RNA AND DNA End structure dependent 5' End Processing during in vitro NHEJ

Dissertation

Zur Erlangung des Grades

Doktor der Naturwissenschaften

Am Fachbereich Biologie

Der Johannes Gutenberg-Universität Mainz

Jayaprakash Srinivasan

geb. am 10/11/1994 in Chennai, Indien

Mainz, 2025

Dekan:

1. Berichterstatter:

2. Berichterstatter:

Tag der mündlichen Prüfung:

Table of Contents

Abstract *5*

Zusammenfassung *6*

Introduction *7*

DNA Double-strand breaks and repair pathways **7**

DSB detection and early signalling for pathway choice **9**

Early steps of NHEJ – Ku binding to ends **9**

DNA-PK structure and activation **11**

End synapsis and ligation **12**

End processing during NHEJ **14**

NHEJ in health and disease **16**

RNA in Double-strand break repair **17**

In vitro assays to study NHEJ **19**

Results *20*

I. RNA-dependent processing of DNA **20**

RNase treatment of NALM6 cell extracts stimulates in vitro NHEJ 20

Addback of RNA doesn't reverse end-joining stimulation 21

The kinetics of end-joining are influenced by RNase treatment 23

Unjoined monomers are not well ligated in further reactions 25

Unjoined monomers show signs of nucleolytic processing 28

5' overhangs are processed by an RNA-dependent nucleolytic factor 30

5' end processing is dependent on the activation of DNA-PK 33

Effect of the 5' end processing on end-joining outcome 36

II. Identification of the processing factor **37**

RNase treatment of the extract causes variance in the DNA-end binding proteome 37

III. Sequence biases in end-joining **43**

Impact of nucleotides at a blunt end 43

Impact of nucleotides upstream of a blunt end 44

Impact of nucleotides upstream of a 5' overhang 45

Discussion *47*

RNA-dependent 5' end processing during NHEJ **47**

The RNA-dependent end-processing factor 47

Effect of end-structure on processing 47

Effect of the end-processing on end-joining 48

End processing and end synapsis 49

NHEJ sequence bias **50**

Future Experiments *52*

Methods *53*

Cell culture **53**

Extract preparation **53**

End-joining assay **53**
Extraction of tRNA and rRNA **53**
Quantification of end-joining **54**
Purification of unjoined monomers **55**
Preparation of ONT sequencing libraries **55**
Sequencing analysis of Quick ligase junctions from unjoined monomers **56**
Design of radio-labelled substrates to study 5'-3' processing **56**
Design of radio-labelled substrates to study 3'-5' processing **57**
End-processing assay **58**
Annealing oligos into duplexes **59**
Preparation of biotinylated DNA substrates for pulldown **59**
Pulldown of proteins bound to DNA ends **59**
Analysis of Mass Spectrometry data **59**
Generation of random DNA end sequence libraries **59**
Sequencing analysis of random nucleotide libraries **61**
Oligonucleotides used in the study **61**
Antibodies used in the study **65**
Acknowledgements ***66***
References ***67***

Abstract

Double-strand breaks (DSBs) are the most hazardous lesions sustained by DNA, with potential for the loss and rearrangement of genetic information, if not repaired correctly. Non-homologous end-joining (NHEJ) and Homologous Recombination (HR) are the 2 major pathways in cells for repairing DSBs. NHEJ processes the broken ends if required, and ligates them to each other. On the other hand, HR requires the presence of a homologous sister chromatid to serve as a template for synthesis, and is thus restricted to the S and G2 phases of the cell cycle.

NHEJ begins with binding of the Ku heterodimer to DNA ends. The binding of DNA-PKcs and the subsequent activation of the DNA-PK holoenzyme through phosphorylation brings the ends together for ligation by the XRCC4-LigIV complex. NHEJ is an iterative process where, if necessary, the ends are processed by a hierarchy of enzymes until the ends are fit for ligation. Besides the repair of DSBs generated by exogenous sources like radiation, NHEJ is crucial for cellular processes like V(D)J recombination during lymphocyte development.

Using human cell extracts that perform in vitro NHEJ on linearized plasmid substrates, I describe a novel 5' end processing activity that is specific for DNA end structures with a protruding 5' end. Treatment of the extract with RNase A or blocking DNA-PK activation inhibited the described processing. The processing occurs at ends that are already compatible for direct joining, implying that it is independent of the typical known NHEJ processing and could act as a mechanism to generate genomic diversity.

