allows for the formation of DNA intra- and interstrand crosslinks. It has been shown that although DNA interstrand crosslinks (ICLs) make up only 1-5% of the total adducts, they are responsible for the cytotoxicity as they provide a complete block to essential processes such as DNA replication and transcription (Noll et al., 2006; Scharer, 2005). The ability of ICLs to induce apoptosis particularly in replicating cells provides some degree of selective cytotoxicity towards rapidly dividing cancer cells in a therapeutic setting (Deans and West, 2011). However, cellular resistance is often observed in patients treated with nitrogen mustards, in large part due to ICL repair pathways that remove NM ICLs from the genome. Several ICL repair pathways exist, and the best understood of these is coupled to replication and has been biochemically reconstituted in *Xenopus* egg extracts (**Fig. 1.4 A**) (Knipscheer et al., 2012; Raschle et al., 2008). Replication-coupled ICL repair makes use of genes involved in a number of pathways, including Fanconi anemia (FA), translesion DNA synthesis (TLS), homologous recombination (HR), and nucleotide excision repair (NER). The coordinated activity of this pathway comprises factors that recognize ICLs at stalled replication forks, remove the ICL and eventually reestablish the replication fork. In another, less understood replication-coupled ICL repair pathway, the replication fork initially bypasses the intact ICL in a process known as replication

traverse (**Fig. 1.4 B**), presumably followed by ICL removal at a later stage (Huang et al., 2013). In addition, ICL repair is also known to occur outside of replication in an NER- and TLS-dependent manner, but the mechanistic details remain poorly understood as well (Clauson et al., 2013).

The major limitation in studying repair of NM ICLs has been the difficulty in generating substrates suitable for biochemical and cell biological studies. Two main challenges exist: first, treatment of a DNA duplex with NMs yields the desired ICLs only as a small fraction of products, with monoadducts and intrastrand crosslinks making up the majority of adducts. This makes the isolation of sufficient amounts of ICLs difficult (Millard et al., 1990; Povirk and Shuker, 1994). Second, the alkylation of guanosine at N7 yields a positive charge on the purine ring (**1**, **Fig. 2.1A**), rendering the base prone to spontaneous depurination, leading to loss of the ICL, formation of abasic sites and eventually strand breaks (**Fig. 2.1B**) (Gates, 2009).

To overcome these limitations, we have developed two different methods to generate stable analogs of NM ICLs that closely mimic the native NM ICL substrates (**Fig. 2.1A**). Our first approach is based on the previously synthesized 7-deaza-guanine phosphoramidite (7CdG, **2**, **Fig. 2.1A**) precursors bearing masked diols that are easily incorporated into DNA oligomers by solid DNA synthesis (Angelov et al., 2009; Guainazzi et al., 2010; Mukherjee et al., 2014). After incorporation into 5'-GNC-3' sequences, two complementary oligonucleotides containing 7CdG-residues with alkyldiol side chains at the 7 position are annealed. The diols are oxidized to aldehydes with sodium periodate

and coupled with hydrazine resulting in a high yield of the NM ICL mimic **2** (**Fig. 2.1A and 2.2B**). By using alkylaldehyde chains and diamines of different lengths, structurally diverse NM-like ICLs can be generated, for example the 5 atom ICL (7CdG-5a, **2**), a close structural mimic of the native NM ICL, or the 8 atom ICL (7CdG-8a, **4**), that has a longer linkage and no distortion in the DNA (**Fig. 2.1A**).

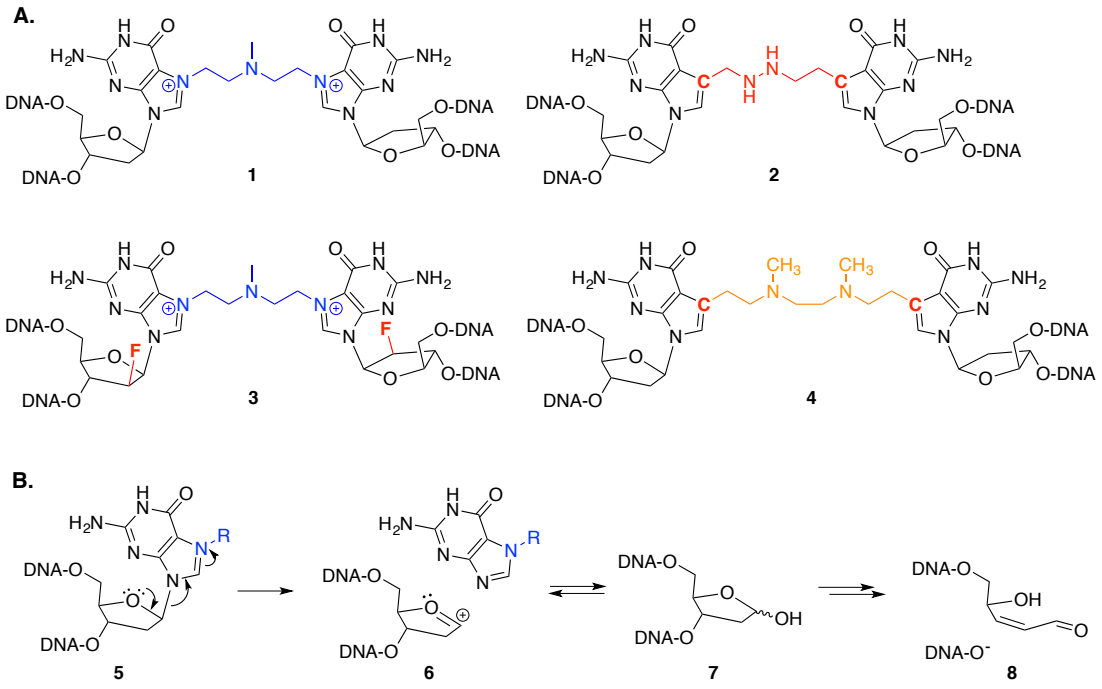


Figure 2.1: Native and stable NM ICLs. A. Nitrogen mustard (NM) interstrand crosslinks (ICLs) formed by reaction of DNA with NM (**1**) and stable analogs in which the glycosidic bond is stabilized by using 7-deazaguanine bases (**2,** 7CdG-5a), to remove the positive charge in the purine ring) or 2'-fluoro-deoxyribose sugars (**3,** 2'FdG), to prevent depurination by destabilizing the positive charge formed during glycosidic bond cleavage). The approach to generate 7CdG ICL additionally allows for the synthesis of structurally diverse ICLs, such as **4** (7CdG-8a). **B.** The positive charge in the native NM ICL destabilizes the glycosidic bond (**5**) and leads to depurination (**6**) resulting in the formation of an abasic site (**7**) and strand cleavage (**8**).

In our second approach, NM ICLs are rendered resistant to glycosydic bond cleavage by incorporation of 2'-deoxy-2'-β-fluoroarabino guanosine (2'FdG,

3, **Fig. 2.1A**) in the 5'-GNC-3' sequences via solid phase DNA synthesis. The electronegative fluorine atom at the 2' carbon of the sugar base is strongly electron withdrawing and dramatically destabilizes the transient positive charge formed during the glycosidic bond cleavage reaction (**6**, **Fig. 2.1B**), thus de facto eliminating spontaneous depurination (Lee et al., 2008; Watts et al., 2009).

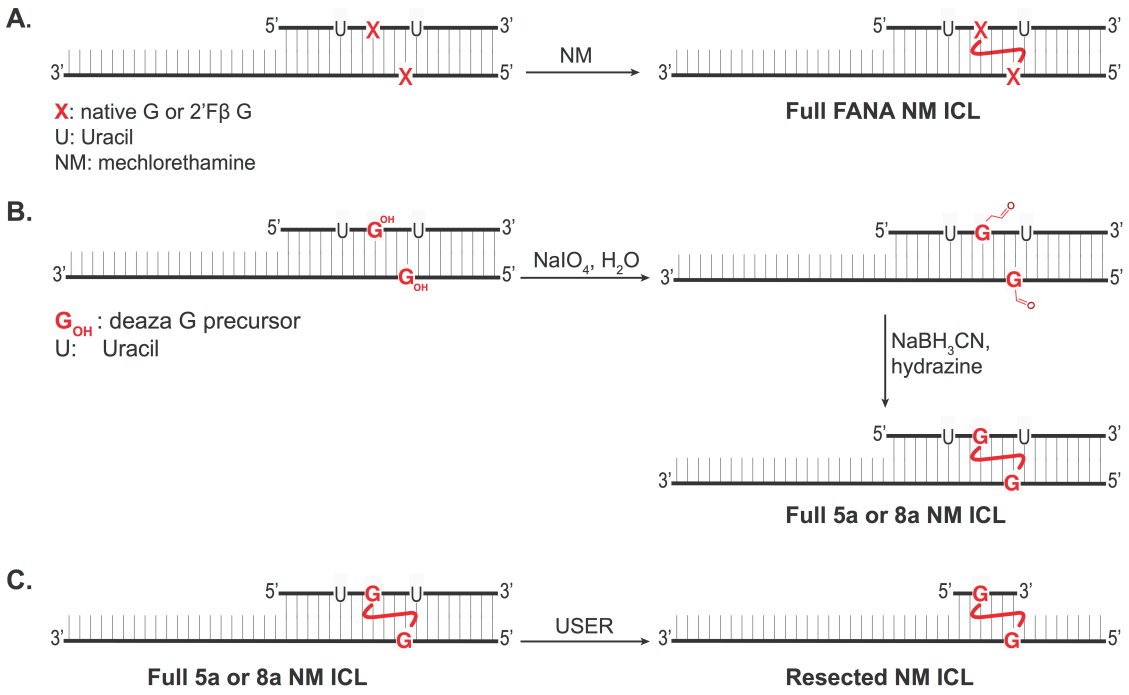


Figure 2.2: Scheme for the formation of NM ICLs. **A.** Native or 2'FdG NM ICLs are formed upon reaction of a duplex DNA containing a site specific 5'-GNC-3' sequence with excess mechlorethamine. **B.** 5a or 8a 7CdG NM ICLs are formed by oxidation of the diol-bearing 7-deaza-guanosines (incorporated within a 5'-GNC-3' sequence in a duplex) with sodium periodate to the aldehydes followed by reductive amination using hydrazine or DMEDA in the presence of sodium cyanoborohydride. **C.** Resected NM ICLs are produced by cleavage of uracil residues flanking the ICL with USER mix (UDG and ENDOVIII).

The generation of various synthetic stable NM ICL analogs and the relative advantages and disadvantages of our approaches are described. We used these ICLs in primer extension assays with Klenow polymerase to assess

the stability of the ICLs. Additionally, we describe a method to generate substrates in which the duplex around the ICL can be resected by incorporation of uracil residues in the oligonucleotide followed by enzymatic cleavage with UDG / EndoVIII (USER, **Fig. 2.2C**). These methods allow the synthesis of a panel of NM-like ICLs for the study of ICL repair mechanisms.

MATERIALS

Reagents and Buffers

- 1 mM DNA oligonucleotides in 1x TE:

  5'-CCCTCTUCTG*TCCUTCTTTC-3' (20mer), and

  5'-GAAAGAAGG*ACAGAAGAGGGTACCATCATAGAGTCAGTG-3' (where G* represents dG, 2'FdG, or 7CdG) (Notes 2 and 4)
- Use ultrapure water 18 MΩ.cm for the preparation of solutions and in all reactions. All the reagents are of analytical grade purity.
- 1x TE (10 mM Tris-HCl pH 7.4, 1 mM EDTA)
- 40 mM sodium cacodylate (pH 7): mix 0.86 g sodium cacodylate trihydrate (Sigma-Aldrich) in 70 mL water. Adjust pH to 7 by addition of 0.2 M HCl, and bring final volume to 100 mL with MilliQ water.
- 5 mM mechlorethamine hydrochloride 98% (5 mM NM): use fresh solution prior reaction with DNA by mixing 1 mg of mechlorethamine (Sigma-Aldrich) in 1 mL 40 mM sodium cacodylate pH 7 (Note 1).
- 50 mM sodium periodate solution: dissolve 0.1 g sodium metaperiodate in 10 mL ultrapure water. Store in the dark at 4°C.

- 0.5 M sodium cyanoborohydride solution: dissolve 31.0 mg sodium cyanoborohydride in 1 mL ultrapure water. Store in the dark at 4°C.
- 5 mM hydrazine solution: Add 2.5 mL of 64-65% hydrazine monohydrate solution (Sigma-Aldrich) to 10 mL ultrapure water. Store in the dark at 4°C.
- 5 mM DMEDA solution: Add 5.4 mL dimethylethylenediamine (Sigma-Aldrich) to 10 mL ultrapure water. Store in the dark at 4°C.
- 1 M sodium phosphate buffer (pH 5.4)
- 5 mM sodium borate buffer (pH 8) for electroelution: mix 1.95 g sodium tetraborate in 800 mL water. Adjust pH to 8 with boric acid and bring up volume to 1 L with milliQ water.
- 9:1 MALDI ToF MS matrix solution: mix 1 mL of 50 mg/mL 3-hydroxypicolinic acid (Protea Bioscience) in 50% acetonitrile/Milli-Q water and 1 mL of 50 mg/mL ammonium citrate (Sigma-Aldrich) in Milli-Q water. Store at 4°C in the dark.
- 50% v/v acetonitrile:water
- 5x Tris–Borate-EDTA buffer (TBE): dissolve 54 g Tris base, 27.5 g boric acid, 20 mL 0.5 M EDTA in 800 mL ultrapure water. Make up volume to 1 L with ultrapure water (pH 8.0).
- 0.5x TBE DPAGE running buffer: dilute 100 mL 5x TBE stock solution with 900 mL H_2O.
- 20% denaturing polyacrylamide gel electrophoresis (20% DPAGE) casting solution: stir vigorously 210.2 g ultrapure urea, 250 mL Acrylamide/Bis 19:1, 40% (w/v) solution, 50 mL 5x TBE in 400 mL ultrapure water. Once all

components have dissolved bring volume up to 500 mL with ultra pure water.

- 0% denaturing polyacrylamide gel electrophoresis (0% DPAGE) casting solution: mix 210.2 g Urea, 50 mL 5x TBE in 400 mL ultrapure water. Once all components have dissolved bring volume up to 500 mL with ultra pure water.
- TEMED
- 10% ammonium persulfate (APS) solution (mix 1 g APS in 10 mL H_2O, store at 4°C)
- 80% formamide/Orange G loading buffer: mix 800 mL formamide, 0.5 mg orange G and 200 mL H_2O). Loading buffer can be stored at room temperature.
- 80% formamide / xylene cyanol / bromophenol blue tracking buffer: mix 800 mL formamide, 0.5 mg xylene cyanol, 0.5 mg bromophenol blue and 200 mL H_2O. Tracking buffer can be stored at room temperature.
- USER enzyme mix (NEB, M5505S)
- 1x SYBR Gold solution (Life Technologies): dissolve 50 mL 10000x SYBR gold into 500 mL 1x TBE in amber plastic bottle. Store at 4°C.

Equipment and Consumables

- NanoDrop or UV spectrophotometer
- Ziptip c18 pipette tips (Millipore)
- MALDI plate (MTP Anchorchip, Bruker Daltonics)

- AutoFlex II MALDI-TOF mass spectrometer (Bruker Daltonics)
- FlexAnalysis 3.0 software
- Thermomixer compact (Eppendorf)
- Heat Block
- Elutrap™ device (Schleicher & Schuell)
- Bio-Trap membranes (BT1 glycerinized and BT2 dry, Whatman)
- Sub-Cell GT Horizontal Elecectrophoresis system, 15 x 25 cm tray (BioRad) to hold the Elutrap device
- Bench-top centrifuge (Sorvall Legend Mach 1.6R with swing bucket rotor)
- 0.5 mL and 1.5 mL microcentrifuge tubes
- 15 mL Falcon tubes
- Amicon Ultra-0.5 3 KDa MWCO (centrifugal filter device, Millipore)
- Amicon Ultra-4 3 KDa MWCO (centrifugal filter device, Millipore)
- Bio Rad PowerPac (400W) power supply
- Electrophoresis tank (model V15.17 Whatman)
- V-series electrophoresis glass sandwich plates (Apogee): long (19.7 cm wide x 19.1 cm long) and short (19.7 cm wide x 16.0 cm long)
- 1.5 mm semi-prep comb (with a single and a long well, and 1.5 mm spacers and 0.75 mm spacers and analytical combs
- TLC Glass Plates with fluorescence indicator for UV shadowing (EMD/Millipore)
- Scalpel, and tweezers
- Hand held UV Lamp (254nm)

- Typhoon 9400 Fluoroimager (GE Healthcare)
- ImageQuant Software to analyze fluorescent images.

METHODS

Preparation of NM ICLs

Native and 2'FdG-containing NM ICL analogs

1. In a 1.5 mL microcentrifuge tube, mix 100 mL 1 mM 20mer 5'-CCCTCTUCTG*TCCUTCTTTC-3' and 100 mL 1 mM 39mer 5'-GAAAGAAGG*ACAGAAGAGGGTACCATCATAGAGTCAGTG-3' (where G* represents dG or 2'FdG) in 200 mL of 40mM sodium cacodylate pH 7 (250 mM final duplex concentration) (Note 2).
2. Heat this 250 mM DNA solution at 95°C for 5 min in a preheated heat block. Switch off heating block and let it cool slowly to allow annealing of the oligos.
3. NM ICL reaction: Add 3 equivalents of freshly made 5 mM NM solution (pH 7) to the annealed DNA solution (Note 1). Incubate the reaction mixture at 37°C for 3 hours in a thermomixer in the dark.
4. The resulting NM ICLs (~7% yield for 2'FdG and ~2% for canonical G) can be purified by denaturing polyacrylamide gel electrophoresis (DPAGE) as described in section 3.3 (Note 3).

7CdG-containing NM ICL analogs

1. In a 1.5 mL microcentrifuge tube, mix 25 nmol each of the 7CdG modified 20mer (5'-CCCTCTUCTG*TCCUTCTTTC-3') and 39mer (5'-GAAAGAAGG*ACAGAAGAGGGTACCATCATAGAGTCAGTG -3') where G* denotes the modified 7CdG, in 125 mL of 100 mM NaCl. Heat this mixture to 95°C for 5 minutes and cool slowly to allow the oligos to anneal. (Note 4)
2. Add 10 mL of 50 mM sodium periodate solution and 15 mL 1 M sodium phosphate buffer (pH 5.4). Incubate this mixture at 4°C overnight in the dark, and allow oxidation to occur.
3. Transfer the mixture to a Millipore Amicon column (3K MWCO) and add 0.1 M sodium phosphate buffer (pH 5.4) up to a volume of 500 mL. Centrifuge at 11,000 rpm for 30 minutes.
4. Discard flow through, and repeat twice.
5. Collect the final solution by inverting the Amicon column into a new collection tube. Centrifuge at 8,000 rpm for 2 minutes.
6. Transfer oxidized oligos from the collection tube to a new 1.5 mL microcentrifuge tube.
7. Add 10 mL 5 mM aqueous solution of the amine (hydrazine or DMEDA) and 10 mL 0.5 M sodium cyanoborohydride. Incubate overnight in the dark at room temperature to allow the crosslinking reaction to take place.

8. The ICL formation can be analyzed by loading 5 pmol of the reaction on a 15% denaturing PAGE gel. The DNA can be visualized by SYBR Gold staining.
9. The ICL can be isolated and purified by denaturing PAGE (Section 3.3).

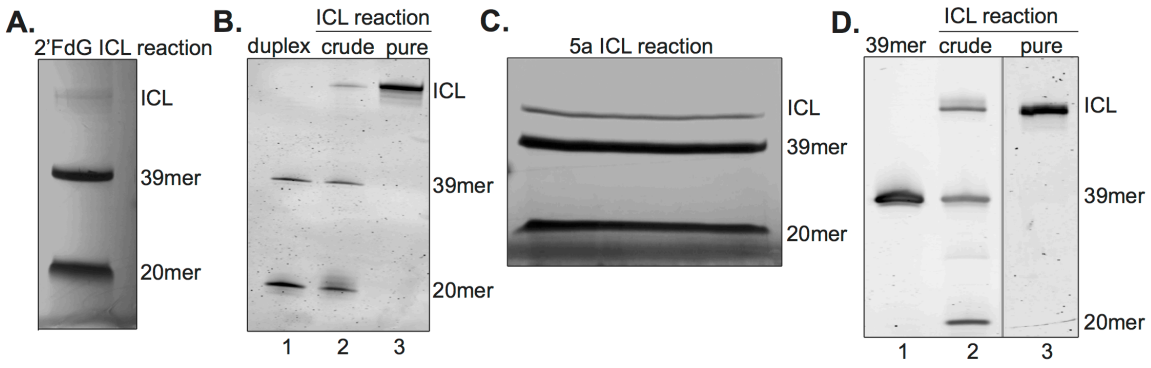


Figure 2.3: Purification of stable NM ICL analogs. **A**. Separation of a 25 nmol 2'FdG NM ICL reaction by 15% DPAGE and visualization by UV shadowing. **B.** 3 pmol of the annealed duplex (lane 1), crude reaction mixture (lane 2) and the purified 2'FdG ICL (lane 3) were resolved by 15% denaturing PAGE and visualized by SYBR gold staining. **C**. Separation of 30 nmol of a 5a 7CdG NM ICL formation reaction by 10% denaturing PAGE and visualization by UV shadowing. **D.** 5 pmol of the 39mer (lane 1), crude ICL reaction (lane 2) and the purified 5a ICL (lane 3) were resolved by 10% denaturing PAGE and visualized by SYBR gold staining. The positions of the ICL, 39mer and 20mer single-stranded oligonucleotides are indicated.

