MOLECULAR MECHANISMS OF FLEXIBILITY IN NONHOMOLOGOUS END JOINING

Michael Patrick Conlin

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Curriculum in Genetics and Molecular Biology in the College of Arts and Sciences.

Chapel Hill
2018

Approved by:

Dale Ramsden

Dorothy Erie

Jeff Sekelsy

Shawn Ahmed

Scott Williams

ABSTRACT

Michael Patrick Conlin: Molecular Mechanisms of Flexibility in Nonhomologous End Joining
(Under the direction of Dale Ramsden)

DNA double strand breaks (DSBs) are highly toxic DNA lesions that play a critical role in human health and disease. The ability to repair these lesions is essential in all kingdoms of life, and in mammals is primarily attributed to the nonhomologous end joining (NHEJ) pathway. NHEJ faces a unique challenge: unlike other forms of DNA damage, DSBs are structurally heterogeneous, varying wildly in end chemistry. To address this problem, NHEJ has evolved uniquely flexible enzymes: DNA polymerases and a DNA ligase that can act on a remarkable variety of substrates, much more so than their counterparts in other pathways. The mechanistic basis of this flexibility, and its significance to biological repair, are unknown.

DNA Ligase IV (LIG4) is the only human DNA ligase that participates in NHEJ, and the only one that can efficiently ligate ends across gaps, or with terminal mispairs. We show by single-molecule analysis that terminal mispairs lead NHEJ complexes to mobilize DNA ends and thereby sample more end alignments. This flexibility is what allows LIG4 to join such ends, since pairing flexibility and ligation both require a LIG4-specific structural motif, insert1. Our work showed that pairing flexibility is what enables LIG4 to tolerate a chemically diverse array of substrates, and that this tolerance is essential for cells to survive exogenous DNA damage such as ionizing radiation.

NHEJ employs two uniquely flexible polymerases to prepare ends for ligation: DNA polymerase μ (pol μ) and terminal deoxynucleotidyl transferase (TdT). These enzymes act on noncanonical substrates that other polymerases cannot engage. We show these polymerases primarily incorporate ribonucleotides (RNA), not deoxynucleotides (DNA), during NHEJ, both during repair of chromosome breaks made by Cas9 and during V(D)J recombination. These ribonucleotides facilitate NHEJ by enabling ligation of ends with adjacent mispairs, and even single strand ligation. Supplementing cells expressing TdT with deoxynucleotides thus blocks repair of Cas9-induced breaks, while ribonucleotide supplementation can improve Cas9-directed mutagenesis. Our results indicate cellular NHEJ often involves transiently embedded ribonucleotides, which promote flexibility in repair at the cost of more fragile intermediates.

ACKNOWLEDGEMENTS

I thank my thesis advisor, Dale, for providing truly exemplary scientific mentorship and training. Dale is the most engaged scientist I have ever met: his passion for the work is truly infectious and has been transformative in my own development both as a scientist and as a person.

I would like to thank each member of the Ramsden lab that I have had the pleasure of working with. I would especially like to thank Dr. David Wyatt, whose friendship and career guidance have been integral to my own success.

I am grateful to the members of my thesis committee, Dr. Dorothy Erie, Dr. R. Scott Williams, Dr. Shawn Ahmed and Dr. Jeff Sekelsky, who have been exceptionally engaged with my projects. Every one of our meetings has been productive, and I've never left one without ideas for new experiments.

I also thank my supportive group of friends, particularly Chris Rounds who has reviewed much of my scientific writing (including this dissertation) and Ryan Hurd who assisted me with bioinformatics analysis of high throughput sequencing data.

Finally, I thank my family: my sister Kelsea, my father Mike Sr., my mother Heather, my fiancée Emma Sternlof and her parents, Kerin and Karl. Emma has been by my side through my entire Ph.D. and I would not have finished without her. Emma has also been indispensable as an editor for this dissertation and all of my professional writing.

Additional acknowledgments specific to each chapter

Chapter 2

This chapter was modified from its original version appearing in Cell Reports in 2017 entitled "DNA Ligase IV Guides End-Processing Choice during Nonhomologous End Joining," in which Dylan Reid and I were co-first authors. I have divided the publication here for the purpose of logical flow and have modified figure layouts for this dissertation. Supplemental figures and tables are included here. I, Dylan Reid, and George Small performed all experiments presented in this chapter and analyzed the data. I, Dale Ramsden, Dylan Reid, and Eli Rothenberg wrote the manuscript with input from Michael Lieber. I performed bioinformatics analysis of DNA sequences. Studies presented in this chapter were supported by CA203156 and GM007092, CA084442, and GM108118. I thank Dr. Eric Hendrickson for providing LIG4$^{+/+}$ and LIG4$^{-/-}$ cell lines used in this work.

Chapter 3

This chapter was modified from its original version which is a manuscript currently under review, in which John Pryor is first author and I am second author. I have divided the publication here for the purpose of logical flow and have modified figure layouts for this dissertation. Supplemental figures and tables are included here. John Pryor, myself, George Small, and Juan Carvajal Garcia performed all experiments presented in this chapter and analyzed the data. In sum, 12 panels of the figures presented here consist of primary data from my experiments. Dale Ramsden and I wrote the manuscript with input from John Pryor and Adam Luthman. I additionally thank Megan Fabry and Adam Luthman for contributing data that, although not

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND SYMBOLS

Bp	base pair
Cas9	CRISPR associated protein 9
CRISPR	Clustered Randomly Interspaced Short Palindromic Repeats
DNA	deoxyribonucleic acid
DNA-PKcs	DNA-dependent Protein Kinase catalytic subunit
dNTP	deoxynucleotide triphosphate
Ds	double stranded
DSB	double strand break
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
E_{FRET}	FRET energy
FRET	Forster resonance energy transfer
G_o	8-oxo-7,8-dihydroguanine
HR	homologous recombination
Igk	immunoglobulin kappa
IR	ionizing radiation
Jk	joining kappa
LIG4	DNA Ligase IV
MEF	mouse embryonic fibroblast
NHEJ	nonhomologous end joining
Nt	nucleotide
NTP	nucleotide triphosphate

PAGE	polyacrylamide gel electrophoresis
PAXX	paralog of XLF and XRCC4
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PEC	paired end complex
Pol	polymerase
qPCR	quantitative polymerase chain reaction
RAG1	recombinase activating gene 1
RAG2	recombinase activating gene 2
RER	ribonucleotide excision repair
RNA	ribonucleic acid
RNP	ribonucleoprotein
rNTP	ribonucleotide triphosphate
Sd	standard deviation
SDS	sodium dodecyl sulfate
sgRNA	CRISPR small guide RNA
sgRosa26	small guide RNA targeting Rosa26
smFRET	single molecule FRET
Ss	single stranded
TBE	tris, boric acid, and ethylenediaminetetraacetic acid
TdT	terminal deoxynucleotidyl transferase
TIDE	Tracking of Indels by DEcomposition
trcrRNA	CRISPR tracr RNA

V(D)J	variable, diversity, and joining
VJk	variable joining junction kappa
Vk	variable kappa
XLF	XRCC4-like factor
XRCC4	X-ray cross complemented 4
μ	mu
λ	lambda

CHAPTER 1: INTRODUCTION

DNA double strand breaks (DSBs) are a ubiquitous and highly toxic form of DNA damage. These lesions arise from a vast array of sources, including normal biological processes and exogenous damaging agents. The frequency and myriad sources of DSBs pose a serious threat to cells because unrepaired DSBs can lead to genome instability and cell death[1,2]. Thus, the ability to repair DSBs is essential in all kingdoms of life. In mammals, the dominant DSB repair pathway is nonhomologous end joining (NHEJ)[3]. A central challenge for this pathway is that DSBs are highly heterogeneous lesions and can have end chemistry that complicates repair[4]. To address this problem, the NHEJ pathway has evolved highly flexible enzymes that work on noncanonical substrates in unorthodox ways[5]. The mechanisms underlying this flexibility and their significance for biological repair are poorly understood.

1.1 DNA Double Strand Break Repair Pathways

In addition to NHEJ, the other main DSB repair pathway is homologous recombination (HR)[6]. HR is a template-based repair pathway: an intact second copy of the broken DNA serves as a template for repair. Mechanistically, HR begins with nucleolytic resection and proceeds through template-directed synthesis using a homologous chromosome or sister chromatid[6].

A third repair pathway, DNA Polymerase Theta Mediated End Joining (TMEJ), typically makes a relatively minor contribution to repair, although it becomes essential in the absence of NHEJ or HR[7,8]. Like HR, TMEJ relies on resection and synthesis to repair breaks[7]. These processing steps effectively convert a DSB into a substrate highly amenable to ligation: a simple set of DNA nicks spread apart by fully complementary, newly synthesized DNA. Because they involve such extensive processing, HR and TMEJ face limitations on when they can be active within the cell[9]. Moreover, HR requires a homologous template and TMEJ requires resected ends with embedded homology. In contrast, NHEJ is free from these requirements; it is active throughout the cell cycle and capable of robustly repairing DSBs irrespective of any homology[9].

The NHEJ pathway has been highly conserved across evolution and is present in bacteria and humans. Hypomorphic mutations in NHEJ genes cause human diseases characterized by severe immunodeficiency, cancer predisposition, neurodevelopmental disorders, and premature aging[10]. Repair by NHEJ is sometimes described as error-prone because it can produce small insertions and deletions at the site of repair[11]. In the event of multiple breaks, NHEJ can also improperly join different chromosomes together, producing chromosomal translocations that can lead to cancer[12]. A key challenge facing the NHEJ pathway is that DSBs are highly diverse lesions, both in terms of their causes and their chemical structures.

1.2 Sources of DNA Double Strand Breaks

Chromosome DSBs arise from a vast array of sources, and are even induced as programmed intermediates in biological processes. In meiosis, cells deliberately

introduce DSBs that are ultimately responsible for much of the genetic diversity of gametes[13]. Similarly, rearrangement of the variable, joining, and diversity gene segments (V(D)J recombination) underlies the vertebrate adaptive immune system and proceeds through a DSB intermediate[14]. However, all DSBs outside of the germline and lymphoid tissue are unintended, with dividing mammalian cells encountering around 10 breaks per cell cycle[15]. Genome replication is a major source of these breaks, which occur when the replication fork encounters DNA damage that cannot be bypassed, or when topoisomerase enzymes fail[16,17].

Outside of those normally encountered throughout biology, the most important sources of DSBs are exogenous damaging agents that include some of the most commonly used cancer therapies. Ionizing radiation (IR), topoisomerase poisons, and many other chemotherapeutics kill cancer cells primarily by inducing DSBs[18,19]. From a biotechnology standpoint, DSBs are an essential tool used to specifically edit the genome through guided nucleases, most notably in the CRISPR-Cas9 system[20,21].

1.3 Heterogeneity of Double Strand Breaks

The many sources of DSBs result in breaks with even more diverse chemical structures. Ionizing radiation (IR) alone produces breaks with a wide variety of associated oxidative lesions including 5,6-thymine glycol, 5-hydroxymethyl-2'-deoxyuridine, 5-formyl-2'-deoxyuridine, 8-oxoguanine, and many chemically distinct abasic sites[22–25]. Each of these forms of damage are common and in sum amount to well over 1000 lesions per cell per Gray of radiation[25]. Moreover, IR produces these lesions in clusters which often occur at or near the ends of DSBs[26,27]. Currently there is

no way to measure the exact fraction of IR-induced DSBs with associated oxidative lesions, but estimates based on lesion frequency suggest the proportion to be around one-third of all breaks[23]. Different sources of ionizing radiation are thought to vary in the number and complexity of the oxidative lesions they produce, and thus create DSBs with different end chemistry as well[28].

In contrast with IR, the structures of breaks induced by V(D)J recombination are well understood. V(D)J recombination occurs in developing lymphocytes and is essential for the adaptive immune system in vertebrates[14,29]. In this pathway, DSBs are intentionally induced by the recombination-activating genes RAG1 and RAG2[30–32]. RAG activity leaves one DSB end with a DNA hairpin that is cleaved imprecisely by the Artemis nuclease resulting in heterogeneous end structures[33]. This heterogeneity is further augmented by terminal deoxynucleotidyl transferase (TdT, see below)[34]. The corresponding sequence heterogeneity in the repaired products of these breaks gives rise to a diverse repertoire of antibody-encoding genes for an effective immune system. Repair in V(D)J recombination is carried out entirely by NHEJ and its enzymes function in this process largely as they do in nonlymphoid cells[35,36]. An exception is TdT, a specialized DNA polymerase that is only expressed in lymphoid tissue and acts almost exclusively during V(D)J recombination[37,38].

Topoisomerase poisons such as etoposide are some of the most commonly used cancer therapy agents. These drugs primarily kill cancer cells by generating DSBs with topoisomerase protein adducts that block repair[19]. In sum, the striking heterogeneity of DSBs poses a key challenge to repair pathways in general and NHEJ in particular.

1.4 Mechanism of NHEJ

The diversity of DSBs makes them unlike any other form of DNA damage, and the way NHEJ addresses these breaks is fundamentally different from any other repair pathway. While HR and TMEJ circumvent the heterogeneity of DSBs by extensively resecting breaks to render them simple repair substrates[6,7], NHEJ attacks complex breaks directly, without resection[11,39]. To repair such a wide range of substrates, NHEJ has evolved extraordinarily flexible enzymes that can act in unique ways and on substrates that other enzymes in their respective classes cannot engage[5].

The mechanism of NHEJ proceeds in 4 steps: 1) recognizing and binding the DSB, 2) bridging of break ends by the NHEJ paired end complex (PEC), 3) optional processing of the ends, and 4) ligation[40]. The first step, end recognition, is carried out by the Ku70/80 heterodimer (Ku)[41,42]. Ku is often one of the first proteins that binds to the ends of a double strand break; its binding then mediates the localization of other NHEJ proteins to the break as well[43,44]. After the NHEJ machinery reaches the break, the ends are juxtaposed and held in place to form a paired end complex (PEC)[45]. At a minimum, formation of stable PECs requires the NHEJ ligase, DNA Ligase IV (LIG4), Ku, and the scaffold proteins XRCC4 and XLF[46–49]. The DNA-dependent protein kinase catalytic subunit (DNA-PKcs) may also participate in PEC formation[50], but its specific role remains controversial[51].

Little is known about how efficiently PECs are formed on different kinds of breaks (i.e. heterogeneous end structures), the dynamics of the DNA ends and the proteins within the complex, and the relationship between PEC formation and ligation[51]. These questions are addressed in Chapter 2. Following PEC formation, NHEJ proceeds either

through direct ligation of the ends or optional processing of ends by polymerases or nucleases[4].

1.5 Rejoining of break ends by DNA Ligase IV

The ligation step of NHEJ is executed by LIG4, one of three DNA ligases in mammals and the only one that participates in this pathway[52]. LIG4 is an ATP-dependent enzyme consisting of a 3-domain N-terminal catalytic core and 2 C-terminal domains that facilitate its interactions with other NHEJ factors[53]. The catalytic core of LIG4, like those of most other ATP-dependent ligases, consists of a DNA binding domain, an adenylation domain containing an active site lysine, and an OB-fold domain[54].

DNA ligases have been observed in two structural conformations: a "closed" conformation in which the catalytic core domains wrap around a DNA substrate and an "open" conformation in which the catalytic core is extended and largely unbound[55–57]. Substrate binding is thought to dictate the transition between these states, and the transition is believed to serve as a checkpoint for ligases to interrogate their substrates for compatible termini to join[58]. This mechanism is highly conserved and is employed by both bacterial and human DNA ligases; thus, LIG4 likely uses it as well, although this remains to be demonstrated structurally. Since LIG4 is required for formation of the NHEJ PEC, its open-closed transition is presumably key to stable PEC formation. Other ligases probably do not form PECs; if and how LIG4 is specialized to act in this role is not understood.

After LIG4 engages the DSB ends, its active lysine covalently binds and then transfers an adenylate group to the 5' phosphate of the DSB, which is then resolved by nucleophilic attack of the 3' hydroxyl to form a phosphodiester bond[54]. LIG4 can carry out this chemistry on a remarkable variety of substrates, including ends with mispairs, oxidative damage, short patches of single strand DNA, and even 1-2 nucleotide gaps, without any processing steps to remove these distortions[11,59–61]. In contrast, the other mammalian ligases, and most ligases throughout biology, are high fidelity enzymes unable to join these kinds of end structures[60,62]. Since the NHEJ ligation step shows low fidelity, it is well adapted to the wide array of substrates it must join to allow for efficient repair[4]. This ligase activity is extremely rare in biology; its only precedents are NHEJ pathways in other organisms and some viral ligases[63,64]. While it is clear that NHEJ is remarkably flexible in accommodating a wide range of substrates, it is unknown if this activity is indeed important for cellular DSB repair in the context of the chromosome. The mechanism underlying this flexibility is also unknown, and both of these issues are addressed in Chapter 2. In addition to a flexible ligation step, NHEJ also addresses the challenge of substrate diversity with a cast of end processing factors that add or subtract nucleotides to generate more suitable ligation substrates[4].

1.6 End Processing During NHEJ

The repair of any single NHEJ substrate can result in a large number of different NHEJ products: heterogeneous substrates are repaired into even more heterogeneous products[11]. This product heterogeneity is intrinsically linked to the complexity of the original break, with breaks more amenable to ligation producing more homogenous

products (e.g. fully complementary overhangs). As the severity of flanking mispairs and damage increases, so do the number and frequency of different products recovered[11]. To some extent, this phenomenon reveals the limitations of LIG4 flexibility: some ends cannot be tolerated even by such a robust ligase and must instead be processed by nucleases and polymerases prior to joining.

The Artemis endonuclease is the only nuclease known to act specifically within NHEJ[33]. In fact, an Artemis fragment has been crystallized in complex with a fragment of LIG4. Since Artemis cleaves V(D)J recombination intermediates, patients with defects in this nuclease are severely immunodeficient[65]. Nonlymphoid cells derived from these patients are sensitive to ionizing radiation, suggesting a role for Artemis in NHEJ outside of V(D)J recombination[66]. Supporting this idea, a recent study showed *in vitro* that Artemis cleaves flaps from 3' overhangs during NHEJ[67].

In addition to specialized nucleases, NHEJ has evolved the ability to utilize 3 X-family DNA polymerases to synthesize nucleotides onto 3' termini of DSB ends in preparation for ligation[68]. DNA polymerases λ (pol λ), μ (pol μ), and TdT all function in this capacity and differ substantially from other polymerases. While replicative polymerases are characterized by high fidelity and processivity, NHEJ polymerases lack proofreading activity and rarely synthesize more than a few nucleotides[69]. Moreover, all three NHEJ polymerases have N-terminal domains that facilitate their interactions with the NHEJ core factors[69]. In addition to these differences, the NHEJ polymerases possess the unique ability to incorporate nucleotides onto noncanonical polymerase substrates[69,70]. Compared with pol λ, pol μ and TdT exhibit activity on an even wider spectrum of substrates and show no requirement for a paired primer terminus[71].

1.7 DNA Polymerase μ and Terminal deoxynucleotidyl Transferase

Pol μ and TdT possess noteworthy biochemical properties that render them completely unique among nucleic acid polymerases. For instance, pol μ is the only eukaryotic polymerase that enables rejoining of non-complementary, broken DNA ends absent a paired primer terminus[72]. The unique ability of pol μ to function in this context is evidently dependent upon loop 1, a conserved structural feature unique to the NHEJ polymerases[73]. Additionally, pol μ carries out repair using a distinctive "skip ahead" mechanism, wherein the polymerase fills gaps larger than one nucleotide by synthesizing only one nucleotide immediately adjacent the 5' DSB terminus[74]. Consistent with this observation, the biological role of pol μ seems to be the incorporation of single nucleotides almost exclusively[72].

In contrast with pol μ and pol λ, TdT is only expressed in lymphoid tissue and participates exclusively in V(D)J recombination[75,76]. Like pol μ, TdT's possible primer and template exhibit a remarkable level of flexibility[71]. Unlike all other DNA polymerases, TdT synthesizes onto fully single stranded overhangs and is thus fully template-independent[71,77]. In accordance with its role in introducing diversity during V(D)J recombination, TdT will polymerize multiple nucleotides onto its primer[78]. Through their unique biochemical activities, pol μ and TdT promote immune system diversity and cellular radioresistance[34,79–82]. Investigations into the biochemical properties of these polymerases have therefore focused on explaining these biological roles.

Perhaps the most intriguing and poorly understood biochemical property of pol μ and TdT is their relatively low sugar selectivity *in vitro*. Generally, DNA polymerases

discriminate against incorporating ribonucleotides into the genome, with at least a 1,000-fold preference for deoxynucleotides[83–85]. This discrimination is vital because genomic ribonucleotides are toxic and the pathway that removes them is essential in mammals[86]. Surprisingly, TdT and pol μ demonstrate sugar selectivity 200-fold lower than other polymerases, and only prefer deoxynucleotides about 5-fold more than ribonucleotides[37,87–90]. It is unclear if these polymerases incorporate ribonucleotides into the genome in cells or if their lack of sugar selectivity serves any biological role whatsoever[88]. An investigation of this phenomenon is the subject of Chapter 3.

1.8 Summary

The NHEJ pathway must successfully repair a wide range of chromosome break substrates to promote genome stability. The DNA ligase and DNA polymerases employed by this pathway are the most flexible enzymes in their respective classes, in terms of their ability to act on unorthodox substrates[5]. The mechanisms underlying this flexibility are poorly understood and it is unclear if they are important for repair in cells. My work has focused on answering these questions, initially focusing on the ligation step (Chapter 2) and then on the interaction between polymerases and the ligation step (Chapter 3).

CHAPTER 2: MECHANISM OF LIGATION FLEXIBILITY IN NHEJ

2.1 Introduction

DNA double-strand breaks (DSBs) are genomic lesions that play an important role in human health and disease. They are frequently generated by exogenous damaging agents (e.g. ionizing radiation) or as programmed intermediates in meiosis and V(D)J recombination[15]. The ends generated by these biological sources of chromosome breaks are often "complex," with DNA helix-distorting nucleotide damage, mismatches, or chemical adducts that pose challenges to the ligases and polymerases needed for DSB repair[91]. This problem is especially relevant to the nonhomologous end joining (NHEJ) pathway since, unlike other DSB repair pathways, these complex ends are not extensively resected prior to synthetic steps (polymerase and ligase activity).

Ligation is the only essential step in NHEJ, and is performed by one of the three mammalian ligases, DNA ligase IV (LIG4)[52]. LIG4 is recruited to broken ends through participation in a complex of core NHEJ factors including XRCC4, the Ku 70/80 heterodimer (Ku)[44], and XLF. This NHEJ core complex is sufficient to physically link a pair of broken ends together, and can thus be termed the paired end complex, or PEC. The PEC is essential for repair of diverse end structures; for example, XLF is required both for stable PEC formation[51] and ligation of complex ends, but only modestly affects ligation of ends with complementary termini[59,61,92]. Recent physical analyses of PECs indicate that they are highly dynamic[51] and that both the flexibility and stability of PECs

11

can be modulated by ligation-compatible DNA end chemistry[93]. However, it is unclear how differences in end structure trigger these changes in dynamics, and whether these changes in dynamics impact cellular repair.

Here we address this problem by assessing the impact of diverse end structures on in vitro functional assays, single-molecule analyses of end-pairing dynamics, and cellular repair and survival. We show that mismatches near strand break termini trigger extensive PEC remodeling. Moreover, a separation-of-function mutation in LIG4 links this mispair-induced PEC remodeling to the sensing of these end structures by LIG4, and argues PEC remodeling is essential to the proficiency of cellular NHEJ in repairing these end structures.

2.2 Methods

DSB Substrates

DSB substrates were made by ligating the 15-30 bp double stranded oligonucleotide "caps" described in Table 2.1 to a 285-bp PCR-generated common DNA "core" segment that had been digested with BsaI to generate appropriate sticky ends. Substrates were purified with the QiaQuick PCR purification kit (Qiagen), 5' phosphorylated with T4 polynucleotide kinase (NEB), and substrate assembly validated by gel electrophoresis.

DNA Constructs and Protein Purification

Constructs for expression after baculovirus delivery of human Ku, XLF, XRCC4, and LIG4^{WT} into Hi-5 insect cells have been previously described[71,94]. LIG4$^{\Delta i}$ was generated by modifying LIG4^{WT} as noted in Figure 2.1A and validated by sequencing.

The LIG^{3+4} chimera was generated by replacing amino acids 1-638 of LIG4^{WT} with a fusion of amino acids 170-862 of human LIG3 to the linker (GGGGS)x3 (Genewiz). Cell pellets were extracted, lysed by sonification, and purified by sequential chromatography on Histrap and MonoQ columns (GE Biosciences). Figure 2.1A structures were prepared in Pymol and include hLIG1 (1X9N)[56], hLIG3 (3L2P)[55], and hLIG4 (3W1B)[95], with disordered insert1 modeled by the SWISSmodel server[96].

in vitro Joining Assays

NHEJ reactions were initiated by incubating 2 nM DSB substrates, 25 nM Ku, 40 nM XLF, and 40 nM XRCC4-LIG4 in a buffer with 25 mM Tris pH 7.5, 100 uM EDTA, 1 mM DTT, 5 mM MgCl, 100 uM ATP, 150 mM KCl, 8.5% polyethelene glycol 3000, and 100 ng of supercoiled plasmid DNA. Reactions were carried out for 10 minutes at 37 C and stopped with 0.1% SDS and 20 mM EDTA. Repair products were purified by phenol-chlorform extraction and recovery was measured by real-time PCR (qPCR) using a QuantStudio 6 system (Applied Biosciences), primers that amplify head-to-tail junctions (Table 2.1), and VeriQuest SYBR Green master mix (Affymetrix). The relative numbers of molecules recovered were quantified by a well characterized qPCR assay[11,72].