In order to identify the processing factor, biotinylated substrates were used to pulldown proteins binding to defined ends from RNase-treated and untreated extracts. Mass spectrometry showed a clear and significant change in the end-binding proteome after RNase treatment. Multiple proteins and complexes were found to not associate with the DNA after RNase treatment of the extract. Further work is required to validate these hits and identify the processing factor.

Parallel to studying RNA-dependent end processing, I also investigated the effect of nucleotide sequence on NHEJ. I generated a library of substrates with random nucleotides at or proximal to the end. Analysis of NHEJ junctions showed a preference towards thymidines in highly joined blunt ended breaks. For substrates with defined ends/overhang, positions far away from the break (>10 bp) exhibited mild bias for certain nucleotides. This altogether suggests that sequence indeed affects repair by NHEJ.

Zusammenfassung

Doppelstrangbrüche (DSBs) sind die gefährlichsten DNA-Schäden, die zum Verlust und zur Umstrukturierung der genetischen Information führen können, wenn sie nicht korrekt repariert werden. Die Nicht-Homologe Endverknüpfung (NHEJ) und die Homologe Rekombination (HR) sind die beiden wichtigsten Wege, auf denen DSBs in Zellen repariert werden. Bei der NHEJ werden die gebrochenen Enden, wenn nötig, wieder zusammengefügt. HR hingegen erfordert die Anwesenheit eines homologen Schwesterchromatids, das als Vorlage für die Synthese dient, und ist daher auf die S- und G2-Phasen des Zellzyklus beschränkt.

NHEJ beginnt mit der Bindung des Ku-Heterodimers an die DNA-Enden. Die Bindung von DNA-PKcs und die anschließende Aktivierung des DNA-PK-Holoenzyms durch Phosphorylierung führen die Enden zur Ligation durch den XRCC4-LigIV-Komplex zusammen. NHEJ ist ein iterativer Prozess, bei dem die Enden, falls erforderlich, durch eine Hierarchie von Enzymen bearbeitet werden, bis sie für die Ligation geeignet sind. Neben der Reparatur von DSBs, die durch exogene Quellen wie Strahlung erzeugt werden, ist NHEJ für zelluläre Prozesse wie die V(D)J-Rekombination während der Lymphozytenentwicklung von entscheidender Bedeutung.

Unter Verwendung von humanen Zellextrakten, die in vitro NHEJ auf linearisierten Plasmidsubstraten durchführen, beschreibe ich eine neue 5'-Endprozessierungsaktivität, die spezifisch für DNA-Endstrukturen mit überstehendem 5'-Ende ist. Die Behandlung des Extrakts mit RNase A oder die Blockierung der DNA-PK-Aktivierung hemmte die beschriebene Prozessierung. Die Prozessierung findet an Enden statt, die bereits für eine direkte Verknüpfung geeignet sind, was bedeutet, dass sie unabhängig von der typischen bekannten NHEJ-Prozessierung ist und als Mechanismus zur Erzeugung genomischer Diversität dienen könnte.

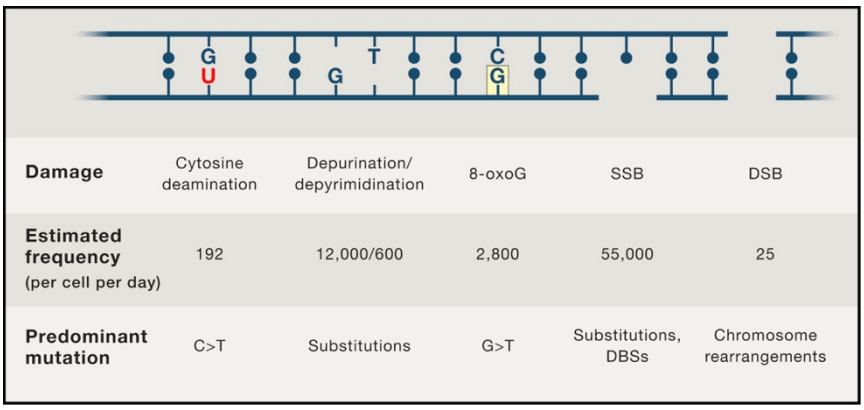
Um den Prozessierungsfaktor zu identifizieren, wurden biotinylierte Substrate verwendet, um endständige Proteine aus RNase-behandelten und unbehandelten Extrakten zu extrahieren. Die Massenspektrometrie zeigte eine deutliche und signifikante Veränderung des endständigen Proteoms nach der RNase-Behandlung. Es wurde festgestellt, dass mehrere Proteine und Komplexe nach der RNase-Behandlung des Extrakts nicht an die DNA binden. Weitere Arbeiten sind erforderlich, um diese Treffer zu validieren und den Prozessierungsfaktor zu identifizieren.