Resection of NM ICLs by USER

1. In a 1.5 mL microcentrifuge tube, dilute 500 pmol of purified ICL in 80 mL ultra pure water. Add 10 mL 0.1 M Tris-Cl (pH 7.6) and 10 mL USER enzyme mix. Incubate at 37°C for 6 hours.
2. The completion of digestion can be checked by loading an aliquot (~5 pmol) on an analytical denaturing PAGE, and visualizing the DNA by

SYBR Gold staining (section 3.3.). Once the digestion is complete, the resected ICL is ready to be purified by denaturing PAGE.

Purification and Characterization of NM ICLs

1. Assemble the DPAGE sandwich using the long and short glass plates with 2 x 1.5 mm spacers and rest horizontally on a solid support (Note 5).
2. In a 100 mL beaker, prepare 60 mL 15% DPAGE by mixing 45 mL 20% DPAGE gel buffer, 14.4 mL 0% DPAGE gel buffer, 0.6 mL 10% APS, and 20 mL TEMED.
3. Using a serological pipette, cast the gel solution into glass sandwich without introducing bubbles. Place the 1.5 mm semi-prep comb into the sandwich preventing the formation of bubbles, and allow the gel solution to polymerize for at least 40 min.
4. Suspend the NM ICL to be purified in an equal volume of 80% formamide/orange G buffer. Denature the NM ICL sample by heating the solution to 95°C for 5 min, and place sample vial on ice immediately for 5 min or until ready to load.
5. Fix the polymerized 15 % DPAGE sandwich gel vertically in electrophoresis tank (longer plate facing outwards), and fill the buffer reservoirs with 0.5x TBE running buffer to completely cover the wells. Carefully remove the comb, and rinse the formed wells with running buffer. Attach the temperature probe to the outer gel glass, and connect the tank to the PowerPac, setting it to 20 W and 50°C. Pre-run the gel for

40 min or until the temperature reaches 50°C. Rinse the wells with 0.5x TBE to remove excess urea.

6. Load the denatured DNA solution into the wide well, and load tracking dye buffer to the single well.
7. Run the PowerPac at 20 W and 50°C until the bromophenol blue dye reaches the bottom of the glass plate and orange G dye exits the gel (Note 6).
8. Disconnect the power supply and the temperature probe, and dislodge the glass sandwich. With a plastic wedge, carefully separate the plates without breaking the gel.
9. Place the gel on top of a TLC plate covered with saran wrap.
10. In a dark room, hold the UV lamp directly above the gel (254 nm wavelength), and slice the resolved bands: NM ICL 59mer (top band), uncrosslinked 39mer (middle band), and uncrosslinked 20mer (bottom band) with a clean scalpel (**Fig. 2.3A, C**) (Note 7). Turn off UV lamp.
11. Cut the gel bands further into ~ 1cm gel pieces and save in labeled falcon tubes until next step.
12. DNA isolation from the gel is done via electroelution utilizing an EluTrap system. Rinse two BT1 glycerinized membranes with ultrapure water and mount tightly at the ends of the Elutrap device. Mount one dry BT2 membrane ~2.5 cm away from positive end of the trap (collection chamber) and another dry BT2 membrane ~ 0.5 cm from positive end (gel piece chamber).

13. Load the purified NM ICL gel pieces into the gel piece chamber, and place the Elutrap device horizontally into the Sub-Cell GT electrophoresis tray. Fill all the reservoirs with 5 mM sodium borate buffer (pH 8) covering the gel pieces, the collection chamber and the outer electrophoresis tank.
14. Cover the electrophoresis tank and connect it to the power supply. Run at 200 V for 15 min. Pipette out the buffer in the collection chamber into a clean-labeled vial and place on ice.
15. Measure DNA content via Nanodrop and record the absorbance. Calculate the ng/uL NM ICL recovered.
16. Refill the collecting chamber with 5 mM sodium borate buffer and repeat steps 14-15 at least 2 more times until no DNA is detected in the collected fraction.
17. Pool all the saved fractions into a Amicon Ultra-3K 4 mL filter device and concentrate the DNA using the bench-top centrifuge at 3.5 K rpm, 4°C for 45 min. Buffer exchange with 1x TE at least twice.
18. Pipette the NM-ICL solution out of the Amicon column and save into a new labeled vial.
19. Take 2 mL of NM-ICL solution and dilute with 48 mL water. Briefly vortex and measure UV absorbance via Nanodrop. Calculate ng/uL DNA in both diluted sample and stock solution.
20. MALDI ToF MS is used to verify the mass of the crosslink. Spot 1 mL of MALDI matrix onto a MALDI plate and let dry at room temperature.

21. ZipTip the diluted NM ICL sample (at least 10 pmol) following the manufacturer's protocol (Millipore).
22. Load 1 mL zip-tipped NM ICL on spot containing MALDI matrix and allow to dry at room temperature.
23. Mount the MALDI plate to MALDI carrier and insert it into the AutoFlex II MALDI-TOF mass spectrometer (previously calibrated with known high molecular weight oligo calibrants).
24. Measure the mass/charge ratio (m/z) in a linear negative mode using ion source acceleration voltage of 20.00 kV at a frequency of 50Hz across a m/z of 7000 to 20000 Da. To achieve a high signal-to-noise ratio (SNR), represent each spectrum integrating at least 600 individual laser shots.
25. The collected spectra is analyzed and visualized using the FlexAnalysis 3.0 software (Table 1).

TABLE 2: MALDI-ToF MS of synthesized NM ICLs

Name	Calculated mass $[M-H]^-$, Da	Experimental mass $[M-H]^-$, Da	Error, %
Full 5a deaza NM ICL	18179.9	18136.4	0.2
Full 8a deaza NM ICL	18194	18226.9	0.2
Full FANA NM ICL	18200.95	18251	0.3

26. To verify the purity of the NM ICL and to quantify any degraded products arising during purification steps, run a 15% analytical DPAGE gel. Use 0.75 mm spacers and a 14 well comb in the DPAGE sandwich. Always

run the appropriate controls (e.g. the 20mer and 39mer used for original reaction) along with the NM ICL. Load 2 pmol in 10 mL 80% formamide/orange G loading dye (Note 6). Run the gel as described above.

27. To visualize the gel via fluorescence detection, suspend the gel in a solution containing 1x SYBR gold for 30 min in the dark.
28. Place gel onto Typhoon fluoroimager plate, and scan gel using the appropriate filter (**Fig. 2.3B, D, Fig. 2.4**).
29. Analyze bands with ImageQuant software, and calculate percentage of degradation (if any).

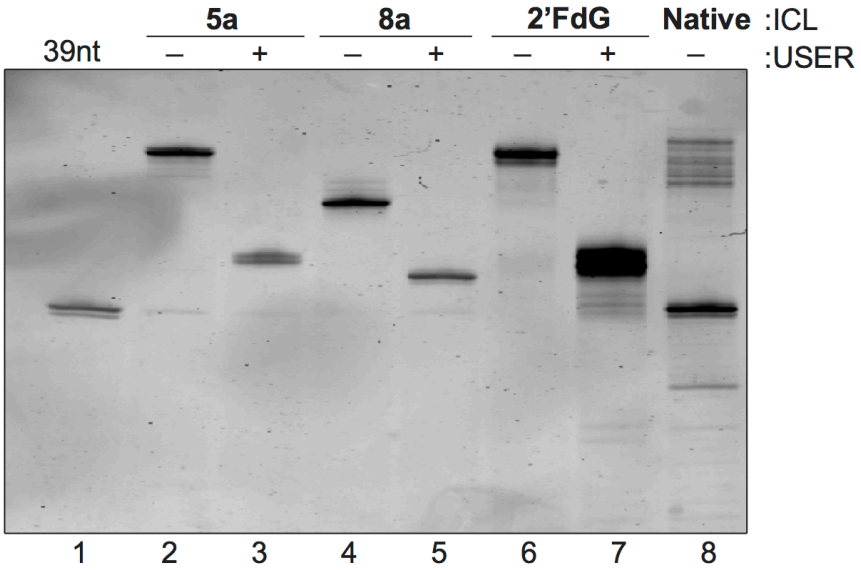


Figure 2.4: **Analysis of purified NM ICLs by 15% PAGE**. 3 pmol 39mer (lane 1), NM ICL analogs (lanes 2-7) and Native NM ICL (lane 8) were resolved by 15% denaturing PAGE and DNA visualized by SYBR Gold staining. The NM ICL analogs were either untreated (lanes 2,4,6) or treated with 0.1 U/mL USER for 6 hours at 37°C to generate the resected forms of the ICL (lanes 3,5,7). Note that the native NM ICL decomposed during purification, isolation and analysis, while the modified NM ICLs show no significant decomposition.

PRIMER EXTENSION ASSAYS WITH NM ICL ANALOGS

1. Dilute the purified NM ICL analogs and FAM labeled primer P15 (5'-FAM-CACTGACTCTATGATG) to 1 mM in 1x TE.
2. Mix 7.5 mL 1 mM ICL, 2.5 mL 1 mM primer P15, 5 mL 10x annealing buffer (100 mM Tris-Cl pH 8, 500 mM NaCl) and 35 mL of ultrapure water. Incubate overnight at room temperature to allow annealing (Note 8).
3. For primer extension assays, mix 1 mL of the annealed mixture, 1 mL NEB2 buffer, 1 mL 1 mM dNTPs and 6 mL ultrapure water.
4. Incubate at 37°C, and add 1 mL 10 nM Klenow (exo-).
5. Incubate for 5 mins at 37°C, then add 10 mL formamide loading buffer.
6. Heat to 95°C for 5 mins and then snap chill on ice.
7. Load on a 10% DPAGE gel and visualize bands by scanning the gel using Typhoon fluoroimager (**Fig. 2.5**).

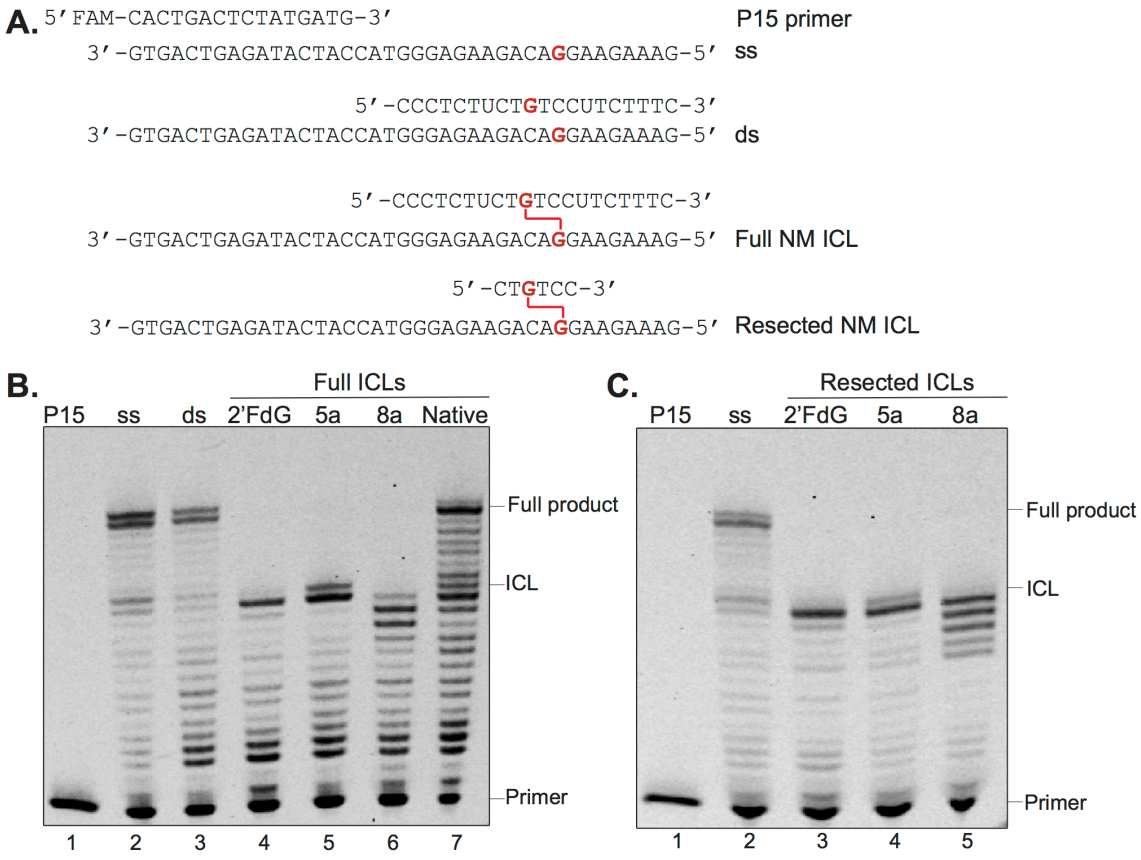


Figure 2.5: **NM ICL analogs block the polymerase reaction by Klenow (exo-) fragment**. **A.** Substrates used for primer extension assay. The modified base and crosslink are highlighted in red. **B.** Primer extension assay of full length NM ICL templates with Klenow using non-crosslinked controls (lanes 2 and 3), and full NM ICL-containing templates (lanes 4-7). Note that decomposition of the ICL in the native NM ICL allow for bypass, while the stable analogs block the polymerase. **C.** Primer extension with resected NM ICLs. Primer (lane 1), non-crosslinked control (lane 2) and resected NM ICLs (lanes 3-5). All substrates were annealed to the FAM labeled primer P15 and incubated with 1 nM Klenow for 5 min at 37°C. Products were resolved by 10% denaturing PAGE.

NOTES

1. CAUTION: Mechlorethamine is a carcinogenic/mutagenic agent! Please refer to its MSDS for proper handling and disposal.
2. Native and 2'FdG-containing 20mer and 39mer DNA oligos (Integrated DNA Technologies) are ordered HPLC purified and diluted to 1 mM stock solutions in 1x TE.

3. DNA mixture can be buffer exchanged with 1x TE utilizing an Amicon Ultra-0.5 column (3K MWCO) at 12K rpm and stored at -20°C if the newly formed NM ICL is not purified immediately.
4. The 7CdG phosphoramidite building block needs to be synthesized as described or can be requested from the authors. (Angelov et al., 2009; Guainazzi et al., 2010; Mukherjee et al., 2014). It is incorporated into the 20mer and 39mer oligos using Expedite DNA synthesizer. The final 7CdG containing oligos used for NM ICL reactions are deprotected by treatment with concentrated NH_4OH solution at 50°C for 12 hrs and purified by Agilent BondElut C18 columns.
5. The sizes of the spacers and combs can vary depending on the amount of DNA to be purified.
6. For analytical gels, (0.75 mm spacers) 30 mL DPAGE solution is sufficient. Running times can be adjusted depending the DNA sizes to be resolved. Here, we are resolving a 20mer, 39 mer and 59mer, thus the bromophenol blue migrates to about 42 bp in a 15% DPAGE. We use orange G in the loading buffer to visualize the loading. Orange G migrates faster than bromophenol blue thus it completely exits the gel.
7. Make sure to wear appropriate personal protection equipment (UV resistant goggles, long sleeve lab coat and nitrile gloves) when shining UV light to the gel. Clean the scalpel every time you slice the different types of gel bands to reduce contamination.

8. Heating of ICLs to 95°C should be avoided to preserve the stability of the ICL. All annealing reactions should be done overnight at room temperature.

SUMMARY AND CONCLUSIONS

This chapter outlines methods to synthesize stable NM ICL analogs for biochemical and cell biological studies. Formation of native NM ICLs by reaction of duplex oligonucleotide with nitrogen mustards typically gives rise to a mixture of products with a very low yield of the desired ICL. Additionally, NM ICLs are hydrolytically unstable, as depurination of the positively charged base in a native NM ICL leads to loss of the ICL and strand breaks (**Fig. 2.1B**), thereby severely limiting their use for functional studies. We compared the stability of our modified analogs to that of native NM ICLs (**Fig. 2.4 and 2.5**) and showed that the modified ICLs are stable under the experimental conditions used, whereas the native NM ICLs are not. We describe two approaches to generate analogs that mitigate the limitations of native NM ICLs. In one, the glycosydic bond is stabilized by introduction of a fluorine substituent at the 2' position (2'FdG) (**3**, **Fig. 2.1A**), which eliminates depurination, yielding a stable ICL during purification and polymerase assays (**Fig. 2.4 and 2.5**). Although the yield of ICL formation is low, the 2'FdG NM ICL can be synthesized from commercially available DNA oligonucleotides and nitrogen mustards without the need to conduct organic synthesis. Importantly, it contains a positively charged purine ring and only differs from the native ICL by addition of a fluorine atom at the 2' position in the β-

orientation, which has been shown to only minimally affect DNA structure (Martin-Pintado et al., 2012). The 7CdG NM ICL requires the synthesis of a specific phosphoramidite precursor (Angelov et al., 2009; Guainazzi et al., 2010), but can be used in any sequence context to generate a site-specific, stable NM ICL analog in a high yielding crosslinking reaction. This approach can furthermore be modified to produce structural variants such as NM ICLs that contain crosslinks of various lengths, allowing for the synthesis of ICLs of different structures for functional studies (Angelov et al., 2009; Mukherjee et al., 2014; Roy et al., 2016). 7CdG NM analogs have also been successfully incorporated into substrates for biochemical and cell based DNA repair studies (Ho et al., 2011; Hodskinson et al., 2014; Pizzolato et al., 2015; Raschle et al., 2008; Roy et al., 2016). A possible drawback of the 7CdG NM ICLs is that there are three substitutions in the crosslink compared to the native NM ICL and the crosslinked purines are not positively charged. Although our modeling studies have shown that this induces only minor effects on the DNA structure (Guainazzi et al., 2010) , the main stalling point of Klenow at the 2'FdG ICL and 7CdG ICL differs by one nucleotide (compare **Fig. 2.5B**, lanes 4 and 5), indicating that the charge may influence interaction with proteins. Thus, we anticipate that the two types of modified NM ICLs will each be useful for specific applications.

In summary, we describe the generation of two types of stable NM ICL analogs that have been and will continue to be of use to the scientific community to study biological pathways triggered by ICLs.

REFERENCES

Angelov, T., Guainazzi, A., and Scharer, O.D. (2009). Generation of DNA interstrand cross-links by post-synthetic reductive amination. *Org Lett 11*, 661-664.

Chabner, B.A., Amrein, P.C., Druker, B.J., Michaelson, M.D., Mistsiades, C.S., Goss, P.E., Ryan, D.P., Ramachandra, S., Richardson, P.G., Supko, J.G.*, et al.* (2005). Antineoplastic Agents. In Goodman & Gilman's The Pharmacological Basis of Therapeutics, L.L. Brunton, J.S. Lazo, and K.L. Parker, eds. (New York: McGraw-Hill), pp. 1315-1403.

Clauson, C., Scharer, O.D., and Niedernhofer, L. (2013). Advances in understanding the complex mechanisms of DNA interstrand cross-link repair. *Cold Spring Harb Perspect Med 3*, a012732.

Deans, A.J., and West, S.C. (2011). DNA interstrand crosslink repair and cancer. *Nat Rev Cancer 11*, 467-480.

Gates, K.S. (2009). An overview of chemical processes that damage cellular DNA: spontaneous hydrolysis, alkylation, and reactions with radicals. *Chem Res Toxicol 22*, 1747-1760.

Guainazzi, A., Campbell, A.J., Angelov, T., Simmerling, C., and Scharer, O.D. (2010). Synthesis and molecular modeling of a nitrogen mustard DNA interstrand crosslink. *Chemistry 16*, 12100-12103.

Ho, T.V., Guainazzi, A., Derkunt, S.B., Enoiu, M., and Scharer, O.D. (2011). Structure-dependent bypass of DNA interstrand crosslinks by translesion synthesis polymerases. *Nucleic Acids Res 39*, 7455-7464.

Hodskinson, M.R., Silhan, J., Crossan, G.P., Garaycoechea, J.I., Mukherjee, S., Johnson, C.M., Scharer, O.D., and Patel, K.J. (2014). Mouse SLX4 is a tumor suppressor that stimulates the activity of the nuclease XPF-ERCC1 in DNA crosslink repair. *Mol Cell 54*, 472-484.

Huang, J., Liu, S., Bellani, M.A., Thazhathveetil, A.K., Ling, C., de Winter, J.P., Wang, Y., Wang, W., and Seidman, M.M. (2013). The DNA translocase FANCM/MHF promotes replication traverse of DNA interstrand crosslinks. *Mol Cell 52*, 434-446.

Knipscheer, P., Raschle, M., Scharer, O.D., and Walter, J.C. (2012). Replication-coupled DNA interstrand cross-link repair in Xenopus egg extracts. *Methods Mol Biol 920*, 221-243.

Lee, S., Bowman, B.R., Ueno, Y., Wang, S., and Verdine, G.L. (2008). Synthesis and structure of duplex DNA containing the genotoxic nucleobase lesion N7-methylguanine. *J Am Chem Soc 130*, 11570-11571.

Martin-Pintado, N., Yahyaee-Anzahaee, M., Campos-Olivas, R., Noronha, A.M., Wilds, C.J., Damha, M.J., and Gonzalez, C. (2012). The solution structure of double helical arabino nucleic acids (ANA and 2'F-ANA): effect of arabinoses in duplex-hairpin interconversion. *Nucleic Acids Res 40*, 9329-9339.

Millard, J.T., Raucher, S., and Hopkins, P.B. (1990). Mechlorethamine cross-links deoxyguanosine residues at 5'-GNC sequences in duplex DNA fragments. *J Am Chem Soc 112*, 2459-2460.