For nick sealing assays, a 5'Cy5 labeled, nicked 41 bp substrate was generated by annealing three oligonucleotides (Table 2.1). 5 nM substrate was incubated with XRCC4-LIG4 at 37 C and products were characterized by denaturing polyacrylamide gel electrophoresis (PAGE). Wild-type XRCC4-LIG4 was titrated to determine that 0.5 nM ligase (1:10 enzyme:substrate) generates sub-saturating (19-21%) amounts of nick sealing in 10 minutes, after which reactions were carried out under these conditions in

triplicate to generate data presented in Figure 2.2B. Reaction velocity was determined by quantifying band intensities using ImageJ software.

Electrophoretic Mobility Shift Assay (EMSA)

Substrates for EMSA were generated by annealing oligonucleotides (Table 2.1) to produce a Cy5 labeled 15 bp substrate to assess DSB end binding (Figure 2.2C), as well as a Cy5 labeled, 60 bp substrate to assess complex formation (Figure 2.2D). To assess intrinsic end binding, the 15 bp substrate was incubated at 10 nM with 125, 250, or 500 nM XRCC4-LIG4. For complex formation, the 60 bp substrate was incubated at 10 nM with 2 nM Ku, 40 nM XLF, and 40 nM XRCC4-LIG4. These samples were incubated for 20 min on ice in EMSA buffer (50 mM NaCl, 75 mM KCl, 25 mM Tris pH 8, 13% Glycerol, and 0.015% X100). Samples were run on 4% (Figure 2.2C) or 6% (Figure 2.2D) polyacrylamide gels in 0.5x TBE buffer and imaged using a Typhoon (GE).

smFRET Assays

smFRET assays and analysis were performed as described previously[51,93]. Briefly, NHEJ reactions composed of 50 nM Ku, LX, XLF gloxy (0.5 mg/mL glucose oxidase and 0.4 μg/mL catylase), and 1 nM dsDNA were added stepwise to NEB4 (20 mM pH 7.5 TricAc, 50 mM KAc, 10 mM MgAc) supplemented with 0.8% glucose, ~5 mM Trolox, 1 mg/mL BSA, and 2 mM DTT. The reaction was immediately flowed into an imaging chamber that had been prepared with surface dsDNA (~250 pM). Movies consisting of 1000 frames (33Hz) were acquired for analysis of PECs. Trajectory analysis, histogram assembly, and autocorrelation of PECs were performed in Matlab[93]. Oligonucleotides used in smFRET experiments are detailed in Table 2.1.

Cell Lines

LIG4[-/-] cells were generated from parental HCT116 human colorectal cancer cells by conventional gene targeting and were the gift of Dr. Eric Hendrickson[97]. We generated additional variants of the parental cells by CRISPR/Cas9 gene targeting. We introduced by electroporation plasmids to express Cas9[98] (Addgene 44758; 5 µg) and an sgRNA[99] (Addgene 51133; 5 µg; guide sequence described in Table 2.1) that targets insert1-encoding sequence from wild-type *LIG4*, as well as a gene-targeting donor plasmid. The donor plasmid was engineered such that it contains 1.1kb of sequence identical to the *LIG4* gene except as modified such that gene targeting ablates the sgRNA target site, generates the *LIG4[Δi]* mutation as described in Figure 2.1A, and introduces synonymous mutations that result in a BsmFI site used for screening. The native *LIG4* sequence in this region and resulting *LIG4[Δi]* alleles are described in Figure 2.6 and Table 2.1. Targeted puromycin resistant clones were identified by amplification of the insert1 region using primers specific to the native locus (i.e. originate outside of donor sequence identity). Two independently generated clones, *LIG4[Δi/Δi]* a and b, were produced that possessed only targeted alleles after sequencing (Figure 2.6A). To generate *LIG4[+r/+r]* reverted cells we repeated gene targeting, but started with *LIG4[Δi/Δi]* a cells and used an sgRNA specific for the *LIG4[Δi]* allele (Table 2.1) as well as a gene-targeting donor with wild-type sequence in this region.

We verified LIG4 expression in all of these cell lines using standard western blot techniques (Figure 2.6B) and antibodies against human LIG4 (Serotec cat no. AHP554) and human Ku70 (Abcam cat no. ab62820). All 5 cell lines were cultured in McCoy's 5A medium (Corning) with 10% fetal bovine serum (Sigma) and determined to be free of

15

mycoplasma contamination by PCR[100]; we additionally employed a third party to validate the absence of mycoplasma by an alternate method for a randomly selected cell line (Hoechst staining)[101].

Cellular NHEJ Assays

Extrachromosomal DNA substrates described above (20 ng) were electroporated into 2×10^5 cells with pMAX-GFP plasmid (600 ng) at 1350 V in one 30 ms pulse in 10 µL (Neon, Invitrogen). Transfected cells were incubated for 30 minutes in antibiotic-free McCoy's 5A media with 10% fetal bovine serum. Cellular repair products were harvested using a QIAamp DNA mini kit (Qiagen). Each electroporation was reproduced in triplicate from 3 independent preparations of cells. Repair efficiency was quantified by qPCR as described above for *in vitro* joining assays.

Repair product structures were determined by restriction digest for the 8-oxoguanine (2-Amino-7,9-dihydro-1H-purine-6,8-dione; G_o) substrate, and by high-throughput sequencing for all other substrates. For the G_o substrate, harvested repair products were amplified with Cy5-labeled primers (Table 2.1) and digested with BstZ17I (New England Biolabs; recognizes transversion mutation after amplification of G_o) and BamHI (New England Biolabs; recognizes accurately amplified G_o) to identify directly ligated products. The intensities of digested and undigested bands were quantified using ImageJ.

To determine repair product structures of all other substrates, sequencing libraries were prepared by PCR amplification of repair products with primers containing 6-nucleotide indices on their 5' ends (Table 2.1). Amplified DNA (40 ng per library) was pooled into groups of 8-12 libraries, 5' phosphorylated, and treated with Klenow exo-

(NEB) to add dA to the 3' termini. Ends were ligated to adapters for paired-end sequencing (Illumina). Pooled libraries were purified from 3% agarose gels to remove unligated adapters using the QIAquick gel extraction kit (Qiagen). Recovered samples were amplified for an additional 9 cycles using enrichment primers (Illumina). Products were again purified using Ampure XP beads (Beckman Coulter). 27.27 ng from each of the 11 pools was combined (for a total of 300 ng of sample), supplemented with PhiX174 (40% final concentration) and submitted to the UNC high throughput sequencing facility for a 2 x 75-bp MiSeq run (Illumina). Genomics Workbench was used to remove PhiX174 DNA, merge read pairs, de-index libraries, and remove low quality sequences (CLC-Bio). Remaining sequences were analyzed using Microsoft Excel.

Colony Formation and Cell Growth Assays

For colony formation assays, seeding densities were determined independently for each dose and cell line such that 50-150 colonies would be produced per 10 cm dish. Cells were plated on 10 cm dishes in fresh McCoy's 5A media with 10% fetal bovine serum, incubated for 4 hours, and then irradiated with indicated doses of X-rays using a RS 2000 irradiator (Rad Source Technologies). Colonies formed after 14 days were stained with a solution of crystal violet (0.5%) and glutaraldehyde (6%). Colonies were manually counted on three plates per dose and cell line. The surviving fraction of *LIG4$^{-/-}$* cells treated with 3 Gy of X-rays was much less than 10^{-3}, and was excluded from analysis because the resulting faint, small colonies could not be reliably discriminated above background debris staining.

For live cell imaging, 2000 cells were plated into 96-well plates in triplicate for each dose and cell line. After overnight incubation, cells were irradiated or treated with etoposide and placed into the IncuCyte live cell imager (Essen biosciences). Four 215 mm^2 images were taken per well at 10x objective every 4 hours for a total of 120 hours. The confluence of each image was determined by generating a confluence mask with IncuCyte software (Essen biosciences).

Statistical Analysis

For all experiments, means were tested for significance against a control (e.g. LIG4^{WT}, *LIG4$^{+/+}$* cells) using two-tailed t-tests for single comparisons, one-way ANOVA for multiple comparisons, and two-way ANOVA for comparisons with multiple variables. Dunnett's correction for testing multiple hypotheses was applied as necessary. For each experiment, the value and definition of n, the representation of error bars, the specific tests used, the specific control tested, and the determination of statistical significance are described in the figure legends.

2.3 Results

LIG4 is specialized to directly ligate mismatched or damaged ends

Activity of all three mammalian ligases requires the encircling of double stranded DNA at a strand break[54]. Structural studies identified 6-10 amino acids inserted in LIG4 orthologs (residues 113-122 in human LIG4), relative to other eukaryotic ligases (Figure 2.1A; Figure 2.2A)[95]. This insert is located within the strand break-bound ligase on the opposite side of the double helix from the strand break and site of catalysis, suggesting a possible function specific to substrates with double helix-distorting mispairs or

damage. Consistent with this idea, we purified LIG4 with this element specifically

deleted (LIG4$^{\Delta i}$) in a complex with XRCC4. We determined that insert1 had no

significant impact on LIG4-XRCC4 intrinsic nick sealing activity (Figure 2.2B), DNA

binding (Figure 2.2C), or ability to form a higher-order complex with NHEJ core factors

Ku and XLF on DNA (Figure 2.2D). *In vitro* NHEJ activity was also similar comparing

LIG4$^{\Delta i}$ to LIG4^{WT} when ends had complementary overhangs (Figure 2.1B; 5' G:C, 3'

G:C). In contrast, when ends had mispairs or damage at strand-break termini, *in vitro*

NHEJ activity using LIG4$^{\Delta i}$ was reduced 21-66-fold relative to LIG4^{WT} (Figure 2.1B; 5'

G$_o$xC, 3' GxT, 3' GxA). LIG4$^{\Delta i}$ is thus specifically defective in supporting *in vitro* NHEJ

when substrates have helix-distorting 8-oxoguanine (Go) damage or mispairs near

strand termini.

      Ends with mispaired nucleotides are critical NHEJ substrates that arise during

V(D)J recombination and after nucleolytic processing of radiation-induced breaks. They

also presumably act as a model for ends with other sources of helical distortion,

including nucleotide damage. To validate this inference, we measured *in vitro* NHEJ of

ends with 8-oxoguanine (Figure 2.1B, 5' G$_o$xC), the most common form of oxidative

base damage. NHEJ activity on this substrate was reduced over 50-fold with LIG4$^{\Delta i}$,

which was comparable to the effect of a terminal G:A mispair. Therefore, insert1 is

required for direct ligation of end structures with flanking helical distortions, whether the

distortions are due to mispairs or nucleotide damage. To further explore the extent to

which ligation of ends with terminal mispairs or damage is specific to wild-type LIG4, we

generated a chimera (LIG^{3+4}) with all three LIG4 catalytic sub-domains replaced with the

equivalent sub-domains from mammalian wild-type LIG3 (Figure 2.2E). Like LIG4$^{\Delta i}$,

LIG^{3+4} physically associates with XRCC4 and was fully competent in Ku- and XLF-dependent ligation of ends with complementary overhangs. However, end joining with this chimera was even more sensitive than LIG4$^{\Delta i}$ to terminal nucleotide damage (activity reduced more than 100-fold, relative to LIG4^{WT}; Figure 2.2E). This result consistent with the argument that LIG4 is unique amongst mammalian ligases in its ability to repair damaged termini. Additionally, the impact of LIG4$^{\Delta i}$ on repair of damaged ends is less severe than that of the LIG^{3+4} chimera, suggesting that insert1 is not entirely responsible for the unique ability of LIG4 to tolerate mispairs and damage. We therefore sought to use the LIG4$^{\Delta i}$ separation-of-function mutation to investigate both the mechanistic basis for the unique ability of NHEJ to tolerate helix-distorting mispairs or damage at the ligation step, as well as its significance to cellular double strand break repair.

Dynamic re-alignment of mismatched ends is required for their ligation

We previously described a single-molecule fluorescent resonance energy transfer (smFRET) assay that reports on pairing of DNA ends as mediated by a complex of Ku, XRCC4, LIG4, and XLF[51,93]. These PECs are apparent as FRET pairs generated when a Cy3 labeled donor dsDNA fragment in solution stably associates with a Cy5 labeled acceptor dsDNA fragment immobilized on a surface (Figure 2.3A). In accord with the *in vitro* ligation assay, LIG4$^{\Delta i}$ and LIG4^{WT} similarly promote stable PEC formation when ends have complementary 4-nucleotide overhangs (G:C, Figure 2.3B). In contrast, pairing of overhangs with 3' terminal G:T mismatches is significantly reduced when comparing LIG4$^{\Delta i}$ to LIG4^{WT}; this reduced pairing efficiency represents a diminished proportion of DNA ends associated in the PEC. The formation of PECs is

even less efficient when termini have a bulkier purine:purine G:A mismatch, but is similarly inefficient for both LIG4^{WT} and LIG4$^{\Delta i}$ (Figure 2.3B). PECs thus form less efficiently with increasing terminal helical distortion, and PECs formed with LIG4$^{\Delta i}$ are more sensitive to this challenge.

Changes in FRET efficiency (E$_{FRET}$) reflect dynamic repositioning of DNA ends relative to each other within individual PECs[51,93]. When using complementary ends (3' G:C; Figure 2.3C), E$_{FRET}$ distributions were not significantly different when comparing PECs formed with LIG4^{WT} (black line) vs. LIG4$^{\Delta i}$ (orange line). LIG4$^{\Delta i}$ PECs also had similar FRET distributions when ends had terminal mispairs (Figure 2.3C); importantly, LIG4^{WT} PECs formed on ends with mispaired termini more often had lower E$_{FRET}$ (DNA labels located further apart; black lines for G:A and G:T mispairs, Figure 2.3C), and consequently overall wider distributions of E$_{FRET}$ (Figure 2.3C; Figure 2.4A) when compared to paired temini (G:C). Ends with terminal distortions thus trigger PECs to sample a wider variety of end-alignment configurations to remain efficiently paired, but only when using LIG4^{WT}.

Examination of smFRET trajectories from individual PECs also shows the transition frequency between FRET states increases when comparing LIG4^{WT} and LIG4$^{\Delta i}$ (Figure 2.3D, Figure 2.4C). We quantified this difference by using autocorrelation of individual FRET trajectories to calculate the average transition times - "lag times" (τ) - between FRET states[93]. For ends with G:T mismatches, these were approximately two-fold lower for PECs formed with LIG4^{WT}, compared to PECs formed with LIG4$^{\Delta i}$ (Figure 2.4B). These values are then used to calculate the relative stability of the DNA ends in the PECs (Figure 2.4C). From these calculated end stabilities we can infer that LIG4^{WT}

PECs have lower energetic barriers in assuming new conformations, compared to

LIG4$^{\Delta i}$ PECs, but again only when ends have terminal mismatches.

PECs containing LIG4$^{\Delta i}$ and mispaired ends are thus formed less efficiently

(Figure 2.3B), and even when formed do not acquire the high degree of conformational

plasticity observed when PECs are formed with LIG4^{WT} (Figure 2.3C-D, Figure 2.4D-E).

We argue the inability of LIG4$^{\Delta i}$ to allow for mispair-induced PEC remodeling accounts

for its specific defect in direct ligation of such end structures (Figure 2.1B). There are

also limits to the extent to which remodeling enables ligation, as even LIG4^{WT} is

inefficient in joining ends with bulky G:A mismatches (Figure 2.1B). PECs formed with

paired termini favor a narrow distribution of high FRET end alignments that more closely

resemble FRET distributions observed with products of ligation[93]; these alignments thus

likely directly juxtapose strand-break termini in anticipation of catalytic steps ("pre-

catalytic", Figure 2.3E). We attribute the LIG4^{WT}-specific, insert1-dependent flexibility in

accommodation of mispaired termini to a favoring of end alignments that both have

lower FRET (more distally-located labels) and are more dynamic. These more dynamic

and lower FRET PECs – "remodeling PECS" - may be catalytically incompetent, but

allow for iterative attempts at the now transient (but occasionally catalytically

competent) high-FRET intermediate (Figure 2.3E).

Cellular NHEJ of complex ends requires remodeling of the PEC

We next addressed whether the differences in PEC flexibility described above

significantly impact cellular NHEJ. We employed scar-free gene targeting to exchange

LIG4^{WT} for *LIG4$^{\Delta i}$* alleles within the native *LIG4* locus of a human cell line (Figure 2.5A).

We independently generated two such cell lines (*LIG4$^{\Delta i/\Delta i}$* a and *LIG4$^{\Delta i/\Delta i}$* b), and

confirmed they express only LIG4$^{\Delta i}$ from endogenous loci (Methods; Figure 2.6A-B). We

then generated a cell line by another round of gene targeting where the *LIG4* locus of

LIG4$^{\Delta i/\Delta i}$ a was reverted back to wild type sequence (*LIG4$^{+r/+r}$*), as a means of assessing

the effects of possible off-target mutations incurred in the original round of gene

targeting (Figure 2.5A, Figure 2.6A-B). Both *LIG4$^{\Delta i/\Delta i}$* subclones acted equivalently in

functional assays below. Similarly, results using parental wild type cells (*LIG4$^{+/+}$*)

matched those from the *LIG4$^{+r/+r}$* reversion, confirming the differences observed in the

LIG4$^{\Delta i/\Delta i}$ cells could be attributed to the 8 amino acids deletion.

DSB substrates with varied end structures were introduced into these cells, after

which efficiency of repair was determined by qPCR and product structures were

characterized by high-throughput sequencing. In accord with *in vitro* results, ends with

complementary overhangs were efficiently joined almost entirely by direct ligation in

both wild-type and *LIG4$^{\Delta i/\Delta i}$* cells (Figure 2.5B). Also in accord with *in vitro* data, ends

with terminal G:T mispairs were efficiently repaired by direct ligation (accounts for 60%

of all repair) in both *LIG4$^{+/+}$* and *LIG4$^{+r/+r}$* cells, while this class of product is rarely

(<10%) seen in *LIG4$^{\Delta i/\Delta i}$* clones (Figure 2.5C). Instead, repair in *LIG4$^{\Delta i/\Delta i}$* cells typically

requires re-alignment of overhangs and gap-repair synthesis prior to ligation of the now

"sticky" end. This alternate pathway is sufficient to fully compensate for the inability of

LIG4$^{\Delta i}$ to directly ligate terminal mispairs, since overall joining efficiency was

comparable for *LIG4$^{+/+}$* vs. *LIG4$^{\Delta i/\Delta i}$* cells. Considering repair of ends with bulkier G:A

mispairs, both wild-type and *LIG4$^{\Delta i/\Delta i}$* cells rely on this compensating pathway (Figure

2.6C), consistent with *in vitro* observations that neither LIG4^{WT} nor LIG4$^{\Delta i}$ can ligate this

substrate in the absence of end processing. Repair of NHEJ substrates was severely

reduced in *LIG4^{-/-}* cells (Figure 2.5C; approximately 0.0005 products per cell), to the extent that we could not recover sufficient repair products to accurately assess product spectra.

Additional substrates were introduced into cells to assess whether barriers to mispair tolerance are routinely bypassed by cellular end processing. Similar to 3' G:T mispairs, ends with 5' G_o terminal damage are primarily repaired by direct ligation in *LIG4^{+/+}* cells. Importantly, joining of 5' G_oxC in *LIG4^{Δi/Δi}* cells is over 10-fold less efficient (Figure 2.5D), even though what little repair does occur is processing-dependent (Figure 2.6D). We also investigated cellular NHEJ of end structures with entirely non-complementary overhangs (TTTT). Joining efficiency was again severely reduced in *LIG4^{Δi/Δi}* cells, relative to wild-type cells (Figure 2.5D). For this substrate, the rare products recovered from *LIG4^{Δi/Δi}* cells only subtly differed from wild-type controls in terms of junction structure (Figure 2.6E). Thus, in contrast to previously tested substrates (Figure 2.5C), end processing was not sufficient to rescue repair of TTTT and 8-oxoguanine substrates in *LIG4^{Δi/Δi}* cells. We initially linked LIG4^{WT} PEC flexibility only to the ability of cellular NHEJ to directly ligate ends with terminal mispairs (Figures 2.1, 2.3, 2.5C); these latter results identify additional important contributions to cellular NHEJ associated with end processing.

PEC remodeling guides end processing choice during cellular NHEJ

We generated the substrate "EC1" (embedded complementarity 1) to further explore the relationship between PEC flexibility, cellular end processing, and ligation. EC1 has long (10-nt) non-complementary overhangs that can plausibly be aligned to juxtapose mispaired 3'OH:5'P termini, in anticipation of direct ligation (Figure 2.7A).

Alternatively, EC1 can be re-aligned to pair complementary sequence embedded within the overhang, where unpaired tails are a presumptive substrate for nucleolytic end processing. These two alignments are readily distinguished by smFRET (Figure 2.8A); PECs formed with a control substrate (with fully complementary 10nt overhangs) had low E_{FRET} ranges, expected for EC1 alignments that juxtapose 3'OH:5'P termini (green lines), while PECs formed with 4-nt complementary overhangs had a clearly distinct population of high E_{FRET} (blue lines) expected for EC1 alignments that pair embedded complementary sequence. Analysis of individual smFRET trajectories of PECs formed with EC1 and NHEJ core factors identified a much larger than typical fraction of transient complexes (lifetimes <5 seconds; Figure 2.7B). Transient PECs had two distinct populations of E_{FRET} distributions, each roughly corresponding to the two alignment classes predicted above (Figure 2.7C). Long-lived PECs (persistent) favor only the high E_{FRET} state, but sample both a wider range of alignment configurations (Figure 2.7C) and are more dynamic (have lower energetic barriers to transition; Figure 2.8B) than PECs formed with complementary overhangs. To further address if the persistent PECs frequently involve pairing at embedded complementary sequence (as suggested by comparison to substrate standards; Figure 2.7C) we used a substrate where the complementary sequence was both reduced and re-located ("EC2"). As expected, PECs formed less efficiently with EC2 (Figure 2.8C), and when formed had mostly lower E_{FRET} (Figure 2.8D). Importantly, LIG4$^{\Delta i}$ was largely unable to form PECs with the EC1 substrate (Figure 2.7D), and the rare PECs that do form primarily have intermediate E_{FRET} states that are inconsistent with either alignment (Figure 2.8E). Therefore, only LIG4^{WT} effectively promotes end-pairing of this substrate. Moreover,

PECs formed with LIG4^{WT} that juxtapose strand termini were "filtered out"; only the most plausibly productive alignments, i.e. those that could lead to ligation after nucleolytic end processing, were stable (Figure 2.7C).

We next assessed how the EC1 substrate was resolved by cellular NHEJ. Nearly all products (>99%) were indeed consistent with ligation after nuclease activity, with the dominant product guided by the alignment at embedded complementary sequence also favored in smFRET analysis (Figure 2.8F). By comparison, direct ligation of EC1 accounted for less than 0.1% of all cellular repairs. Importantly, joining efficiency of this substrate was reduced over 60-fold in $LIG4^{\Delta i/\Delta i}$ cells, relative to $LIG4^{+/+}$ cells (Figure 2.7E), even though $LIG4^{\Delta i/\Delta i}$ cells are fully proficient at ligating the inferred product of alignment-guided nuclease activity (a 4bp complementary overhang; Figure 2.5B). This result suggests that for this substrate, LIG4$^{\Delta i}$ fails to efficiently mediate repair because it is defective at an earlier step than ligation – specifically, stable accommodation of end-alignments required for nucleolytic end processing (Figure 2.7D).

Cellular radioresistance requires tolerance of complex ends by LIG4

LIG4^{WT} thus uniquely accommodates diverse end structures during end pairing. However, there is wide variation in how this flexibility impacts cellular NHEJ. Depending on the starting end structure it can be dispensable (Figure 2.5B, Figure 2.6C), alter product spectra (Figure 2.5C), or can be critical for efficient repair (Figure 2.5D-E, Figure 2.7E). We therefore addressed the extent to which the inability of LIG4 to tolerate structural diversity impacts cell growth and survival after ionizing radiation. Using both standard colony forming assays and real-time imaging of cell growth, $LIG4^{+/+}$ and $LIG4^{+r/+r}$ cells were similarly resistant to increasing dose of ionizing radiation. By

comparison, *LIG4^{Δi/Δi}* cells were radiosensitive to a degree intermediate between *LIG4^{+/+}*

and *LIG4^{-/-}* cells (Figure 2.9, Figure 2.10), a result strikingly similar to joining efficiencies

described for the majority of substrates with complex ends (Figure 2.5D-E, Figure 2.7E).

In contrast with ionizing radiation, *LIG4^{Δi/Δi}* and *LIG4^{+/+}* cells are equally resistant to

etoposide (Figure 2.10). This is consistent with specific requirement for insert1 in repair

of ends with mispairs or damage, since etoposide induced breaks can be processed by

tyrosine phosphodiesterase 2 such that overhangs are undamaged and fully

complementary[102]. These results show that the ability of LIG4 to sense distortions

facilitates cell survival following treatment with ionizing radiation.

2.4 Discussion

Repair by NHEJ implicitly requires the pairing together of broken chromosome

ends. A complex of Ku, XRCC4, DNA Ligase IV, and XLF (paired end complex, or PEC)

is necessary and sufficient for this purpose[51]; we describe here dynamic changes in this

complex that are triggered by differences in end structure, and show that this response

is essential for efficient cellular repair.