Parallel zur Untersuchung der RNA-abhängigen Prozessierung habe ich auch den Einfluss der Nukleotidsequenz auf NHEJ untersucht. Ich erstellte eine Substratbibliothek mit zufälligen Nukleotiden am Ende oder in der Nähe des Endes. Die Analyse der NHEJ-Verbindungen zeigte eine Präferenz für Thymidine in stark verknüpften Fragmenten mit stumpfem Ende. Bei Substraten mit definierten Enden/Überhängen zeigten Positionen weit entfernt vom Bruch (>10 bp) eine leichte Präferenz für bestimmte Nukleotide. Insgesamt deutet dies darauf hin, dass die Sequenz tatsächlich die Reparatur durch NHEJ beeinflusst.

Introduction

DNA Double-strand breaks and repair pathways

Maintaining a cell's genomic integrity is crucial both for the maintenance of the cell's functions and for transmission of the genome to successive generations. DNA is constantly challenged leading to the formation of various DNA lesions *(Figure 1)*. Double-strand breaks (DSBs) are by far the most dangerous lesions DNA can sustain, because loss of genetic information and chromosomal rearrangements are among the extreme consequences of double-strand breaks.



Damage	Cytosine deamination	Depurination/ depyrimidination	8-oxoG	SSB	DSB
Estimated frequency (per cell per day)	192	12,000/600	2,800	55,000	25
Predominant mutation	C>T	Substitutions	G>T	Substitutions, DBSs	Chromosome rearrangements

Figure 1 – Frequency of the major DNA lesions in a cell. From Tubbs & Nussenzweig (2017).

Despite the danger associated with DSBs, they are an essential part and side-effect of various cellular processes (Khan & Ali, 2017). The introduction of DSBs is fundamental for genetic recombination events – V(D)J recombination during lymphocyte development (Gellert, 2002; Schatz & Ji, 2011), Immunoglobulin Class switching in mature B cells (Stavnezer & Schrader, 2014) and meiotic recombination (Dereli et al., 2024). DNA replication and transcription, for example, generate torsional stress on the DNA in the form of positive and negative supercoiling. Class II topoisomerases TOP2A and TOP2B help relieve that stress through the generation of a double-strand break (Pommier et al., 2022). The dominant Class I topoisomerase TOP1 relieves supercoiling through a single-strand break (SSB), but these can act as primers for double-strand breaks when there is a nearby SSB on the complementary strand (Cristini et al., 2019). DSBs can also arise by replication across a nick, or during the repair of other lesions like Nucleotide excision repair (NER) of UV-induced damage (Jeggo & Löbrich, 2007; Wakasugi et al., 2014).

Non-homologous end-joining (NHEJ, or cNHEJ – classical NHEJ) and homologous recombination (HR) are the major pathways in cells that repair DSBs (Scully et al., 2019) *(Figure 2)*. HR requires the presence of extensive sequence homology through a homologous sequence such as a sister chromatid. Hence, HR is considered to be an 'error-free' pathway for the repair of DSBs. NHEJ, on the other hand, generally uses no homology, but can use up to 4 bp of 'microhomology' in rare cases (Pannunzio et al., 2014). The coding joints ligated by NHEJ during V(D)J recombination often show 1-2 nt of microhomology, but only in the absence of Terminal deoxyribonucleotide transferase (TdT) (Gauss & Lieber, 1996). In the absence of classical NHEJ factors, a subpathway of NHEJ called alternative end-joining (a-EJ) can occur after mild resection and the presence of 2-20 nt of microhomology (Chang et al., 2017).