Mukherjee, S., Guainazzi, A., and Scharer, O.D. (2014). Synthesis of structurally diverse major groove DNA interstrand crosslinks using three different aldehyde precursors. *Nucleic Acids Res.*

Noll, D.M., Mason, T.M., and Miller, P.S. (2006). Formation and repair of interstrand cross-links in DNA. *Chem Rev 106*, 277-301.

Pizzolato, J., Mukherjee, S., Scharer, O.D., and Jiricny, J. (2015). FANCD2-associated nuclease 1, but not exonuclease 1 or flap endonuclease 1, is able to unhook DNA interstrand cross-links in vitro. *J Biol Chem 290*, 22602-22611.

Povirk, L.F., and Shuker, D.E. (1994). DNA damage and mutagenesis induced by nitrogen mustards. *Mutat Res 318*, 205-226.

Raschle, M., Knipscheer, P., Enoiu, M., Angelov, T., Sun, J., Griffith, J.D., Ellenberger, T.E., Scharer, O.D., and Walter, J.C. (2008). Mechanism of replication-coupled DNA interstrand crosslink repair. *Cell 134*, 969-980.

Roy, U., Mukherjee, S., Sharma, A., Frank, E.G., and Scharer, O.D. (2016). The structure and duplex context of DNA interstrand crosslinks affects the activity of DNA polymerase eta. *Nucleic Acids Res 44*, 7281-7291.

Scharer, O.D. (2005). DNA interstrand crosslinks: natural and drug-induced DNA adducts that induce unique cellular responses. *Chembiochem 6*, 27-32.

Watts, J.K., Katolik, A., Viladoms, J., and Damha, M.J. (2009). Studies on the hydrolytic stability of 2'-fluoroarabinonucleic acid (2'F-ANA). *Org Biomol Chem 7*, 1904-1910.

CHAPTER 3

PREFACE

This chapter has been adapted from the article "*The structure and duplex context of DNA interstrand crosslinks affects the activity of DNA polymerase η*" by Upasana Roy, Shivam Mukherjee, Anjali Sharma, Ekaterina G. Frank and Orlando D. Schärer published in *Nucleic Acids Research* (2016), **44** (15), 7281 – 91.

Using a strategy developed in our lab to generate stable, site-specific, 7-deaza-dG (7CdG) NM ICL analogs, we synthesized a panel of structurally diverse NM ICL substrates to investigate the effect of ICL distortion and duplex context on translesion synthesis by pol η. The single nucleotide ICL (5a/1nt), and all modified 7CdG containing oligos were synthesized by Shivam Mukherjee, the 6a' ICL was synthesized by Anjali Sharma and the human pol η used in these experiments was purified by Ekaterina G. Frank. I synthesized all the other crosslinks used from the precursors and perfomed all the experiments with Klenow (exo-) and pol η.

Previous work from our lab has shown that pol η activity is influenced by structure of an ICL substrate (Ho et al., 2011), and this chapter investigates these structure-function relationships further. We quantified the effect of duplex context and distortion on the efficiency and accuracy of bypass by pol η. We also describe a new synthetic strategy to generate the most processed form of an ICL with a single nucleotide crosslink (5a/1nt), and study the effect of extensive duplex resection on bypass by Klenow (exo-) and pol η.

INTRODUCTION

Interstrand crosslinks (ICLs) are highly cytotoxic DNA lesions formed by a number of bifunctional alkylating agents used in cancer chemotherapy, including cisplatin, nitrogen mustards and mitomycin C. ICLs covalently link the two strands of a DNA duplex, preventing strand separation and blocking essential processes such as replication and transcription (Noll et al., 2006; Scharer, 2005). The cytotoxic effect of blocking DNA metabolism in tumor cells with high proliferation rates is the basis of the therapeutic value of ICLs as anticancer agents. One of the limitations of using ICLs in the clinic is that the complex cellular pathways that remove ICLs from the genomes of tumor cells, lead to the occurrence of resistance to such treatment (Deans and West, 2011).

In vertebrates, the predominant pathways for ICL repair are coupled to replication and involve multiple cellular pathways including Fanconi anaemia (FA), translesion DNA synthesis (TLS), homologous recombination (HR) as well as the activity of endo- and exonucleases (Clauson et al., 2013; Kottemann and Smogorzewska, 2013). Although multiple pathways for ICL repair exist, a pathway defined in replication competent *Xenopus* egg extracts using plasmids containing site-specific ICLs has provided a mechanistic framework for understanding ICL repair (**Fig. 1.4A**) (Räschle et al., 2008). In this system, two replication forks converge on an ICL, with one leading strand extending up to 1nt before the ICL, and the other leading strand stalling 20-40 nt before the ICL (Räschle et al., 2008; Zhang et al., 2015) (**Fig 1.4A, ii**). Arrival of the leading strand at the ICL triggers the FA pathway and FANCD2/FANCI ubiquitylation,

leading to dual incisions around the ICL on the opposing parental strand to generate an 'unhooked ICL' that still remains attached to one strand (**Fig. 1.4A, iii** and **iv**) (Knipscheer et al., 2009). The endonuclease ERCC1-XPF has been shown to be required for these incisions (Klein Douwel et al., 2014) and this step is believed to involve other endo or exonucleases, possibly SNM1A or SLX1 (Castor et al., 2013; Wang et al., 2011; Zhang and Walter, 2014). One of the open questions is at what distance from the ICL the incisions occur. The position of the incisions influences the subsequent step, the extension of the leading strand past the unhooked ICL by TLS polymerases (**Fig. 1.4A, v**). Following full extension past the ICL, the newly synthesized strand is ligated to the downstream Okazaki fragments, restoring one of the daughter duplexes (**Fig. 1.4A, vii**), therefore providing a template to repair the other sister chromatid by HR. NER is believed to remove the remnant of the unhooked ICL, completing the repair process. In the *Xenopus* system, the ICL remnant has been observed still attached to the parent strand after completion of the replication of both strands of the plasmid, most likely as a single cross-linked nucleotide (Räschle et al., 2008).

A critical step in ICL repair is the bypass of the unhooked ICL by DNA polymerases. This step may lead to the introduction of mutations at or around the ICL site as it is mediated by error-prone TLS polymerases. These enzymes have been furthermore implicated in mediating chemoresistance to crosslinking drugs (Doles et al., 2010; Srivastava et al., 2015; Xie et al., 2010). Although it is unknown where the incisions are made during the unhooking step in ICL repair and how many nucleotides surround the ICL (Zhang and Walter, 2014) (**Fig.**

1.4A, iii and **iv**), it is thought that unhooked ICLs can be accommodated in the enlarged active sites and bypassed by TLS polymerases (Ho and Schärer, 2010). Evidence from genetic and functional assays suggests a key role for pol ζ and Rev1 in ICL repair (Budzowska et al., 2015; Nojima et al., 2005; Räschle et al., 2008; Sarkar et al., 2006; Shen et al., 2006; Sonoda et al., 2003). Additional enzymes, including pol η (Albertella et al., 2005; Chen et al., 2006; Misra and Vos, 1993), pol κ (Minko et al., 2008a; Williams et al., 2012) or pol υ (Moldovan et al., 2010; Zietlow et al., 2009) have also been implicated in ICL repair, suggesting that the choice of polymerase may depend on the structure of ICLs and the pathways used. *In vitro* studies have demonstrated that pol η, pol κ, pol ι and pol υ can bypass a variety of different ICLs (Ho et al., 2011; Klug et al., 2012; Minko et al., 2008a; Yamanaka et al., 2010; Zietlow et al., 2009). These studies have shown that the efficiency of bypass depends on the structure of the ICL itself, and in particular also on the length of the duplex surrounding the crosslink. While ICLs in long duplexes (~20 base pairs) were hardly bypassed by any of the polymerases, several enzymes were able to bypass ICLs in duplexes of 2-5 base pairs.

Here, we report a more detailed structure-function relationship of pol η on a panel of nitrogen mustard-like ICLs. The most important biological role of pol η is the error-free bypass of UV-induced CPD and mutations of the protein in humans are associated with the cancer-prone syndrome xeroderma pigmentosum variant (XP-V) (Johnson et al., 1999; Masutani et al., 1999). However, pol η deficiency also causes sensitivity to crosslinking agents

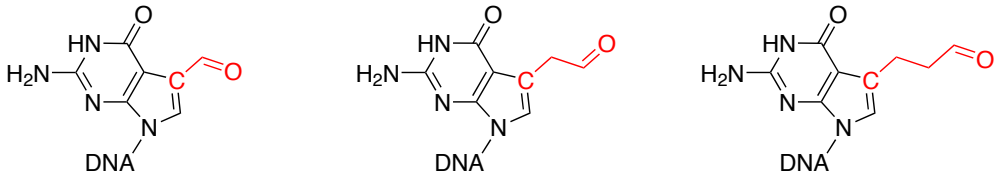
(Albertella et al., 2005; Chen et al., 2006; Misra and Vos, 1993) and pol η upregulation in tumors is associated with resistance to treatment with cisplatin and nitrogen mustards (Srivastava et al., 2015; Tomicic et al., 2014). Our studies dissect and quantify the effect of DNA distortion, flexibility of the ICL and position of incisions on translesion synthesis by pol η. We furthermore report a new strategy to synthesize the most extensively resected form of an ICL possible, a putative single nucleotide cross-linked intermediate and find that it only represents a minimal obstacle for DNA polymerases.

MATERIALS AND METHODS

Synthesis of 2-aminoethyl 7-deazaguanosine

Full experimental procedures and analytical data for the synthesis of 2-aminoethyl-7-deazaguanosine are available in the in the published supplementary information.

A.



C1 **C2** **C3**

B.

ICL	Modified base 1 (39mer)	Modified base 2 (20mer)	Amine
5 atom (5a)	C2	C1	Hydrazine (HY)
6 atom (6a)	C2	C2	Hydrazine (HY)
8 atom (8a)	C2	C2	Dimethylethylenediamine (DA)
6 atom (6a')	C3	C1	Ammonia

Figure 3.1. Scheme for generation of NM-like ICLs with various structures. **(A)** C1, C2 and C3 7-deaza-dG aldehyde ICL precursors used for the synthesis of ICLs. **(B)** Combination of oligonucleotide precursors and amines used for synthesis of 5a, 6a and 8a ICLs.

Oligonucleotides and Primers

The following oligonucleotides were synthesized for the generation of ICLs as described (Angelov et al., 2009; Guainazzi et al., 2010) containing 7-deaza-dG aldehyde ICL precursors denoted as 'X' (see **Fig. 3.1** for the three 7-deaza-dG crosslink precursors used):

T39 (39mer): 5'-GAAAGAAG**X**ACAGAAGAGGGTACCATCATAGAGTCAGTG-3'

C20 (20mer): 5'-CCCTCTUCT**X**TCCUTCTTTC-3'.

The following primers for the polymerase reactions containing a 5' FAM fluorescent label were purchased from IDT technologies:

P15 : 5'-(6-FAM)CACTGACTCTATGATG-3';

P0 : 5'-(6-FAM)GACTCTATGATGGTACCCTCTTCTGT-3'

Preparation of 20bp and 6bp ICLs

ICLs were generated as described (Ho et al., 2011; Mukherjee et al., 2014) (**Fig. 3.2**). The T39 and C20 oligonucleotides containing the ICL precursors were annealed, oxidized with 50 mM $NaIO_4$ in 100mM sodium phosphate buffer (pH 5.4), and excess oxidizing agent removed by washing with 10mM sodium phosphate buffer (pH 5.4) using Amicon Ultracel 3K columns (cat. no. UFC500396). ICLs were formed by treatment with either 5mM hydrazine (HY) or 5 mM N,N-dimethylethylenediamine (DA) in the presence of $NaBH_3CN$. The coupling reaction was incubated overnight in the dark at room temperature, and the ICLs were purified by 12%-15% denaturing PAGE. ICLs were extracted from the gel by electroelution using the Scheicher & Schuell BT1000 Biotrap system according to manufacturer's instructions in 5 mM $Na_2B_4O_7$ (pH 8.0). To generate

the resected 6bp ICLs, the purified 20 bp ICLs were digested with the USER enzyme mix (uracil DNA glycosylase and endonuclease VIII, NEB M5505), which cleaved the phosphodiester backbone at the position of the uracil residues. The 6 bp ICLs were purified by 12% -15% denaturing PAGE followed by electroelution as described above.

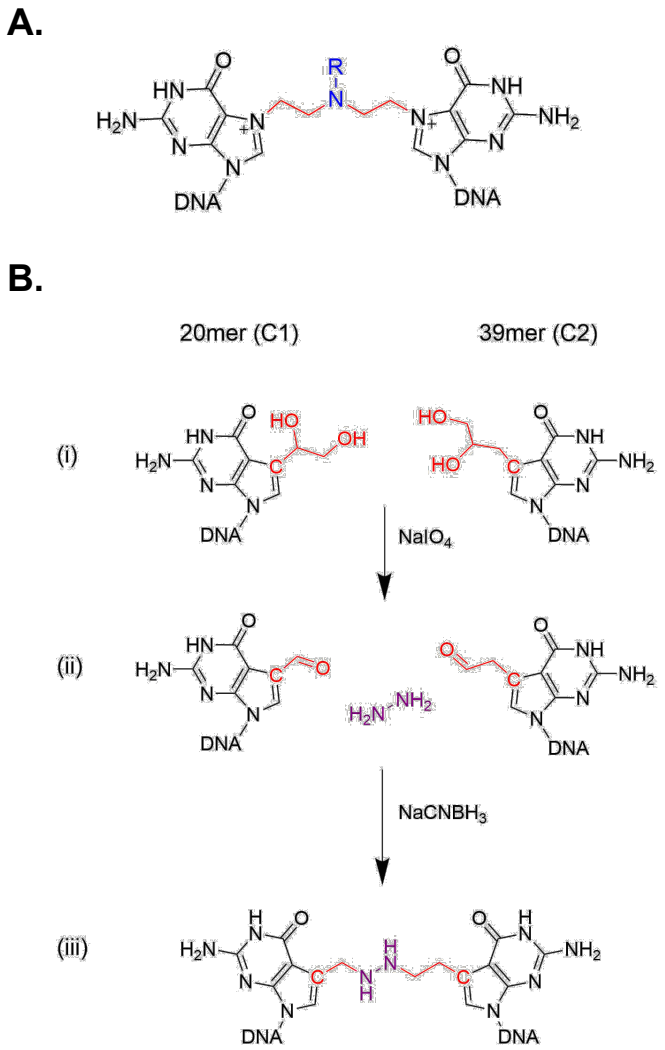


Figure 3.2. Strategy for synthesis of the 5a NM ICL mimic. (A) Structure of a native NM ICL. **(B)** Strategy for generation of a NM ICL mimic (5a ICL) using oligonucleotides containing C2 and C1 diol 7-deaza-dG precursors. The oligos with the diol precursors were (i) annealed and oxidized to the corresponding aldehydes and (ii) crosslinked using a double reductive amination reaction using hydrazine and $NaBH_3CN$ (iii) (Angelov et al., 2009; Guainazzi et al., 2010).

Preparation of Single Nucleotide (1 nt) ICLs

A solution of an 11mer oligonucleotide (20 nmols, 100 μL) (5'-GAAAGAAG**X**AC-3') containing the C2 ICL precursor (**Fig. 3.1**) was treated with 10 μL of 50 mM $NaIO_4$ and allowed to stand overnight in the dark at 4°C. Excess $NaIO_4$ was removed by washing with 10mM sodium phosphate buffer (pH 5.4) using Amicon Ultracel 3K columns (cat. no. UFC500396). The ICL was formed by adding 10 μL of 0.5 M solution of 7-(2-aminoethyl)-7-deazaguanosine (**3**) and 10 μL of 0.5 M $NaCNBH_3$ and incubation overnight in the dark at room temperature. The product and starting material were separated on a 20% denaturing PAGE gel and the ICL band was excised under UV-light. The band was extracted with 0.5 M NH_4OAc using the crush and soak method. The identity of the ICL band was confirmed using MALDI-TOF mass spectrometry (m/z calculated: 3683; found 3676) (**Fig. 3.3**).

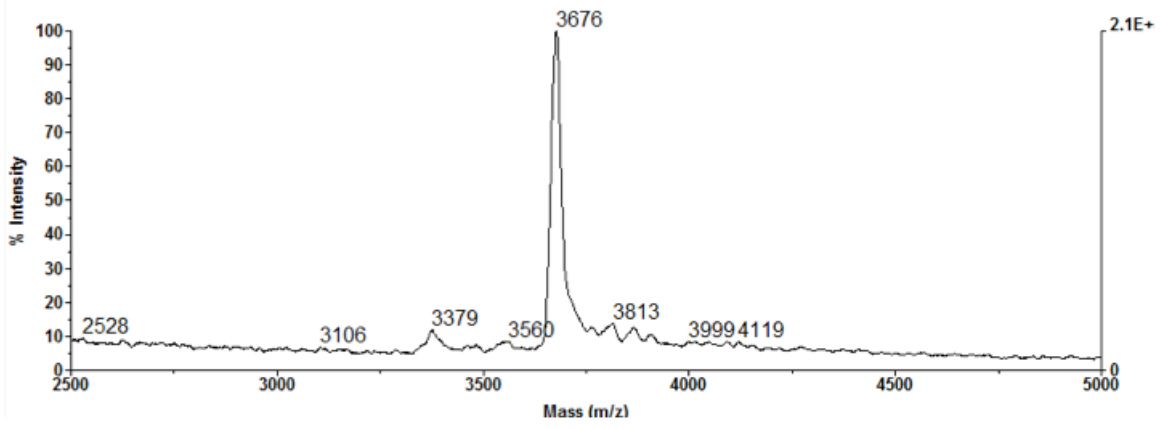


Figure 3.3: MALDI-TOF spectrum of the single nucleotide ICL formed by reductive amination between the 11-mer oligonucleotide and 7-(2-aminoethyl)-7-deazaguanine. The calculated m/z is 3683.

To generate T39 containing the single nucleotide ICL, the purified 11-mer with the single nucleotide ICL was ligated to a 5'-phosphorylated 28-mer (5'P-AGAAGAGGGTACCATCATAGAGTCAGTG-3'). The two oligos (500 nM each) were annealed to a complementary 51-mer splint (1 μM, 5'-TTGGAACACTGACTCTATGATGGTACCCTCTTCTGTCCTTCTTTCGTTAAC-3) in 10 mM Tris-HCl pH 8.0, 50 mM NaCl overnight at room temperature. The annealed oligonucleotides were incubated with T4 DNA ligase (NEB M0202S) for 30 minutes at 37°C. Products were resolved by 15% denaturing PAGE and the 39-mer containing single nucleotide ICL was extracted from the gel by electroelution as described above.

Enzymes

Klenow (exo-) enzyme (5 U/μL equivalent to 3.6 μM) was purchased from NEB (M0212). The protein was diluted in 25 mM Tris-Cl (pH 7.4), 1 mM DTT, 0.1 mM EDTA and 50% glycerol to the indicated concentrations for use in polymerase assays. Human pol η (with a C terminal His tag) was prepared as described previously yielding a preparation with a concentration of 0.3 mg/ml (~ 4 μM) (Frank and Woodgate, 2007). The protein was diluted in 40 mM Tris-Cl (pH 8.0), 10 mM DTT, 0.1 mg/mL BSA and 30% glycerol to the indicated concentrations for use in polymerase assays.

Polymerase assays

ICL substrates (150 nM) and 6-FAM labeled primer P15 (50 nM) were annealed in 10 mM Tris-HCl pH 8.0, 50 mM NaCl, overnight at room temperature to ensure the stability of the ICLs. The ICL substrates/primers (5nM, with respect to the

primer) and 100 μM dNTPs were incubated with DNA polymerase in a reaction volume of 10 μL. For assays with Klenow (exo-), 1 nM enzyme was used in reaction buffer NEB2 (50 mM NaCl, 10 mM Tris-HCl, 10 mM $MgCl_2$,1mM DTT). 40nM pol η was used in a reaction buffer containing 40 mM Tris-Cl pH 8.0, 50 mM NaCl, 5 mM $MgCl_2$, 10 mM DTT and 2.5% glycerol. Reactions were incubated for 10 minutes at 37°C and stopped by addition of 10μL of formamide buffer (80% formamide, 1mM EDTA, 1mg/mL Orange G), denatured at 95°C for 2 minutes and chilled on ice. The products of the reaction were resolved on a 10% 7 M Urea PAGE and FAM labeled DNA was visualized using a Typhoon 9400 scanner (GE Healthcare). Images were analyzed and quantified using ImageQuant software (Molecular Dynamics).

Single nucleotide insertion assays

ICL substrates were annealed to 6-FAM labeled P0 primer as described above. All reactions were incubated for 5 minutes at 37°C with 1 nM Klenow (exo-) or 20 nM pol η using the reaction buffers described above. Reactions were stopped by addition of 10 μL of formamide buffer (80% formamide, 1mM EDTA, 1mg/mL Orange G) and products analyzed by denaturing PAGE as described for the polymerase assays.

Analysis of single nucleotide insertion assays

Since the efficiency of primer extension is different for our various substrates (high for undamaged DNA and low for 6bp ICLs), to compare fidelity of insertion across these substrates, a 'normalized intensity' value was used instead of 'percent primer extension'. This value was calculated for each nucleotide taking

in to account the efficiency of nucleotide insertion for that substrate, such that 'normalized intensity' for a single nucleotide = (band intensity of lane A/T/C/G) ÷ (band intensity of lane 'N') for each substrate. For dCTP incorporation, the bands at position 0 and +1 was combined for the measurement of insertion, as the sequence contained two consecutive G residues.