Mechanistic basis for repair of complex ends by NHEJ

Ends with complementary ("sticky") overhangs are aligned efficiently and with

little mobility, to the extent that pairing E_{FRET} more closely resemble the narrow

distributions observed in products of ligation, relative to other end structures tested

here. We suggest these PECs describe "pre-catalytic" end alignments, where strand

break termini are directly juxtaposed in anticipation of ligase-mediated catalytic steps

(Figure 2.11). In contrast, ends with helix distorting mispairs or damage near strand

termini – complex ends - induce the sampling of a much wider variety of alignment configurations, most or all of which no longer juxtapose strand termini.

We use a LIG4 separation of function mutation (LIG4$^{\Delta i}$) to identify an essential role for this second, more dynamic "remodeling" class of PECs in cellular NHEJ for the repair of complex ends. We show LIG4$^{\Delta i}$ is specifically unable to accommodate PEC remodeling in response to complex ends. As a consequence, PECs formed with LIG4$^{\Delta i}$ are unable to directly ligate such substrates, but are also – with rare exception (Figure 2.5C) – unable to couple ligation to end processing when end complexity is sufficient to block direct ligation.

Notably, the exceptions are restricted to contexts where alignment-directed synthesis generates a fully complementary 6-nt overhang, a substrate expected to be especially permissive for the ligation step.

By comparison, insert1 is dispensable for the XLF-, XRCC4-, and Ku-dependent alignment of ends with complementary overhangs, as well as catalytic activity on this conventional ligase substrate. Moreover, a chimeric ligase with all three LIG4 catalytic subdomains replaced with LIG3 counterparts is equally effective in ligation of "sticky" ends (and is similarly stimulated by Ku and XLF), but is even less able to repair complex ends. Prior work emphasized the importance of a variety of NHEJ proteins, including PAXX[103,104], end processing factors[33,67,72], and especially the end-bridging filament of XRCC4-XLF[59,105,106] in repairing complex breaks. Indeed, we previously reported that these filaments form on bleomycin-induced DSBs and orchestrate their repair[51]. Here we identify a critical role for specialization of LIG4 catalytic subdomains in repair of complex ends, and show this role is attributable to insert1-dependent PEC remodeling.

How does insert1 contribute to PEC remodeling? The three subdomains of eukaryotic ligases are extended in the absence of DNA ("open" conformation), and engage substrates by forming a ring around double stranded DNA ("closed" conformation)[55–57]. In the closed conformation, the central catalytic subdomain is bound to strand break termini while insert1 is located in the N-terminal subdomain on the opposite side of the double helix[95]. Though not resolved in current apo-enzyme crystal structures, its location suggests that insert1 helps LIG4 maintain a closed conformation, either by stabilizing the ring-closing interactions between N and C terminal catalytic subdomains or by interacting with DNA[107]. We suggest stable end-pairing is dependent on LIG4 maintaining a closed conformation, even if LIG4 can directly interact with only the 5' phosphate side of a strand break ("half-site" binding). LIG4$^{\Delta i}$ instead transitions to an open conformation in this context (like conventional ligases), which leads to failure of end pairing.

Significance of LIG4 sensing complex ends

Prior work indicates that LIG4 has functions in NHEJ distinct from the ligation step, most clearly in promoting end pairing[51,93,108–110]. Data presented here identify a much more sophisticated function. Differences in how LIG4 catalytic domains interact with different end structures trigger dramatic changes in the dynamics of the entire paired-end complex – i.e. including Ku, XRCC4, and XLF paired ends – and these altered dynamics determine the steps taken to complete repair. This role is distinct from critical LIG4 roles in catalysis and end pairing, since both of the latter functions are fully intact in PECs formed with LIG4$^{\Delta i}$. LIG4 can thus be identified as the PEC "sensor," helping tailor the path to repair as is appropriate to end structure, possibly even to the

extent that how LIG4 interacts with aligned ends may dictate the identity of the end processing factor that next engages the end.

Inhibitors of LIG4 are being explored for their potential to sensitize tumors to radiation therapy[111]. Here we identify a role of LIG4 that is specific to the ability of cells to repair complex damage, identify a structural element required for this role, then show deletion of this element leads to cellular sensitivity to ionizing radiation. Since this structural element is unique to LIG4 and required for radioresistance, it presents a promising therapeutic target, as it is less likely to engender the off-target effects observed with current LIG4 inhibitors[112].
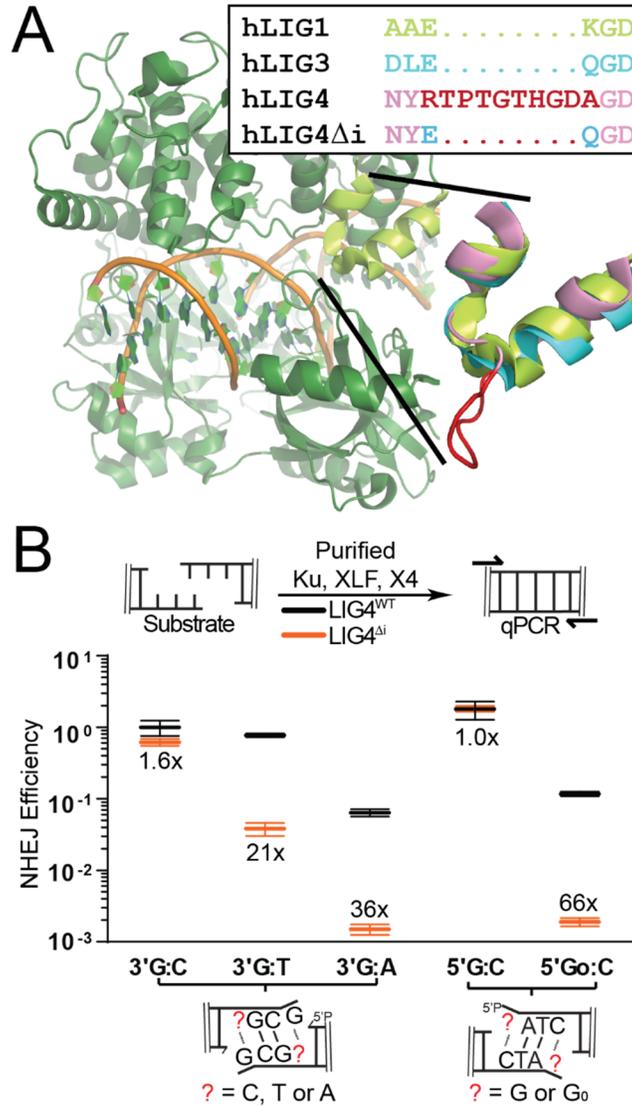
Figure 2.1: Effect of LIG4 insert1 on NHEJ of complex ends *in vitro*

(A) Structure of human LIG1 bound to DNA (1X9N; green), with inset emphasizing sequence and structural alignments of human LIG1 α helices 5-6 to human LIG3 (3L2P; blue) and human LIG4 (based on 3W1B; pink), with sequence and a modeled location of LIG4 insert1 (disordered in the 3W1B apoenzyme) in red. **(B)** Ku, XLF, and either XRCC4-LIG4^{WT} (gray) or XRCC4-LIG4$^{\Delta i}$ (orange) were incubated with substrates containing different complementary (5' G:C, 3' G:C) or non-complementary (5' GoxC, 3' GxT, 3' GxA) overhangs as noted. Joining efficiency is expressed as a fraction of the total junctions recovered using the 5' G:C substrate with NHEJ reactions containing LIG4^{WT}. Ligation reactions were performed in triplicate and the mean joining efficiencies are shown, along with the fold difference between LIG4^{WT} and LIG4$^{\Delta i}$ for each substrate. Error bars represent the range of observed values for each set of experiments.

A Species AA LIG4　　insert1

H. sapiens	105	DALKLLNY	RTPTGTHGDA	--GDFAMIAYF-V
M. musculus	105	DAQKLLNY	RTPSGARTDA	--GDFAMIAYF-V
X. laevis	110	DAIKLLNY	RSPTSSHSDA	--GDFAAIAYF-V
D. rerio	110	EANKLLNY	RAPTTSQGEA	--GDFALMAYF-V
A. thaliana	84	DAVRLLNW	RKGGTAKAGANA	GNFSLIAAE-V
S. cerevisiae	108	TEQRLKDW	KQRVG-K-G	---GNLSSLLVEEI
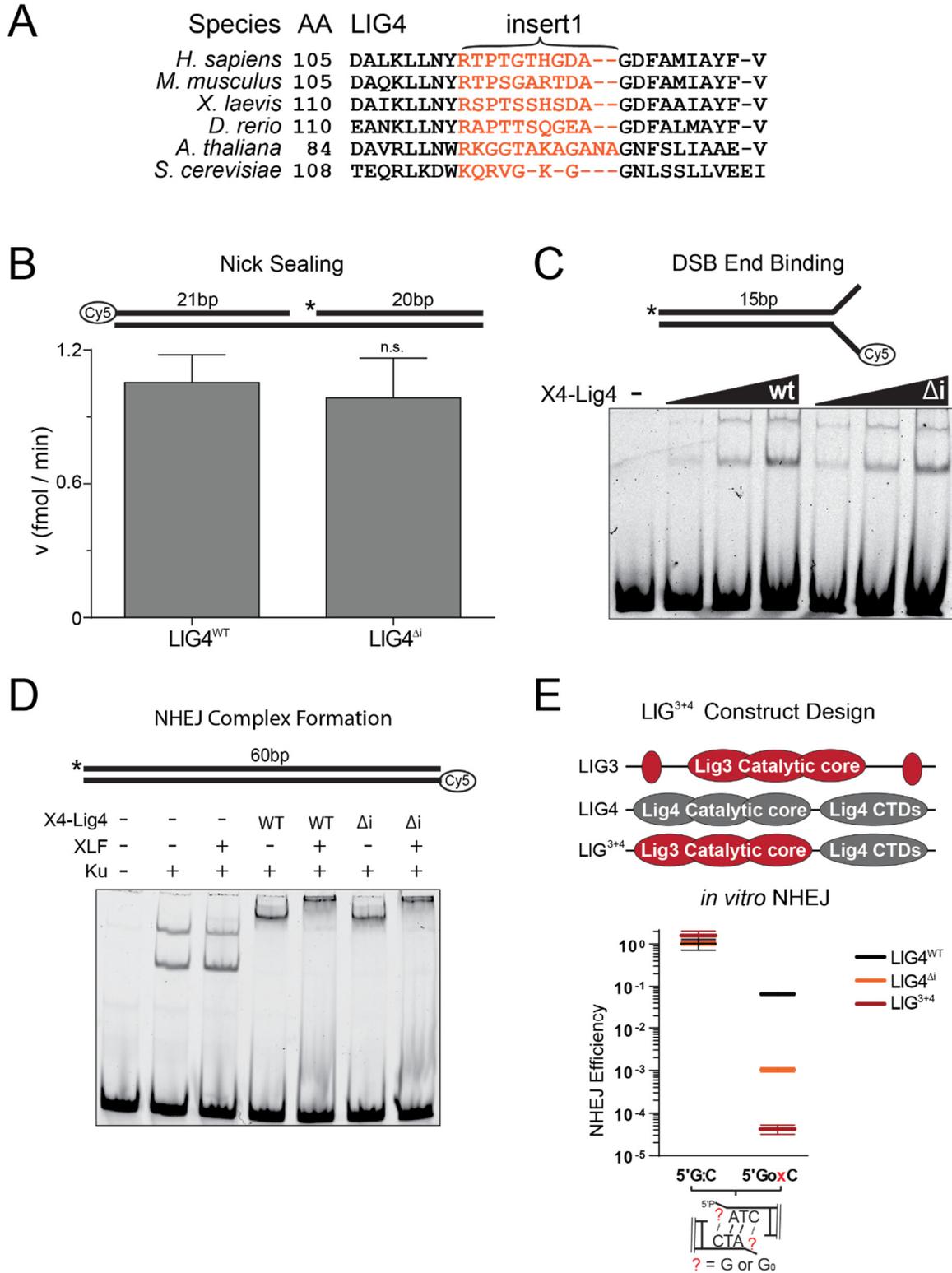
Figure 2.2: Biochemical Characterization of LIG4 variants

Figure 2.2: Biochemical Characterization of LIG4 variants

(A) Amino acid sequence alignment of LIG4 orthologs, with insert1 in orange. **(B)** A nicked, Cy5-labeled 41 bp substrate was incubated with XRCC4-LIG4^{WT} or XRCC4-LIG4$^{\Delta i}$ in triplicate and joining was assessed by stand-denaturing gel electrophoresis. Error bars represent standard error of the mean for 3 experiments. The mean for LIG4$^{\Delta i}$ was assessed by t-test as not statistically significantly different (ns) from control (LIG4^{WT}). **(C)** A 15 bp Cy5-labeled substrate was incubated with XRCC4-LIG4^{WT} or XRCC4-LIG4$^{\Delta i}$ and substrate binding was assessed by native gel electrophoresis. **(D)** A 60 bp Cy5 labeled substrate was incubated with indicated NHEJ factors and NHEJ complex formation was assessed by native gel electrophoresis. **(E)** LIG^{3+4} chimera was generated by fusing catalytic domains of LIG3 with C-terminal domains of LIG4 and purified after co-expression with XRCC4. NHEJ reactions were performed *in vitro* as in Figure 2.1B using undamaged (5' G:C) and damaged (5' GoxC) substrates. Joining efficiency is expressed as a fraction of the total junctions recovered using the 5' G:C substrate with LIG4^{WT}. Ligation reactions were performed in triplicate and error bars represent the range of observed data points.
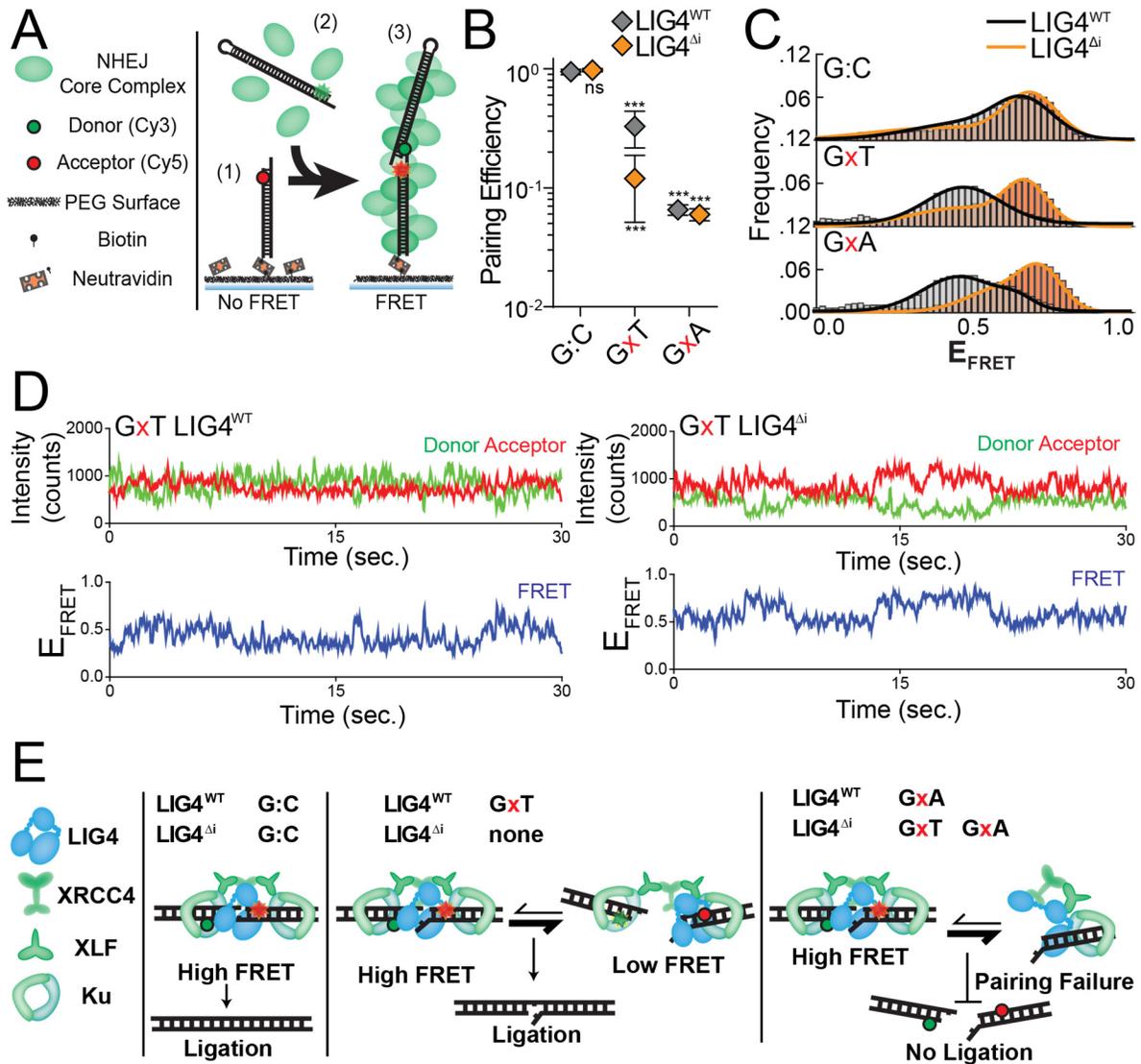
Figure 2.3: Effect of complex end structures on pairing dynamics of single molecule complexes with LIG4^{WT} or LIG4$^{\Delta i}$

(A) smFRET NHEJ assay: (1) dsDNA with a Cy5 acceptor is tethered to a bitonylated PEG surface via a biotin-neutravadin linkage, (2) dsDNA with a Cy3 donor and NHEJ proteins (green) are added to the chamber, and (3) ends are paired and FRET is observed. **(B)** Quantitation of pairing efficiency of ends with complementary (G:C) or mismatched (GxT, GxA) overhangs by Ku, XLF, XRCC4 and either LIG4^{WT} (gray) or LIG4$^{\Delta i}$ (orange). Error bars represent standard error of the mean for 3 experiments. Means were assessed by two-way ANOVA as significantly different from control (LIG4^{WT} on G:C substrate) with confidence $p<0.001$ (***). **(C)** Histograms of observed E$_{FRET}$ for PECs formed as in (B). **(D)** Representative smFRET trajectory for LIG4^{WT} and LIG4$^{\Delta i}$ PECs formed with GxT ends demonstrating altered transition frequency and FRET states **(E)** LIG4^{WT} enables PECs to oscillate between high and low E$_{FRET}$ states in response to distortions, and this flexibility is essential for joining distorted breaks.
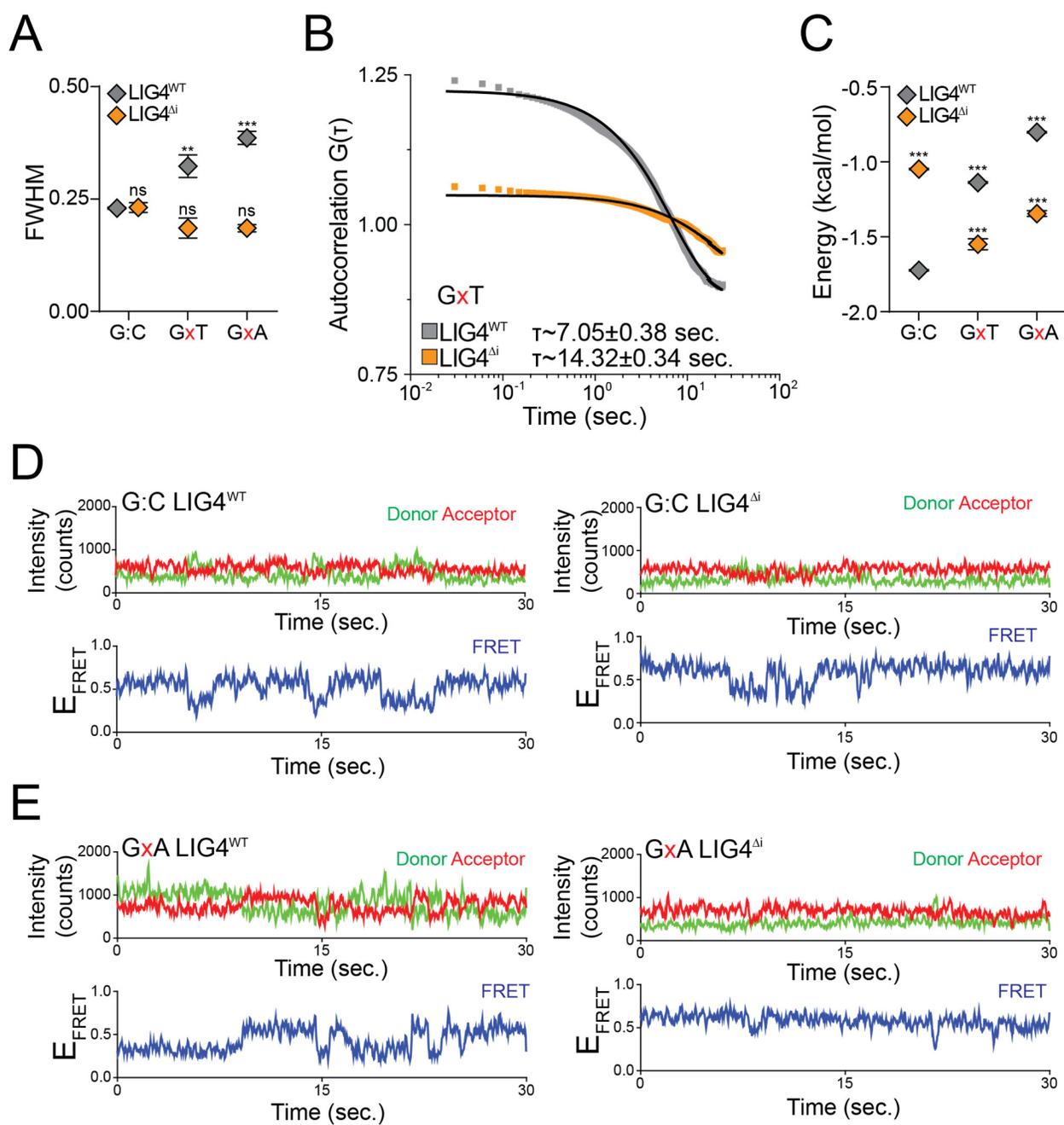
34

Figure 2.4: Effect of distorted ends on pairing dynamics of single molecule complexes with LIG4^{WT} or LIG4$^{\Delta i}$

Figure 2.4: Effect of distorted ends on pairing dynamics of single molecule complexes with LIG4^{WT} or LIG4$^{\Delta i}$

(A) Full width at half maximum (FWHM) of peaks was calculated from E$_{FRET}$ histograms for G:C, GxT, and GxA substrates. For **(A)** and **(C)**, error bars represent standard error of the mean for 3 experiments, and means were assessed for significance as in Figure 2.3B with confidence $p<0.01$ (**), $p<0.001$ (***) or not significantly different (ns). **(B)** Autocorrelation of individual FRET trajectories was used to calculate average transition times (τ) between FRET states of PECs formed on the GxT substrate with LIG4^{WT} or LIG4$^{\Delta i}$. **(C)** Transition energy between FRET states calculated from autocorrelation. **(D)** Representative smFRET trajectory for LIG4^{WT} and LIG4$^{\Delta i}$ PECs formed with G:C complementary ends **(E)** Representative smFRET trajectories of LIG4^{WT} and LIG4$^{\Delta i}$ PECs formed with GxA ends.
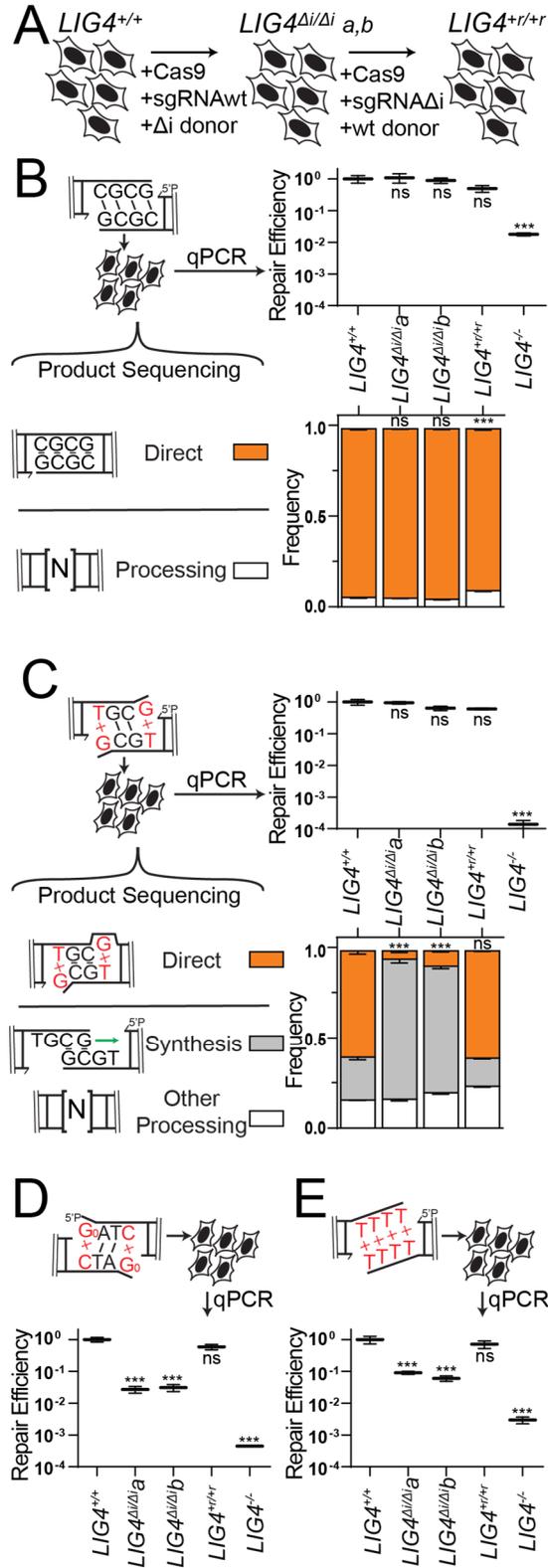
Figure 2.5: Effect of LIG4 insert1 on cellular joining of complex end structures