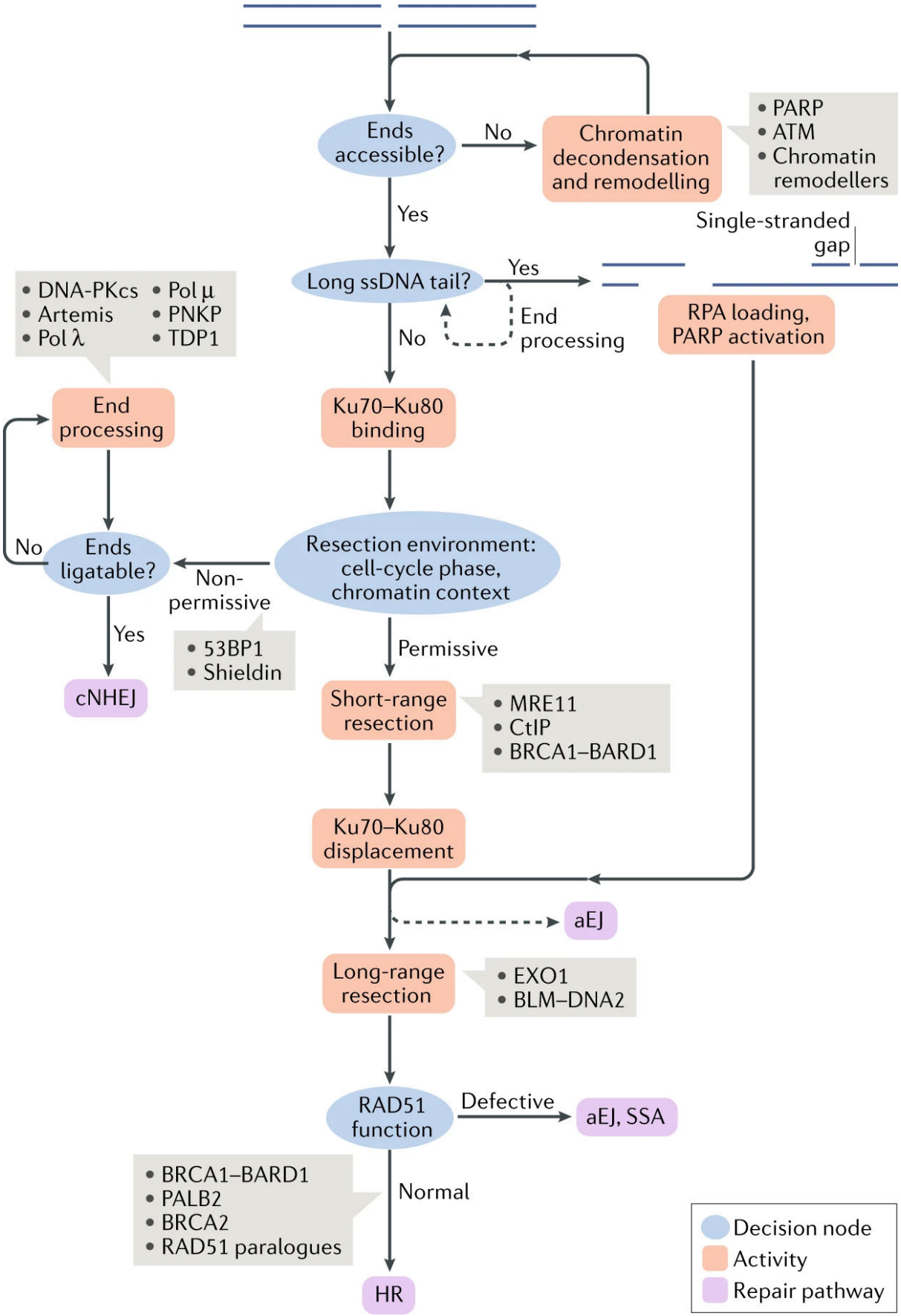


Figure 2 – Pathway choice in the repair of DSBs. From Scully et al., (2019).

The NHEJ pathway is the dominant one for repairing DSBs and is active throughout the cell cycle (Kakarougkas & Jeggo, 2014; Karanam et al., 2012), whereas HR is restricted to S and G2 due to the requirement of homology (Takata et al., 1998). Even in the G2 phase of the cell cycle where sister chromatids for HR are available, >80% of DSBs induced by irradiation are repaired by NHEJ (Beucher et al., 2009). This number can vary across different sites in the genome depending on the chromatin state (Aymard et al., 2014; Chen & Tyler, 2022). NHEJ is a fast process for DSB repair, taking on average 30 minutes to complete, whereas HR can take upwards of 7 hours (Mao et al., 2008).