RESULTS

Design of NM ICL substrates mimicking unhooked repair intermediates

Nitrogen mustards preferentially crosslink two guanines in a duplex within a -GNC- sequence via their N7 positions (**Fig. 3.4 A**) (Millard et al., 1990). This major groove ICL induces a bend of about 20° in the DNA duplex, as the length of the ICL is shorter than the distance between the two N7 of the guanine bases it connects (Dong et al., 1995; Fan and Gold, 1999; Guainazzi et al., 2010; Rink and Hopkins, 1995). Using a strategy previously developed in our lab (Angelov et al., 2009; Guainazzi et al., 2010; Mukherjee et al., 2014) (**Fig. 3.2**), we synthesized a substrate containing a stable site-specific NM ICL mimic (which we denote 5a for a crosslink with 5 atoms) along with variants containing 6 atom (6a) and 8 atom (8a) ICLs (**Fig. 3.4 B**). Our molecular modeling studies have shown that the 5a ICL, like its native NM equivalent, induce a bend of about 20° in the DNA duplex (36). Based on the length of the ICLs, we expect the 6a ICL to have less distortion than the 5a ICL and our preliminary NMR studies have shown that the 8a ICL is intact B-form DNA and free of distortion (Guanazzi, A, de los Santos C, ODS, unpublished observations).

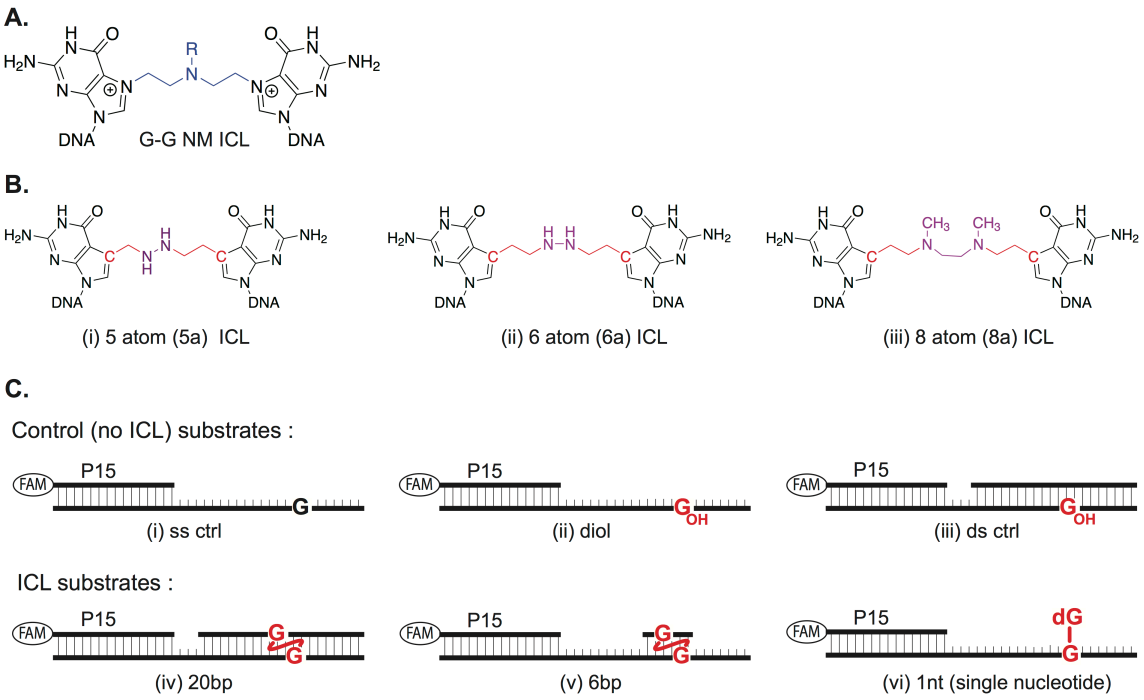


Figure 3.4: ICL substrates used in this study. (**A**) Structure of a nitrogen mustard (NM) ICL linking two guanine bases. (**B**) Structure of the 5 atom (i, 5a), 6 atom (ii, 6a) and 8 atom (iii, 8a) NM ICL. (**C**) Substrates used for polymerase assay substrates reactions. 5'-FAM labeled primer P15 was annealed to various templates. (i) single stranded DNA undamaged control (ii) single stranded substrate with ICL precursor 'diol' (Fig. 3.2 B, i, C2) (iii) double stranded substrate with ICL precursor 'diol', (iv) ICL substrate within a 20bp duplex, (v) ICL substrate within a 6bp duplex (vi) single nucleotide ICL substrate. Crosslinked or adducted bases are highlighted in red.

We and others have previously shown that the length of the dsDNA around an ICL dramatically influences the efficiency of bypass and that a long duplex (~10 bps) on either side of the ICL prevents bypass by TLS polymerases (Ho et al., 2011; Klug et al., 2012; Minko et al., 2008a; Yamanaka et al., 2010; Zietlow et al., 2009). We generated substrates with the 5a, 6a and 8a ICLs embedded in a 20mer duplex (5a/20 bp, 6a/20 bp and 8a/20 bp respectively, **Fig. 3.4 C**, **iv**). As in our previous study, we also generated a resected ICL, by partially degrading the duplex around the ICL at uracil residues incorporated into

the duplex to yield an ICL with a 6 mer duplex around it (6 bp ICLs) (**Fig. 3.4 C, v**). In our reactions, unmodified ssDNA (**Fig. 3.4 C**, **i**), ssDNA containing our ICL precursor (**Fig. 3.4 C**, **ii,** see **Fig. 3.2 B**, **i** for structure) and dsDNA with our ICL precursor on the template strand (**Fig. 3.4 C**, **iii**). The most completely processed form of an unhooked ICL is a "single nucleotide ICL" (1 nt ICL), in which exonucleases resect an unhooked ICL down to a single nucleotide (**Fig. 3.4 C, vi**) (Smeaton et al., 2009). The strategy to generate such an ICL by incorporation and cleavage of uracil residues was unsuccessful due to the failure of UNG to cut immediately adjacent to the ICL. Therefore, we developed a synthetic route based on our double reductive amination approach to generate the single nucleotide ICLs.

Synthesis of the Single Nucleotide NM ICL

The single nucleotide NM ICL (1nt ICL) was synthesized by coupling 7-(2-aminoethyl)-deazaguanosine **3** to our aldehyde-containing ICL precursor T39. The synthesis of **3** began with the allyl **1** (Angelov et al., 2009) (**Fig. 3.5**), in which the double bond was oxidized to the diol with osmium tetroxide, oxidized to the aldehyde with sodium periodate and trapped with O-methyl-hydroxylamide **2** to form the oxime. Zinc reduction and removal of the protecting groups yielded amine **3**, which was reacted with a 11mer single-stranded oligonucleotide containing a C2 aldehyde ICL precursor **4** under reductive amination conditions (**Fig. 3.5 B**). Analysis of the reaction products by denaturing PAGE revealed the formation of a slower moving band, indicating the formation of the desired product (**Fig. 3.5 C**). Isolation and analysis of the product by mass spectrometry

revealed it to be the target single nucleotide ICL (**Fig. 3.3**). In our polymerase assays we compared the single nucleotide ICL to that of our stable C2 ICL precursor – a deazaguanine residue substituted with a dihydroxypropyl group (diol) (**Fig. 3.2 B, i**) - in the template strand to asses the effect of a single nucleotide ICL (**Fig. 3.4C, vi**) versus a smaller lesion (**Fig. 3.4 C, ii**).

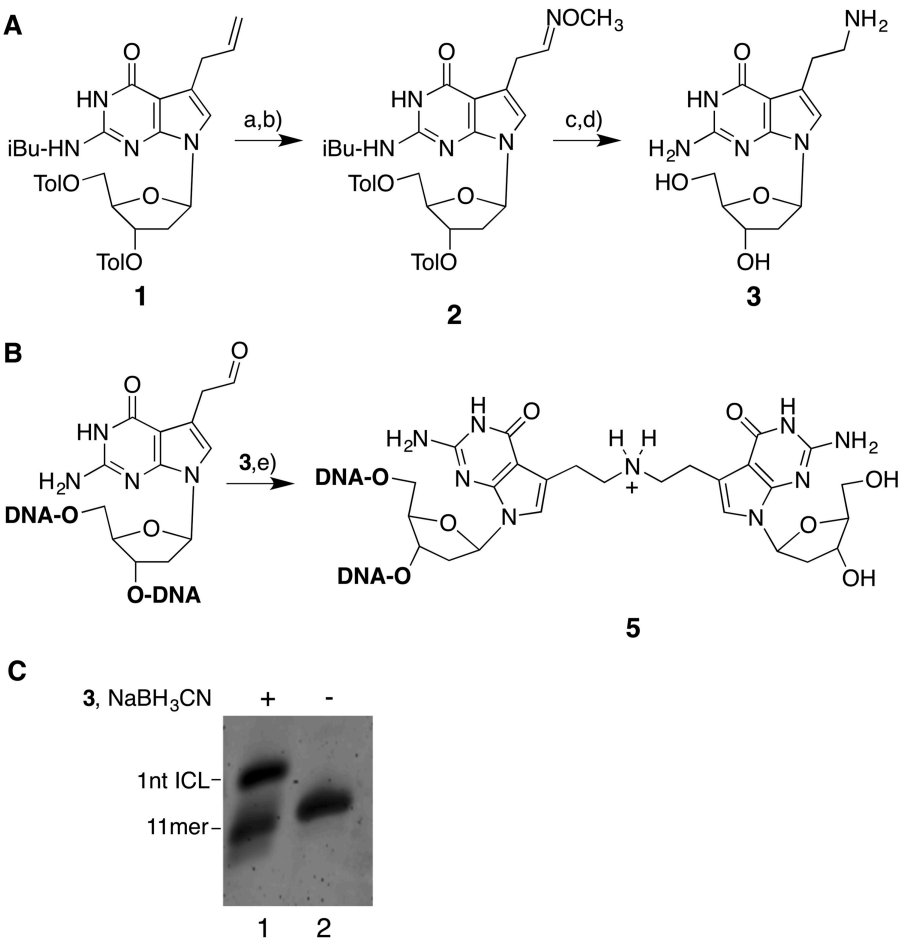


Figure 3.5: Synthesis of the single nucleotide ICL. (**A**) Reaction conditions: a) OsO_4, NMM, THF, 0°C, 72%; b) i. $NaIO_4$, MeOH, THF; ii. NH_2OMe, 86%; c) Zn/HCl, MeOH, HOAc, 85%; d) NH3, MeOH. (**B**) e) i) $NaIO_4$, H_2O ii) $NaBH_3CN$. (**C**) 20% denaturing PAGE analysis of reaction of the 11 mer (GAAAGAAG**4**AC) with amine **3**. DNA was visualized with SYBR gold.

Reducing duplex around an ICL facilitates bypass

TLS is believed to occur in three stages: approach of the replicative or TLS polymerase up to the lesion, insertion of a dNTP across the lesion and extension past the lesion, often with different polymerases carrying out the insertion and extension steps (Lehmann et al., 2007; Prakash et al., 2005; Shachar et al., 2009). Keeping this in mind, we evaluated our primer extension assays in three parts (**Fig. 3.6 A**): 'Approach' (extension of primers up to 1nt before the crosslinked base (-1)), 'Insertion' (insertion opposite the ICL up to 3nt past it (0 to +3)) and 'Extension' (all products from +4 to the full length product). We chose to include 3 nucleotides in the "insertion" category as we previously observed that some TLS polymerases have prominent stalling points at and within a few nucleotides of the insertion site (Ho et al., 2011).

In a first set of experiments, we aimed to understand how the length of duplex around an ICL - reflecting the position of incisions during unhooking - would affect translesion DNA synthesis. Using the 5a NM ICL mimic (**Fig. 3.4 B, i**) in the 20 bp, 6 bp and 1 nt substrates (**Fig. 3.4 C, iv-vi**), we first used the bacterial replicative polymerase, exonuclease deficient Klenow fragment as a benchmark. As we have found previously (Ho et al., 2011), Klenow stalled predominantly at -1 in the 20 bp ICL (**Fig. 3.6 B**, lane 5). In the 6 bp ICL, resection of the duplex allowed ~30% insertion opposite the ICL, introducing a stalling at position 0 in addition to the main stalling point at -1, without any further extension to full product (**Fig. 3.6 B**, compare lanes 4 and 5).

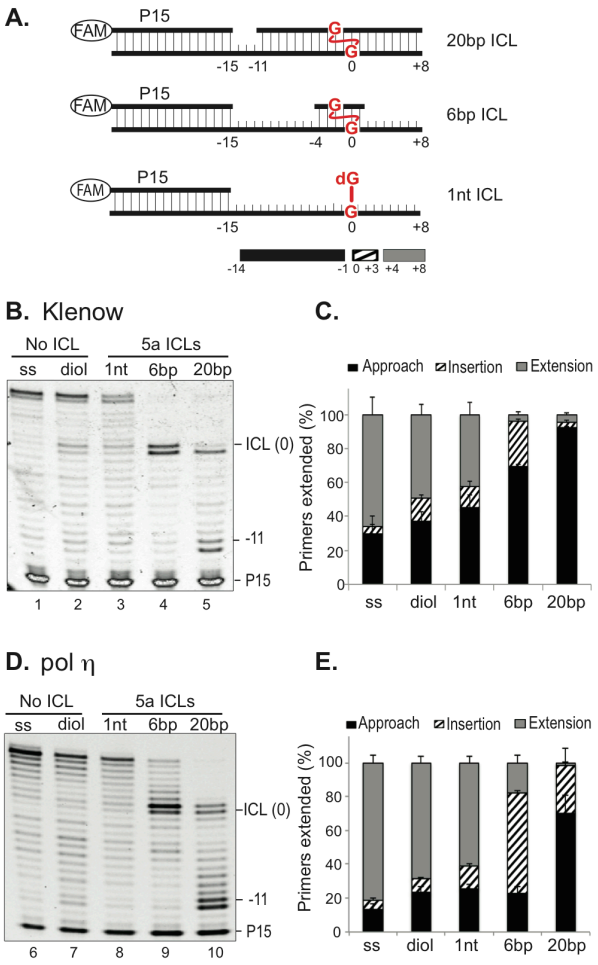


Figure 3.6: Shortening of duplex around the ICL facilitates bypass. (**A**) 20bp, 6bp and 1 nt ICL substrates used. The crosslinked base in the template strand was designated '0', and all primer extension products up to -1 were evaluated as 'approach', from 0 to +3 as 'insertion' and beyond +4 as 'extension'. (**B-E**) Translesion synthesis assay of 5a ICL templates with Klenow and pol η. Unmodified (lanes 1 & 6), diol (monoadduct, lanes 2 & 7) and 5a ICL-containing templates (lanes 3-5, 8-10) were annealed to the FAM labeled primer P15 and incubated with (**B**) 1 nM Klenow, (**D**) 40 nM pol η for 10 mins at 37°C. Products were resolved by 10% denaturing PAGE. Quantification and analysis of primer extension products with Klenow (**C**) and pol η (**E**). Each lane was divided into approach, insertion and extension segments, and corresponding band intensities expressed as a percentage of the total products combined. Data represent the mean of three experiments and error bars indicate S.D.

As expected, there was no stalling at the G residue on the ssDNA template (**Fig. 3.6 B**, lane 1). With the single stranded C2 ICL precursor (diol) (**Fig. 3.4 C**, ii **and Fig. 3.2 B**, i), there was a pausing at the 0 and -1 positions, while most of the primer was extended to the full length product (**Fig. 3.6 B**, lane

2), suggesting that a substitution at the N7 position is only a minor impediment for Klenow, consistent with the frequent modification at that position for DNA sequencing and other applications (Prober et al., 1987). Interestingly, a similar pattern was observed for the single nucleotide (1 nt) ICL (**Fig. 3.6 B**, lane 3), with only a pausing site at the crosslinked base, accompanied by efficient bypass of the ICL (**Fig. 3.6 C**). This result raises the possibility that some forms of unhooked ICLs may be bypassed by a replicative polymerase and that a TLS polymerase may not always be absolutely required for ICL repair.

We then assayed the activity of TLS polymerase pol η with the various ICLs. Pol η also stalled during approach to the ICL in the 5a/20 bp substrate, but was able to insert a nucleotide opposite the ICL and at the +1 position (**Fig. 3.6 D** and **E**). The approach and insertion by pol η was significantly facilitated in the 5a/6bp ICL, with the +1 and +2 insertion products making up close to 70% of the products. The amount of fully extended product was however still limited. Interestingly, the 5a/1 nt ICL, similar to the single stranded undamaged DNA (ss) and C2 ICL precursor (diol) was bypassed by pol η with high efficiency, resulting primarily in extension of the primers to full length products. These findings show that the amount of duplex around an ICL greatly affected the efficiency of pol η to bypass ICLs, and that at least in the case of a fully processed 1 nt ICL, pol η could carry out both insertion and extension steps alone.

NM ICL-induced distortion facilitates approach and insertion, but inhibits extension by pol η

NM ICLs cause a slight local distortion in the surrounding duplex by introducing a bend of about 20° in the DNA helix (Guainazzi et al., 2010), and we were interested to understand how this influenced the ability of pol η to bypass the ICL. We addressed this by using NM ICL variants with longer linkers (6a and 8a vs. 5a of the NM ICL mimic) that are expected to have less or no distortion, respectively (**Fig. 3.4 B**). We generated the 5a, 6a and 8a ICLs embedded in 6 or 20 bp duplexes and annealed them to fluorescently labeled primers for the analysis of bypass by pol η. Pol η was able to insert a nucleotide at the +1 position in the 5a/20bp ICL (**Fig. 3.7 A**, lane 3). By contrast, the enzyme stalled at the -1 position with the 6a and 8a/20 bp ICLs with no detectable extension to full products (**Fig. 3.7 A**, lanes 4 and 5). Our data suggests that the distortion caused by the 5a ICL facilitates insertion by pol η. We speculate that the greater stability of the crosslinked duplexes without distortion (6a and 8a ICLs) resists strand displacement and therefore approach and insertion. Interestingly, the increased flexibility of the 8a linker compared to the 6a linker had no additional effect on the efficiency of insertion, suggesting that the relief of the distortion and not the additional flexibility is key to the outcome of the reaction (**Fig. 3.7 B**, compare 6a and 8a).

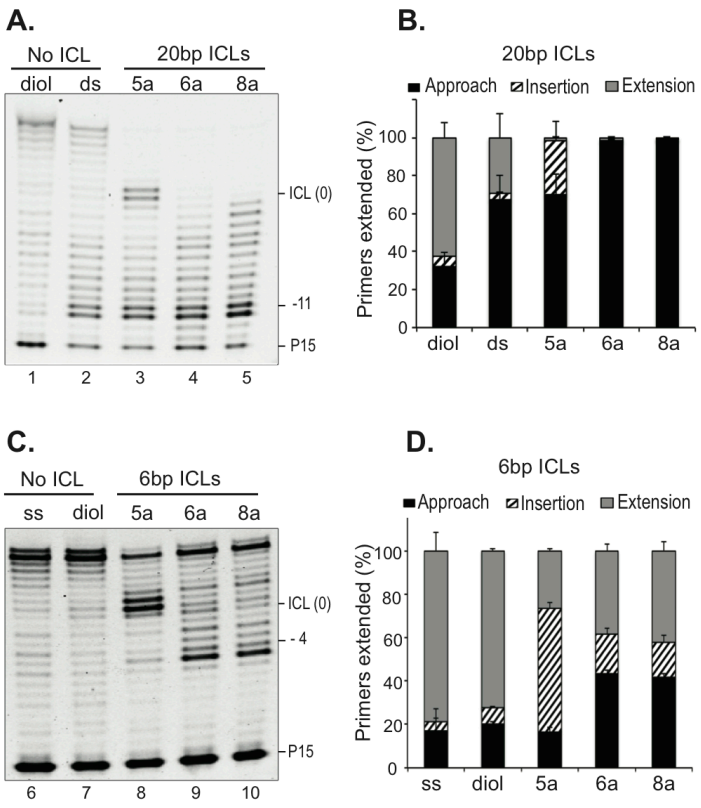


Figure 3.7: Duplex distortion facilitates approach and insertion by pol η at the ICL. Control and monoadduct (lanes 1-2, 6-7) and 5a, 6a and 8a ICL templates in a (**A**) 20 bp duplex (lanes 3-5) or (**C**) 6bp duplex (lanes 8-10) were annealed to primer P15 and incubated with 40 nM pol η for 10 mins at 37°C. The products were resolved by 10% denaturing PAGE. Quantification and analysis of primer extension products with 20 bp ICLs (**B**) and 6 bp ICLs (**D**). Each lane was divided into approach, insertion and extension segments, and corresponding band intensities expressed as a percentage of all the products combined. Data represent the mean of three experiments and error bars indicate S.D.

We then asked whether duplex destabilization would play an equally important role in promoting bypass when the amount of duplex surrounding the ICL is reduced to 6 base pairs. We found that also with the ICLs within a shorter duplex, distortion facilitated the approach to the ICL. In the reaction with the 5a/6 bp ICL, the main stalling points were at the +1 and +2 positions (**Fig. 3.7 C**, lane 8), while for both the non-distorting 6a/6bp and 8a/6bp ICLs, primer extension stalled at the beginning of the duplex (**Fig. 3.7 C**, lanes 9 and 10). Interestingly,

there was no stalling point for the 6a and 8a ICLs at or around the crosslinked base, and once pol η was able to initiate the strand displacement reaction, most of the primer was extended to the full length product (**Fig. 3.7 C**, compare lane 8 to lane 9 & 10). Therefore, pol η is able to efficiently insert dNTPs opposite the non-distorting ICL and extend the primer to the full length product.