Figure 2.5: Effect of LIG4 insert1 on cellular joining of complex end structures

(A) Cells were engineered to express LIG4$^{\Delta i}$ from the native LIG4 locus by CRISPR/Cas9-based gene targeting, then the LIG4$^{\Delta i}$-a clone was reverted back to wild-type (*LIG4$^{+r/+r}$*) by a second round of gene targeting **(B-E)** Substrates with varied end structures were introduced in the cell types described in (A). Joining efficiency was assessed by qPCR, and product structure by sequencing or diagnostic restriction digestion and defined as directly ligated or ligated after end processing as noted. Cellular NHEJ was assessed for **(B)** complementary ends, **(C)** ends with 3' G:T terminal mismatches, **(D)** ends with 5' terminal 8-oxoguanine (Go; product structures reported in Figure 2.6D), and **(E)** ends with fully non-complementary overhangs (product structures reported in Figure 2.6E). Error bars represent the standard error of the mean for 3 experiments. Means of linearized qPCR data and direct joining products were assessed by one-way ANOVA as significantly different from control (*LIG4$^{+/+}$* cells) with confidence $p < 0.001$ (***) or not significantly different (ns).
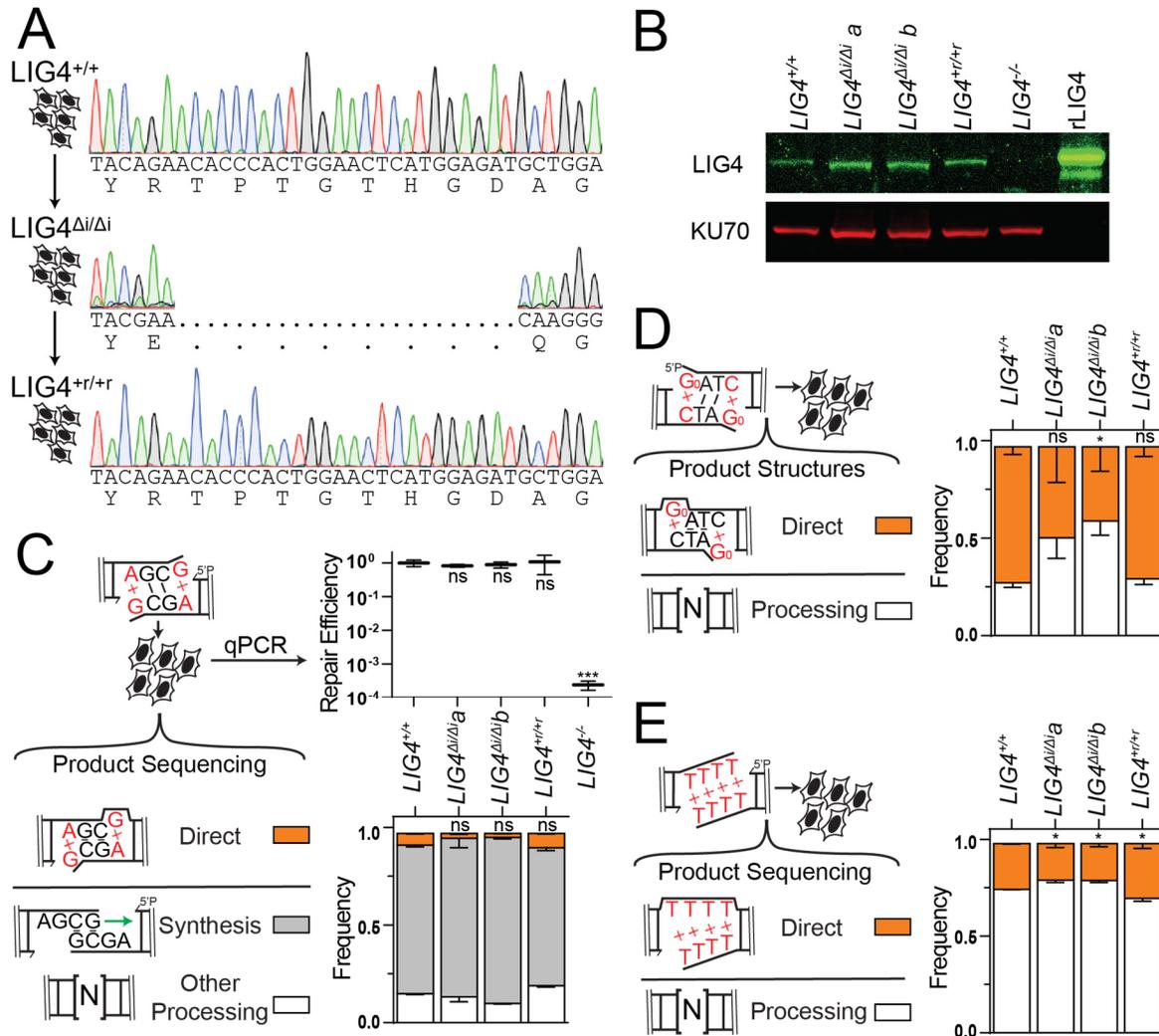
Figure 2.6: Effect of LIG4 insert1 on cellular joining of complex ends

(A) Sequences of targeted region of genomic DNA harvested from *LIG4+/+*, *LIG4Δi/Δi*, and *LIG4+r/+r* cells. **(B)** Western blot was performed to validate similar LIG4 expression in the indicated cell lines **(C-E)** Substrates with varied end structures were introduced in the noted cell types. **(C)** G:A mispaired overhangs was electroporated into cells. Repair efficiency was quantified by qPCR and repair product structures were determined by sequencing. Product structures were classified as directly repaired (orange), gap fill-in synthesis (gray), or other processing (white) **(D)** A substrate with radiomimetic terminal 8-oxoguanine damage was electroporated into cells. Repair product structures were chacterized by diagnostic restriction digests and classified as either directly ligated (orange) or ligated after end processing (white) **(E)** A substrate with fully mispaired TTTT overhangs was electroporated into cells. Repair product structures were characterized by sequencing and classified as either directly repaired (orange) or processed (white). Error bars for **(C-E)** represent standard error of the mean for 3 experiments. Means were assessed for significance as in Figure 2.5B-E with confidence $p < 0.05$ (*), $p < 0.001$ (***) or not significantly different (ns).
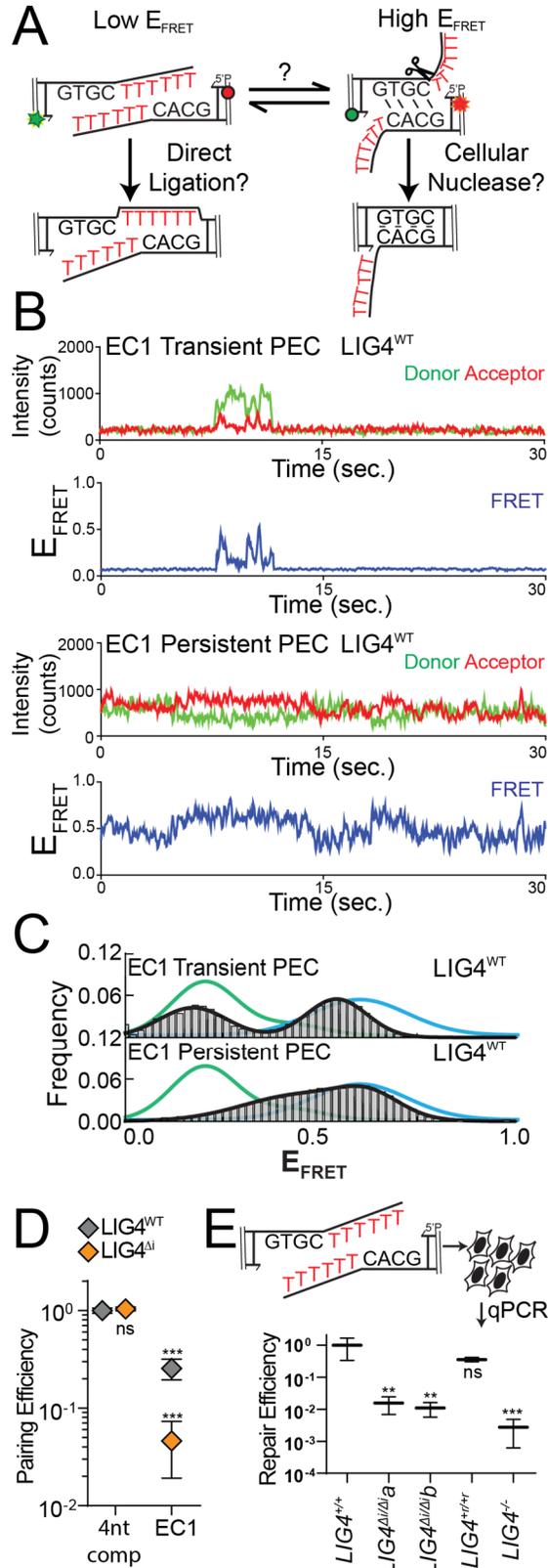
Figure 2.7: Effect of PEC flexibility on nucleolytic end processing

Figure 2.7: Effect of PEC flexibility on nucleolytic end processing

(A) A substrate was designed with embedded complementarity (EC1) that can be aligned in a high-FRET conformation guided by base pairing (a presumptive nuclease substrate) or in a low-FRET conformation with juxtaposition of 5' and 3' termini. **(B)** Representative smFRET trajectories for transient (short lived; <30sec) and persistent (long lived; 30+ sec) PECs formed with the EC1 substrate **(C)** Histograms of E_{FRET} of transient (top) or persistent (bottom) LIG4^{WT} PECs formed on the EC1 substrate (black), compared to FRET standards with complementary overhangs either 4 nt (blue) or 10 nt (green) in length **(D)** Quantitation of pairing efficiency of EC1 substrate by LIG4^{WT} (gray) or LIG4$^{\Delta i}$ (orange). Error bars represent standard error of the mean for 3 experiments. Means were assessed for significance as in Figure 2.3B with confidence $p<0.001$ (***) or not significantly different (ns). **(E)** The EC1 substrate was transfected into cells and repair efficiency was quantitated by qPCR. Error bars represent standard error of the mean for 3 experiments. Linearized means were assessed for significance as in Figure 2.5B-E with confidence $p<0.01$ (**), $p<0.001$ (***) or not significantly different (ns).
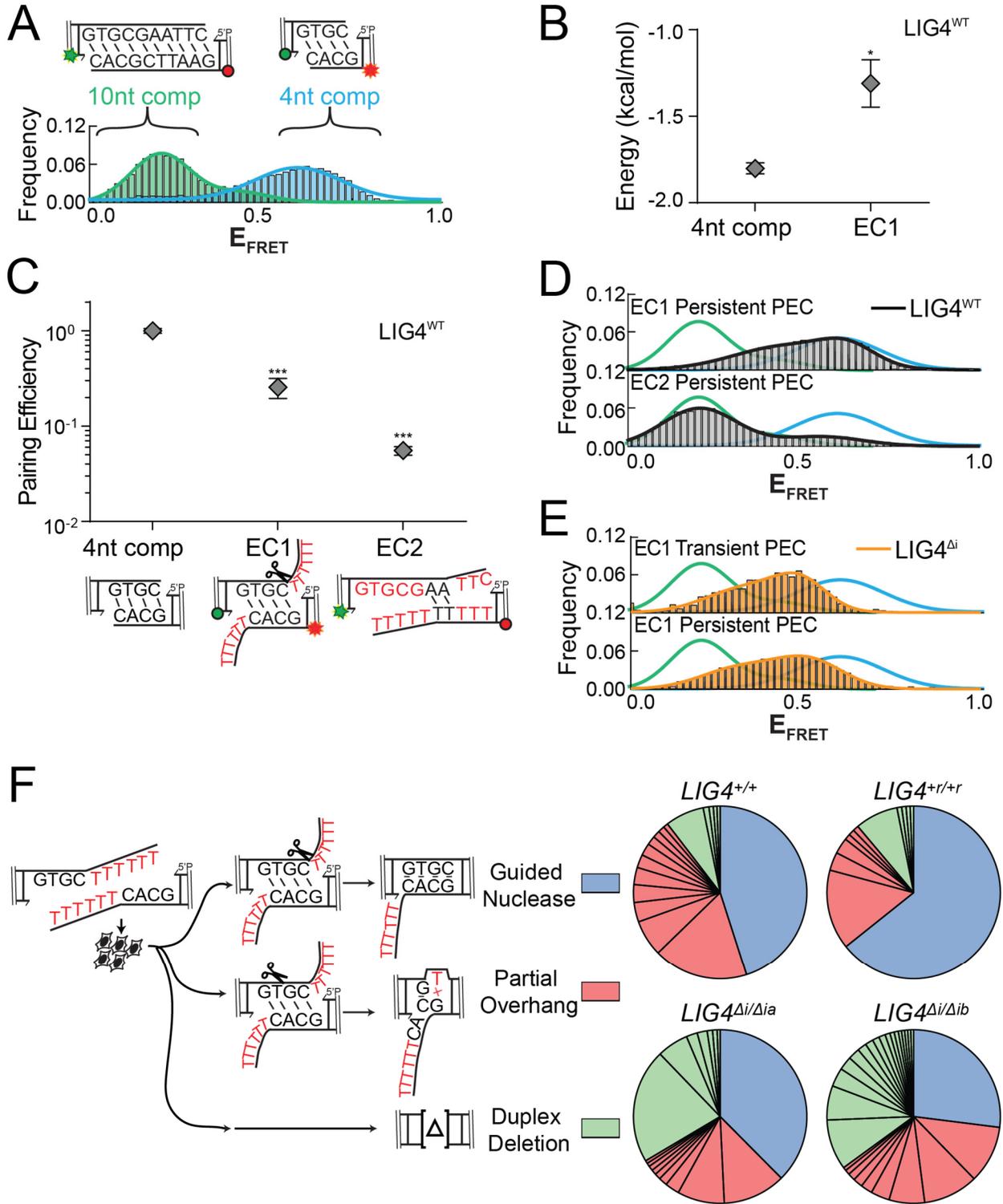
**Figure 2.8: Effect of PEC flexibility on nucleolytic end processing**

Figure 2.8: Effect of PEC flexibility on nucleolytic end processing

(A) Histogram of E_{FRET} of PECs formed with FRET standards containing fully complementary overhangs either 4 nt (blue) or 10 nt (green) in length **(B)** Transition energies calculated for 4nt complementary and EC1 substrates with LIG4^{WT}. Error bars represent standard error of the mean for 3 experiments. The mean energy for the EC1 substrate was assessed by t-test as significantly different from control (4nt complementary) with confidence $p < 0.05$ (*). **(C)** Quantitation of pairing efficiency of a substrate where embedded complementarity was reduced and relocated (EC2), relative to 4 nt complementary overhangs. Error bars represent standard error of the mean for 3 experiments. Means were assessed by one-way ANOVA as significantly different from control (4nt complementary) with confidence $p < 0.001$ (***). **(D)** Histograms of E_{FRET} of PECs formed on the EC1 (top) and EC2 (bottom) substrates (black), compared to FRET standards with complementary overhangs either 4 nt (blue) or 10 nt (green) in length **(E)** Histograms of E_{FRET} of PECs formed with LIG4$^{\Delta i}$ on the EC1 substrate (orange), compared to FRET standards with complementary overhangs either 4 nt (blue) or 10 nt (green) in length **(F)** The EC1 substrate was electroporated into cells. Repair product structures were determined by sequencing and classified as either deletion guided by embedded complementarity (blue), other deletions limited to single stranded overhang (red), or deletions that extended into double stranded flanking DNA (green).
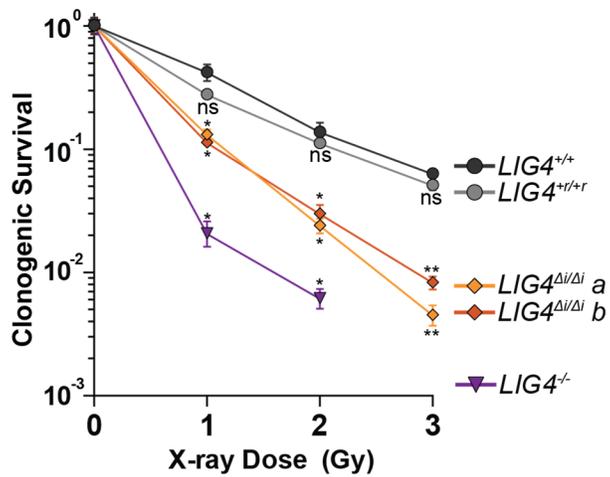
Figure 2.9: Effect of LIG4 insert1 on cellular sensitivity to ionizing radiation

Cells were exposed to indicated doses of X-rays and assessed for colony formation. Data represents the mean and standard deviation of 3 experiments. Mean surviving fractions were assessed by one-way ANOVA as significantly different from control (*LIG4*$^{+/+}$ cells) independently for each dose with confidence $p<0.05$ (*), $p<0.01$ (**) or not significantly different (ns).
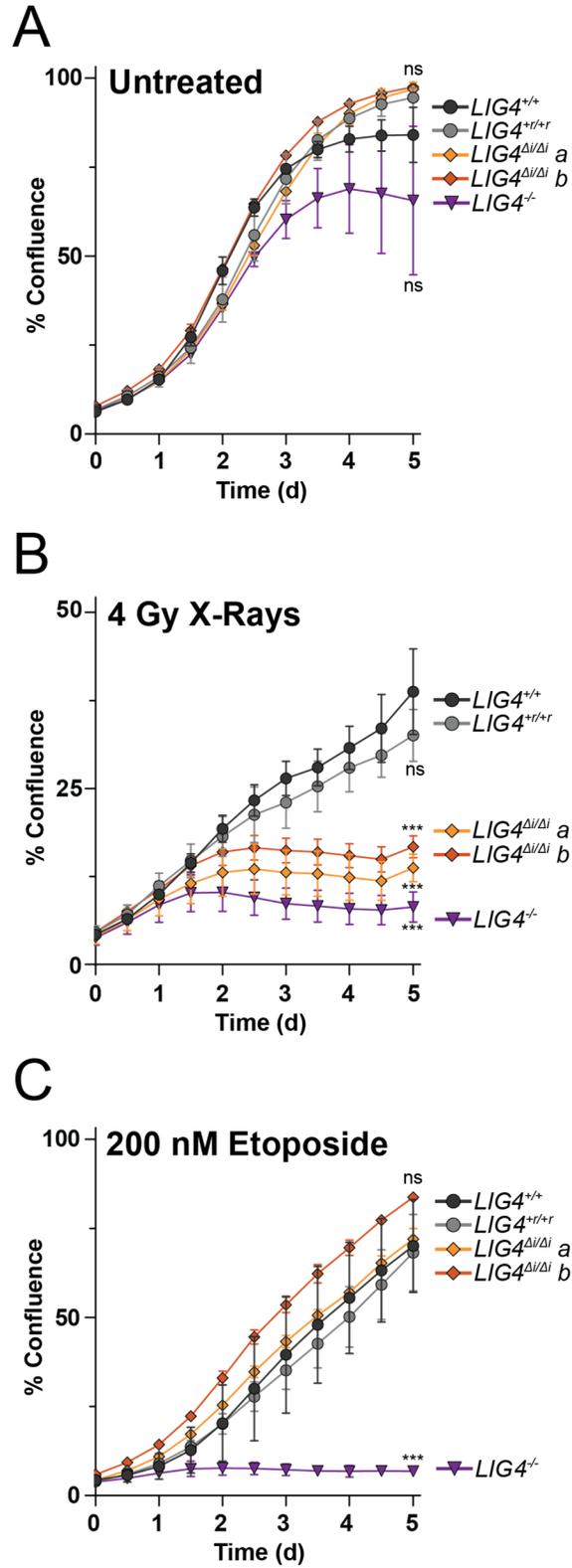
Figure 2.10: Sensitivity of LIG4 cell lines to damaging agents

Figure 2.10: Sensitivity of LIG4 cell lines to damaging agents

(A-C) Cell growth was assessed by live cell imaging every 4 hours for 5 days after seeding cultures. Cells were either **(A)** left untreated, **(B)** irradiated with 4 Gy X-rays, or **(C)** treated with 200 nM etoposide. Mean growth after 5 days was assessed by one-way ANOVA as significantly different from control ($LIG4^{+/+}$ cells) with confidence p<0.001 (***) or not significantly different (ns). For all live cell imaging experiments, error bars represent the standard deviation of a triplicate.
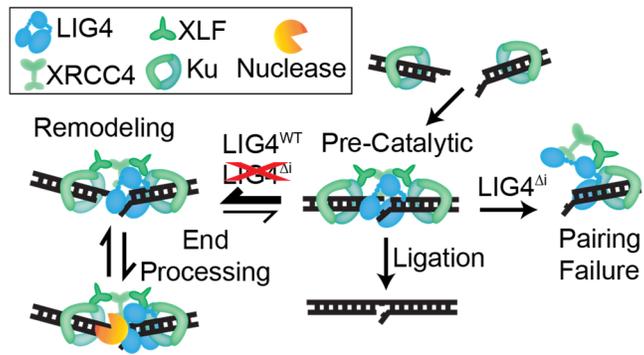
Figure 2.11: Sensing of differences in end structure by LIG4 guides repair

Model for LIG4-dependent remodeling of paired end complexes in response to complex ends.

Substrate Construction Oligonucleotides	
Core	
CAAGTGGTCTCAGACTGGCTACCCTGCTTCTTTGAGCATTTCTGAAACTATCACTTGTGTTTATT ATTACACTGGCATTCATTCTCCAGAGAACATGTCTAGCCTATTCCCAGCTTTGCTTACGGAGTTA CTCTGTATCTTTGCCTTGGAGAGTGCCAGAATCTGGTTTCAGAGTAAGATTTTATACATCATTTT TAGACATAGAAGCCACAGACATAGACAACGGAAGAAAGAGACTTTGGATTCTACTTACGTTTGAT TTCCCTGACGGAGACCTCGGC	
5' G:C Left Cap Top Strand	GATCCTCACACCCATCTCA
5' G:C Left Cap Bottom Strand	/5PHOS/AGTCTGAGATGGGTGTGAG
5' G:C Right Cap Top Strand	GATCCTCGCTTAGCTGTATA
5' G:C Right Cap Bottom Strand	/5PHOS/TGACTATACAGCTAAGCGAC
5' GoxC Left Cap Top Strand	G_OATCCTCACACCCATCTCA
5' GoxC Left Cap Bottom Strand	/5PHOS/AGTCTGAGATGGGTGTGAG
5' GoxC Right Cap Top Strand	G_OATCCTCGCTTAGCTGTATA
5' GoxC Right Cap Bottom Strand	/5PHOS/TGACTATACAGCTAAGCGAC
3' G:C Left Cap Top Strand	/5PHOS/AGTCTGAGATGGGTGTGCGGCGCG
3' G:C Left Cap Bottom Strand	CCGCACACCCATCTCA
3' G:C Right Cap Top Strand	/5PHOS/TGACTATACAGCTAAGCGGCGCG
3' G:C Right Cap Bottom Strand	CCGCTTAGCTGTATA
3' GxT Left Cap Top Strand	/5PHOS/AGTCTGAGATGGGTGTGCTGTGCG
3' GxT Left Cap Bottom Strand	CAGCACACCCATCTCA
3' GxT Right Cap Top Strand	/5PHOS/TGACTATACAGCTAAGCTGTGCG
3' GxT Right Cap Bottom Strand	CAGCTTAGCTGTATA
3' GxA Left Cap Top Strand	/5PHOS/AGTCTGAGATGGGTGTGCAGAGCG
3' GxA Left Cap Bottom Strand	CTGCACACCCATCTCA
3' GxA Right Cap Top Strand	/5PHOS/TGACTATACAGCTAAGCAGAGCG
3' GxA Right Cap Bottom Strand	CTGCTTAGCTGTATA
3' TTTT Left Cap Top Strand	/5PHOS/AGTCTGAGATGGGTGTGCTGTTTT
3' TTTT Left Cap Bottom Strand	CAGCACACCCATCTCA
3' TTTT Right Cap Top Strand	/5PHOS/TGACTATACAGCTAAGCTGTTTT
3' TTTT Right Cap Bottom Strand	CAGCTTAGCTGTATA
EC1 Left Cap Top Strand	/5PHOS/AGTCTGAGATGGGTGTGTTTGTGCTTTTTT
EC1 Left Cap Bottom Strand	AAACACACCCATCTCA
EC1 Right Cap Top Strand	/5PHOS/TGACTATACAGCTAAGCGTGCACTTTTTT
EC1 Right Cap Bottom Strand	ACGCTTAGCTGTATA
Nicked Substrate Top Strand Left Fragment	/5CY5/AGAAAACTGGCCCTTGCCATT
Nicked Substrate Top Strand Right Fragment	/5PHOS/CTCGGTGAGAGCATCGCTTA
Nicked Substrate Bottom Strand	TAAGCGATGCTCTCACCGAGAATGGCAAGGGCCAGTTT TCT
DSB End Substrate Top Strand	/5PHOS/TCACACACGCACGCATTTTT
DSB End Substrate Bottom Strand	/5CY5/TTTTTTGCGTGCGTGTGTGA
Complex Formation Substrate Top Strand	/5PHOS/CTCAGCTGGGAATTCCATATGAGTACTGCAG ATGCACTTGCTCGATAGATCTAACATGAG
Complex Formation Substrate Bottom Strand	/5CY5/GTAGGGCTCATGTTAGATCTATCGAGCAAGTG CATCTGCAGTACTCATATGGAATTCCCAGCTGAG