DSB detection and early signalling for pathway choice

The Mre11-Rad50-Nbs1 (MRN) complex is one of the early sensors of double-strand breaks (D'Amours & Jackson, 2002; van den Bosch et al., 2003) that recruits ATM to the break site and activates it by phosphorylation (Bakkenist & Kastan, 2003). ATM is part of the Phosphatidylinositol 3-kinase (PI3K) family, which also includes ATR (activated by ssDNA bound protein RPA) (Haahr et al., 2016; Zou & Elledge, 2003) and DNA-PKcs (activated by the Ku heterodimer and DNA) (Gottlieb & Jackson, 1993). The activated kinases phosphorylate Serine139 of the histone variant H2A.X (Burma et al., 2001; Friesner et al., 2005; Stiff et al., 2004) forming γH2AX foci around the DSB site. An adaptor protein, MDC1, interacts directly with pS139 of H2A.X and helps in the amplification of DDR signalling (Stucki & Jackson, 2006).

The E3 ubiquitin ligase RNF8 binds to MDC1 and catalyses the K63-linked polyubiquitination of H2A. This is essential for the further recruitment of downstream repair factors BRCA1 and 53BP1 (p53-binding protein 1) that promote HR and NHEJ respectively (Huen et al., 2007; Kolas et al., 2007; Mailand et al., 2007). A complex of BRCA1 and BARD1 promotes end-resection for HR (Salunkhe et al., 2024), while 53BP1 promotes NHEJ and acts as an antagonist against resection through its interaction with Rif1 (Rap1 interacting factor) and the Shieldin complex (Callen et al., 2020; Noordermeer et al., 2018). The competition between BRCA1 and 53BP1 is influenced by chromatin architecture (Michelena et al., 2021), histone post-translational modifications (Jacquet et al., 2016; Tang et al., 2013), cell cycle stage (Chapman et al., 2012) and this ultimately determines the repair pathway *(Figure 2)*.

Early steps of NHEJ – Ku binding to ends

It is generally considered that DSB repair through the classical NHEJ pathway starts with the binding of the Ku70/80 heterodimer to the DNA ends *(Figure 3)*. The Ku complex is extremely stable and a deficiency of one of the proteins reduces the levels of the other (Gu, Jin, et al., 1997), implying that the proteins only function as a heterodimer. Ku is an extremely tight binder to DNA ends with a very low dissociation constant (K_d of $2.4*10^{-9}$ M^{-1}) (Blier et al., 1993). It binds specifically to the ends of dsDNA over uncut plasmid DNA and ssDNA independent of sequence (Mimori & Hardin, 1986). Ku binds to DNA ends in an ATP-independent manner (Blier et al., 1993), and is able to translocate along the DNA molecule (de Vries et al., 1989). Ku can even transfer directly from one molecule to another with non-homologous ends (Chiu et al., 2001).

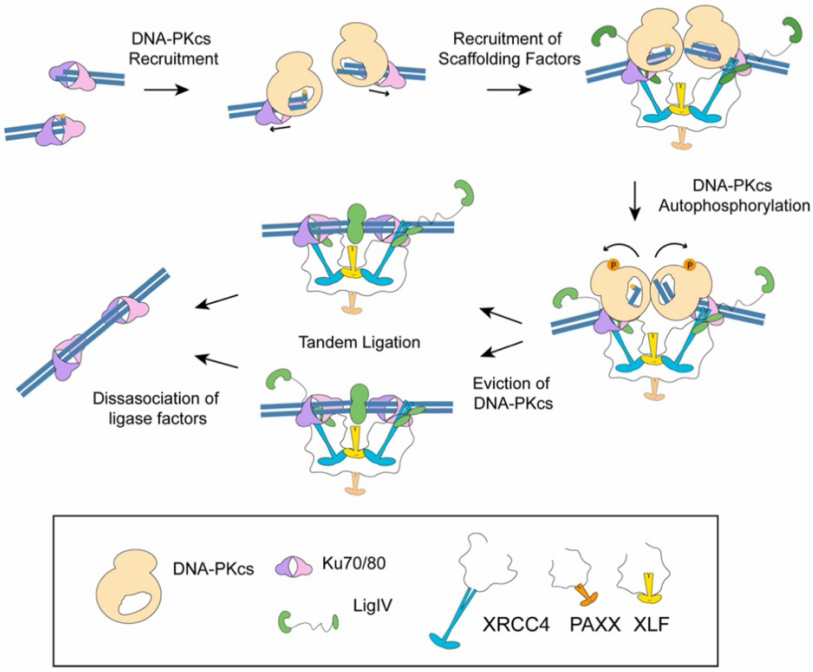


Figure 3 – Overview of the cNHEJ pathway. From Vogt & He (2023).