One reason for the lower insertion activity of pol η on non-distorting ICLs could be a relatively weak strand displacement ability of the enzyme. Duplex destabilization could therefore facilitate approach and insertion across the ICL. This led us to ask whether the duplex destabilization would be equally important for a polymerase with a stronger strand displacement activity, such as Klenow. We found that there is also a significant difference in approach between distorting and non-distorting ICLs with Klenow (**Fig. 3.8**). While the initial strand displacement was similar for the 5a, 6a and 8a/20 bp ICLs, Klenow was able to extend the primer to the 0 and -1 position for the 5a ICL, while it stalled at the -3, -2 and -1 positions for the 6a and 8a ICLs (**Fig. 3.8A**, lanes 3-5). Very similar observations were made for the equivalent 6bp ICLs, where Klenow stalled primarily at -1 and 0 for the 5a/6bp ICL and at -2 and -1 for the 6a and 8a/6 bp ICLs (**Fig. 3.8A**, lanes 8-10). This suggests that duplex destabilization by a distorting crosslink is important in determining the efficiency of the approach to, and bypass of ICLs by polymerases with widely different strand displacement abilities.

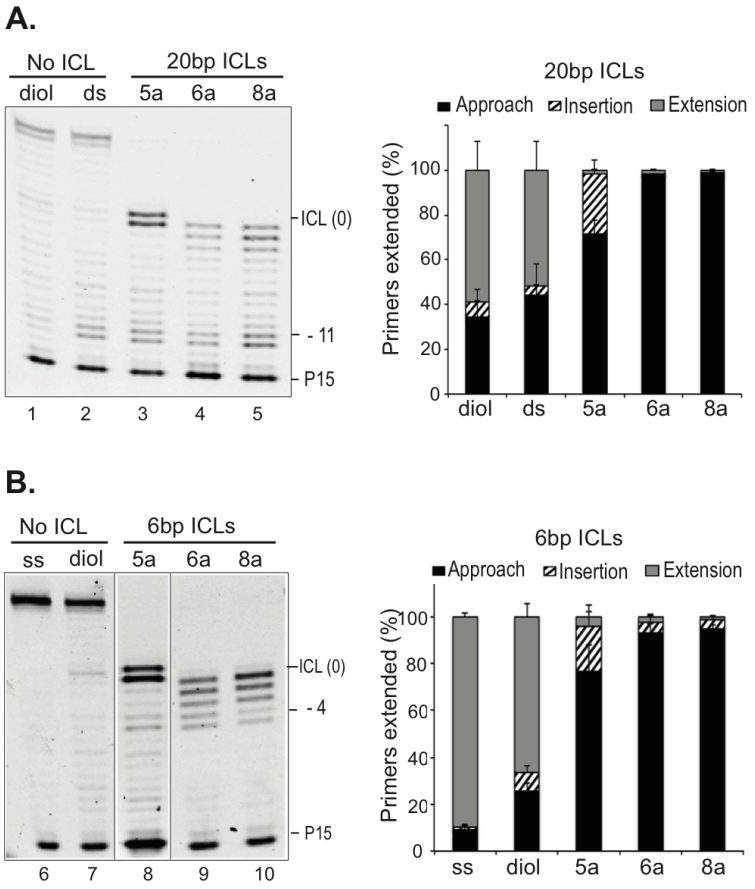


Figure. 3.8: Duplex distortion by NM ICLs facilitates approach by Klenow
Uncrosslinked templates (lanes 1-2, 6-7) and ICL templates with varying linkers within a **(A)** 20bp duplex (lanes 3-5) or **(B)** 6bp duplex (lanes 8-10) were annealed to FAM labeled primer P15 and incubated with 1 nM Klenow for 10 mins at 37°C. The products were resolved by denaturing PAGE and quantified by Image Quant. The crosslinked base in the template strand was designated '0', and all primer extension products up to -1 were evaluated as 'approach', from 0 to +3 as 'insertion' and beyond +4 as 'extension'. Each lane was divided into approach, insertion and extension segments, and corresponding band intensities expressed as a percentage of the total band intensity for all primer extension products combined. Data are represented as the mean of three experiments and error bars indicate S.D.

Finally, we tested whether our observed effects on bypass depend only on the length of a crosslink or whether they are also influenced by its chemical composition. For this purpose, we used two ICLs with 6 atom linkers: 6a and 6a', which differ in that 6a has a hydrazine and 6a' has an amine linkage (**Fig. 3.9 A**). The reactions with the two 6 atom ICLs were very similar for the 6 bp and 20 bp

ICLs with Klenow (**Fig. 3.9 B**), and the 20 bp ICL with pol η (**Fig. 3.9 C**). The only minor difference observed was with the 6 bp ICL and pol η, where the ICL with the amine linkage seemed to be bypassed more efficiently, primarily as the strand displacement reaction appeared to be more efficient (**Fig. 3.9 C**). However quantification of the bypass reactions indicated that this may not be a significant difference.

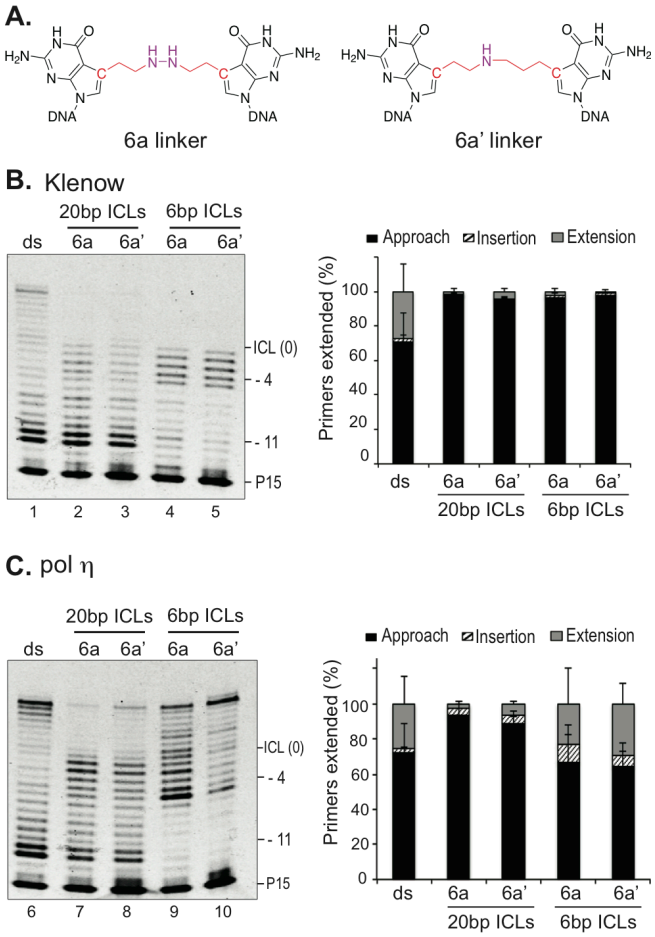


Figure. 3.9: Reaction of Klenow and pol η with two distinct 6a ICLs . (A) Structure of two 6 atom ICLs linkers with slightly different chemical compositions used as substrates for TLS assays. **(B)** Substrates with 6a or 6a' ICL were annealed to FAM labeled P15 primer and incubated with 1nM Klenow or 40nM pol η at 37°C for 10 minutes. The products were resolved by denaturing PAGE and bands visualized using a fluorescence gel scanner.

Pol η is more accurate on NM ICLs than on undamaged DNA

Like all TLS polymerases, pol η is a low fidelity enzyme, yet it has the ability to efficiently and accurately replicate past UV-induced CPD lesions thereby preventing UV-induced mutations (Biertumpfel et al., 2010; McCulloch et al., 2004). We therefore investigated how NM ICL structure affected the fidelity of dNTP insertion by pol η at the lesion site. We carried out single nucleotide incorporation assays with pol η by annealing primer P0 to the undamaged control, single nucleotide (5a/1 nt), and 5a, 6a and 8a/6 bp substrates using different concentrations of the individual dNTPs (**Fig. 3.10 A**).

As already observed in our bypass assays, the efficiency of primer extension was lower for the ICLs than for undamaged DNA (**Fig. 3.10 B**, lane 'N'). To control for different efficiencies of dNTP incorporation for the various substrates, we quantified and used the ratio of incorporation of each individual nucleotide (A/T/C/G) to that of the four dNTPs (N) as our 'normalized relative intensity'. At 1 μM dNTP concentration, pol η incorporated primarily dCTP – at least 2-fold more than the incorrect dNTPs, opposite the ICLs or the undamaged control G residue. (**Fig. 3.10 B,C**). At the next higher dNTP concentration, 10 μM, incorporation opposite the control G became more promiscuous, with only about 1.2 -1.4 fold higher incorporation of dCTP (**Fig. 3.10 B,D**). Interestingly, incorporation opposite the ICLs was more accurate than opposite an undamaged dG residue, with at least 4-fold lower incorporation of dATP and dGTP than dCTP. The misincorporation of purines opposite the ICLs was significantly lower even at 100 μM dNTP concentrations, (**Fig. 3.10 B,E**), where all four dNTPs

were incorporated with similar efficiency opposite the control G and 5a / 1 nt ICL. Although differences were minor, longer ICL linkages (8a and 6a versus 5a) allowed for higher fidelity of incorporation (compare **Fig. 3.10 D,E**). It is interesting to note that the fidelity of incorporation was significantly greater for the ICL surrounded by a 6 bp duplex compared to the 1 nt ICL. Collectively, our data suggest that pol η has a higher fidelity of dNTP incorporation opposite an NM ICL than unmodified DNA.

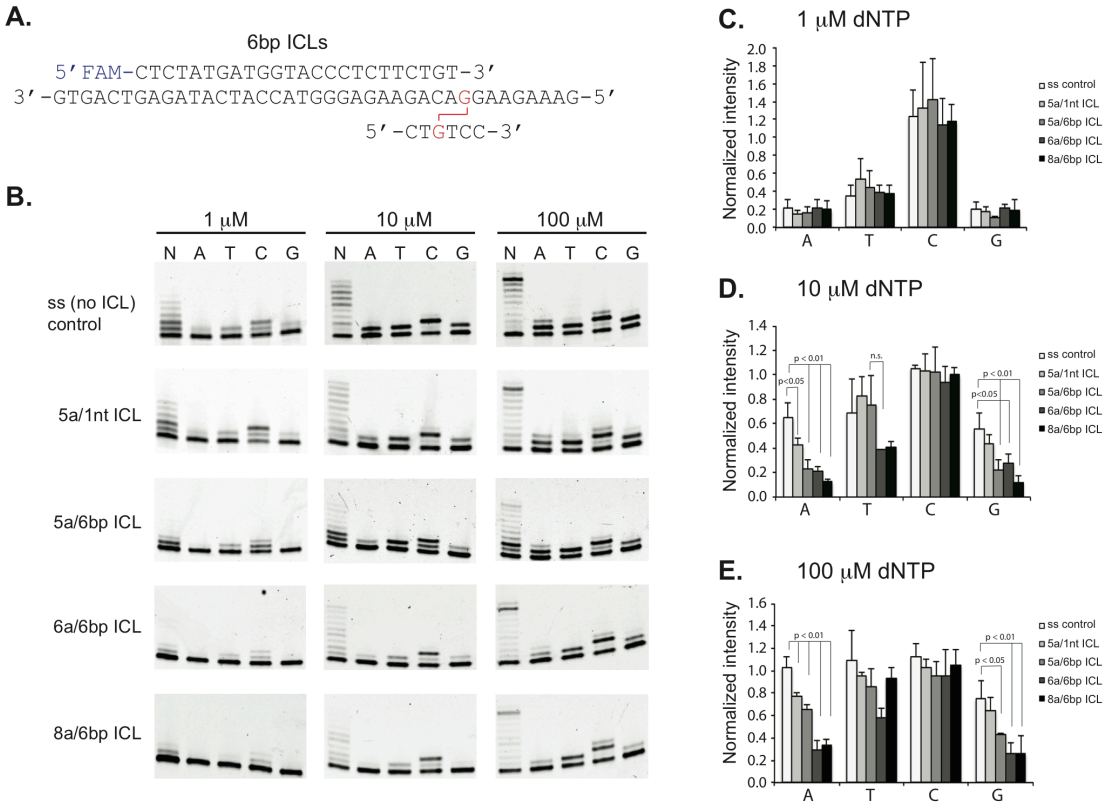


Figure 3.10: pol η is more accurate across NM ICLs than undamaged DNA. (**A**) Sequence of ICL substrate for single base insertion assays. (**B**) Templates were annealed to FAM labeled P0 primer and incubated with 20nM pol η at 37°C for 5 minutes with 1, 10 or 100 μM of individual dNTPs (A/T/C/G) or all four dNTPs combined (N). The products were resolved by 10% denaturing PAGE and band intensities quantified using ImageQuant. Quantification and analysis of single nucleotide incorporation with (**C**) 1 μM dNTPs (**D**) 10 μM dNTPs and (**E**) 100 μM dNTPs. The ratio of band intensity for each individual nucleotide (lane A/T/C/G) to that of the four dNTPs combined (lane N) was calculated and expressed as the 'normalized intensity'. Data are represented as the mean of three experiments and error bars indicate S.D. p-values were calculated by a one-way ANOVA with the Bonferroni correction for multiple comparisons.

DISCUSSION

Multiple pathways for ICL repair have been described which share as one common feature the bypass of an unhooked ICL by DNA polymerases. This step restores one of the two strands modified by the ICL as a template for repair synthesis. To date it has not been possible to determine what the structures of unhooked ICLs look like. These structures are determined by several factors including the ICL repair pathway used and the positions of the incisions at the ICL (Zhang and Walter, 2014). As a result, replicative and translesion synthesis polymerases are likely to encounter a variety of unhooked ICL structures. In this study we investigated the reaction of the Y-family TLS polymerase pol η with a set of diverse model unhooked NM ICL structures that reflect different points of incision during unhooking, and ICL structures with different degrees of helix distortions and flexibilities.

The influence of helix distortion and ICL resection on polymerase activity

We and others have shown that resection of the duplex around an ICL greatly facilitates bypass across various ICLs (Ho et al., 2011; Klug et al., 2012; Minko et al., 2008a; Yamanaka et al., 2010; Zietlow et al., 2009). This is likely due to a reduced need for strand displacement synthesis during approach to the ICL and due to the increased flexibility of a shorter duplex during insertion and extension of the ICL. Our data (**Figs. 3.7** and **3.8**) show that a duplex-distorting crosslink facilitates strand displacement, while non-distorting ICLs inhibit it. Similarly, in studies with *Xenopus* egg extracts, the approach to the ICL was more efficient for a highly distorting cisplatin ICL compared to a non-distorting

nitrogen mustard-like ICL (Räschle et al., 2008). These observations suggest that the need for resection of the duplex around an ICL may be especially important for the repair of non-distorting crosslinks. Mechanistically, this resection could be performed by hSNM1A- an exonuclease implicated in ICL repair with demonstrated ability to digest duplex DNA around an ICL (Allerston et al., 2015; Wang et al., 2011). Cells lacking hSNM1A were more sensitive to exposure to the crosslinking agent MMC and SJG-136, which form non-distorting ICLs, than to nitrogen mustards, which form more distorting ICLs (Wang et al., 2011). This suggests that the exonuclease activity of hSNM1A is more important for repair of non-distorting crosslinks. Our data provide a possible explanation for why exonucleolytic processing of the duplex around an ICL would be more important for such non-distorting ICLs. Taken together this suggests duplex distortion and/or destabilization by an ICL is an important determinant of how lesions are processed, approached and bypassed in ICL repair.

Bypass of a single nucleotide NM ICL

One of the open questions in ICL repair is where incisions occur in the unhooking step and how much duplex surrounds the ICL when it is encountered by DNA polymerases. Two observations suggest that resection to a single nucleotide ICL might occur: 1) In experiments with *Xenopus* egg extracts an ICL remnant has been observed upon completion of replication and repair of an ICL-containing plasmid with a single nucleotide cross-linked to the template strand (Räschle et al., 2008). 2) The hSNM1A exonuclease, implicated in ICL repair, has been shown to be capable of digesting a duplex across an ICL down to a

singe nucleotide, providing a mechanism for how such intermediates may be generated (Allerston et al., 2015; Wang et al., 2011). Similarly, FAN1 has been shown to be able to digest a DNA duplex across an ICL (Pizzolato et al., 2015; Wang et al., 2014).

We devised a strategy to synthesize single nucleotide NM ICLs to study how such an intermediate interacts with polymerases. Intriguingly, this ICL did not provide an obstacle for pol η, alleviating the characteristic stalling points of pol η especially after dNTP insertion opposite the ICL (**Fig. 3.6 D**). This observation suggests that pol η can completely bypass such structures on its own, and that ICL repair may not always require other TLS polymerases such as REV1/pol ζ to carry out the extension step. It will be interesting to see if this would also apply to ICLs such as those formed by cisplatin or psoralen that would more severely constrain the structures of single nucleotide ICL intermediates.

Interestingly, we found that the bacterial replicative polymerase Klenow was also able to bypass the single nucleotide NM ICL (5a / 1 nt ICL) (**Fig. 3.6 B**). An earlier study using artificial model ICLs which link two strands through the exocyclic amine groups of dC or through the N3 positions of dT residues found that single nucleotide ICLs capable of forming Watson-Crick base pairs could be bypassed by Klenow (Smeaton et al., 2009). Together these observations raise the intriguing possibility that some ICLs, in the most processed form, may not require the activity of a TLS polymerase to be repaired. Given that the approach to the ICL is most likely carried out by replicative polymerases (Budzowska et al.,

2015; Long et al., 2011), it will be interesting to see whether the mammalian replicative polymerases δ and ε similarly have the ability to bypass our single nucleotide NM ICL and whether it will be possible to discover a pathway of ICL repair that does not require TLS polymerases.

What could the role of pol η in ICL repair be?

Given the variety of ICL structures formed, the possible redundancy among various TLS polymerases, and the limited mechanistic resolution of current assays available for the study of ICL repair, it has been challenging to identify the polymerase(s) carrying out the insertion and extension steps on the unhooked ICLs template. Based on genetic and biochemical considerations, a prime candidate for insertion across crosslinked guanines, the base most frequently modified by crosslinking agents, is the dCMP transferase REV1 (Haracska et al., 2002; Nelson et al., 1996). However, experiments in *Xenopus* demonstrated that REV1 and pol ζ were only required for the extension step and not the insertion across the cisplatin ICL (Budzowska et al., 2015) and that they were dispensable altogether for the repair of non-distorting 8 atom nitrogen mustard-like ICL (Räschle et al., 2008). Of the alternative candidate polymerases that may instead act on these ICLs, we focused on pol η in this study. Genetic and biochemical experiments have implicated pol η in the bypass of a variety of ICLs (Biertumpfel et al., 2010; Chen et al., 2006; Ho et al., 2011; Shachar et al., 2009). Structural features of pol η show this enzyme may be particularly well suited to accommodate ICLs (Biertumpfel et al., 2010; Reissner et al., 2010; Silverstein et al., 2010; Zhao et al., 2012). Pol η is the Y-family polymerase with

the largest active site, and detailed structural studies of the enzyme bypassing CPDs have shown that it operates as a molecular splint holding on to the damaged DNA until the primer is extended 3 nucleotides past the lesion (Biertumpfel et al., 2010). Remarkably, we found that pol η extends a primer efficiently up to 2 nucleotides past an ICL in the 5a ICL, consistent with the idea that rigid binding to the primer template would allow extension by a few nucleotides past the site of insertion opposite the lesion. We think that this feature also allows for insertion and complete bypass of the more flexible 6 and 8 atom ICLs. Our studies furthermore are similar to what is observed with CPDs (McCulloch et al., 2004), in that the bypass of NM ICLs by pol η is more accurate than that of non-damaged DNA (**Fig. 3.10**). These observations warrant more detailed studies of the role of pol η in the repair of different ICLs.

Conclusion

Using and expanding a synthetic approach in our laboratory, we generated a number of structurally diverse NM-like ICLs that differ in their degree of distortion induced in the duplex (by varying the length of the crosslink) and the amount of duplex surrounding the ICLs. Our studies indicate that more distorting (shorter) ICLs facilitate strand displacement and approach to the ICL, while less distorting (longer) ICLs are extended more efficiently after insertion. Importantly, we showed that unhooked ICL intermediates that have been processed down to a single nucleotide only pose a minimal obstacle for DNA polymerases, suggesting that the polymerase reaction past ICLs may be more facile than

commonly assumed and may not always require the activity of a TLS polymerase.

ACKNOWLEDGEMENTS: We are grateful to Robert Rieger for help with the HR-MS and MALDI-TOF spectral analyses supported by grant NIH/NCRR 1 S10 RR023680-1, to Todor Angelov for early studies on the synthesis of 7-(2-aminoethyl)-deazaguanosine **3,** and to Roger Woodgate (NIH) for helpful discussions and comments on the manuscript.

REFERENCES

Albertella, M.R., Green, C.M., Lehmann, A.R., and O'Connor, M.J. (2005). A role for polymerase η in the cellular tolerance to cisplatin-induced damage. *Cancer Res 65*, 9799-9806.