FRET Acceptor Top Strand	`/5PHOS/CGTG/ICY5/AGAGGAGACAGAGTGCGGGCG` `AACAACATAAATCGTACCCTCGTATGTATCGTATGGCT` `CATGCTTATCAGATGCT/3BIO/`
FRET Acceptor Bottom Strand (G:C)	`AGCATCTGATAAGCATGAGCCATACGATACATACGAGG` `GTACGATTTATGTTGTTCGCCCGCACTCTGTCTCCTCT` `CACGCGCG`
FRET Acceptor Bottom Strand (GxT)	`AGCATCTGATAAGCATGAGCCATACGATACATACGAGG` `GTACGATTTATGTTGTTCGCCCGCACTCTGTCTCCTCT` `CACGTGCG`
FRET Acceptor Bottom Strand (GxA)	`AGCATCTGATAAGCATGAGCCATACGATACATACGAGG` `GTACGATTTATGTTGTTCGCCCGCACTCTGTCTCCTCT` `CACGAGCG`
FRET Acceptor Bottom Strand (4 nt comp)	`AGCATCTGATAAGCATGAGCCATACGATACATACGAGG` `GTACGATTTATGTTGTTCGCCCGCACTCTGTCTCCTCT` `CACGGCAC`
FRET Acceptor Bottom Strand (EC1)	`AGCATCTGATAAGCATGAGCCATACGATACATACGAGG` `GTACGATTTATGTTGTTCGCCCGCACTCTGTCTCCTCT` `CACGGCACTTTTTT`
FRET Acceptor Bottom Strand (EC2)	`AGCATCTGATAAGCATGAGCCATACGATACATACGAGG` `GTACGATTTATGTTGTTCGCCCGCACTCTGTCTCCTCT` `CACGGAATTCGCAC`
FRET Donor Bottom Strand	`/5PHOS/TCTG/ICY3/ATAAGCATGAGCCATACGATA` `CATACGAGGGTACGATTTATGTTGTTCGCCCGCACTCT` `GTCTCCTCTCACGTTTTCGTGAGAGGAGACAGAGTGC`
FRET Donor Top Strand (G:C)	`GGGCGAACAACATAAATCGTACCCTCGTATGTATCGTA` `TGGCTCATGCTTATCAGACGCG`
FRET Donor Top Strand (GxT)	`GGGCGAACAACATAAATCGTACCCTCGTATGTATCGTA` `TGGCTCATGCTTATCAGATGCG`
FRET Donor Top Strand (GxA)	`GGGCGAACAACATAAATCGTACCCTCGTATGTATCGTA` `TGGCTCATGCTTATCAGAAGCG`
FRET Donor Top Strand (4 nt comp)	`GGGCGAACAACATAAATCGTACCCTCGTATGTATCGTA` `TGGCTCATGCTTATCAGAGTGC`
FRET Donor Top Strand (10 nt comp)	`GGGCGAACAACATAAATCGTACCCTCGTATGTATCGTA` `TGGCTCATGCTTATCAGAGTGCGAATTC`
FRET Donor Top Strand (EC1)	`GGGCGAACAACATAAATCGTACCCTCGTATGTATCGTA` `TGGCTCATGCTTATCAGAGTGCTTTTTT`
FRET Donor Top Strand (EC2)	`GGGCGAACAACATAAATCGTACCCTCGTATGTATCGTA` `TGGCTCATGCTTATCAGATTTTTTTTTT`

PCR Primer Oligonucleotides	
qPCR NHEJ Assay Forward	`CTTACGTTTGATTTCCCTGACTATACAG`
qPCR NHEJ Assay Reverse	`GCAGGGTAGCCAGTCTGAGATG`
Illumina Amplification Forward	`NNNNNNCTTACGTTTGATTTCCCTGACTATACAG`
Illumina Amplification Reverse	`NNNNNNGCAGGGTAGCCAGTCTGAGATG`
Illumina Adapter Top Strand	`/5PHOS/GATCGGAAGAGCGGTTCAGCAGGAATGCCGA` `G`
Illumina Adapter Bottom Strand	`ACACTCTTTCCCTACACGACGCTCTTCCGATCT`
Illumina Enrichment Forward	`AATGATACGGCGACCACCGAGATCTACACTCTTTCCCT` `ACACGACGCTCTTCCGATCT`

Illumina Enrichment Reverse	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTC CTGCTGAACCGCTCTTCCGATCT
CRISPR Screening Forward	TGAGTTGGAATCGAATGCTG
CRISPR Screening Reverse	GAGGGGGCTTCTCTGCTACT
GoxC Amplification Forward	/5CY5/CCTTGGAGAGTGCCAGAATC
GoxC Amplification Reverse	CTGGAGAATGAATGCCAGTG

Oligonucleotides to generate sgRNAs (guide italicized and underlined)	
Target *LIG4^WT* Top Strand	ccgg*GCATCTCCATGAGTTCCAGT*
Target *LIG4^WT* Bottom Strand	aaac*ACTGGAACTCATGGAGATGC*
Target *LIG4^Δi* Top Strand	ccgg*ACTTTTAAACTACGAACAAG*
Target *LIG4^Δi* Bottom Strand	aaac*CTTGTTCGTAGTTTAAAAGT*

Gene Targeting Regions
LIG4^WT and *LIG4^+r/+r* sequence (guide italicized and underlined; BpmI site bolded)
AAACTTTTAAACTACAGAACA*CCCACTGGAACTCATGGAGATGC***CTGGAG**ACTTTGC
LIG4^Δi sequence (guide italicized and underlined; BsmFI site bolded)
AA*ACTTTTAAACTACGAACAA*---------------------*GGG***GAC**TTTGC

Table 2.1: Sequences of DNA Reagents

CHAPTER 3: RIBONUCLEOTIDES ENABLE FLEXIBILITY IN NHEJ

3.1 Introduction

Nonhomologous end joining (NHEJ) is the primary pathway for repairing chromosomal double-strand breaks (DSBs) in mammals, and is required for genome stability in all cell types as well as assembly of antigen specific receptors by V(D)J recombination in lymphocytes[113]. NHEJ employs specialized nucleases and polymerases, including the widely-expressed Pol μ (gene name, *Polm*) and lymphocyte-specific terminal deoxynucleotidyl transferase (TdT), to modify broken end structures in preparation for ligation[4]. Accordingly, loss of Pol μ or TdT interferes with normal cell growth, resistance to DNA damage, and development of robust immune responses[14,72,80,81,114–116]. Pol μ and TdT are exceptionally poor at discriminating against ribonucleotide incorporation *in vitro*[37,87–90]*,* favoring deoxynucleotides 1.4 to 11-fold (depending on nucleotide base) more than ribonucleotides[88]. By comparison, closely related Pol λ and Pol β incorporate deoxynucleotides several 1000-fold more efficiently than ribonucleotides[83–85] (similar to most other polymerases that maintain DNA genomes). However, it is unknown whether ribonucleotide incorporation occurs during cellular NHEJ, and if ribonucleotide incorporation occurs, if it significantly impacts NHEJ function.

3.2 Methods

Cell Lines

WT (C57BL/6) or $Polm^{-/-}$ murine fibroblast cells (generously provided by Dr. L. Blanco) were derived from E14.5-d embryos and immortalized by the introduction of SV-40 large T-antigen as described previously[117]. The ts-AbMLV pre-B cells were a generous gift from Dr. Y. Chang (Arizona State University). These lines and variants described below were confirmed to be free of mycoplasma contamination by PCR[118]; cell lines were additionally selected at random for third party validation of PCR results using Hoechst staining[119]. Variants of MEFs and pre-B cell lines with frameshift mutations in Exon 2 of the *Rnaseh2a* gene or Exon 1 of the *Polm* gene were generated by transient expression of nickase Cas9 D10A and a pair sgRNAs (Table 3.3). Cell lines were engineered to express either Myc-tagged Pol µ or TdT by infection with retrovirus derived from pBabe-puro constructs containing the appropriate murine cDNAs. All cell lines were sub cloned by limiting dilution and verified by western blot analysis and allele sequencing (where applicable). Wild-type and LIG4-deficient HCT116 cells were a gift from Dr. Eric Hendrickson. Western blots were performed using standard techniques and antibodies against murine RNaseH2A (ProSci, 4979), Actin (Sigma, A2066), TdT (Sigma, 14.9739.80), or the Myc affinity tag (Santa Cruz, sc409E10). MEF and HCT116 cells were maintained in DMEM supplemented with 10% (vol/vol) FBS (Sigma), 5 mM N-acetyl-L-cysteine (Sigma), 2 mM L-glutamine (Gibco), 100U/ml penicillin, and (in variant lines expressing puromycin resistance markers) 2 µg/mL puromycin, at 37°C and 5% CO_2. The ts-AbMLV cell lines were maintained in RPMI 1640 (Corning) supplemented with 10% (vol/vol) FBS (Sigma), 2 mM L-glutamine (Gibco), 55 µM 2-

mercaptoethanol (Gibco), 100U/ml penicillin, and (in variant lines expressing puromycin resistance markers) 2 μg/mL puromycin, at 33°C and 5% CO_2. Recombination in SP9 cells was induced by culture at 40°C.

Double strand break repair assays

Substrates described in Figure 3.1 were generated by PCR amplification of a common 280-bp DNA segment with primer pairs containing embedded restriction enzyme digest sites chosen to generate the desired end structures (Table 3.3). Substrates were digested, purified by agarose gel electrophoresis, and purified substrate recovered using the QiaQuick gel-extraction kit (Qiagen). Substrates described in Figure 3.7 were assembled by ligation of ~15 bp double-stranded DNA caps (oligonucleotide pairs annealed to generate gaps in Table 3.3) containing the desired overhang sequence to a 280 bp core fragment digested with BsaI-HF (New England Biolabs), with the caps in 3 fold excess. Substrates were then purified using a QiaQuick PCR purification kit (Qiagen), and resolved on a native acrylamide gel to ensure substrates preparations were free of detectable unappended core and excess cap.

Extrachromosomal substrate electroporations were carried out using the NEON transfection system (Invitrogen) with 20 ng of substrate, 600 ng pMAX-GFP carrier plasmid, and 2×10^5 cells by a 1350 V, 30 ms pulse in a 10 μl chamber. Following transfection, cells were incubated at 37°C in HBSS (Gibco) supplemented with 1 mM $MgCl_2$ and Benzonase Nuclease (Sigma) for the indicated amount of time. Cellular DNA was harvested using a QiaAmp DNA mini kit (Qiagen). *Rosa26* locus-targeting Cas9 ribonucleoproteins (RNPs) were assembled from purified Cas9 (derived from Addgene

#69090), and annealed Alt-R modified crRNA (Table S2) and trcrRNA (IDT). The Cas9-sgRosa26 RNP complex was introduced at 1.8 µM into $2×10^6$ cells and a 100 uL chamber using a 1350 V, 30 ms pulse and incubated at 37°C before cell harvesting and purification of genomic DNA (QiaAmp DNA mini kit). For experiments involving nucleotide electroporation into cells, the transfection mixture was supplemented with either 10 mM NaCl, 10 mM rGTP, or 10 mM dGTP.

In vitro NHEJ assays were performed initiated by incubating 2 nM DSB substrates with either NHEJ proteins (25 nM Ku, 40 nM XLF, and 40 nM XRCC4-LIG4) or 160 units of T4 DNA ligase (New England Biolabs) in a buffer with 25 mM Tris pH 7.5, 100 uM EDTA, 1 mM DTT, 1 mM MgCl, 100 uM ATP, 150 mM KCl, 7.5% polyethelene glycol 3000, and 100 ng of supercoiled plasmid DNA. Reactions were carried out for 10 minutes at 37 C and stopped with 0.1% SDS and 20 mM EDTA.

Repair product analysis

Recovery of joined products was quantified by real-time PCR (qPCR) using a QuantStudio 6 System (Applied Biosystems), primers that amplify head-to-tail junctions (Table 3.3), and VeriQuest Probe qPCR Master Mix (Affymetrix). We validated our assays as efficient, reproducible, and linear over the range relevant to these experiments (Figures 3.2A, 3.2B, 3.5A, 3.6A). The fraction of products with embedded ribonucleotides (%ribo.) was measured by comparison of samples digested with 2.5 units of RNaseHII (New England Biolabs) at 37 C for 16 hours in the manufacturer's provided buffer to mock treated samples; these conditions were sufficient to cleave to completion sites of embedded ribonucleotides regardless of opposite strand structure, with no obvious digestion of DNA-only controls. Restriction enzymes that cut outside the

amplicons (NlaIII and MseI for Rosa26 samples, and HaeIII and MseI for substrate samples) were also included in RNaseHII and mock treated samples to ensure the initial denaturation of template duplexes occurred with equal efficiency. As an alternative method of ribonucleotide cleavage, samples were treated with 300 mM NaOH for 2 hours at 55 ºC (or mock treated with 300 mM NaCl) and neutralized by a 10x dilution into 65 mM Tris-HCl buffer (pH 8.0; Figure 3.3A). For Figures 3.1E, 3.7A, and 3.7B, products were amplified with a Cy5-labelled primer and amplified products further characterized by digestion with restriction enzymes diagnostic for specific products (Table 3.1; NsiI for G3' and CG3', SalI for CAG3', AatII for CGCAG3', and FspI for TTTTTTTTGC3'). Digestion products were resolved on a 5% non-denaturing polyacrylamide gel, visualized using a Typhoon Imager (GE Healthcare), and quantified using ImageJ software.

Mutations in CRISPR-Cas9 products were assessed by amplifying the targeted locus and either by restriction digesting the non-mutated products with XbaI, or by sanger sequencing and Tracking of Indels by DEcomposition (TIDE)[120].

Next-generation Sequencing

Template DNA for each sequencing library (~5 × 10^5 input molecules) was amplified using Phusion DNA polymerase (New England Biolabs) and PCR primers with six-nucleotide index sequences appended to their 5′-ends (Table 3.3). Amplified DNA was 5′phosphorylated with T4 Polynucleotide kinase then treated with Klenow exo- to add dA to the 3′ termini (New England Biolabs). Sequencing adapters for paired-end reads were appended to the amplicons by treatment with T4 DNA ligase, and free adapter removed by agarose gel purification. After a final enrichment PCR amplification

the products were purified by Agencourt AMPure XP beads (Beckman Coulter).

Libraries were submitted for a 2 × 150-bp sequencing run (MiSeq; Illumina) with a

PhiX174 DNA "spike". Data analysis with Genomics workbench v7.5.1 (CLC-Bio) and

Microsoft Excel was carried out as described[72].

3.3 Results

Cellular NHEJ products contain ribonucleotides

      We initially investigated whether ribonucleotides are incorporated during NHEJ of

a linear DNA substrate introduced into transformed mouse embryonic fibroblasts

(MEFs; Figure 3.1A). We optimized this assay to allow for rapid harvesting of repair

products, in anticipation that ribonucleotides were only transiently present, and

employed qPCR to quantify products with embedded ribonucleotides (Figure 3.2A-C).

Embedded ribonucleotides were present in 60% (standard deviation = sd; 4.2%) of

NHEJ products (Figure 3.1B) when products were assessed within the first minute, and

were dependent on either Pol μ or TdT (Figure 3.1C and Figure 3.2D-E).

      The fraction of NHEJ products with embedded ribonucleotides decayed until

ribonucleotides were undetectable after 20 minutes (Figure 3.1B, black line). To

determine if this reduction was due to replacement of incorporated ribonucleotides with

deoxynucleotides (Ribonucleotide excision repair; RER), we generated a MEF variant

deficient in *Rnaseh2a* (Figure 3.2F), which initiates RER[86]. Levels of embedded

ribonucleotides were initially similar in *Rnaseh2a*-deficient cells, before stabilization at

levels approximately two-fold less than was initially observed (Figure 3.1B, orange line).

The substrate used above possessed a single nucleotide 3' overhang (3'G). Pol

μ-dependent addition of a complementary C (+C product) accounts for approximately

half of repair of this substrate[72] (Figure 3.1D-E, Figure 3.3), while remaining products

have heterogeneous 1-5 bp deletions of flanking sequence and variable dependency on

Pol μ activity[72]. To focus on Pol μ-dependent NHEJ, we digested products with a

restriction enzyme specific for the +C product (NsiI; Figure 3.1D). Embedded

ribonucleotides were present in 91% (sd 8%) of NsiI-sensitive products after 1 minute

(Figure 3.1E). Similar results were observed when using a different method to detect

ribonucleotide-containing products (Figure 3.3A). Moreover, embedded ribonucleotides

were only modestly less frequent (75%, sd 8) using a substrate with a different

overhang template (C3') and Pol μ-dependent added nucleotide (+G product; Figure

3.3B). Importantly, our ability to detect repair products (and thus %RNA) is fully

dependent on NHEJ, as LIG4-deficient cells show 330-fold defective repair (Figure

3.3C). We conclude most Pol μ and TdT-dependent NHEJ products contain embedded

ribonucleotides, and that the modest preference of these polymerases for addition of

deoxynucleotides *in vitro*[88] is overwhelmed by higher concentrations of ribonucleotides

in cells[121].

As also informed by data presented in subsequent figures, we suggest these

early products involve one ligated strand only. Subsequent repair of the complementary

strand with deoxynucleotides accounts for the two-fold dilution of products with

embedded ribonucleotides that is independent of *Rnaseh2a* (Figure 3.1B, orange line),

while complete removal of ribonucleotides requires *Rnaseh2a*-dependent ribonucleotide

excision repair.

RNA is integrated into the genome during chromosome break repair

We determined if ribonucleotides are similarly incorporated during repair of

breaks in the chromosome by NHEJ. We used a pre-B cell line (SP9) that can be

induced to undergo V(D)J recombination at the immunoglobulin kappa locus (*Igk*; Figure

3.4A), since Pol μ is efficiently engaged by the 3' overhang intermediates in this

process[71,81]. Embedded ribonucleotides were undetectable 24 hours after induction

when cells were proficient in RER. By comparison, 35% of *Igk* recombination products

possessed embedded ribonucleotides in an *Rnaseh2a*-deficient variant (Figure 3.4B

and Figure 3.5A-B). This frequency is approximately half of the frequency of *Igk*

products where Pol μ is active[71,81], consistent with the model proposed above (where

only the first strand of a chromosome double strand break is repaired with

ribonucleotides). Embedded ribonucleotides were again largely dependent on either Pol

μ or TdT (Figure 3.4B and Figure 3.5C-D). We additionally confirmed that V(D)J

recombination was induced in these cell lines by qPCR across the VJ_K junction (Figure

3.5E).

We sought to assess the frequency of Polymerase-dependent ribonucleotide

incorporation during chromosomal NHEJ shortly after repair, and also to extend analysis

to a non-lymphoid model. Cas9-induced chromosome breaks in TdT-expressing MEFs

(Figure 3.2E) results in a repair product - addition of two or more Gs, with no loss of

flanking DNA ("+GG product") –that was abundant (18% of chromosomal repair

products, Table 3.2), unambiguously polymerase-dependent, and detectable by qPCR

(Figure 3.4C-D and Figure 3.6A-B). Additionally, direct introduction of *Rosa26* locus-

targeted Cas9 nuclease allowed for assessment of repair products immediately after

they were generated (Figure 3.6C). In *Rnaseh2a*-deficient MEFs, 84% and 77% of +GG

NHEJ products had embedded ribonucleotides when analyzed 1 and 4 hours after

introduction of Cas9 (Figure 3.4D). Embedded ribonucleotides were reduced two-fold in

these RER-deficient cells over the next 20 hours, consistent with repair of

complementary strands with deoxynucleotides. As expected, embedded ribonucleotides

did not accumulate to high levels in cells with intact RER (23% after 1 hour; Figure

3.6D). RNA was not detected at significant levels by amplification an in-tact genomic

locus in RER-deficient MEFs, consistent with an expected ~3% rate of genomic

ribonucleotides in a 100 bp amplicon in these cells, based on RER studies (Figure

3.6E).

Ribonucleotides enable ligation of mismatched ends

Both Pol μ and TdT primarily incorporate ribonucleotides during cellular NHEJ,

thus we investigated the consequences of ribonucleotide addition on the next step of

cellular NHEJ, ligation. Variants of substrates previously associated with Pol μ and TdT

activity were made such that polymerase-dependent nucleotide additions were already

included, and varied such that the included nucleotide(s) terminated with a

ribonucleotide vs. a deoxynucleotide (Figure 3.7A). We then assessed effects of

differing termini on the NHEJ ligation step by introducing these simulated polymerase

products into cells that express neither Pol μ nor TdT (*Polm⁻/⁻* MEFs). When the

opposite strand was complementary (i.e. cohesive ends or nicks), there was little impact

of a ribonucleotide vs. deoxynucleotide terminus on the ligation step of cellular NHEJ

(Figure 3.7B)[88,90]. In contrast, a ribonucleotide terminus was 48-fold more effective in

promoting direct ligation on a candidate product of TdT activity, where the opposite

strand possessed a 9-nucleotide gap (Table 3.1 and Figure 3.7B). Ribonucleotide

termini were also 17-fold more effective in promoting direct ligation when using typical

products of Pol μ activity[72], where the opposite strand was short (2 or 4 nucleotides

long) and largely mispaired (Table 3.1 and Figure 3.7B). By comparison, the same

substrates with deoxynucleotide termini resulted in both less efficient repair, and more

frequent deletion of flanking DNA in the repair products that were recovered (Table 3.1

and Figure 3.8A-C). We also showed that ribonucleotides stimulate LIG4 activity on

such substrates *in vitro*. In contrast, T4 ligase gains no benefit from a terminal

ribonucleotide, suggesting that LIG4 is specifically suited to ligate ends with a terminal

ribonucleotide (Figure 3.8D). The contribution of these polymerases to cellular NHEJ is

thus largely reliant on their ability to incorporate ribonucleotides.

We substituted the 3' terminal ribonucleotide with different nucleotide analogs to

clarify the chemical basis for stimulation of the NHEJ ligation step.

Arabinofuranosylcytidine (AraC) is a stereoisomer of (ribo)cytidine (rC), differing from rC

only with respect to orientation of the 2'OH relative to its chiral center. As a

consequence, AraC favors a C2' endo sugar pucker, relative to the C3' endo sugar

pucker favored by rC (Figure 3.8E). Significant amounts of direct ligation product were

not detectable when employing an AraC terminus (Figure 3.8E and Table 3.1) – indeed,

direct ligation was even less efficient than observed when using a deoxynucleotide

terminus (which, while favoring C2' endo, can adopt both sugar puckers; Figure 3.8E).

Substitution of a fluorine for the 2' OH of rC (2-FC; favors C3' endo) led to levels of

direct ligation similar to those observed with rC (Figure 3.8E and Table 3.1). Stimulation

of the NHEJ ligation step by ribonucleotides is thus best attributed to the presence of a

C3' endo sugar pucker in the 3' terminal nucleotide.

Deoxynucleotides block mutagenic repair of Cas9 breaks

To address the biological significance of ribonucleotide-dependent NHEJ, we directly introduced nucleotide triphosphates into TdT-expressing cells alongside Cas9-sg*Rosa26* RNP and harvested genomic DNA one hour after transfection. We initially assessed accumulation of the +GG product and found that adding riboguanosine triphosphate (rGTP) stimulates +GG recovery, while adding deoxyguanosine triphosphate (dGTP) impairs it 5-fold. (Figure 3.7D). These effects are attributable to the stimulation of direct ligation by ribonucleotides, since araGTP mimics dGTP in blocking +GG recovery (Figure 3.8F).

We next used two different methods, Tracking of Indels by DEcomposition (TIDE)[120] and simple restriction enzyme screening to assess the recovery of all Cas9-mediated insertions and deletions, instead of focusing on a single product. In both cases, we found that dGTP transfection reduces the frequency of indels, while rGTP transfection had no significant effect (Figure 3.8G-H). The inhibitory effect of dGTP is restricted to insertions, and thus attributable to its incorporation by these polymerases (Figure 3.8H). As observed with the +GG product, araGTP has a similar effect on indel recovery as dGTP (Figure 3.8G-H), implying the reduction of insertion products is attributable to a block of the NHEJ ligation step.

3.4 Discussion

Our results argue mammalian NHEJ often requires three sequential strand-break repair reactions (Figure 3.7E). In the first reaction, Pol μ and TdT primarily add one or

more ribonucleotides to one 3' terminus of a double strand break, which is necessary for efficient ligation of the first strand at end structures where these polymerases are uniquely active (Table 3.1). The ligase required for NHEJ (LIG4) may even be alone amongst mammalian ligases in the ability to take advantage of ribonucleotide termini, analogous to the *in vitro* activity of bacterial *Pseodomonas* LigD when compared to that of e.g. *E. coli* LigA[122].

The now-ligated, ribonucleotide-containing first strand can then be used as a template for a straightforward gap repair reaction of the opposite (second) strand, explaining the two-fold dilution of embedded ribonucleotides that is independent of RER/RNaseH2A. Alternative models – where ribonucleotides are incorporated in both strands, or if RNaseH2 incises the first strand before second strand repair is complete – risk re-breakage of the chromosome. Additionally, the transient nature of the ribonucleotide embedded intermediate ($t_{1/2}$ is likely less than 5 minutes; Figure 3.1B) suggests that RER is directly coupled to NHEJ. Much of double strand break repair may thus require three distinct but coupled repair reactions, with NHEJ-specific factors definitively implicated only in the first.