Both Ku proteins contain an N-terminal von Willebrand A (vWA) domain, a core domain and a C terminal DNA-binding domain *(Figure 4)*. The Ku70 C terminus contains an additional SAP domain whereas Ku80 contains the DNA-PKcs interacting domain at the C terminus (Downs & Jackson, 2004; Gell & Jackson, 1999; Gottlieb & Jackson, 1993). Though the proteins are largely similar, the regions in both proteins that interact with each other (Ku70 residues 449-578 and Ku80 residues 439-532) have little sequence homology (Gell & Jackson, 1999). This reduces chances of Ku70/70 and Ku80/80 homodimers forming and increases the specificity of the heterodimer complex formation. The Ku heterodimer forms a central ring for DNA binding with a base containing the core domains of the 2 subunits that 'cradle' the DNA. The ring structure contains a concentration of positive charged amino acids on its inner face and the formation of the ring itself is independent of the presence of DNA. These 2 features help with the sequence-independent binding nature of Ku. The cradle fits around 20 bp of DNA in it, with the major groove located in its centre (J. R. Walker et al., 2001).

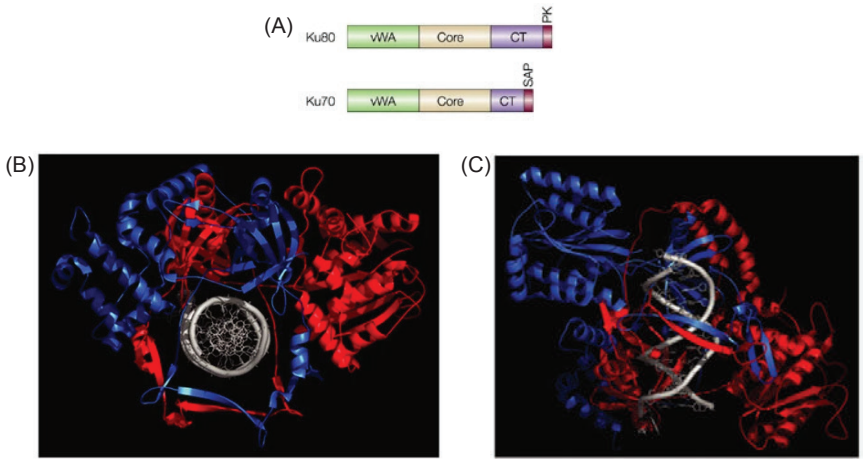


Figure 4 – Structure of the Ku heterodimer bound to DNA. (A) Domain architecture of Ku70 and Ku80. vWA – von Willebrand A domain; PK – DNA-PKcs binding domain. (B,C) Crystallographic

structure of the Ku heterodimer bound to a 14 bp long dsDNA. Ku70 in blue, Ku80 in red and DNA in grey. From Downs & Jackson (2004).

Early binding of Ku to DNA ends accomplishes a few things. It prevents DNA ends from drifting away and keeps them in proximity to each other through the interaction between the opposing Ku heterodimers (Cary et al., 1997; Ribes-Zamora et al., 2013; Rinaldi et al., 2023). Ku binding also downregulates HR by inhibiting end resection through Exo1 (Krasner et al., 2015). Despite protecting ends from degradation, the complex still allows processing and ligation through the side opposite to the cradle (Walker et al., 2001). Ku is also essential for the recruitment of downstream factors XRCC4-Ligase IV (McElhinny et al., 2000) and XLF (Yano et al., 2008).

DNA-PK structure and activation

The catalytic subunit of DNA-PK (DNA-PKcs), in complex with the Ku heterodimer and DNA ends, forms the active DNA-PK holoenzyme (DNA dependent Protein Kinase). DNA-PKcs is a massive protein at 460 kDa (Hartley et al., 1995) and potentially phosphorylates more than 700 substrates (Yue et al., 2020). The discovery of DNA-PK was accidental – RNA samples used for studying protein synthesis were contaminated with dsDNA (Walker et al., 1985). The experiments indicated the presence of a novel kinase that was activated by the presence of DNA, and later studies were able to purify the DNA-dependent kinase activity (Carter et al., 1990; Jackson et al., 1990; Lees-Miller et al., 1990).