Allerston, C.K., Lee, S.Y., Newman, J.A., Schofield, C.J., McHugh, P.J., and Gileadi, O. (2015). The structures of the SNM1A and SNM1B/Apollo nuclease domains reveal a potential basis for their distinct DNA processing activities. *Nucleic Acids Res 43*, 11047-11060.

Angelov, T., Guainazzi, A., and Scharer, O.D. (2009). Generation of DNA interstrand cross-links by post-synthetic reductive amination. *Org Lett 11*, 661-664.

Biertumpfel, C., Zhao, Y., Kondo, Y., Ramon-Maiques, S., Gregory, M., Lee, J.Y., Masutani, C., Lehmann, A.R., Hanaoka, F., and Yang, W. (2010). Structure and mechanism of human DNA polymerase eta. *Nature 465*, 1044-1048.

Budzowska, M., Graham, T.G., Sobeck, A., Waga, S., and Walter, J.C. (2015). Regulation of the Rev1-pol zeta complex during bypass of a DNA interstrand cross-link. *EMBO J 34*, 1971-1985.

Castor, D., Nair, N., Declais, A.C., Lachaud, C., Toth, R., Macartney, T.J., Lilley, D.M., Arthur, J.S., and Rouse, J. (2013). Cooperative control of holliday junction resolution and DNA repair by the SLX1 and MUS81-EME1 nucleases. *Mol Cell 52*, 221-233.

Chen, Y.-w., Cleaver, J.E., Hanaoka, F., Chang, C.-f., and Chou, K.-m. (2006). A novel role of DNA polymerase η in modulating cellular sensitivity to chemotherapeutic agents. *Mol Cancer Res 4*, 257-265.

Clauson, C., Scharer, O.D., and Niedernhofer, L. (2013). Advances in understanding the complex mechanisms of DNA interstrand cross-link repair. *Cold Spring Harbor Perspect Biol 5*, a012732.

Deans, A.J., and West, S.C. (2011). DNA interstrand crosslink repair and cancer. *Nat Rev Cancer 11*, 467-480.

Doles, J., Oliver, T.G., Cameron, E.R., Hsu, G., Jacks, T., Walker, G.C., and Hemann, M.T. (2010). Suppression of Rev3, the catalytic subunit of Pol{zeta}, sensitizes drug-resistant lung tumors to chemotherapy. *Proc Natl Acad Sci U S A 107*, 20786-20791.

Dong, Q., Barsky, D., Colvin, M.E., Melius, C.F., Ludeman, S.M., Moravek, J.F., Colvin, O.M., Bigner, D.D., Modrich, P., and Friedman, H.S. (1995). A structural

basis for a phosphoramide mustard-induced DNA interstrand cross-link at 5'-d(GAC). *Proc Natl Acad Sci U S A 92*, 12170-12174.

Fan, Y.H., and Gold, B. (1999). Sequence-specificity for DNA interstrand cross-linking by alpha,omega-alkanediol dimethylsulfonate esters: Evidence for DNA distortion by the initial monofunctional lesion. *J Am Chem Soc 121*, 11942-11946.

Frank, E.G., and Woodgate, R. (2007). Increased catalytic activity and altered fidelity of human DNA polymerase iota in the presence of manganese. *J Biol Chem 282*, 24689-24696.

Guainazzi, A., Campbell, A.J., Angelov, T., Simmerling, C., and Scharer, O.D. (2010). Synthesis and molecular modeling of a nitrogen mustard DNA interstrand crosslink. *Chemistry 16*, 12100-12103.

Haracska, L., Prakash, S., and Prakash, L. (2002). Yeast Rev1 protein is a G template-specific DNA polymerase. *J Biol Chem 277*, 15546-15551.

Ho, T.V., Guainazzi, A., Derkunt, S.B., Enoiu, M., and Schärer, O.D. (2011). Structure-dependent bypass of DNA interstrand crosslinks by translesion synthesis polymerases. *Nucleic Acids Res 39*, 7455-7464.

Ho, T.V., and Schärer, O.D. (2010). Translesion DNA synthesis polymerases in DNA interstrand crosslink repair. *Environ Mol Mutagen 51*, 552-566.

Johnson, R.E., Prakash, S., and Prakash, L. (1999). Efficient bypass of a thymine-thymine dimer by yeast DNA polymerase, Poleta. *Science 283*, 1001-1004.

Klein Douwel, D., Boonen, R.A., Long, D.T., Szypowska, A.A., Raschle, M., Walter, J.C., and Knipscheer, P. (2014). XPF-ERCC1 acts in Unhooking DNA interstrand crosslinks in cooperation with FANCD2 and FANCP/SLX4. *Mol Cell 54*, 460-471.

Klug, A.R., Harbut, M.B., Lloyd, R.S., and Minko, I.G. (2012). Replication bypass of N^2-deoxyguanosine interstrand cross-links by human DNA polymerases η and ι. *Chem Res Toxicol 25*, 755-762.

Knipscheer, P., Räschle, M., Smogorzewska, A., Enoiu, M., Ho, T.V., Schärer, O.D., Elledge, S.J., and Walter, J.C. (2009). The Fanconi anemia pathway promotes replication-dependent DNA interstrand cross-link repair. *Science 326*, 1698-1701.

Kottemann, M.C., and Smogorzewska, A. (2013). Fanconi anaemia and the repair of Watson and Crick DNA crosslinks. *Nature 493*, 356-363.

Lehmann, A.R., Niimi, A., Ogi, T., Brown, S., Sabbioneda, S., Wing, J.F., Kannouche, P.L., and Green, C.M. (2007). Translesion synthesis: Y-family polymerases and the polymerase switch. *DNA Repair (Amst) 6*, 891-899.

Long, D.T., Räschle, M., Joukov, V., and Walter, J.C. (2011). Mechanism of RAD51-dependent DNA interstrand cross-link repair. *Science 333*, 84-87.

Masutani, C., Kusumoto, R., Yamada, A., Dohmae, N., Yokoi, M., Yuasa, M., Araki, M., Iwai, S., Takio, K., and Hanaoka, F. (1999). The XPV (xeroderma pigmentosum variant) gene encodes human DNA polymerase eta. *Nature 399*, 700-704.

McCulloch, S.D., Kokoska, R.J., Masutani, C., Iwai, S., Hanaoka, F., and Kunkel, T.A. (2004). Preferential cis-syn thymine dimer bypass by DNA polymerase eta occurs with biased fidelity. *Nature 428*, 97-100.

Millard, J.T., Raucher, S., and Hopkins, P.B. (1990). Mechlorethamine cross-links deoxyguanosine residues at 5'-GNC sequences in duplex DNA fragments. *J Am Chem Soc 112*, 2459-2460.

Minko, I.G., Harbut, M.B., Kozekov, I.D., Kozekova, A., Jakobs, P.M., Olson, S.B., Moses, R.E., Harris, T.M., Rizzo, C.J., and Lloyd, R.S. (2008). Role for DNA polymerase κ in the processing of N^2-N^2-guanine interstrand cross-links. *J Biol Chem 283*, 17075-17082.

Misra, R.R., and Vos, J.M. (1993). Defective replication of psoralen adducts detected at the gene-specific level in xeroderma pigmentosum variant cells. *Mol Cell Biol 13*, 1002-1012.

Moldovan, G.-L., Madhavan, M.V., Mirchandani, K.D., McCaffrey, R.M., Vinciguerra, P., and D'Andrea, A.D. (2010). DNA polymerase POLN participates in cross-link repair and homologous recombination. *Mol Cell Biol 30*, 1088-1096.

Mukherjee, S., Guainazzi, A., and Scharer, O.D. (2014). Synthesis of structurally diverse major groove DNA interstrand crosslinks using three different aldehyde precursors. *Nucleic Acids Res 42*, 7429-7435.

Nelson, J.R., Lawrence, C.W., and Hinkle, D.C. (1996). Deoxycytidyl transferase activity of yeast REV1 protein. *Nature 382*, 729-731.

Nojima, K., Hochegger, H., Saberi, A., Fukushima, T., Kikuchi, K., Yoshimura, M., Orelli, B.J., Bishop, D.K., Hirano, S., Ohzeki, M.*, et al.* (2005). Multiple repair pathways mediate tolerance to chemotherapeutic cross-linking agents in vertebrate cells. *Cancer Res 65*, 11704-11711.

Noll, D.M., Mason, T.M., and Miller, P.S. (2006). Formation and repair of interstrand cross-links in DNA. *Chem Rev 106*, 277-301.

Pizzolato, J., Mukherjee, S., Scharer, O.D., and Jiricny, J. (2015). FANCD2-associated nuclease 1, but not exonuclease 1 or flap endonuclease 1, is able to unhook DNA interstrand cross-links in vitro. *J Biol Chem 290*, 22602-22611.

Prakash, S., Johnson, R.E., and Prakash, L. (2005). Eukaryotic translesion synthesis DNA polymerases: specificity of structure and function. *Ann Rev Biochem 74*, 317-353.

Prober, J.M., Trainor, G.L., Dam, R.J., Hobbs, F.W., Robertson, C.W., Zagursky, R.J., Cocuzza, A.J., Jensen, M.A., and Baumeister, K. (1987). A system for rapid DNA sequencing with fluorescent chain-terminating dideoxynucleotides. *Science 238*, 336-341.

Räschle, M., Knipscheer, P., Enoiu, M., Angelov, T., Sun, J., Griffith, J.D., Ellenberger, T.E., Schärer, O.D., and Walter, J.C. (2008). Mechanism of replication-coupled DNA interstrand crosslink repair. *Cell 134*, 969-980.

Reissner, T., Schneider, S., Schorr, S., and Carell, T. (2010). Crystal structure of a cisplatin-(1,3-GTG) cross-link within DNA polymerase eta. *Angew Chem Int Ed Engl 49*, 3077-3080.

Rink, S.M., and Hopkins, P.B. (1995). A mechlorethamine-induced DNA interstrand cross-link bends duplex DNA. *Biochemistry 34*, 1439-1445.

Sarkar, S., Davies, A.A., Ulrich, H.D., and McHugh, P.J. (2006). DNA interstrand crosslink repair during G1 involves nucleotide excision repair and DNA polymerase zeta. *EMBO J 25*, 1285-1294.

Scharer, O.D. (2005). DNA interstrand crosslinks: natural and drug-induced DNA adducts that induce unique cellular responses. *Chembiochem 6*, 27-32.

Shachar, S., Ziv, O., Avkin, S., Adar, S., Wittschieben, J., Reissner, T., Chaney, S., Friedberg, E.C., Wang, Z., Carell, T.*, et al.* (2009). Two-polymerase mechanisms dictate error-free and error-prone translesion DNA synthesis in mammals. *EMBO J 28*, 383-393.

Shen, X., Jun, S., O'Neal, L.E., Sonoda, E., Bemark, M., Sale, J.E., and Li, L. (2006). REV3 and REV1 play major roles in recombination-independent repair of DNA interstrand cross-links mediated by monoubiquitinated proliferating cell nuclear antigen (PCNA). *J Biol Chem 281*, 13869-13872.

Silverstein, T.D., Johnson, R.E., Jain, R., Prakash, L., Prakash, S., and Aggarwal, A.K. (2010). Structural basis for the suppression of skin cancers by DNA polymerase eta. *Nature 465*, 1039-1043.

Smeaton, M.B., Hlavin, E.M., Noronha, A.M., Murphy, S.P., Wilds, C.J., and Miller, P.S. (2009). Effect of cross-link structure on DNA interstrand cross-link repair synthesis. *Chem Res Toxicol 22*, 1285-1297.

Sonoda, E., Okada, T., Zhao, G.Y., Tateishi, S., Araki, K., Yamaizumi, M., Yagi, T., Verkaik, N.S., van Gent, D.C., Takata, M.*, et al.* (2003). Multiple roles of Rev3, the catalytic subunit of pol[zeta] in maintaining genome stability in vertebrates. *EMBO J 22*, 3188-3197.

Srivastava, A.K., Han, C., Zhao, R., Cui, T., Dai, Y., Mao, C., Zhao, W., Zhang, X., Yu, J., and Wang, Q.E. (2015). Enhanced expression of DNA polymerase eta contributes to cisplatin resistance of ovarian cancer stem cells. *Proc Natl Acad Sci U S A 112*, 4411-4416.

Tomicic, M.T., Aasland, D., Naumann, S.C., Meise, R., Barckhausen, C., Kaina, B., and Christmann, M. (2014). Translesion polymerase eta is upregulated by cancer therapeutics and confers anticancer drug resistance. *Cancer Res 74*, 5585-5596.

Wang, A.T., Sengerová, B., Cattell, E., Inagawa, T., Hartley, J.M., Kiakos, K., Burgess-Brown, N.A., Swift, L.P., Enzlin, J.H., Schofield, C.J.*, et al.* (2011). Human SNM1A and XPF–ERCC1 collaborate to initiate DNA interstrand cross-link repair. *Genes & Dev 25*, 1859-1870.

Wang, R., Persky, N.S., Yoo, B., Ouerfelli, O., Smogorzewska, A., Elledge, S.J., and Pavletich, N.P. (2014). DNA repair. Mechanism of DNA interstrand cross-link processing by repair nuclease FAN1. *Science 346*, 1127-1130.

Williams, Hannah L., Gottesman, Max E., and Gautier, J. (2012). Replication-independent repair of DNA interstrand crosslinks. *Mol Cell 47*, 140-147.

Xie, K., Doles, J., Hemann, M.T., and Walker, G.C. (2010). Error-prone translesion synthesis mediates acquired chemoresistance. *Proc Natl Acad Sci U S A 107*, 20792-20797.

Yamanaka, K., Minko, I.G., Takata, K.-i., Kolbanovskiy, A., Kozekov, I.D., Wood, R.D., Rizzo, C.J., and Lloyd, R.S. (2010). Novel enzymatic function of DNA polymerase v in translesion DNA synthesis past major groove DNA−peptide and DNA−DNA cross-links. *Chem Res Toxicol 23*, 689-695.

Zhang, J., Dewar, J.M., Budzowska, M., Motnenko, A., Cohn, M.A., and Walter, J.C. (2015). DNA interstrand cross-link repair requires replication-fork convergence. *Nat Struct & Mol Bio 22*, 242-247.

Zhang, J., and Walter, J.C. (2014). Mechanism and regulation of incisions during DNA interstrand cross-link repair. *DNA Repair (Amst) 19*, 135-142.

Zhao, Y., Biertumpfel, C., Gregory, M.T., Hua, Y.J., Hanaoka, F., and Yang, W. (2012). Structural basis of human DNA polymerase eta-mediated chemoresistance to cisplatin. *Proc Natl Acad Sci U S A 109*, 7269-7274.

Zietlow, L., Smith, L.A., Bessho, M., and Bessho, T. (2009). Evidence for the involvement of human DNA polymerase N in the repair of DNA interstrand cross-links. *Biochemistry 48*, 11817-11824.

CHAPTER 4

INTRODUCTION

During replication-coupled ICL repair, replication fork proteins are the first to encounter crosslinks, activating downstream repair events. A typical eukaryotic replication fork complex consists of multiple essential proteins including the replicative CMG helicase, replicative polymerases, PCNA and the clamp loader RFC (Burgers and Kunkel, 2017). The replicative CMG helicase encircles the leading strand and unwinds the double stranded DNA in a 5' to 3' direction releasing two single stranded DNA strands, which are replicated asymmetrically as the leading and lagging strands, respectively. Each DNA strand is encircled by the ring shaped homotrimeric DNA clamp PCNA which slides along the DNA with replicative polymerases pol ε and pol δ on the leading and lagging strands respectively. The leading strand is replicated in the same direction as helicase unwinding and fork progression, allowing for continuous DNA synthesis by pol ε. On the lagging strand, DNA synthesis takes place in a direction opposite to fork progression, leading to discontinuous DNA synthesis by pol δ. These discontinuous DNA segments, termed Okazaki fragments, are processed and ligated to produce a continuous DNA strand during replication. Both pol ε and pol δ belong to the B-family of polymerases, and have a very high accuracy of DNA synthesis due to their low tolerance for an incorrect base in the active site, as well as a 3'-5' proofreading exonuclease activity (Johansson and Dixon, 2013). Due to their constrained active site and low tolerance for damaged templates, the

prevalent model assumes replicative polymerases are not capable of carrying out replication of DNA containing ICLs. It is thought that upon stalling of the replication fork at the ICL, the lagging strand is incised on either side of the ICL releasing it from one strand (unhooking, **Fig. 1.4A iii & iv**) (Knipscheer et al., 2009; Raschle et al., 2008). The exact structure of the unhooked ICL is unknown, but is believed to contain a short piece of incised DNA attached to the leading strand through two covalently linked bases (**Fig. 1.4A iv**). As studies have shown that such structures are not readily bypassed by replicative polymerases, it is likely that error-prone TLS polymerases, recruited to the ICL after stalling of the replicative polymerases, carry out insertion and extension of the unhooked ICL template (Budzowska et al., 2015; Clauson et al., 2013).

The replication fork pauses at multiple positions when approaching an ICL (Raschle et al., 2008). Leading strands stall around 20-40 nt away from the crosslink as the progression of the CMG helicase is blocked by the ICL (**Fig. 1.4 Ai**), The CMG helicase is then removed in a BRCA1-dependent manner to allow one of the leading strands to “approach” to within 1nt of the ICL to (**Fig. 1.4 A ii**) (Long et al., 2014). There is strong evidence to suggest that the replicative polymerases, and not TLS polymerases, carry out this approach. First, after fork stalling at the -20 position, inhibition of replicative polymerases by aphidicolin treatment inhibits further approach (Long et al., 2014). Second, ChIP experiments show that PCNA, pol ε and pol δ are preferentially retained at the ICL locus during the approach (Budzowska et al., 2015). However, which of the replicative polymerases carries out the approach has not been demonstrated

conclusively. Pol ε is the more likely candidate since it already occupies the leading strand, however it has been suggested that due to it's strong interaction with CMG helicase, pol ε may be removed along with the helicase before approach (Budzowska et al., 2015; Georgescu et al., 2014). In this scenario PCNA could recruit pol δ to the leading strand to carry out DNA synthesis for the approach (Budzowska et al., 2015).

Pol ε is a highly processive enzyme, and this activity is only moderately stimulated by interaction with PCNA (Chilkova et al., 2007; Ganai et al., 2016; Maga et al., 1999). However, pol ε has a low strand displacement activity and is only able to displace 1nt in a double stranded DNA region due to a strong 3'-5' exonuclease activity. Suppression of the exonuclease activity leads to a significant increase in strand displacement abilities (Ganai et al., 2016; Garg et al., 2004). By contrast, the inherently very low processivity of pol δ as well as rate of DNA synthesis is drastically increased by interaction with PCNA (Chilkova et al., 2007; Stodola and Burgers, 2016). During Okazaki fragment maturation, pol δ carries out strand displacement synthesis in the presence of PCNA, creating short 1nt flaps which are cleaved by FEN1 to produce a ligatable nick (Jin et al., 2003). However, pol δ is unable to carry out extensive strand displacement synthesis to produce longer flaps, and instead stalls due to it's 3'-5' exonuclease activity in a process called "idling" (Garg et al., 2004). Consistent with this observation, pol δ mutants lacking the exonuclease activity have a much stronger strand displacement activity and can produce longer flaps (Garg et al., 2004). The strand displacement activities of pol ε and pol δ are likely to play an

important role in the approach to an ICL, since there may be some amount of duplex DNA surrounding the ICL during this step.

There is some evidence to suggest that ICL structure influences the nature of the approach. Studies in *Xenopus* egg extracts showed that a highly distorting cisplatin ICL is approached more efficiently than a non-distorting NM-like 8a ICL (Raschle et al., 2008). We also find this to be true in our studies with pol η and Klenow polymerase (discussed in Chapter 3), where we observed that the approach varied based on the structure of the ICL. Consistent with the results from *Xenopus*, approach was facilitated when the crosslink was distorting, and inhibited when it was non-distorting. Since the approach likely involves replicative polymerases, the effect of ICL structure on the activity of pol ε and pol δ is of interest, but has not been investigated so far. Using our unhooked ICL substrates we studied how the structure of an ICL influences the approach and translesion synthesis by the yeast replicative polymerases, pol ε and pol δ. We also discuss the synthesis of ICL substrates with blocked ends to study the effect of PCNA on this process.

MATERIALS AND METHODS

Oligonucleotides used for synthesis of biotinylated substrates

The following oligos were used to generate the biotinylated ICLs:

Extension oligos:

5'ext (24nt): 5'biotin – CACTAGACGAAGCTTGATATGGGC– 3'OH

3'ext (30nt): 5'P – GGCATTCAAGTGACGGGTACCATAGTCACG– 3'biotin

Splint oligos:

5'splint (22nt) : 5'-TCTTTCGCCCATATCAAGCTTC-3'

3'splint (21nt) : 5'-CACTTGAATGCCCACTGACTC- 3'

Oligos for ICL synthesis:

The following oligonucleotides were synthesized for the generation of ICLs as described (Angelov et al., 2009; Guainazzi et al., 2010) containing 7-deaza-dG aldehyde ICL precursors denoted as 'X':

T39 (39mer): 5'-GAAAGAAG**X**ACAGAAGAGGGTACCATCATAGAGTCAGTG-3'

C20 (20mer): 5'-CCCTCTUCT**X**TCCUTCTTTC-3'.