Pol μ is widely expressed and participates in repair of 16% of Cas9-induced blunt ended structures and the majority of end structures with mispaired 3' overhangs[72]. In lymphocytes, either TdT or Pol μ is active in 65% of NHEJ events required for V(D)J recombination[71,81,115,116]. Therefore, the model proposed here – three obligatorily ordered strand break repair reactions, with a ribonucleotide-embedded intermediate that is central to flexible and efficient repair – is relevant to a large fraction of mammalian NHEJ. This mechanism is likely also important for NHEJ in species with polymerases

similarly prone to ribonucleotide incorporation (e.g. fungi and bacteria)[123–125]. Our work

argues the important roles of mammalian Pol μ and TdT in promoting long term cellular

proliferative capacity, the development of adaptive immunity, and radioresistance are

reliant on their preference for incorporating ribonucleotides[14,80,81,114].
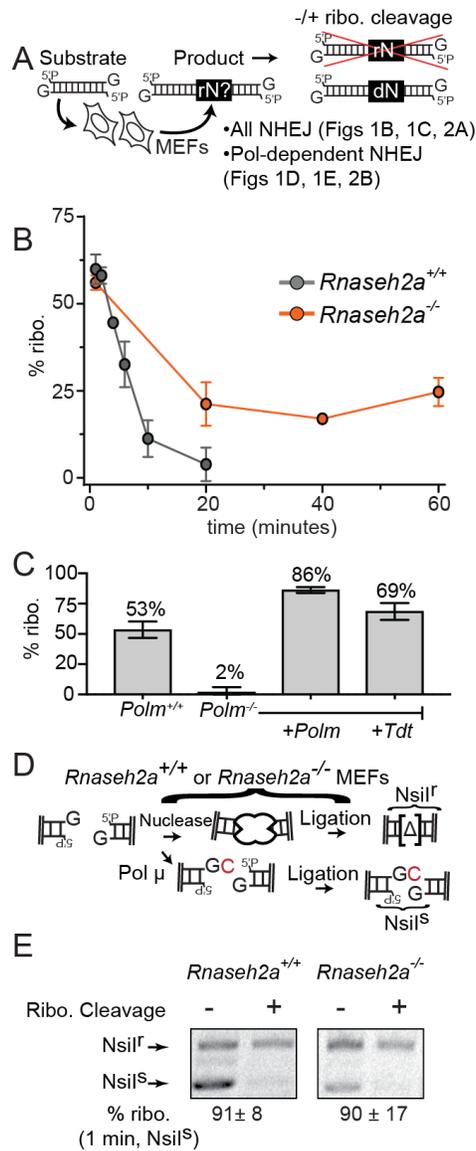
Figure 3.1: Ribonucleotide incorporation during cellular NHEJ

(A) DNA fragments with 3'G overhangs at ends were introduced into MEFs, and % cellular NHEJ products with embedded ribonucleotides (% ribo.) quantified by comparing amplification efficiency with and without prior cleavage at sites of ribonucleotide incorporation (also Fig. S1B). **(B)** % ribo. after introduction of substrate into *Rnaseh2a*$^{+/+}$ (gray) or *Rnaseh2a*$^{-/-}$ (orange) MEFs. Data points are the mean of 3 transfections, and error bars represent sd. **(C)** %ribo. in products recovered after 1 minute. Data points are the mean of 3 transfections, and error bars represent sd. **(D-E)** Digestion of amplified products with NsiI and electrophoresis distinguishes Pol m-dependent +C products from products with deletions of flanking sequence (Δ). **(E)** The mean % ribo. in NsiI^s products recovered after 1 minute from three independent transfections, ± sd, is noted below gel.

64
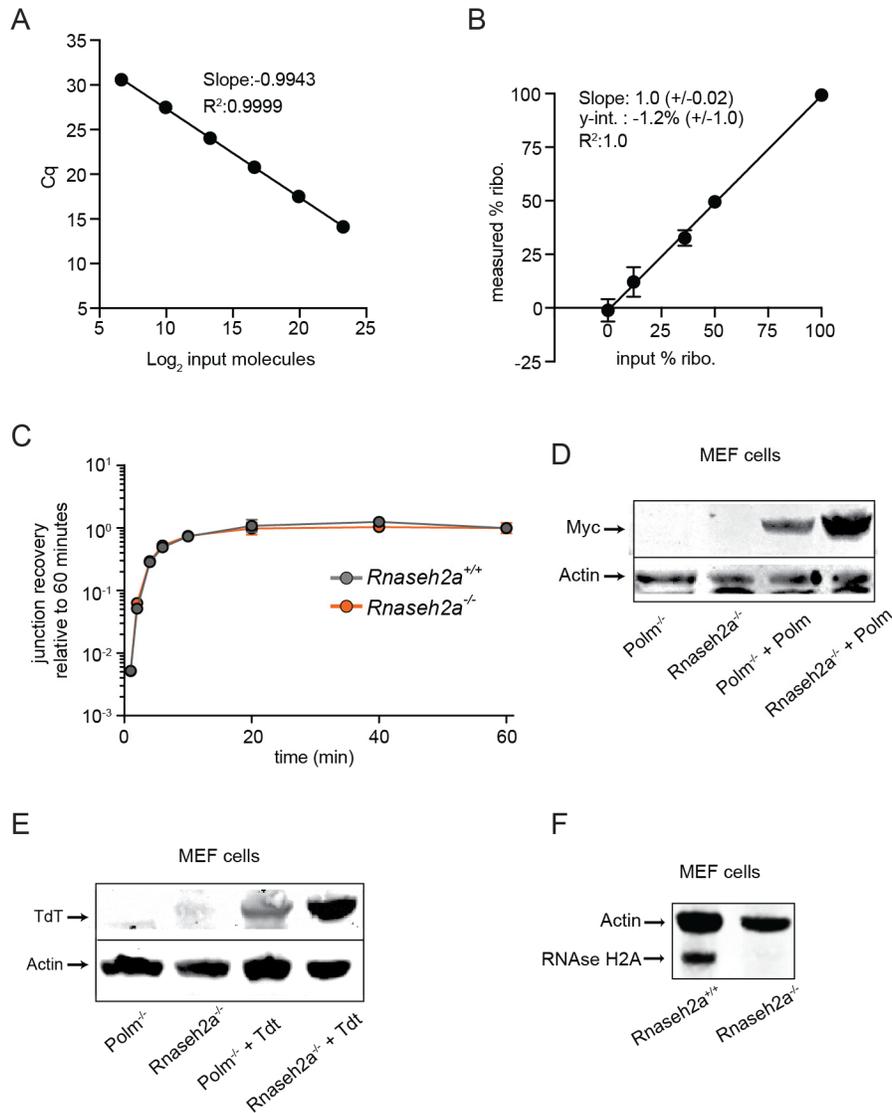
Figure 3.2: Assays and cell lines to assess ribonucleotide content during NHEJ

(A) A serial dilution of an oligonucleotide model repair product from the extrachromosomal NHEJ assay (Fig. 1) was used as a template in a qPCR reaction. Mean qualification cycle from 3 independent experiments is plotted with error bars representing sd. **(B)** Model amplicons containing a single embedded ribonucleotide were mixed with fully DNA amplicons at known proportions, and diluted into genomic DNA at a target to genomic DNA ratio similar to experimental conditions. These mixtures were mock or RNaseHII digested before qPCR as described in methods. Mean detected ribonucleotides from 3 experiments is plotted with error bars representing sd. **(C)** Extrachromosomal substrate assay was performed as in Fig. 1A for the indicated amounts of time. Mean repair efficiency of 3 independent transfections is plotted with error bars representing sd. **(D-F)** Western blots were performed against the indicated affinity tag and mouse proteins (Myc, TdT, Actin or RNaseH2A) in the indicated MEF cell lines.
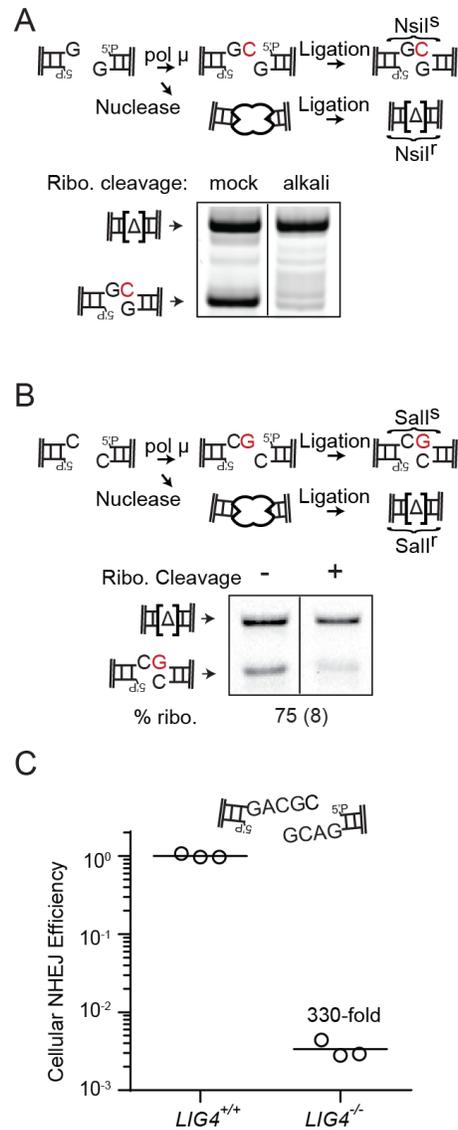
Figure 3.3: Detection of ribonucleotides in cellular NHEJ products

(A-B) Repair products were recovered after 1 minute as in Fig. 1E. **(A)** %ribo. determined by comparison of samples treated with alkali and heat to mock treated. **(B)** %ribo. of +G product determined as in Fig. 1E. Ribonucleotide detection is shown as the mean ± sd of 3 independent transfections. **(C)** The indicated NHEJ substrate was introduced into cells either proficient or deficient in LIG4. Repair products were amplified by qPCR. All data points and the mean repair efficiency are shown.
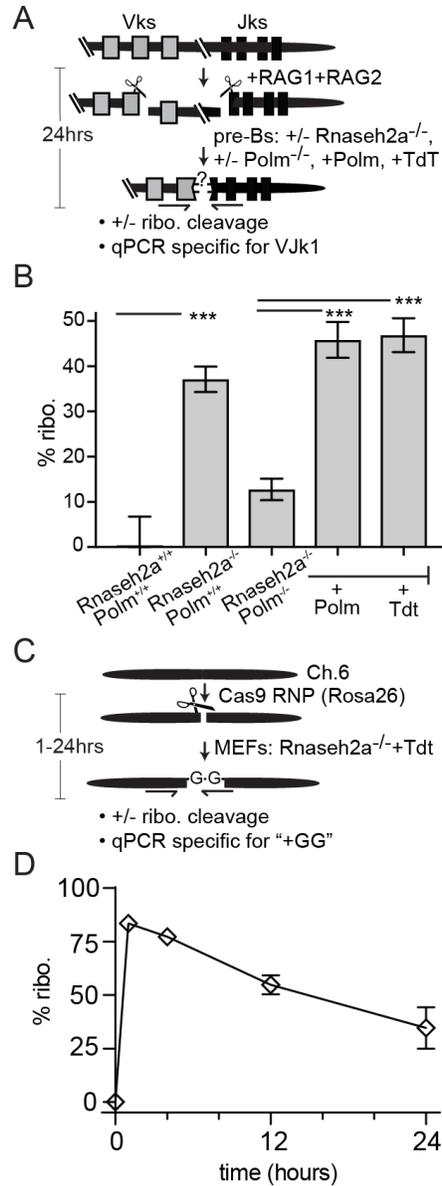
Figure 3.4: Ribonucleotide incorporation during repair of chromosomal breaks

(A-B) SP9 pre-B cells were induced for 24 hours, resulting in expression of RAG1+2 nuclease and introduction of chromosome breaks adjacent to V_K and J_K coding segments (boxes). % ribo. in VJ_K coding junctions was measured as in Fig. 1A. **(B)** Data points are the mean of five independent inductions, and error bars represent sd. **(C-D)** *Rosa26* locus-targeting Cas9 ribonucleoprotein was introduced into MEFs deficient in *Rnaseh2a* and expressing TdT. % ribo. was detected as in Fig. 1A using a qPCR specific for the TdT-dependent +GG product (also Fig. S4). **(D)** Data points are the mean of 3 independent transfections, and error bars represent sd.

A qPCR Standards: SP-9 pre-B cell V_κ locus

Slope:-1.007
R^2:0.9999

B SP-9 pre-B cells

C SP-9 pre-B cells

D SP-9 pre-B cells

E
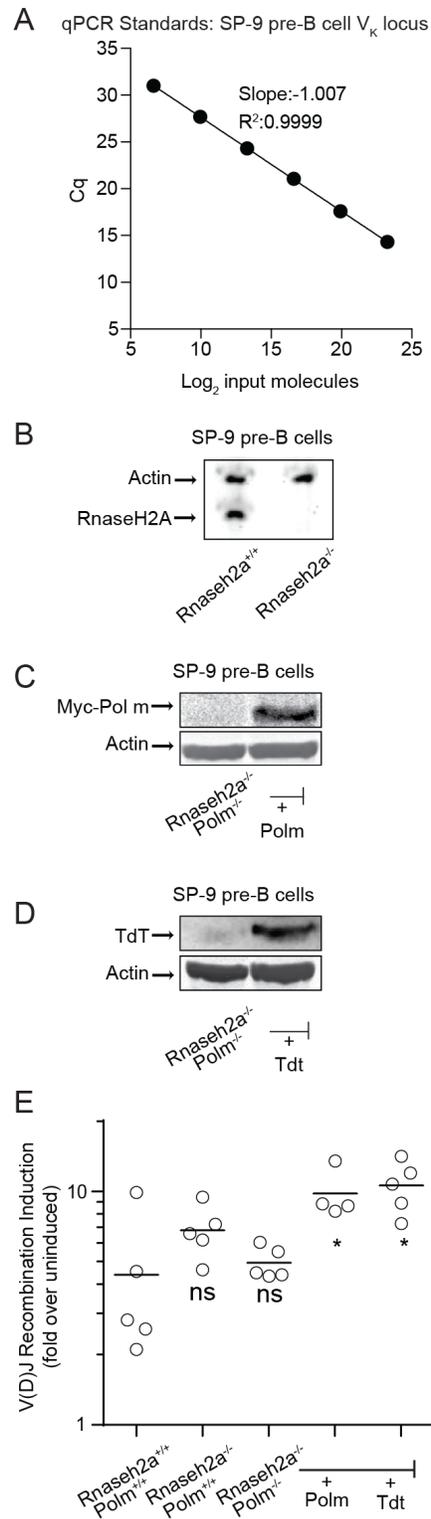
Figure 3.5: Detecting ribonucleotides in V(D)J recombination

Figure 3.5: Detecting ribonucleotides in V(D)J recombination

(A) A model amplicon of V(D)J recombination products at the murine V$_K$ locus was amplified by qPCR. Data represent the mean quantitation cycle (Cq) from 3 independent experiments and error bars represent the sd. **(B-D)** Western blots were performed against the indicated affinity tag and murine proteins (Myc, TdT, Actin or RNaseH2A) in the indicated SP9 cell lines. **(E)** V(D)J recombination was induced for 24 hours in SP9 pre-B cells of the indicated genotypes and induction was measured by qPCR across the VJ$_K$ junction. All data points and the mean induction level are shown. Experiments were compared to the wild type parental line by ANOVA with p values corrected for multiple comparisons, and are reported as ns (no significant difference from wild type parental line), or * (significantly different from parental line, $p < 0.05$).

A qPCR Standards: MEF Rosa26 +GG product

Slope:-1.044
R^2:0.9987
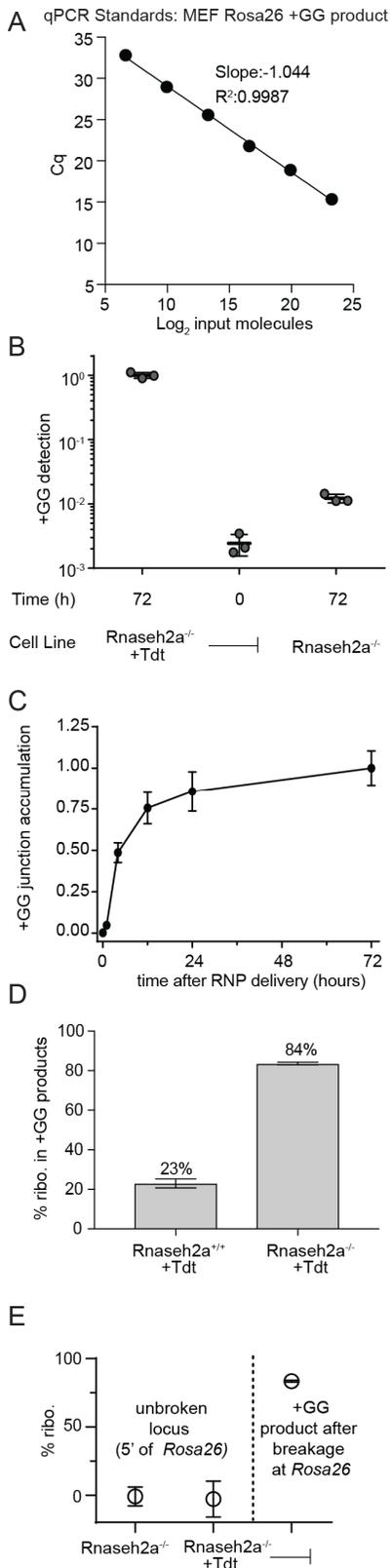
Figure 3.6: Ribo-NHEJ facilitates genome engineering.

Figure 3.6: Ribo-NHEJ facilitates genome engineering.

(A) A serial dilution of a model amplicon of the TdT-dependent +GG repair CRISPR repair product was used as a template for qPCR. Data represent the mean Cq from 3 independent experiments and error bars represent the sd. **(B-E)** CRISPR break repair assay was performed as described in Fig. 2C and **(B)** +GG repair products were assessed 0 or 72 hours after introduction of Cas9 RNP, with or without TdT expression. Results from each of triplicate transfections are shown, and error bars represent the sd from the mean. **(C)** Accumulation of +GG repair products over the indicated amounts of time using TDT-expressing *RnaseH2a*$^{-/-}$ MEF cells. Mean product recovery efficiency, relative to 72 hours, for 3 independent transfections is plotted with error bars representing sd. **(D)** %ribo. in +GG products recovered after 1 hour from TDT-expressing MEF cells either proficient or deficient in *Rnaseh2a*. **(E)** %ribo. detected at an uncut genomic control locus or in +GG products as in Fig. 2D. **(D-E)** Data represent the mean ribonucleotide detection in 3 independent experiments and error bars represent the sd.
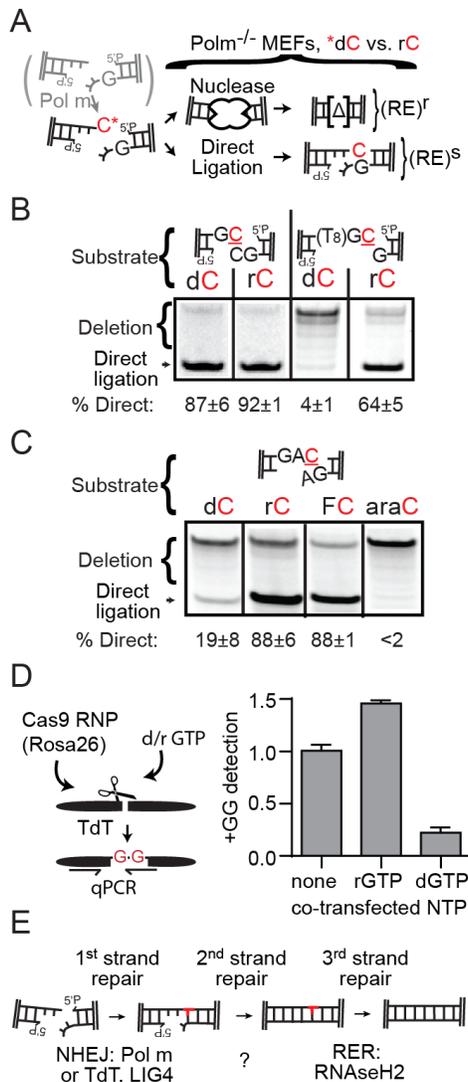
Figure 3.7: Impact of ribonucleotide termini on the NHEJ ligation step

(A-B) Termini of NHEJ substrates were varied to be consistent with polymerase-dependent addition of a ribonucleotide vs. a deoxonucleotide, and introduced into *Polm-/-* MEFs. Sensitivity of amplified products to a diagnostic restriction enzyme (RE) was used to identify examples of direct head-to-tail ligation. **(B)** The mean % directly ligated products for three independent transfections, ± sd, is noted below. (**C**) Substrates with terminal nucleotides varied to have dC, rC, araC, or FC were introduced into *Polm-/-* MEFs, and directly ligated products quantified as in Fig. 3A-B. **(D)** Deoxy- or ribo-guanine triphosphate were added to *Rosa26* Cas9-sgRNP transfections performed as in Fig. 2C-D, and genomic DNA was harvested after 1 hour. Data are the mean +GG recovery from 4 transfections and error bars are standard deviations. **(E)** Triple strand break repair model.
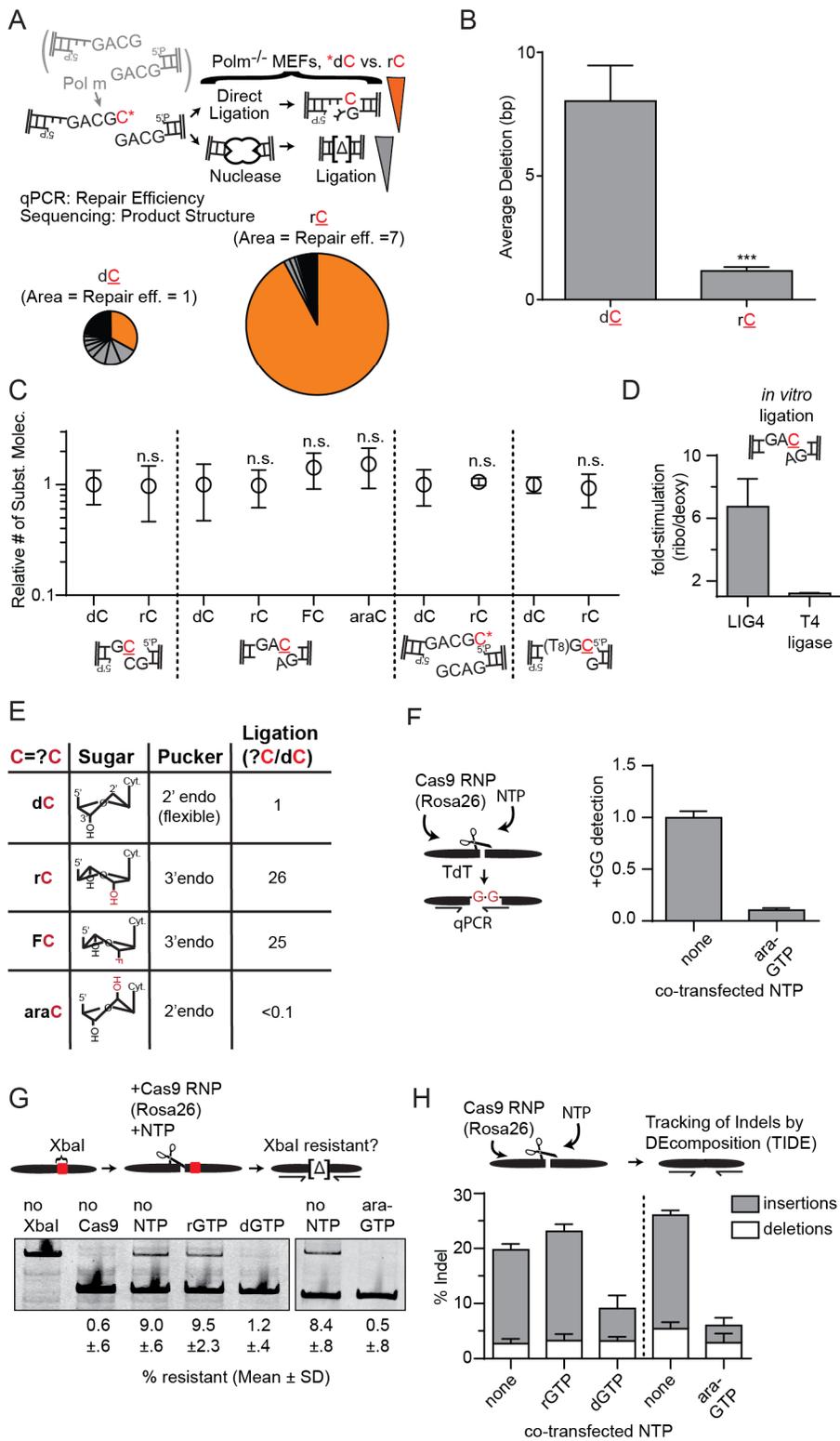
Figure 3.8: Ribonucleotides enable direct ligation of complex end structures.

Figure 3.8: Ribonucleotides enable direct ligation of complex end structures.