The 20bp ICL and 6bp ICL were synthesized as described previously in Chapter 2 and Chapter 3. The resulting 6bp ICL had the following sequence, with the crosslinked bases denoted as 'X':

3'-CCT**X**TC-5'

5'-GAAAGAAG**X**ACAGAAGAGGGTACCATCATAGAGTCAGTG-3

Synthesis of biotinylated substrates

500 pmol of purified the oligonucleotide 39mer 6bp ICL was annealed to 6 µL of 100 µM extension oligos and 10 µL of 100 µM splint oligos, in 10 mM Tris-HCl pH 8.0, 50mM NaCl overnight at room temperature in a total volume of 200 µL. The ligation reaction was performed by addition of 6 µL of T4 PNK (NEB M0201S), 12 µL T4 DNA ligase (NEB 0202S) in NEB DNA ligase buffer (50mM Tris-Cl pH 7.5, 10mM $MgCl_2$, 10mM DTT and 1mM ATP) in a total reaction volume of 300 µL. The ligation reaction was incubated for 1 hour at 37°C, and the biotinylated 6bp ICL purified by 10% denaturing PAGE. The bio-6bp ICL was extracted from the

gel by electroelution using the Scheicher & Schuell BT1000 Biotrap system according to manufacturer's instructions in 5mM Na_2B4O_7 (pH 8), and desalted by ethanol precipitation.

To synthesize the biotinylated 93mer 20bp ICLs, 1 nmol of the T39 oligo containing the C2 ICL precursor was annealed to 12 µL 100 µM extension oligos and 20 µL 100 µM splint oligos in 10 mM Tris-HCl pH 8.0, 50 mM NaCl overnight at room temperature in a total volume of 200 µL. The ligation reaction was performed by addition of 6 µL of T4 PNK (NEB M0201S), 12 µL T4 DNA ligase (NEB 0202S) in NEB DNA ligase buffer (50mM Tris-Cl pH 7.5, 10mM $MgCl_2$, 10mM DTT and 1mM ATP) in a total volume of 300 µL. The ligation reaction was incubated for 1 hour at 37°C, and the biotinylated 93mer oligo containing ICL precursor (bio-T39) purified by 10% denaturing PAGE and extracted from the gel by electroelution using the Scheicher & Schuell BT1000 Biotrap system according to manufacturer's instructions in 5 mM $Na_2B_4O_7$ (pH 8), and desalted by ethanol precipitation. ICLs were generated as described previously (Ho et al., 2011; Mukherjee et al., 2014). A solution of 1 nmol purified bio-T39 and 1 nmol C20 oligonucleotide (containing C2 ICL precursor) was prepared in 50 mM NaCl in a total volume of 100 µL, heated to 95°C and cooled slowly to allow annealing to take place. The annealed oligonucleotides were oxidized by addition of 8 µL 50 mM $NaIO_4$ and 12 µL 1M sodium phosphate buffer (pH 5.4) overnight at 4°C, and excess oxidizing agent removed by washing with 10 mM sodium phosphate buffer (pH 5.4) using Amicon Ultracel 3K columns. ICLs were formed by treatment with 2 µL either 5 mM hydrazine or 5 mM N,N-

dimethylethylenediamine, and 2 μL $NaBH_3CN$. The coupling reaction was incubated overnight in the dark at room temperature, and the bio-20bp ICLs were purified by 10% denaturing PAGE, and extracted from the gel by electroelution as described above.

Enzymes

Yeast pol ε (WT) and (exo-) with a concentration of 237 nM were a gift from Erik Johansson and Pia Osterman (Umea University) (Ganai et al., 2016). The protein was diluted in 25 mM HEPES (pH 7.6), 0.005% NP-40, 1 mM EDTA, 0.5 mM EGTA, 1 mM DTT, 400 mM NaOAc, 0.2 mg/mL BSA and 10% glycerol for use in polymerase assays. Yeast pol δ (WT) and (exo-) with a concentration of 6 μM and 1.9 μM, respectively, was a gift from Peter Burgers (Washington University, St. Louis) (Koc et al., 2015) . The enzymes were diluted in 30 mM HEPES (pH 7.4), 200 mM NaCl, 1 mM DTT and 10% glycerol.

Polymerase assays

7.5 μL 1 μM purified ICL and 2.5 μL 1 μM P15 primer were annealed in 10 mM Tris-HCl pH 8.0, 50 mM NaCl, in a volume of 50 μL overnight at room temperature to ensure the stability of the ICLs. The ICL substrates/primers (5 nM, with respect to the primer) and 100 μM dNTPs were incubated with DNA polymerase in a reaction volume of 10 μL for the indicated amounts of time. For assays with pol ε, the reaction was carried out in 20 mM Tris-HCL (pH 7.8), 100 μg/mL BSA, 1 mM DTT and 8 mM $MgCl_2$; the reactions with pol δ were carried out in 20 mM Tris-HCl (pH 7.8), 8 mM $MgOAc_2$, 1 mM DTT, 100 μg/mL BSA and 20 mM NaCl. Reactions were incubated at 37°C and stopped by addition of 10

μL of formamide buffer (80% formamide, 1 mM EDTA, 1 mg/mL Orange G), denatured at 95°C for 2 minutes and chilled on ice. The products of the reaction were resolved on a 10% 7M Urea denaturing PAGE and the FAM labeled DNA was visualized using a Typhoon 9400 scanner (GE Healthcare). Images were analyzed and quantified using ImageQuant software (Molecular Dynamics).

RESULTS

Duplex resection and distortion facilitates approach by pol ε and pol δ

Replicative polymerases have been implicated in carrying out the approach to an ICL during replication-dependent ICL repair (Budzowska et al., 2015; Long et al., 2014). Once the replication forks converge on an ICL, leading strands stall at a distance of 20-40 nt from the ICL, and removal of the CMG helicase allows one of the leading strands to approach the ICL. There has been some speculation as to the state of the duplex around the ICL after helicase removal (Zhang and Walter, 2014). The duplex around an ICL may re-anneal to some extent after CMG helicase removal, requiring the replicative polymerases to carry out strand displacement through a long duplex to extend the leading strand to -1 position. Using our ICL substrates, we addressed how ICL structure and duplex context influence this step. We used purified yeast polymerase ε with our 5a, 6a and 8a ICLs in the 20bp duplex substrate (**Fig. 3.4**). The FAM-labeled primer P15 was annealed to the substrates, incubated with pol ε and the extension of the primer monitored at two different enzyme concentrations by denaturing PAGE analysis. Pol ε stalled predominantly at the start of the dsDNA region in all of the substrates – including the ds control – consistent with its low

strand displacement activity (**Fig. 4.1B**). We also observed degradation of the primer due to the 3'-5' exonuclease activity of the wild-type polymerase. We next evaluated whether pol ε can approach the ICL more easily in the substrates with a shorter, 6bp duplex. Interestingly, although pol ε was unable to carry out strand displacement of the 6a/6bp and 8a/6bp substrates, it readily extended the primer in the 5a/6bp substrate to within 1nt of the ICL, stalling at the -1 position (**Fig. 4.1C**). No stalling at the beginning of the duplex was observed, indicating that the increased distortion of the 5a/6bp substrate facilitated the strand displacement reaction.

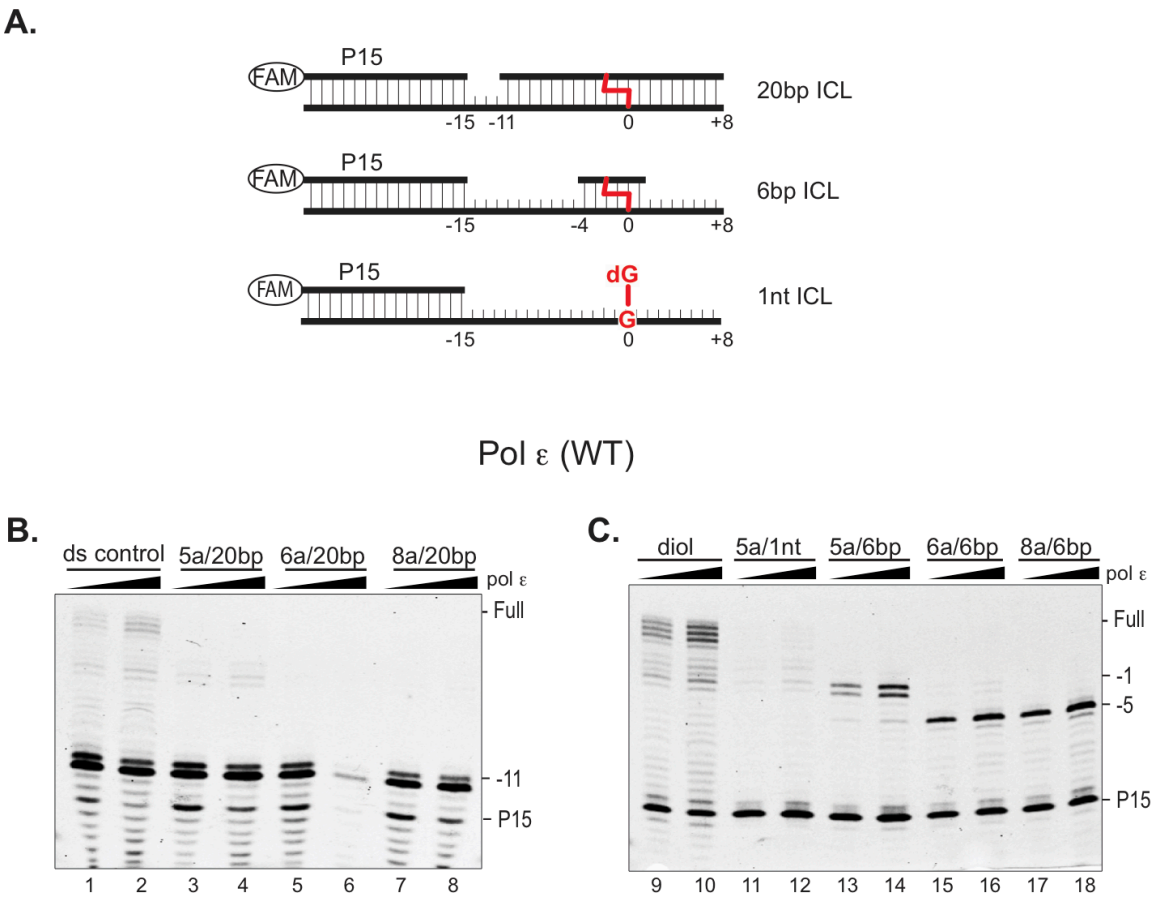


Figure 4.1 Duplex shortening and distortion facilitates approach by pol ε (WT). A. 20 bp, 6 bp and 1nt ICL substrates used. The crosslinked base is shown in red. **B,C**. Control and monoadduct (lanes 1-2, 9-10), and 5a 6a and 8a ICLs in a (**B.**) 20 bp duplex (lanes 3-8) or (**C.**) 6 bp duplex (lanes 13-18) or 1 nt substrate (lanes 11-12) were annealed to primer P15 and incubated with 3 or 6 nM pol ε (WT) for 5 mins at 37°C. The products were resolved by 10% denaturing PAGE.

Pol ε and pol δ mutants that retain polymerase activity, but lack exonuclease activity are more efficient at strand displacement (Ganai et al., 2016; Garg et al., 2004). We therefore wondered whether exonuclease deficient pol ε (exo-) could carry out approach to the 5a, 6a and 8a ICLs in the 20bp substrate. In contrast to pol ε (WT), pol ε (exo-) was able to carry out strand displacement and approach the ICL to within a few nucleotides in all ICL substrates (**Fig. 4.2 A**). There appeared to be a progressive increase in the extent of approach by pol ε (exo-) from the 8a/20bp to 6a/20bp and 5a/20bp ICLs (**Fig. 4.2A**, compare lanes 10-12 to lanes 7-9 and lanes 4-6). In the 8a ICL the main stalling was at the -3 position, in the 6a ICL an additional stalling point was observed closer to the ICL at the -1 position, and for the 5a ICL the stalling was mainly at the -1 position, indicating that the distortion induced by the 5a ICL facilitates approach to the crosslink. In the short 6bp duplex the enzyme stalled at -1 position for all three ICLs (5a/6bp, 6a/6bp and 8a/6bp) (**Fig. 4.2 B**, lanes 17-22) and unexpectedly, was also able to carry out insertion at 0 and +1 positions in the 5a/6bp ICL substrate (**Fig. 4.2 B** lanes 17-18). This shows that replicative polymerase ε may be able to insert a dNTP opposite certain unhooked NM ICL structures.

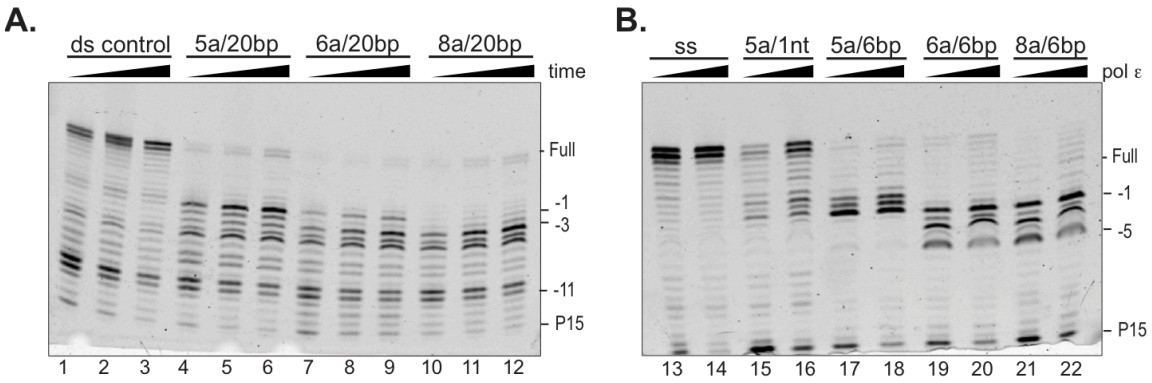


Figure 4.2 Duplex shortening and distortion facilitates approach by pol ε (exo-). (A,B): Control (lanes 1-3, 13-14), and 5a 6a and 8a ICLs in a (**A.**) 20 bp duplex (lanes 4-12) or (**B.**) 6 bp duplex (lanes 17-22) or 1 nt substrate (lanes 15-16) were annealed to primer P15 and incubated with (**A**) 3 nM pol ε (exo-) for 5, 10 and 20 mins at 37°C or (**B**) 3 or 6 nM pol ε (exo-) for 10 mins at 37°C. The products were resolved by 10% denaturing PAGE.

The lagging strand polymerase pol δ has a lower processivity and strand displacement activity than pol ε, and but its strand displacement activity is stimulated significantly by PCNA (Chilkova et al., 2007; Garg et al., 2004; Johansson and Dixon, 2013). Moreover, inactivation of the 3'-5' exonuclease activity increases its strand displacement activity even in the absence of PCNA (Garg et al., 2004; Koc et al., 2015). Given that we observed strand displacement in the 20bp ICL substrates with only pol ε (exo-) and not pol ε (WT), we decided to study the approach by exonuclease deficient pol δ (exo-). In contrast to pol ε (exo-), pol δ (exo-) was not able to carry out efficient strand displacement in the 20bp ICL substrates under our experimental conditions, stalling primarily at the beginning of the duplex (**Fig. 4.3 A**). However, the use of shorter duplexes facilitated approach and stalling points were observed at the -5 and -3 positions for the 6a/6bp and 8a/6bp ICL substrates (**Fig. 4.3B**, lanes 21-24, and at the -1

position for the 5a/6bp ICL substrate (**Fig. 4.3 B**, lanes 19-20). Together, the data indicates that duplex length and distortion are important factors that determine the nature of approach by pol ε and pol δ.

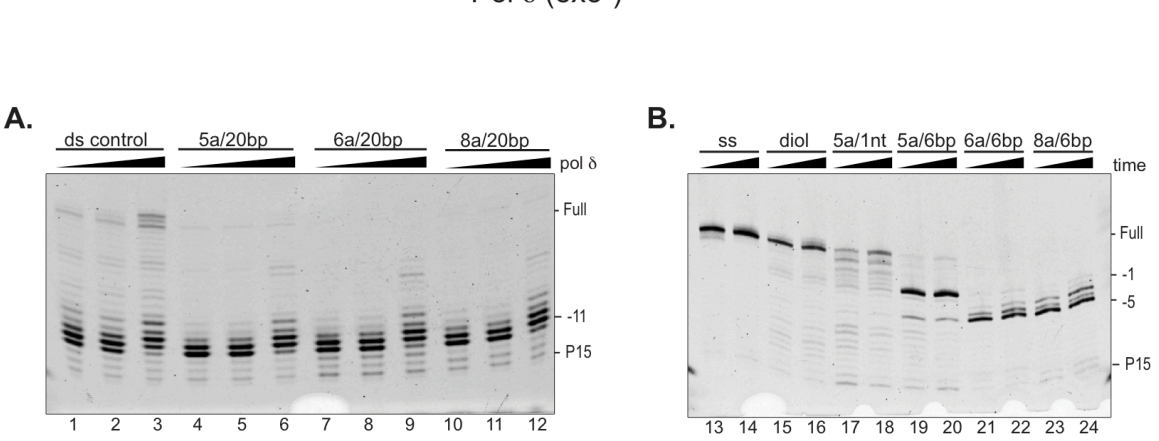


Figure 4.3 Duplex shortening and distortion facilitates approach by pol δ (exo-). (A,B). Control (lanes 1-3, 13-14), monoadduct (lanes 15-16) and 5a, 6a and 8a ICLs in a (**A**) 20 bp duplex (lanes 4-12) or (**B**) 6 bp duplex (lanes 19-24) or 1 nt substrate (lanes 17-18) were annealed to primer P15 and incubated with (**A**) 2, 4 or 8 nM pol δ (exo-) for 10 mins at 37°C or (**B**) 6 nM pol δ (exo-) for 10 or 20 mins at 37°C. The products were resolved by 10% denaturing PAGE.

Pol ε and pol δ can fully bypass the 1nt ICL

Having observed that resection of the duplex around an ICL facilitates bypass, we then asked how pol ε and pol δ interact with the most processed form of an ICL- the 5a/1nt substrate containing a single crosslinked nucleotide (**Fig. 3.4C vi**). This type of ICL structure has been observed after completion of ICL repair in *Xenopus* egg extracts, and is believed to be a possible ICL repair intermediate produced by unhooking and trimming of the duplex by exonucleases such as SNM1A or FAN1 to facilitate translesion synthesis (Allerston et al., 2015; Raschle et al., 2008; Wang et al., 2011; Wang et al., 2014). Remarkably, pol ε (exo-) was not only able to approach and insert efficiently across the 5a/1nt ICL,

but carried out complete bypass of the crosslink, extending the insertion product to full product (**Fig. 4.2 B** lanes 15-16). Similarly, pol δ (exo-) was also able to fully bypass the 5a/1nt ICL, carrying out the approach, insertion as well as extension (**Fig. 4.3 B** lanes 17-18). Our data demonstrate that both leading and lagging strand replicative polymerases are able to fully bypass the completely processed form of an NM ICL, raising the possibility that at least in some situations and for certain ICLs, a TLS polymerase may not be absolutely essential for ICL repair.

Synthesis of 93mer ICL substrates with biotinylated ends for studies with PCNA

During replication pol ε and pol δ are bound to PCNA on the leading and lagging strands respectively, although the strength of this interaction and effect on the individual polymerase activity varies (Burgers and Kunkel, 2017). Pol δ binds strongly to PCNA, and this interaction leads to a significant increase in the processivity and catalytic activity of pol δ (Chilkova et al., 2007; Garg et al., 2004). By contrast, the interaction of PCNA with pol ε is weaker and induces only a moderate increase in the processivity of pol ε (Chilkova et al., 2007; Maga et al., 1999).

To better simulate the *in vivo* activity of the replicative polymerases, we wanted to study how accessory factors such as PCNA influence pol ε and pol δ during translesion synthesis of ICL substrates. We devised a strategy to synthesize a 93-mer substrate with biotin at the 5' and 3' ends, allowing for

blockage of the ends with streptavidin to prevent PCNA from sliding off the linear substrate, after being loaded by the clamp loader RFC. Using our 6bp 39mer ICL substrate, we ligated 5' and 3' biotinylated "extension" oligos using two "splint" oligos (**Fig. 4.4**). The splints were designed to align the ends of the extension oligos to the 6bp ICL overhangs for an efficient ligation reaction. This strategy generated the biotinylated 6bp ICLs in a 93mer template (bio-6bp ICLs) with high efficiency, yielding the doubly ligated oligo (ligated on both sides) as the major product, with only small amounts of partially ligated products remaining (**Fig. 4.5 A**). The biotinylated substrates were gel purified for use in polymerase assays (**Fig. 4.5 B**).