(A-B) Substrates with complex ends and either a terminal ribonucleotide or deoxynucleotide were introduced into polymerase-deficient MEF cells and repair product structures were analyzed by high throughput sequencing. **(A)** Repair products were categorized as either directly joined (orange) or deletion products (gray). The area of the rC pie graph was made proportional to the mean joining efficiency for this substrate relative to the joining efficiency for the dC substrate, as determined by qPCR. **(B)** The deletion size was determined for each recovered product and averaged for rC vs. dC substrates. Data represent the means of three independent experiments, and error bars represent sd. **(C)** Indicated substrates were introduced into MEF cells as in Table 1. The total amount of substrate introduced into cells was quantified by qPCR. Data represent the means of 3 independent transfections and error bars represent s.d. **(D)** *in vitro* ligation reactions were performed on the indicated substrate using either NHEJ proteins (Ku, XLF, XRCC4, LIG4) or T4 DNA ligase. Ligation was quantified by qPCR and data represent the mean ligation stimulation conferred by a ribonucleotide, from 3 independent ligation reactions with error bars representing s.d. **(E)** Structural models and sugar puckers of substrates used in Fig. 3C to assess the mechanism by which ribonucleotides benefit complex end ligation. **(F-H)** *Rosa26*-targeted Cas9 RNP was introduced into MEF cells as in Fig. 2C-D, along with nucleotide triphosphates as indicated. Genomic DNA was harvested 1 hour after transfection. **(F)** TdT-specific +GG product was detected and analyzed as in Fig. 2C-D. **(G)** The targeted locus was amplified by PCR and digested with XbaI to remove non-mutated products. Mutated products were quantified as the mean ratio of XbaI-resistant amplicon in 3 independent transfections with error bars representing s.d. **(H)** Tracking of Indels by DEcomposition (TIDE) was performed on recovered repair products. Briefly, the targeted locus was amplified and sanger sequenced, and sequence degeneration was used to quantify insertions and deletions from 3 independent transfections.

Substrate	Terminal **C**	Relative Joining efficiency[#]	% Direct ligation	Ligation Stimulation[+]
--G**C** [-- --] CG--	deoxy**C** (ribo)**C**	1 1.2 ± 0.7	87 ± 6 92 ± 1	1 1.3
--GA**C** [-- --] AG--	deoxy**C** (ribo)**C** 2'fluoro**C** arabino**C**	1 5.8 ± 1.9 5.4 ± 3.9 1.1 ± 0.6	19 ± 8 88 ± 6 88 ± 1 < 2*	1 26 25 < 0.1*
--GCAG**C** [-- --] GCAG--	deoxy**C** (ribo)**C**	1 6.3 ± 3.1	32 ± 3 88 ± 3	1 17
--TTTTTTTG**C** [-- --] G--	deoxy**C** (ribo)**C**	1 3.0 ± 0.9	4 ± 1 64 ± 5	1 48

[#] Joining efficiency r<u>C</u>/joining efficiency dC, as measured by qPCR that amplifies all NHEJ products.

*Direct ligation product was undetectable

[+]Ligation stimulation=Relative joining efficiency multiplied by %Direct ligation r<u>C</u>/%Direct ligation d<u>C</u>

Table 3.1: Stimulation of NHEJ repair pathway by a terminal ribonucleotide

5' flank	Insertion	3' flank	Frequency % (sd)	
TCTTTCTAGA		AGATGGGCGG	+TdT (sd)	-TdT (sd)
TCTTTCTAGA	GG	AGATGGGCGG	13 (1.1)	0.4 (0.16)
TCTTTCTAGA	CC	AGATGGGCGG	5.2 (0.15)	0.2 (0.04)
TCTTTCTAG-		AGATGGGCGG	4.2 (1.0)	32 (1.7)
TCTTTCTAGA	G	AGATGGGCGG	4.4 (0.52)	0.3 (0.05)
TCTTTCTAGA	A	AGATGGGCGG	3.3(0.31)	5.5 (0.51)
TCTTTCTAGA		AGATGGGCGG	2.9 (0.15)	8.9 (3.5)
TCTTTCTAG-	GG	AGATGGGCGG	2.4 (0.20)	0.1 (.007)
TCTTTCT---		AGATGGGCGG	1.9 (0.61)	12.3 (0.6)
TCTTTCTAGA	AA	AGATGGGCGG	1.7 (0.02)	0.09 (0.02)
TCTTTCTAGA	GGG	AGATGGGCGG	1.6 (0.01)	0.05 (0.02)

Table 3.2: Frequencies of Cas9-sgRosa26 repair products in TdT-expressing MEFs

Substrate Construction Oligonucleotides (5' to 3') *_Italics_ = deoxy/ribo*	
Core	
CAAGTGGTCTCAGACTGGCTACCCTGCTTCTTTGAGCATTTCTGAAACTATCACTTGTGTTTATTATTACACTGGCATTCATTCTCCAGAGAACATGTCTAGCCTATTCCCAGCTTTGCTTACGGAGTTACTCTGTATCTTTGCCTTGGAGAGTGCCAGAATCTGGTTTCAGAGTAAGATTTTATACATCATTTTTAGACATAGAAGCCACAGACATAGACAACGGAAGAAAGAGACTTTGGATTCTACTTACGTTTGATTTCCCTGACGGAGACCTCGGC	
Substrates Generated by PCR and Restriction Digest	
G Substrate Forward	GTACCAAGTGGACCACATGTCTTAGCTGTATAGTCAGGGA
G Substrate Reverse	GTACGCCGCCGACGCCATGTCACACCCATCTCAGACTGGC
C Substrate Forward	CAAGTGGACCAGACGTCTTAGCTGTATAGTCAGGGAAATC
C Substrate Reverse	CCGCCGACGCGACGTCACACCCATCTCAGACTGGCTACCC
Substrates Generated by Cap Annealing and Ligation	
CG / CG Left Cap Top Strand	AGTCTGAGATGGGTGTCATG*C*
CG / CG Left Cap Bottom Strand	ATGACACCCATCTCA
CG / CG Right Cap Top Strand	TGACTATACAGCTAAGGTCATGC
CG / CG Right Cap Bottom Strand	ATGACCTTAGCTGTATA
CAG / AG Left Cap Top Strand	AGTCTGAGATGGGTGTGTCGA*C*
CAG / AG Left Cap Bottom Strand	GACACACCCATCTCA
CAG / AG Right Cap Top Strand	TGACTATACAGCTAAGGTGTCGA
CAG / AG Right Cap Bottom Strand	GACACCTTAGCTGTATA
CGCAG / GCAG Left Cap Top Strand	AGTCTGAGATGGGTGCCACGACG*C*
CGCAG / GCAG Left Cap Bottom Strand	GTGGCACCCATCTCA
CGCAG / GCAG Right Cap Top Strand	TGACTATACAGCTAAGCCCACGACG
CGCAG / GCAG Right Cap Bottom Strand	GTGGGCTTAGCTGTATA
CGTTTTTTTT / G Left Cap Top Strand	AGTCTGAGATGGGTGTGCCATTTTTTTTG*C*
CGTTTTTTTT / G Left Cap Bottom Strand	TGGCACACCCATCTCA
CGTTTTTTTT / G Right Cap Top Strand	TGACTATACAGCTAAGTGCG
CGTTTTTTTT / G Right Cap Bottom Strand	GCACTTAGCTGTATA
PCR Primer Oligonucleotides	
PCR-Digest assay to detect extrachromosomal substrate repair products	
Forward Primer	CTTACGTTTGATTTCCCTGACTATACAG

Reverse Primer	GCAGGGTAGCCAGTCTGAGATG
TaqMan assay to measure repair efficiency of extrachromosomal substrates	
Forward Primer	CCACAGACATAGACAACGGAAG
Reverse Primer	ACACAAGTGATAGTTTCAGAAATGC
Probe (FAM-ZEN)	TCTCAGACTGGCTACCCTGCTTCT
TaqMan assay for V(D)J recombination products at murine IgK locus	
Forward Primer	GGTTTAGTGGCAGTGGGTCTGGGAC
Reverse Primer	CTTTGCCTTGGAGAGTGCCAGAATC
Probe (FAM-ZEN)	AGCCACAGACATAGACAACGGAAGA
TaqMan assay for TdT-dependent CRISPR repair product (+GG)	
Forward Primer	TCAGTTGGGCTGTTTTGGAG
Reverse Primer	GAAGACTCCCGCCCATCACC
Probe (FAM-ZEN)	TCAGTAAGGGAGCTGCAGTGGAGTA
High-throughput sequencing library preparation	
Substrate Forward Primer	(INDEX)CTTACGTTTGATTTCCCTGACTATACAG
Substrate Reverse Primer	(INDEX)GCAGGGTAGCCAGTCTGAGATG
Adapter Top Strand	GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
Adapter Bottom Strand	ACACTCTTTCCCTACACGACGCTCTTCCGATCT
Enrichment Forward Primer	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT
Enrichment Reverse Primer	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCT
SYBR Green assay to detect unrepaired and repaired extrachromosomal substrates	
Forward Primer	GGCACTCTCCAAGGCAAAGA
Reverse Primer	ACATGTCTAGCCTATTCCCGGCTT
PCR Primers to amplify targeted *Rosa26* locus for product structure analysis	
Forward Primer	GGCGGATCACAAGCAATAAT
Reverse Primer	TCAGTTGGGCTGTTTTGGAG
sgRNA Targets *lower case = PAM	
Mouse *Rnaseh2a* Target Guide 1	GCCACTTTCCCCACGGGCCTagg
Mouse *Rnaseh2a* Target Guide 2	TTTCTGCAGCCTGGGCAGACagg
Mouse *Polm* Target Guide 1	CAAGGTAGATGGCCACATCCggg
Mouse *Polm* Target Guide 2	CGCGAATGGGCCGCAGCCGCcgg
Mouse *Rosa26* Target Guide	ACTCCAGTCTTTCTAGAAGAtgg

Table 3.3: Sequences of DNA Reagents

CHAPTER 4: DISCUSSION

Chromosomal double strand breaks are highly toxic genomic lesions that play a central role in human health and disease. These highly heterogeneous breaks often have associated distortions that block repair. The NHEJ pathway overcomes this daunting challenge in a unique way, as it is the only repair pathway that does not utilize extensive resection to bypass the problem of substrate diversity[4]. Instead, NHEJ enzymes have evolved remarkable mechanistic flexibility. The NHEJ ligase and polymerases are the most flexible enzymes of their respective classes, in terms of their ability to engage a wide range of unorthodox substrates that other ligases and polymerases cannot act upon[5]. The basis of this flexibility and its importance for biological DSB repair were unknown.

My research has uncovered two mechanisms employed by NHEJ enzymes to address the challenge of substrate diversity: a DNA ligase that promotes a more mobile repair complex, allowing for flexible repair, and two "DNA" polymerases that actually use RNA to allow ligation of otherwise irreparable breaks. In Chapter 2, my work shows that LIG4 is specialized to directly ligate broken ends with adjacent damage and mispairs; that this specialization is attributable to an increase in the repair complex's mobility; and that this flexibility promotes cellular resistance to ionizing radiation. I present data in Chapter 3 revealing that pol μ and TdT are actually RNA polymerases in cells; that their RNA incorporation confers a ligation advantage to LIG4; and that this mechanism alters

the repair efficiency and product structures of CRISPR-Cas9 induced breaks. Taken together, my findings provide a clear picture of two novel mechanisms that facilitate repair of diverse chromosome breaks, establishing these mechanisms as central to the critical biological functions of the NHEJ pathway.

4.1 Tolerance of Mispairs and Damage at the Ligation Step

The final step in NHEJ is the rejoining of ends by LIG4, which is the only DNA ligase that acts in this pathway[52]. By directly comparing joining activity of different ligases on complementary and incompatible ends, previous *in vitro* studies have demonstrated that LIG4 is more tolerant of mispaired ends than other ligases[60]. My work identifies the mechanistic basis of this specialized activity: I report that LIG4's ability to join damaged and mispaired ends requires a conserved structural element, insert1, that is fully unique to this ligase (Figure 2.1B, Figure 2.5, Figure 2.6).

A major limitation of previous studies of DNA ligases in end joining is that these studies have compared ligases in the absence of other factors[60,126]. However, biological activity of LIG4 requires Ku, XRCC4, and XLF; other ligases do not interact with these proteins. To address this shortcoming, I generated a chimeric DNA ligase consisting of the catalytic core of LIG3 and the interaction domains of LIG4 (Figure 2.2E). I found that this chimera interacts with other NHEJ proteins and maintains full activity on complementary ends, but that it is even more impaired than LIG4$^{\Delta i}$ on ends with associated damage (Figure 2.2E). My work is thus the first to compare the NHEJ activity of different DNA ligases.

4.2 Structural Basis of Damage Tolerance

My data suggest that other structural elements of LIG4 contribute to damage tolerance and provide a starting point for a detailed interrogation of ligase biochemistry (Figure 2.2E). The clear next step in this line of investigation is the generation of additional chimeric ligases from each of the 3 catalytic domains of LIG1, LIG3, and LIG4. Instead of substituting the entire catalytic core, individual domains can be fused to isolate each potentially significant part of the enzyme. Studies of LIG1 further suggest that activity can be reconstituted by adding the domains of ligases *in trans*[56].

Of particular interest are the adenylation domain that contains the active site and the DNA binding domain of LIG4 that contains insert1, since I show that this insert is particularly important for joining complex end structures. With respect to insert1, I propose that it acts as a hinge to help LIG4 maintain a closed, substrate-bound conformation via protein-protein interactions between insert1 and the OB-fold domain. This proposition generates testable hypotheses about mutating residues opposite insert1 in the OB-fold domain, or substituting the OB-fold domain of LIG4 with that of another ligase.

Our understanding of LIG4 has been limited by the lack of a substrate-bound crystal structure. My work predicts that such a structure would confirm an important role for insert1. Arginine 113 is the most conserved residue in insert1 and is identical in LIG4 from humans to plants (Figure 2.2A); it could be solely responsible for the effects of the LIG4$^{\Delta i}$ mutant.

4.3 Mobilization of Ends within Repair Complexes

Formation of the NHEJ paired end complex (PEC) implicitly involves the bridging of DSB ends by NHEJ proteins. Previous studies have identified core NHEJ factors as essential for this end bridging, including Ku, LIG4, XRCC4, and XLF[46–49]. Super-resolution imaging of fixed cells has confirmed these studies and additionally identified the role of multiprotein filaments that form along the break site and assist in end bridging[51]. Interestingly, the role of LIG4 in end bridging is distinct from its ligation activity, as a catalytic dead LIG4 which cannot ligate can still bridge ends in pulldown assays[109].

My work builds upon these foundational studies to establish LIG4 as a "sensor" of end chemistry within PECs. In response to differences in end structure, LIG4 mobilizes the entire PEC to induce dramatic changes in the conformation of the ends (Figure 2.3C-D, Figure 2.4). This mobilization also enables LIG4 to join damaged and mispaired end structures, and both activities require insert1 (Figure 2.1B, Figure 2.5, Figure 2.6). Unexpectedly, I found that LIG4 enables nucleolytic processing of 3' flaps, indicating that LIG4 can tailor processing steps based on end structure (Figure 2.7, Figure 2.8). Moreover, FRET states observed in single molecule analysis suggest that LIG4 physically extrudes these flaps so that a nuclease can clip them (Figure 2.7C). A previous *in vitro* study suggests Artemis may be the nuclease responsible for this clipping in cells[67]. My data provide a clear substrate context for testing this hypothesis.

I proposed that insert1 allows LIG4 to maintain a closed, substrate-bound conformation, and that this closed conformation allows for increased PEC mobility in response to differences in end structure. Further analysis of single molecule dynamics

within PECs could test this hypothesis directly. The DNA binding domain and OB-fold domain are the domains furthest apart within the ligase catalytic core. Each could be tagged with FRET probes to discern between open and closed conformations, in the presence or absence of insert1, and these conformations could then be directly linked with PEC mobility.

4.4 Biological Significance of Ligation Flexibility

Since NHEJ is the only pathway that repairs damaged and mispaired ends without resection, there has been significant interest in identifying factors required for this unique activity. This prior work has implicated NHEJ polymerases, nucleases, PAXX, and filaments formed by XRCC4 and XLF in repair of complex ends[33,59,67,72,103–106]. Attempts to test the role of LIG4 in complex end repair have focused on *in vitro* systems that fail to recapitulate the genetic requirements for NHEJ[60,126]. My work is the first to assess the biological contribution of LIG4 specialization to complex end repair in cells.

The biopharmaceutical research community has significant interest in drugging DNA ligases as cancer therapy agents, including LIG4[111]. In fact, targeting DNA repair machinery has already shown success in treating cancers with inborn DNA repair defects, such as BRCA-defective diseases[127]. Development efforts have focused on the potential of these drugs to sensitize tumors to ionizing radiation. To this end, I have shown that cells lacking LIG4 insert1 are extremely sensitive to ionizing radiation (Figure 2.9, Figure 2.10). Insert1 thus presents a promising therapeutic target; since it is

completely unique to LIG4, it is less likely to engender the off-target effects seen with existing ligase inhibitors[112].

Single molecule dynamics revealed that LIG4 enables mobilization of ends in response to differences in end chemistry, and analysis of cellular sensitivity to damaging agents revealed that this mobilization is critical for repair in biological contexts. My work suggests that the dynamic alignment of two DSB ends within the paired end complex can ultimately affect whether a cell will live or die as a result of a break (Figure 2.11).

4.5 Pol µ and TdT are RNA polymerases in cells

In addition to using a ligase specialized to join diverse end structures, NHEJ has evolved the ability to employ 3 X-family DNA polymerases to process ends and thereby render substrates more amenable to ligation[69]. Traditional RNA-DNA transactions adhere strictly to the central dogma of molecular biology: DNA is transcribed into RNA via templated synthesis. The most noteworthy exception is synthesis in the opposite direction, synthesizing DNA from an RNA template[128,129]. Other exceptions, including use of RNA as a primer for DNA synthesis, carefully avoid incorporating RNA into the genome[130].

Two of NHEJ's specialized polymerases, pol µ and TdT, have highly unusual biochemical properties, including uniquely poor discrimination against ribonucleotides *in vitro*[37,87–90]. The studies presented here show that pol µ and TdT are primarily RNA polymerases during the repair of chromosome breaks in cells (Figure 3.4). This finding

is significant because it is the first example of RNA being preferentially incorporated into a DNA genome.

4.6 NHEJ and Ribonucleotide Excision Repair

The preferential incorporation of RNA into the mammalian genome by NHEJ polymerases was unanticipated because genomic ribonucleotides are toxic to cells[86]. RNA is more reactive than DNA; its labile 2' hydroxyl group leaves the genome vulnerable to hydrolytic breakage and thus embedded RNA contributes to genome instability[86]. To address this problem, evolution has produced the ribonucleotide excision repair (RER) pathway mediated by a heterotrimeric ribonuclease, RNase H2[86].

The primary cellular role of the RER pathway is thought to be repair of RNA incorporated by replicative polymerases during genome synthesis. Our data identify a novel role for RER in the removal of ribonucleotides incorporated by pol μ and TdT during chromosome break repair (Figure 3.1B, Figure 3.4B, Figure 3.6D). RNase H2 carries out its role in NHEJ rapidly; RNA-containing products of extrachromosomal repair are undetectable after only 20 minutes (Figure 3.1B). The speed of RNA removal could be attributable to a physical interaction between NHEJ complexes and RNase H2. Such interactions could readily be assessed with pulldown assays in cells and electrophoretic mobility shifts *in vitro.* It also remains unclear if RER functions in this new context as it does outside of NHEJ: in canonical RER, strand displacement synthesis and ligation follow the RNase H2-mediated nicking of the ribonucleotide 5' terminus[86].

Perhaps the most significant implication of the NHEJ-RER interaction is the potential for a triple strand break repair model in which the first strand is repaired using RNA, the second (opposite) strand is repaired with only DNA, and the original first strand is then repaired again by RER (Figure 3.7E). We propose that second strand repair precedes RER because RNase H2 would re-introduce the double strand break unless the second strand is repaired before RER. This model generates testable hypotheses about the regulation of RNase H2: its activity must be stalled by some mechanism until the second strand is repaired.

4.7 Ribonucleotides Enable Ligation

Since ribonucleotide incorporation can potentially destabilize the genome, we reasoned that it might carry some benefit to NHEJ that offsets its potential costs. Indeed, we found that terminal ribonucleotides enable ligation of ends with adjacent mispairs both *in vitro* and in cells (Figure 3.7B, Figure 3.8C, Table 3.1). Furthermore, if a deoxynucleotide is incorporated instead of a ribonucleotide, such ends are repaired less efficiently and less accurately (Figure 3.8A-B, Table 3.1). I further demonstrated that this effect is specific to LIG4: T4 bacteriophage ligase does not receive any ligation benefit from a terminal ribonucleotide *in vitro* (Figure 3.8C). Future work will identify the mechanistic basis of LIG4 specialization for 3' terminal ribonucleotides. To this end, my experiments from Chapter 2 provide a framework for interrogating structural elements and activities unique to this ligase.

In existing crystal structures of DNA ligases, nucleotides around the DNA nick are twisted from their normal B-form (B-DNA) into the noncanonical, RNA-like A-form

(A-DNA)[131]. This structural conformation closely positions the unjoined 3' and 5' termini and may be required for the chemistry of ligation. The transition from B-DNA to A-DNA is attributable to the sugar pucker of the nucleotides involved: an RNA-like C3'-endo sugar pucker promotes A-DNA, while a DNA-like C2'-endo pucker promotes B-DNA[131]. To test if this mechanism is involved in ribonucleotide stimulation of LIG4, we generated substrates with terminal cytarabine (Ara-C) nucleotides (Figure 3.8E). Ara-C differs from ribo-cytidine (ribo-C) only in the chirality of its 2' center, and thus adopts a DNA-like C2'-endo pucker. In accordance with structural studies, we found that Ara-C termini fully block direct ligation of NHEJ substrates (Figure 3.7C). Ara-C is a commonly used cancer therapy agent and future studies will elucidate whether its incorporation into NHEJ substrates is a potential mechanism of action[132].

4.8 Ribonucleotides in CRISPR-Cas9 Repair Products

Targeted endonucleases have proven useful for precisely editing the genome. In particular, there has been tremendous interest in modifying the CRISPR-Cas9 system to alter DSB induction by Cas9 and repair by the cellular machinery[20,21]. Because the majority of Cas9-induced breaks are repaired by NHEJ in mammalian cells, inhibition of the NHEJ pathway has been investigated as a way of modulating genome engineering outcomes[133]. I found that adding deoxy- or ribonucleotides to cells expressing TdT can alter the efficiency and accuracy of repair of Cas9-induced breaks (Figure 3.7D, Figure 3.8F-H). Specifically, ribonucleotides increase the recovery of specific TdT-dependent insertions, while deoxynucleotides impair recovery of mutated repair products altogether.

I assessed repair at early time points because nucleotide pool perturbations are necessarily transient in mammalian cells[134]. A deeper kinetic profile of the effect of nucleotide supplementation on Cas9 repair products would yield useful insight. Deletion products accumulate more slowly than insertion products in NHEJ[11]; do deletions thus rescue the inhibitory effect of deoxynucleotides at later time points, does nucleotide pool stabilization diminish the effect of nucleotide supplementation, or is the inhibitory effect durable over time? Sequencing of repair products recovered from cells with or without NHEJ polymerases supplemented with different nucleotides would also help assess the significance of this Cas9-induced break repair phenomenon.

4.9 Biotechnology Applications

Further experimentation will also focus on the development of this strategy as a biotechnology tool for genome engineering. Of particular interest is the fact that both pol μ and TdT recombinant protein can be purified in milligram quantities and electroporated into cells intact and active, along with nucleotide triphosphates. My experiments were performed in cells expressing TdT, warranting further experimentation with recombinant protein electroporation. These proteins can also be introduced at levels vastly exceeding what can be expressed in cells. The impact of ribonucleotides and deoxynucleotides on repair may be responsive to increases in polymerase concentration.

I showed that supplementation with ribonucleotides increases the recovery of a single, specific, polymerase-dependent repair product (Figure 3.7D). Therefore, TdT and ribonucleotides could potentially be used to site-specifically engineer 1-3 base pair

insertions of a desired nucleotide into the genome. If this proves effective, it could become simple to introduce targeted early stop codons. TdT is also appreciated in molecular biology for its ability to incorporate derivatized nucleotides *in vitro*[135,136]. Harnessing my results, these experiments can potentially be done in cells to incorporate a desired derivatized nucleotide into a specific genomic site.

4.10 Concluding Remarks

To preserve the integrity of the genome, the NHEJ pathway must repair double strand breaks with a variety of end structures. The work presented in this dissertation identifies two mechanisms that allow this pathway to carry out repair with remarkable flexibility. I found that the NHEJ ligase is uniquely adapted to work on highly unusual substrates, and that this adaptation is essential for cells to survive genomic damage. I also showed that NHEJ—bizarrely—incorporates RNA into the genome to achieve repair of otherwise irreparable breaks. Both of these mechanisms are integral for NHEJ to overcome its central challenge: the repair of complex breaks. In addition to advancing the field of DNA repair, my findings have clear and direct biotechnology applications.

REFERENCES

1.   Kaplan, H. S. DNA-strand scission and loss of viability after x irradiation of normal and sensitized bacterial cells. *Proc. Natl. Acad. Sci. U. S. A.* **55,** 1442–6 (1966).

2.   Toczyski, D. P., Galgoczy, D. J. & Hartwell, L. H. CDC5 and CKII control adaptation to the yeast DNA damage checkpoint. *Cell* **90,** 1097–106 (1997).

3.   Beucher, A. *et al.* ATM and Artemis promote homologous recombination of radiation-induced DNA double-strand breaks in G2. *EMBO J.* **28,** 3413–3427 (2009).

4.   Waters, C. A., Strande, N. T., Wyatt, D. W., Pryor, J. M. & Ramsden, D. A. Nonhomologous end joining: a good solution for bad ends. *DNA Repair (Amst).* **17,** 39–51 (2014).

5.   Gu, J. & Lieber, M. R. Mechanistic flexibility as a conserved theme across 3 billion years of nonhomologous DNA end-joining. *Genes Dev.* **22,** 411–5 (2008).

6.   San Filippo, J., Sung, P. & Klein, H. Mechanism of Eukaryotic Homologous Recombination. *Annu. Rev. Biochem.* **77,** 229–257 (2008).