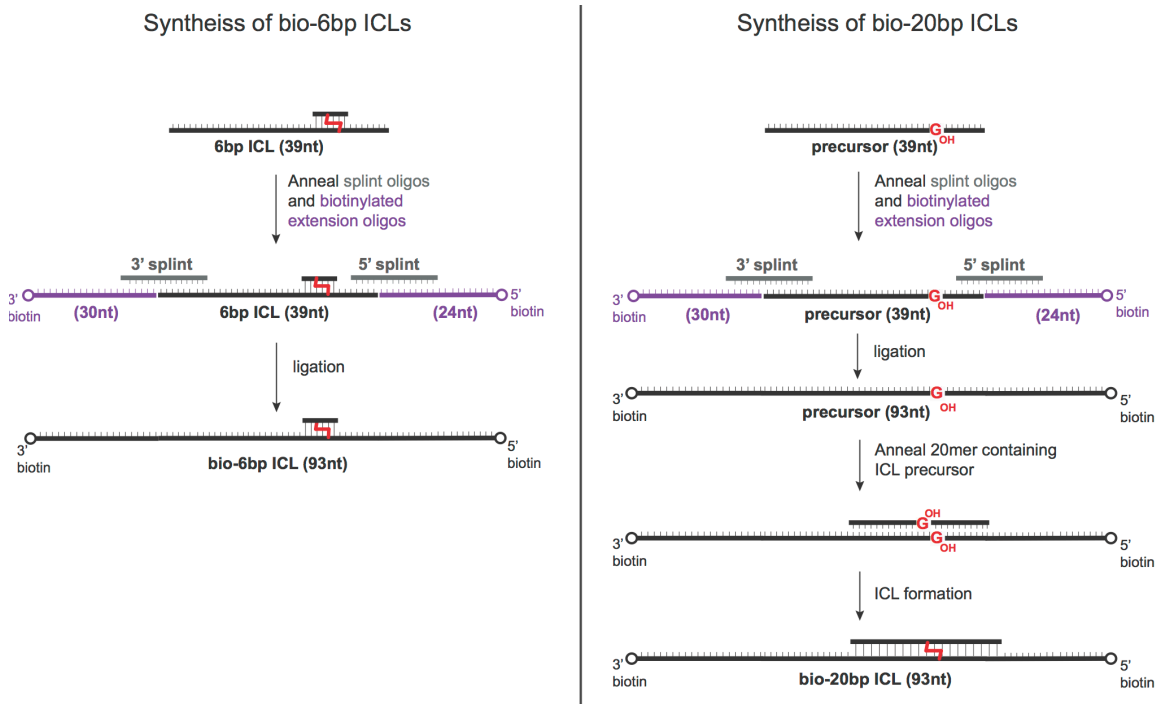


Figure 4.4 Scheme for synthesis of biotinylated ICL substrates

We were unable to use the same strategy to generate biotinylated 20bp ICLs due to lack of an overhang at the 5' end needed for annealing of the splint oligo. We therefore ligated our extension oligos to the single stranded 39mer containing the ICL precursor (T39) to generate a single stranded biotinylated 93mer. The purified 93mer was annealed to a complementary 20mer containing the ICL precursor (C20), and annealed oligos were crosslinked using oxidation of the precursors followed by reductive amination. This strategy allowed for the generation of the biotinlyated 20bp ICL in a 93mer template (bio-20bp ICLs) in high yield, and it was gel purified for use in polymerase assays (**Fig. 4.5B**, lane 11). These substrates will allow us to study the effect of PCNA on the approach and translesion synthesis of diverse ICL structures.

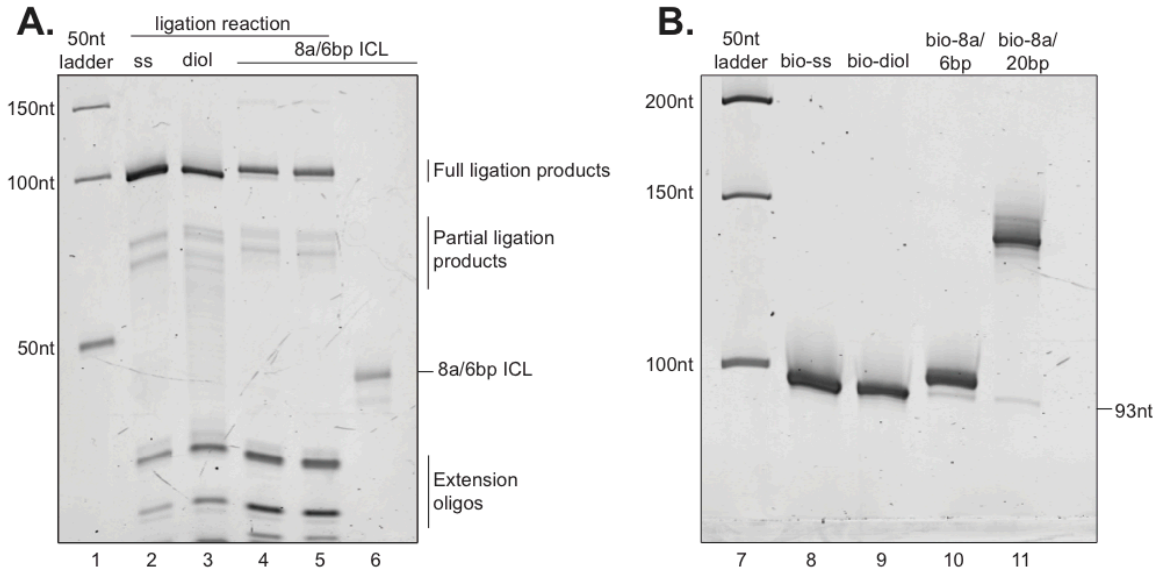


Figure 4.5: Synthesis and purification of longer substrates with biotinylated ends. **A.** 5 pmol purified 8a/6bp ICL (lane 6) or ligation reactions with biotinylated oligos (lanes 2-5) were resolved by 10% denaturing PAGE and visualized by SYBR Gold staining. Single stranded 39 mer (lane 2), diol (lane 3) or 8a/6bp ICL (lanes 4-5) was annealed to biotinylated extension oligos and incubated with 0.2U/uL T4 PNK and 0.4U/uL T4 DNA ligase for 1 hour at 37°C to generate the longer biotinylated substrates. The positions of the ICL, extension oligos and ligation products are indicated. **B.** 10 pmol purified 93mer bio-ss control (lane 8), 93mer bio-diol (lane 9), bio-8a/6bp ICL (lane 10) and bio-8a/20bp ICL (lane 11) were resolved by 10% denaturing PAGE and DNA visualized by SYBR Gold staining.

DISCUSSION

A common step in replication-coupled repair of all ICLs is the approach to the ICL, likely carried out by replicative polymerases. Evidence for this comes from experiments in *Xenopus* egg extracts, where it was observed that pol ε, pol δ and PCNA were enriched at the ICL during the approach, and inhibition of the replicative polymerases by aphidicolin blocked the approach (Budzowska et al., 2015; Long et al., 2014). Replicative polymerases therefore appear to carry out the approach to a variety of crosslinks, however the effect of ICL structure on the activity of replicative polymerases and the nature of approach has not been studied so far. Here we investigated this process *in vitro* using yeast pol ε and pol δ, and our panel of diverse unhooked ICL structures.

Effect of duplex resection and distortion on polymerase activity

Multiple biochemical studies have shown that shortening the duplex around an ICL significantly facilitates bypass by TLS polymerases across a variety of ICL structures (**Fig. 1.6**) [reviewed in (Roy and Scharer, 2016)]. However the amount of duplex surrounding an ICL during the approach is unknown, and could vary depending on the type of ICL. Studies in *Xenopus* egg extracts have shown that the approach and CMG unloading occur concurrently, with the helicase unloading being facilitated by the approach (Fu et al., 2011; Long et al., 2014). It has been speculated that once the helicase is removed, the parental DNA strands could re-anneal to some extent around the ICL (Zhang and Walter, 2014). Depending on the degree of re-annealing, the approach may

therefore require some amount of strand displacement. Our data with pol ε (WT) suggest that the polymerase alone cannot carry out the approach to an ICL in a long duplex (**Fig. 4.1B**), and would likely need additional factors to facilitate strand displacement. We mimicked this scenario using pol ε (exo-), which has a stronger strand displacement activity (Ganai et al., 2016) and found it was able to approach ICLs in the long duplex (**Fig. 4.2 A**). Resection of the duplex also significantly facilitated the approach (**Fig. 4.1 C, 4.2 B**) demonstrating that this step is heavily dependent on the need for strand displacement. This highlights the importance of a) mechanisms that may exist to prevent extensive re-annealing after helicase removal, such as binding of RPA to single stranded DNA and b) additional regulatory factors that could assist the polymerase in strand displacement.

The amount of distortion induced by an ICL also affects the efficiency of the approach by various DNA polymerases. Studies with *Xenopus* egg extracts have shown that a highly distorting cisplatin ICL was approached more efficiently than a non-distorting NM-like 8a ICL (Raschle et al., 2008). Our experiments provide biochemical data on how the duplex distortion influences the approach by replicative polymerases, and supports the observations that a distorting crosslink is approached more effectively than a non-distorting crosslink (**Figs. 4.1, 4.2 and 4.3**), likely due to a reduced need for strand displacement. Our data also raise the possibility that a replicative polymerase may not only carry out the approach, but also insertion across a distorting NM ICL (**Fig. 4.2 B** lanes 17-18). It would be interesting to see whether this insertion is error-free as the identity of

the polymerase that carries out the insertion step will have significant implications for the mutagenicity of the process. However, determining whether a replicative polymerase is responsible for insertion *in vivo* will be challenging, given the redundancy among various polymerases.

We were unable to observe sufficient strand displacement in our assays with pol δ, despite using the exonuclease deficient mutant with higher strand displacement activity (Garg et al., 2004; Koc et al., 2015). However, a recent study reported that the strand displacement activity of pol δ (exo-) is inhibited by excess single stranded DNA (Koc et al., 2015). It is therefore possible that the excess ICL substrate in our assays is inhibiting strand displacement by pol δ (exo-). However, interaction with PCNA alleviates this inhibitory effect and therefore including PCNA in our assays could stimulate strand displacement by pol δ (exo-), and will be important for studying the activity of pol δ on our ICL substrates.

Bypass of the single nucleotide NM ICL

As discussed in Chapter 3, there is some evidence for the generation of an ICL intermediate where the duplex is resected down to a single crosslinked nucleotide (Raschle et al., 2008; Wang et al., 2011). Our experiments with the bacterial replicative polymerase Klenow suggested that in fact, replicative polymerases may be able to carry out full bypass of the single nucleotide ICL structure (5a/1nt ICL) (**Fig. 3.6B**). Following up on our observations with Klenow, we found that the eukaryotic replicative polymerases pol ε and pol δ are also able to completely bypass the single nucleotide ICL. This raises the intriguing

possibility that bypass of at least certain ICL structures may not absolutely require TLS polymerases. Whether bypass of this ICL structure by replicative polymerases is error-free remains to be seen. It would also be important to determine during which step of ICL repair this intermediate is generated.

Reconstituting translesion synthesis at unhooked ICLs

In our studies so far we have addressed how ICL structure specifically influences the activity of DNA polymerases. *In vivo* however, translesion synthesis involves the coordination of multiple regulatory proteins. One such important factor is the DNA clamp PCNA that interacts with replicative polymerases, and can mediate polymerase switching due to its interactions with TLS polymerases (Burgers and Kunkel, 2017; Choe and Moldovan, 2017). To better understand how factors such as PCNA influence the bypass of structurally diverse ICLs by replicative polymerases, we have synthesized ICL substrates that can be blocked at the 5' and 3' ends with Streptavidin, allowing for stable loading of PCNA onto the ICL substrate (**Figs. 4.4 and 4.5**). Further studies with this substrate will provide insight into how accessory factors influence the approach and bypass of structurally diverse ICLs.

REFERENCES:

Allerston, C.K., Lee, S.Y., Newman, J.A., Schofield, C.J., McHugh, P.J., and Gileadi, O. (2015). The structures of the SNM1A and SNM1B/Apollo nuclease domains reveal a potential basis for their distinct DNA processing activities. *Nucleic Acids Res 43*, 11047-11060.

Angelov, T., Guainazzi, A., and Scharer, O.D. (2009). Generation of DNA interstrand cross-links by post-synthetic reductive amination. *Org Lett 11*, 661-664.

Budzowska, M., Graham, T.G., Sobeck, A., Waga, S., and Walter, J.C. (2015). Regulation of the Rev1-pol zeta complex during bypass of a DNA interstrand cross-link. *EMBO J 34*, 1971-1985.

Burgers, P.M., and Kunkel, T.A. (2017). Eukaryotic DNA Replication Fork. *Ann Rev Biochem*.

Chilkova, O., Stenlund, P., Isoz, I., Stith, C.M., Grabowski, P., Lundstrom, E.B., Burgers, P.M., and Johansson, E. (2007). The eukaryotic leading and lagging strand DNA polymerases are loaded onto primer-ends via separate mechanisms but have comparable processivity in the presence of PCNA. *Nucleic Acids Res 35*, 6588-6597.

Choe, K.N., and Moldovan, G.L. (2017). Forging Ahead through Darkness: PCNA, Still the Principal Conductor at the Replication Fork. *Mol Cell 65*, 380-392.

Clauson, C., Scharer, O.D., and Niedernhofer, L. (2013). Advances in understanding the complex mechanisms of DNA interstrand cross-link repair. *Cold Spring Harb Perspect Med 3*, a012732.

Fu, Y.V., Yardimci, H., Long, D.T., Ho, T.V., Guainazzi, A., Bermudez, V.P., Hurwitz, J., van Oijen, A., Scharer, O.D., and Walter, J.C. (2011). Selective bypass of a lagging strand roadblock by the eukaryotic replicative DNA helicase. *Cell 146*, 931-941.

Ganai, R.A., Zhang, X.P., Heyer, W.D., and Johansson, E. (2016). Strand displacement synthesis by yeast DNA polymerase epsilon. *Nucleic Acids Res 44*, 8229-8240.

Garg, P., Stith, C.M., Sabouri, N., Johansson, E., and Burgers, P.M. (2004). Idling by DNA polymerase delta maintains a ligatable nick during lagging-strand DNA replication. *Genes & Dev 18*, 2764-2773.

Georgescu, R.E., Langston, L., Yao, N.Y., Yurieva, O., Zhang, D., Finkelstein, J., Agarwal, T., and O'Donnell, M.E. (2014). Mechanism of asymmetric polymerase assembly at the eukaryotic replication fork. *Nat Struct & Mol Bio 21*, 664-670.

Guainazzi, A., Campbell, A.J., Angelov, T., Simmerling, C., and Schärer, O.D. (2010). Synthesis and molecular modeling of a nitrogen mustard DNA interstrand crosslink. *Chemistry 16*, 12100-12103.

Ho, T.V., Guainazzi, A., Derkunt, S.B., Enoiu, M., and Scharer, O.D. (2011). Structure-dependent bypass of DNA interstrand crosslinks by translesion synthesis polymerases. *Nucleic Acids Res 39*, 7455-7464.

Jin, Y.H., Ayyagari, R., Resnick, M.A., Gordenin, D.A., and Burgers, P.M. (2003). Okazaki fragment maturation in yeast. II. Cooperation between the polymerase and 3'-5'-exonuclease activities of Pol delta in the creation of a ligatable nick. *J Biol Chem 278*, 1626-1633.

Johansson, E., and Dixon, N. (2013). Replicative DNA polymerases. *Cold Spring Harb Perspect Biol 5*.

Koc, K.N., Stodola, J.L., Burgers, P.M., and Galletto, R. (2015). Regulation of yeast DNA polymerase delta-mediated strand displacement synthesis by 5'-flaps. *Nucleic Acids Res 43*, 4179-4190.

Long, D.T., Joukov, V., Budzowska, M., and Walter, J.C. (2014). BRCA1 promotes unloading of the CMG helicase from a stalled DNA replication fork. *Mol Cell 56*, 174-185.

Maga, G., Jonsson, Z.O., Stucki, M., Spadari, S., and Hubscher, U. (1999). Dual mode of interaction of DNA polymerase epsilon with proliferating cell nuclear antigen in primer binding and DNA synthesis. Journal of molecular biology *285*, 259-267.

Mukherjee, S., Guainazzi, A., and Scharer, O.D. (2014). Synthesis of structurally diverse major groove DNA interstrand crosslinks using three different aldehyde precursors. *Nucleic Acids Res*.

Raschle, M., Knipscheer, P., Enoiu, M., Angelov, T., Sun, J., Griffith, J.D., Ellenberger, T.E., Scharer, O.D., and Walter, J.C. (2008). Mechanism of replication-coupled DNA interstrand crosslink repair. *Cell 134*, 969-980.

Roy, U., and Scharer, O.D. (2016). Involvement of translesion synthesis DNA polymerases in DNA interstrand crosslink repair. *DNA Repair (Amst) 44*, 33-41.

Stodola, J.L., and Burgers, P.M. (2016). Resolving individual steps of Okazaki-fragment maturation at a millisecond timescale. *Nat Struct & Mol Bio 23*, 402-408.

Wang, A.T., Sengerová, B., Cattell, E., Inagawa, T., Hartley, J.M., Kiakos, K., Burgess-Brown, N.A., Swift, L.P., Enzlin, J.H., Schofield, C.J.*, et al.* (2011). Human SNM1A and XPF–ERCC1 collaborate to initiate DNA interstrand cross-link repair. *Genes & Dev 25*, 1859-1870.

Wang, R., Persky, N.S., Yoo, B., Ouerfelli, O., Smogorzewska, A., Elledge, S.J., and Pavletich, N.P. (2014). DNA repair. Mechanism of DNA interstrand cross-link processing by repair nuclease FAN1. *Science 346*, 1127-1130.

Zhang, J., and Walter, J.C. (2014). Mechanism and regulation of incisions during DNA interstrand cross-link repair. *DNA Repair (Amst) 19*, 135-142.

CHAPTER 5

CONCLUSIONS AND PERSPECTIVES

The aim of this project was to understand how ICL structure influences the activity of DNA polymerases during bypass – a key step in replication-dependent and -independent ICL repair pathways. During this step, DNA polymerases carry out translesion synthesis, using an unhooked ICL intermediate as a template. Although the structure of this intermediate is not known, the outcome of a bypass reaction has implications for the efficiency as well as the accuracy of repair. Using a panel of structurally diverse nitrogen mustard ICL analogs, we investigated the structure-function relationships of ICL structure and DNA polymerase activity. In particular, we studied the effect of ICL distortion, and position of unhooking incisions (duplex context), and evaluated the effect on approach, insertion, and extension abilities of replicative and TLS polymerases.

We carried out detailed studies with pol η, and performed some preliminary experiments with replicative polymerases pol δ and pol ε. We demonstrated that duplex distortion facilitates the approach and insertion by polymerases, but inhibits full extension. By contrast, approach to non-distorting crosslinks is inhibited, but the extension to the full product is easier. Consistent with previous studies, we also showed that resection of the duplex around an ICL greatly facilitates bypass. Together our results suggest that resection of the duplex by nucleases may be more important for the bypass and repair of non-distorting crosslinks, and future studies will be required to investigate this further.

In addition, using a new strategy developed in our lab to synthesize the most resected form of a crosslink – the 5a/1nt ICL, we found that it was fully bypassed by replicative polymerases alone. This observation raises the exciting possibility that a TLS polymerase may not be absolutely essential for bypassing certain unhooked ICL structures, although the stage at which such an intermediate is produced remains to be determined. We also laid the groundwork for future studies with proteins such as PCNA and RFC, by synthesizing a 93mer ICL substrate with biotin at the 5' and 3' ends. It will be interesting to expand on our results using a bottom-up approach, and see how these regulatory factors influence polymerases during the bypass of ICLs.

We are also excited about studying the translesion synthesis activity of REV1-pol ζ with our ICL substrates, in collaboration with Dr. Rachel Bezalel-Buch and Dr. Peter Burgers at Washington University (St. Louis, MO). Among all the polymerases tested, cells with mutations in REV1 and pol ζ show the highest sensitivity to crosslinking agents, and are strongly implicated in playing a critical role in ICL repair. However, the preeminent role of these two polymerases in the bypass of ICLs has so far not been recapitulated in biochemical assays. Therefore, there is considerable interest in the field to bridge the gap between genetic evidence and biochemical data for the role of pol ζ –REV1 in ICL repair. It is likely that additional factors are required for efficient activity on ICL substrates, and this is supported by the recent report showing that pol ζ functions with two additional subunits *in vivo* – polD2 and polD3- which stimulate the activity of pol ζ on *intra*strand crosslinks. The Burgers lab has successfully

purified the four-subunit pol ζ containing polD2 and polD3 subunits, and we are excited to see the activity of this complex on our ICL substrates.

A few other “next steps” will greatly advance our understanding of the structure-function relationships between ICL structure and DNA polymerase activity. First, crystal structures of DNA polymerases bound to ICL substrates will be important for understanding the structural basis for their translesion synthesis abilities. This will be particularly interesting in the case of replicative polymerases carrying out insertion in 5a/6bp ICL, and full bypass in the 5a/1nt ICL, since replicative polymerases are not expected to accommodate bulky lesions in their active site. Second, sequencing the products of a bypass reaction will provide crucial information about the fidelity of this reaction. Although we have some useful insight into the accuracy of bypass based on our single nucleotide insertion assays, sequencing the products will allow us to better understand the situation *in vivo*. It would also be valuable for determining the mutational signatures of individual polymerases, which could then be compared with depletion/knock-down studies to identify the polymerases involved in carrying out translesion synthesis in cells.

We have also contributed to the study of other aspects of ICL repair through a rich network of collaborations. In collaboration with Dr. Alice Miller and Dr. Mitch McVey at Tufts University, we investigated the role of *Drosophila* pol θ helicase-domain in ICL repair, and this work led to a manuscript accepted for publication in *PLoS Genetics*. Dr. Ian Macara and colleagues at Vanderbilt University identified a new nuclease important for ICL repair, and we synthesized

ICL substrates used to study the biochemical activity of this novel nuclease. This work has led to a manuscript that is currently in preparation. We also collaborated with Niyo Kato and Dr. Jean Gautier at Columbia University, interested in understanding the mechanisms of replication-independent ICL repair. Experiments in *Xenopus* egg extracts with our ICL substrates have provided important insight into this pathway, and the results are discussed in a manuscript submitted to *Nature Structural and Molecular Biology*. Our ICL substrates have served as crucial tools for understanding ICL repair mechanisms so far, and will be extremely valuable for future studies as well.