7.   Wyatt, D. W. *et al.* Essential Roles for Polymerase θ-Mediated End Joining in the Repair of Chromosome Breaks. *Mol. Cell* **63,** 662–673 (2016).

8.   Ceccaldi, R. *et al.* Homologous-recombination-deficient tumours are dependent on Polθ-mediated repair. *Nature* **518,** 258–62 (2015).

9.   Chapman, J. R., Taylor, M. R. G. & Boulton, S. J. Playing the end game: DNA double-strand break repair pathway choice. *Mol. Cell* **47,** 497–510 (2012).

10.   Woodbine, L., Gennery, A. R. & Jeggo, P. A. The clinical impact of deficiency in DNA non-homologous end-joining. *DNA Repair (Amst).* **16,** 84–96 (2014).

11.   Waters, C. a *et al.* The fidelity of the ligation step determines how ends are resolved during nonhomologous end joining. *Nat. Commun.* **5,** 4286 (2014).

12.   Guirouilh-Barbat, J. *et al.* Impact of the KU80 pathway on NHEJ-induced genome rearrangements in mammalian cells. *Mol. Cell* **14,** 611–23 (2004).

13.   Keeney, S., Giroux, C. N. & Kleckner, N. Meiosis-specific DNA double-strand breaks are catalyzed by Spo11, a member of a widely conserved protein family. *Cell* **88,** 375–84 (1997).

14.   Tonegawa, S. Somatic generation of antibody diversity. *Nature* **302,** 575–81 (1983).

15.   Mehta, A. & Haber, J. E. Sources of DNA double-strand breaks and models of

recombinational DNA repair. *Cold Spring Harb. Perspect. Biol.* **6,** a016428 (2014).

16. Costanzo, V. *et al.* Mre11 protein complex prevents double-strand break accumulation during chromosomal DNA replication. *Mol. Cell* **8,** 137–47 (2001).

17. Pommier, Y., Leo, E., Zhang, H. & Marchand, C. DNA Topoisomerases and Their Poisoning by Anticancer and Antibacterial Drugs. *Chem. Biol.* **17,** 421–433 (2010).

18. Iliakis, G. The role of DNA double strand breaks in ionizing radiation-induced killing of eukaryotic cells. *Bioessays* **13,** 641–8 (1991).

19. Wu, C.-C. *et al.* Structural Basis of Type II Topoisomerase Inhibition by the Anticancer Drug Etoposide. *Science (80-.).* **333,** 459–462 (2011).

20. Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339,** 823–6 (2013).

21. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339,** 819–23 (2013).

22. Kapp, D. S. & Smith, K. C. Chemical nature of chain breaks produced in DNA by x-irradiation in vitro. *Radiat. Res.* **42,** 34–49 (1970).

23. Ward, J. F. Biochemistry of DNA lesions. *Radiat. Res. Suppl.* **8,** S103-11 (1985).

24. Ward, J. F. & Kuo, I. Strand breaks, base release, and postirradiation changes in DNA gamma-irradiated in dilute O2-saturated aqueous solution. *Radiat. Res.* **66,** 485–98 (1976).

25. Lomax, M. E., Folkes, L. K. & O'Neill, P. Biological Consequences of Radiation-induced DNA Damage: Relevance to Radiotherapy. *Clin. Oncol.* **25,** 578–585 (2013).

26. Goodhead, D. T. Initial events in the cellular effects of ionizing radiations: clustered damage in DNA. *Int. J. Radiat. Biol.* **65,** 7–17 (1994).

27. Ward, J. F. The complexity of DNA damage: relevance to biological consequences. *Int. J. Radiat. Biol.* **66,** 427–32 (1994).

28. Tilly, N., Brahme, A., Carlsson, J. & Glimelius, B. Comparison of cell survival models for mixed LET radiation. *Int. J. Radiat. Biol.* **75,** 233–43 (1999).

29. Hozumi, N. & Tonegawa, S. Evidence for somatic rearrangement of immunoglobulin genes coding for variable and constant regions. *Proc. Natl. Acad. Sci. U. S. A.* **73,** 3628–32 (1976).

30. McBlane, J. F. *et al.* Cleavage at a V(D)J recombination signal requires only

RAG1 and RAG2 proteins and occurs in two steps. *Cell* **83,** 387–95 (1995).

31.   Schatz, D. G., Oettinger, M. A. & Baltimore, D. The V(D)J recombination activating gene, RAG-1. *Cell* **59,** 1035–48 (1989).

32.   Oettinger, M. A. *et al.* RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination. *Science* **248,** 1517–23 (1990).

33.   Ma, Y., Pannicke, U., Schwarz, K. & Lieber, M. R. Hairpin opening and overhang processing by an Artemis/DNA-dependent protein kinase complex in nonhomologous end joining and V(D)J recombination. *Cell* **108,** 781–94 (2002).

34.   Cabaniols, J. P., Fazilleau, N., Casrouge, A., Kourilsky, P. & Kanellopoulos, J. M. Most alpha/beta T cell receptor diversity is due to terminal deoxynucleotidyl transferase. *J. Exp. Med.* **194,** 1385–90 (2001).

35.   Frank, K. M. *et al.* Late embryonic lethality and impaired V (D)J recombination in mice lacking DNA ligase IV. *Nature* **396,** 173–177 (1998).

36.   Gao, Y. *et al.* A critical role for DNA end-joining proteins in both lymphogenesis and neurogenesis. *Cell* **95,** 891–902 (1998).

37.   Kato, K. I., Gonçalves, J. M., Houts, G. E. & Bollum, F. J. Deoxynucleotide-polymerizing enzymes of calf thymus gland. II. Properties of the terminal deoxynucleotidyltransferase. *J. Biol. Chem.* **242,** 2780–9 (1967).

38.   Motea, E. A. & Berdis, A. J. Terminal deoxynucleotidyl transferase: the story of a misguided DNA polymerase. *Biochim. Biophys. Acta* **1804,** 1151–66 (2010).

39.   Lee, S. E. *et al.* Saccharomyces Ku70, mre11/rad50 and RPA proteins regulate adaptation to G2/M arrest after DNA damage. *Cell* **94,** 399–409 (1998).

40.   Lieber, M. R. The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annu. Rev. Biochem.* **79,** 181–211 (2010).

41.   Falzon, M., Fewell, J. W. & Kuff, E. L. EBP-80, a transcription factor closely resembling the human autoantigen Ku, recognizes single- to double-strand transitions in DNA. *J. Biol. Chem.* **268,** 10546–52 (1993).

42.   Cary, R. B. *et al.* DNA looping by Ku and the DNA-dependent protein kinase. *Proc. Natl. Acad. Sci. U. S. A.* **94,** 4267–72 (1997).

43.   Aleksandrov, R. *et al.* Protein Dynamics in Complex DNA Lesions. *Mol. Cell* **69,** 1046–1061.e5 (2018).

44.   Nick McElhinny, S. A., Snowden, C. M., McCarville, J. & Ramsden, D. A. Ku recruits the XRCC4-ligase IV complex to DNA ends. *Mol. Cell. Biol.* **20,** 2996–

3003 (2000).

45. Weterings, E. & van Gent, D. C. The mechanism of non-homologous end-joining: a synopsis of synapsis. *DNA Repair (Amst).* **3,** 1425–35 (2004).

46. Mari, P.-O. *et al.* Dynamic assembly of end-joining complexes requires interaction between Ku70/80 and XRCC4. *Proc. Natl. Acad. Sci.* **103,** 18597–18602 (2006).

47. Ahnesorg, P., Smith, P. & Jackson, S. P. XLF Interacts with the XRCC4-DNA Ligase IV Complex to Promote DNA Nonhomologous End-Joining. *Cell* **124,** 301–313 (2006).

48. Grawunder, U., Zimmer, D., Kulesza, P. & Lieber, M. R. Requirement for an interaction of XRCC4 with DNA ligase IV for wild-type V(D)J recombination and DNA double-strand break repair in vivo. *J. Biol. Chem.* **273,** 24708–14 (1998).

49. Yano, K. & Chen, D. J. Live cell imaging of XLF and XRCC4 reveals a novel view of protein assembly in the non-homologous end-joining pathway. *Cell Cycle* **7,** 1321–5 (2008).

50. Graham, T. G. W., Walter, J. C. & Loparo, J. J. Two-Stage Synapsis of DNA Ends during Non-homologous End Joining. *Mol. Cell* **61,** 850–858 (2016).

51. Reid, D. A. *et al.* Organization and dynamics of the nonhomologous end-joining machinery during DNA double-strand break repair. *Proc. Natl. Acad. Sci.* **112,** E2575–E2584 (2015).

52. Lieber, M. R., Wilson, T. E. & Grawunder, U. Yeast DNA ligase IV mediates non-homologous DNA end joining. *Nature* **388,** 495–498 (1997).

53. Ellenberger, T. & Tomkinson, A. E. Eukaryotic DNA Ligases: Structural and Functional Insights. *Annu. Rev. Biochem.* **77,** 313–338 (2008).

54. Ellenberger, T. & Tomkinson, A. E. Eukaryotic DNA Ligases: Structural and Functional Insights. *Annu. Rev. Biochem.* **77,** 313–338 (2008).

55. Cotner-Gohara, E. *et al.* Human DNA ligase III recognizes DNA ends by dynamic switching between two DNA-bound states. *Biochemistry* **49,** 6165–76 (2010).

56. Pascal, J. M., O'Brien, P. J., Tomkinson, A. E. & Ellenberger, T. Human DNA ligase I completely encircles and partially unwinds nicked DNA. *Nature* **432,** 473–8 (2004).

57. Nair, P. A. *et al.* Structural basis for nick recognition by a minimal pluripotent DNA ligase. *Nat. Struct. Mol. Biol.* **14,** 770–778 (2007).

58. Schellenberg, M. J., Tumbale, P. P. & Williams, R. S. Molecular underpinnings of Aprataxin RNA/DNA deadenylase function and dysfunction in neurological

disease. *Prog. Biophys. Mol. Biol.* **117,** 157–165 (2015).

59. Gu, J., Lu, H., Tsai, A. G., Schwarz, K. & Lieber, M. R. Single-stranded DNA ligation and XLF-stimulated incompatible DNA end ligation by the XRCC4-DNA ligase IV complex: influence of terminal DNA sequence. *Nucleic Acids Res.* **35,** 5755–5762 (2007).

60. Gu, J. *et al.* XRCC4:DNA ligase IV can ligate incompatible DNA ends and can ligate across gaps. *EMBO J.* **26,** 1010–23 (2007).

61. Tsai, C. J., Kim, S. A. & Chu, G. Cernunnos/XLF promotes the ligation of mismatched and noncohesive DNA ends. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 7851–6 (2007).

62. Husain, I. *et al.* Purification and characterization of DNA ligase III from bovine testes. Homology with DNA ligase II and vaccinia DNA ligase. *J. Biol. Chem.* **270,** 9683–90 (1995).

63. Aniukwu, J., Glickman, M. S. & Shuman, S. The pathways and outcomes of mycobacterial NHEJ depend on the structure of the broken DNA ends. *Genes Dev.* **22,** 512–27 (2008).

64. Sriskanda, V. & Shuman, S. Specificity and fidelity of strand joining by Chlorella virus DNA ligase. *Nucleic Acids Res.* **26,** 3536–41 (1998).

65. Li, L. *et al.* A founder mutation in Artemis, an SNM1-like protein, causes SCID in Athabascan-speaking Native Americans. *J. Immunol.* **168,** 6323–9 (2002).

66. Musio, A. *et al.* Damaging-agent sensitivity of Artemis-deficient cell lines. *Eur. J. Immunol.* **35,** 1250–6 (2005).

67. Chang, H. H. Y. *et al.* Different DNA End Configurations Dictate Which NHEJ Components Are Most Important for Joining Efficiency. *J. Biol. Chem.* **291,** 24377–24389 (2016).

68. Moon, A. F. *et al.* The X family portrait: Structural insights into biological functions of X family polymerases. *DNA Repair (Amst).* **6,** 1709–1725 (2007).

69. Yamtich, J. & Sweasy, J. B. DNA polymerase Family X: Function, structure, and cellular roles. *Biochim. Biophys. Acta - Proteins Proteomics* **1804,** 1136–1150 (2010).

70. Ramadan, K., Shevelev, I. V, Maga, G. & Hübscher, U. De Novo DNA Synthesis by Human DNA Polymerase λ, DNA Polymerase μ and Terminal Deoxyribonucleotidyl Transferase. *J. Mol. Biol.* **339,** 395–404 (2004).

71. Nick McElhinny, S. A. *et al.* A Gradient of Template Dependence Defines Distinct Biological Roles for Family X Polymerases in Nonhomologous End Joining. *Mol.*

Cell **19,** 357–366 (2005).

72.   Pryor, J. M. *et al.* Essential role for polymerase specialization in cellular nonhomologous end joining. *Proc. Natl. Acad. Sci.* **112,** E4537–E4545 (2015).

73.   Moon, A. F. *et al.* Sustained active site rigidity during synthesis by human DNA polymerase μ. *Nat. Struct. Mol. Biol.* **21,** 253–60 (2014).

74.   Moon, A. F., Gosavi, R. A., Kunkel, T. A., Pedersen, L. C. & Bebenek, K. Creative template-dependent synthesis by human polymerase mu. *Proc. Natl. Acad. Sci. U. S. A.* **112,** E4530-6 (2015).

75.   Bollum, F. J. Terminal deoxynucleotidyl transferase as a hematopoietic cell marker. *Blood* **54,** 1203–15 (1979).

76.   Desiderio, S. V *et al.* Insertion of N regions into heavy-chain genes is correlated with expression of terminal deoxytransferase in B cells. *Nature* **311,** 752–5

77.   Bollum, F. J. Chemically Defined Templates and Initiators for Deoxypolynucleotide Synthesis. *Science (80-.).* **144,** 560–560 (1964).

78.   Kunkel, T. A., Gopinathan, K. P., Dube, D. K., Snow, E. T. & Loeb, L. A. Rearrangements of DNA mediated by terminal transferase. *Proc. Natl. Acad. Sci. U. S. A.* **83,** 1867–71 (1986).

79.   Gilfillan, S., Dierich, A., Lemeur, M., Benoist, C. & Mathis, D. Mice lacking TdT: mature animals with an immature lymphocyte repertoire. *Science* **261,** 1175–8 (1993).

80.   Lucas, D. *et al.* Altered Hematopoiesis in Mice Lacking DNA Polymerase μ Is Due to Inefficient Double-Strand Break Repair. *PLoS Genet.* **5,** e1000389 (2009).

81.   Bertocci, B., De Smet, A., Berek, C., Weill, J.-C. & Reynaud, C.-A. Immunoglobulin kappa light chain gene rearrangement is impaired in mice deficient for DNA polymerase mu. *Immunity* **19,** 203–11 (2003).

82.   Komori, T., Okada, A., Stewart, V. & Alt, F. W. Lack of N regions in antigen receptor variable region genes of TdT-deficient lymphocytes. *Science* **261,** 1171–5 (1993).

83.   Cavanaugh, N. A., Beard, W. A. & Wilson, S. H. DNA polymerase beta ribonucleotide discrimination: insertion, misinsertion, extension, and coding. *J. Biol. Chem.* **285,** 24457–65 (2010).

84.   Brown, J. A. *et al.* A novel mechanism of sugar selection utilized by a human X-family DNA polymerase. *J. Mol. Biol.* **395,** 282–90 (2010).

85.   Gosavi, R. A., Moon, A. F., Kunkel, T. A., Pedersen, L. C. & Bebenek, K. The

catalytic cycle for ribonucleotide incorporation by human DNA Pol λ. *Nucleic Acids Res.* **40,** 7518–27 (2012).

86.   Williams, J. S., Lujan, S. A. & Kunkel, T. A. Processing ribonucleotides incorporated during eukaryotic DNA replication. *Nat. Rev. Mol. Cell Biol.* **17,** 350–363 (2016).

87.   Boulé, J. B., Rougeon, F. & Papanicolaou, C. Terminal deoxynucleotidyl transferase indiscriminately incorporates ribonucleotides and deoxyribonucleotides. *J. Biol. Chem.* **276,** 31388–93 (2001).

88.   Nick McElhinny, S. A. & Ramsden, D. A. Polymerase mu is a DNA-directed DNA/RNA polymerase. *Mol. Cell. Biol.* **23,** 2309–15 (2003).

89.   Ruiz, J. F. *et al.* Lack of sugar discrimination by human Pol mu requires a single glycine residue. *Nucleic Acids Res.* **31,** 4441–9 (2003).

90.   Martin, M. J., Garcia-Ortiz, M. V, Esteban, V. & Blanco, L. Ribonucleotides and manganese ions improve non-homologous end joining by human Polμ. *Nucleic Acids Res.* **41,** 2428–36 (2013).

91.   Breen, A. P. & Murphy, J. A. Reactions of oxyl radicals with DNA. *Free Radic. Biol. Med.* **18,** 1033–77 (1995).

92.   Andres, S. N., Modesti, M., Tsai, C. J., Chu, G. & Junop, M. S. Crystal Structure of Human XLF: A Twist in Nonhomologous DNA End-Joining. *Mol. Cell* **28,** 1093–1101 (2007).

93.   Reid, D. A. *et al.* Bridging of double-stranded breaks by the nonhomologous end-joining ligation complex is modulated by DNA end chemistry. *Nucleic Acids Res.* **45,** 1872–1878 (2016).

94.   Roberts, S. A. *et al.* Ku is a 5′-dRP/AP lyase that excises nucleotide damage near broken ends. *Nature* **464,** 1214–1217 (2010).

95.   Ochi, T., Gu, X. & Blundell, T. L. Structure of the Catalytic Region of DNA Ligase IV in Complex with an Artemis Fragment Sheds Light on Double-Strand Break Repair. *Structure* **21,** 672–679 (2013).

96.   Biasini, M. *et al.* SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.* **42,** W252–W258 (2014).

97.   Oh, S., Wang, Y., Zimbric, J. & Hendrickson, E. A. Human LIGIV is synthetically lethal with the loss of Rad54B-dependent recombination and is required for certain chromosome fusion events induced by telomere dysfunction. *Nucleic Acids Res.* **41,** 1734–49 (2013).

98. Shen, B. *et al.* Generation of gene-modified mice via Cas9/RNA-mediated gene targeting. *Cell Res.* **23,** 720–723 (2013).

99. Shen, B. *et al.* Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects. *Nat. Methods* **11,** 399–402 (2014).

100. Uphoff, C. C. & Drexler, H. G. Detection of mycoplasma contaminations in cell cultures by PCR analysis. *Hum. Cell* **12,** 229–36 (1999).

101. Battaglia, M. F., Balducci, L. & Finocchiaro, M. Use of Hoechst 33258 fluorochrome for detection of mycoplasma contamination in cell cultures: development of a technique based on simultaneous fixation and staining. *Boll. Ist. Sieroter. Milan.* **59,** 155–8 (1980).

102. Gó mez-Herreros, F. *et al.* TDP2–Dependent Non-Homologous End-Joining Protects against Topoisomerase II–Induced DNA Breaks and Genome Instability in Cells and In Vivo TDP2–Dependent Non-Homologous End-Joining Protects against Topoisomerase II–Induced DNA Breaks and Genome Instability in Cells and. *PLoS Genet* **9,** (2013).

103. Ochi, T. *et al.* PAXX, a paralog of XRCC4 and XLF, interacts with Ku to promote DNA double-strand break repair. *Science (80-.).* **347,** 185–188 (2015).

104. Xing, M. *et al.* Interactome analysis identifies a new paralogue of XRCC4 in non-homologous end joining DNA repair pathway. *Nat. Commun.* **6,** 6233 (2015).

105. Andres, S. N. *et al.* A human XRCC4-XLF complex bridges DNA. *Nucleic Acids Res.* **40,** 1868–1878 (2012).

106. Roy, S. *et al.* XRCC4/XLF Interaction Is Variably Required for DNA Repair and Is Not Required for Ligase IV Stimulation. *Mol. Cell. Biol.* **35,** 3017–3028 (2015).

107. Ochi, T. *et al.* Structural insights into the role of domain flexibility in human DNA ligase IV. *Structure* **20,** 1212–22 (2012).

108. Budman, J., Kim, S. A. & Chu, G. Processing of DNA for nonhomologous end-joining is controlled by kinase activity and XRCC4/ligase IV. *J. Biol. Chem.* **282,** 11950–9 (2007).

109. Cottarel, J. *et al.* A noncatalytic function of the ligation complex during nonhomologous end joining. *J. Cell Biol.* **200,** 173–186 (2013).

110. Davis, B. J., Havener, J. M. & Ramsden, D. A. End-bridging is required for pol mu to efficiently promote repair of noncomplementary ends by nonhomologous end joining. *Nucleic Acids Res.* **36,** 3085–94 (2008).

111. Tomkinson, A. E., Howes, T. R. L. & Wiest, N. E. DNA ligases as therapeutic targets. *Transl. Cancer Res.* **2,** (2013).

112. Greco, G. E. *et al.* SCR7 is neither a selective nor a potent inhibitor of human DNA ligase IV. *DNA Repair (Amst).* **43,** 18–23 (2016).

113. Taccioli, G. E. *et al.* Impairment of V(D)J recombination in double-strand break repair mutants. *Science* **260,** 207–10 (1993).

114. Cabaniols, J. P., Fazilleau, N., Casrouge, A., Kourilsky, P. & Kanellopoulos, J. M. Most alpha/beta T cell receptor diversity is due to terminal deoxynucleotidyl transferase. *J. Exp. Med.* **194,** 1385–90 (2001).

115. Komori, T., Okada, A., Stewart, V. & Alt, F. W. Lack of N regions in antigen receptor variable region genes of TdT-deficient lymphocytes. *Science* **261,** 1171–5 (1993).

116. Gilfillan, S., Dierich, A., Lemeur, M., Benoist, C. & Mathis, D. Mice lacking TdT: mature animals with an immature lymphocyte repertoire. *Science* **261,** 1175–8 (1993).

117. Lucas, D. *et al.* Polymerase mu is up-regulated during the T cell-dependent immune response and its deficiency alters developmental dynamics of spleen centroblasts. *Eur. J. Immunol.* **35,** 1601–11 (2005).

118. Uphoff, C. C. & Drexler, H. G. in *Current Protocols in Molecular Biology* **106,** 28.4.1-28.4.14 (John Wiley & Sons, Inc., 2014).

119. Battaglia, M., Pozzi, D., Grimaldi, S. & Parasassi, T. Hoechst 33258 staining for detecting mycoplasma contamination in cell cultures: a method for reducing fluorescence photobleaching. *Biotech. Histochem.* **69,** 152–6 (1994).

120. Brinkman, E. K., Chen, T., Amendola, M. & van Steensel, B. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res.* **42,** e168–e168 (2014).

121. Traut, T. W. Physiological concentrations of purines and pyrimidines. *Mol. Cell. Biochem.* **140,** 1–22 (1994).

122. Zhu, H. & Shuman, S. Bacterial nonhomologous end joining ligases preferentially seal breaks with a 3'-OH monoribonucleotide. *J. Biol. Chem.* **283,** 8331–9 (2008).

123. Bebenek, K., Garcia-Diaz, M., Patishall, S. R. & Kunkel, T. A. Biochemical Properties of *Saccharomyces cerevisiae* DNA Polymerase IV. *J. Biol. Chem.* **280,** 20051–20058 (2005).

124. Della, M. *et al.* Mycobacterial Ku and ligase proteins constitute a two-component NHEJ repair machine. *Science* **306,** 683–5 (2004).

125. Zhu, H. & Shuman, S. A primer-dependent polymerase function of pseudomonas aeruginosa ATP-dependent DNA ligase (LigD). *J. Biol. Chem.* **280,** 418–27

(2005).

126. Bauer, R. J. *et al.* Comparative analysis of the end-joining activity of several DNA ligases. *PLoS One* **12,** e0190062 (2017).

127. Lord, C. J. & Ashworth, A. PARP inhibitors: Synthetic lethality in the clinic. *Science (80-.).* **355,** 1152–1158 (2017).

128. BALTIMORE, D. Viral RNA-dependent DNA Polymerase: RNA-dependent DNA Polymerase in Virions of RNA Tumour Viruses. *Nature* **226,** 1209–1211 (1970).

129. TEMIN, H. M. & MIZUTANI, S. Viral RNA-dependent DNA Polymerase: RNA-dependent DNA Polymerase in Virions of Rous Sarcoma Virus. *Nature* **226,** 1211–1213 (1970).

130. Chapados, B. R. *et al.* Structural biochemistry of a type 2 RNase H: RNA primer recognition and removal during DNA replication11Edited by K. Morikawa. *J. Mol. Biol.* **307,** 541–556 (2001).

131. Pascal, J. M. DNA and RNA ligases: structural variations and shared mechanisms. *Curr. Opin. Struct. Biol.* **18,** 96–105 (2008).

132. Bolwell, B. J., Cassileth, P. A. & Gale, R. P. High dose cytarabine: a review. *Leukemia* **2,** 253–60 (1988).

133. Li, G. *et al.* Small molecules enhance CRISPR/Cas9-mediated homology-directed genome editing in primary cells. *Sci. Rep.* **7,** 8943 (2017).

134. Rampazzo, C. *et al.* Regulation by degradation, a cellular defense against deoxyribonucleotide pool imbalances. *Mutat. Res.* **703,** 2–10 (2010).

135. Roychoudhury, R., Jay, E. & Wu, R. Terminal labeling and addition of homopolymer tracts to duplex DNA fragments by terminal deoxynucleotidyl transferase. *Nucleic Acids Res.* **3,** 101–16 (1976).

136. Winz, M.-L., Linder, E. C., André, T., Becker, J. & Jäschke, A. Nucleotidyl transferase assisted DNA labeling with different click chemistries. *Nucleic Acids Res.* **43,** e110 (2015).