Structurally, DNA-PKcs also contains a ring motif, similar to the Ku heterodimer, formed by the N-terminal HEAT repeats, with the kinase domain at the head of the ring *(Figure 5)* (Sibanda et al., 2010). 4 patches within the N-HEAT repeats (DB1-DB4; DB – DNA binding) are shown binding to the DNA end (X. Chen et al., 2021), and another putative DNA-end binding region, the DNA-end blocking helix (DEB) has been recently identified (S. Chen et al., 2021). The binding of DNA-PKcs to the DNA end also triggers a change in the Ku-DNA interaction. In the absence of DNA-PKcs, Ku remains bound at the DNA end. After DNA-PK holoenzyme formation, Ku translocates inward by about one helical turn (Yoo & Dynan, 1999).

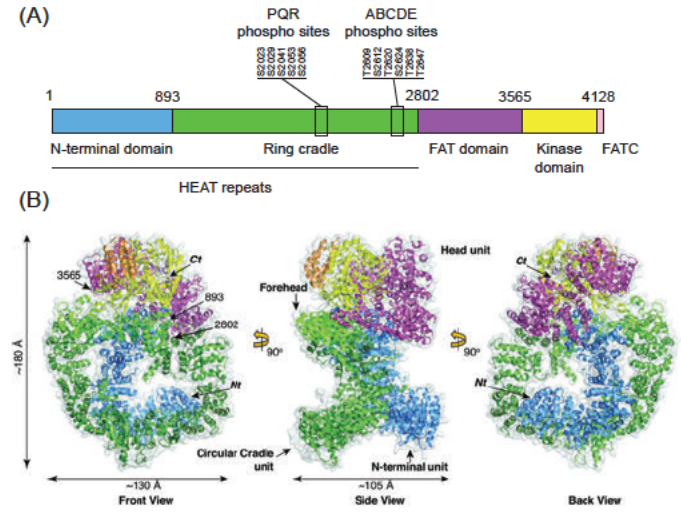


Figure 5 – Structure of DNA-PKcs. (A) Domain architecture (B) Crystal structure from Sibanda et al., (2017).

The kinase activity of DNA-PK is essential for NHEJ as shown by the defective NHEJ and increased radiosensitivity following the inhibition of DNA-PK's kinase activity (Zhao et al., 2006). Although DNA-PK phosphorylates multiple proteins in vitro, not all of them appear to be essential for NHEJ activity (Douglas et al., 2005; Yu et al., 2003, 2008). DNA-PK itself gets phosphorylated by ATM (Chen et al., 2007) and has prolific autophosphorylation tendencies (Dobbs et al., 2010; Douglas et al., 2002) which are essential for NHEJ (Chan et al., 2002; Reddy et al., 2004; Soubeyrand et al., 2003).

There are 2 major clusters of phospho-sites on the DNA-PK catalytic subunit *(Figure 5)* – the ABCDE cluster (six conserved serine/threonine residues between 2609 and 2647) (Douglas et al., 2002) and the PQR cluster (five serine residues between 2023 and 2056) (Cui et al., 2005). The PQR cluster is less well-studied between the two, except for S2056 - which gets phosphorylated in trans in response to double-strand breaks induced by ionizing radiation (Chen et al., 2005). DNA-PK autophosphorylation mediates a conformational change that regulates access to the DNA ends for processing factors (Block et al., 2004; Ding et al., 2003). Trans-phosphorylation in the ABCDE cluster inactivates the kinase activity of DNA-PK (Chan & Lees-Miller, 1996) and is essential for the disassembly of the DNA-PK complex along with S2056 phosphorylation (Jiang et al., 2015; Merkle et al., 2002).

Independent of its kinase activity, DNA-PK juxtaposes the 2 broken ends against each other, forming a synaptic complex. This appears to be an important step in the activation of DNA-PK through trans-phosphorylation, since complete kinase activity could only be observed after the formation of the synaptic complex (DeFazio et al., 2002; Meek et al., 2007).

End synapsis and ligation

DNA ends are initially held in long-range synapsis by the DNA-PK complex. The activation of DNA-PK's catalytic activity then leads to the formation of the short-range synaptic complex which includes additional NHEJ proteins XLF and XRCC4-Ligase IV that facilitate alignment of the broken ends *(Figure 6)* (Graham et al., 2016). Though DNA-PKcs is generally considered to be part of the synaptic complex, it is not universally essential for the formation of synaptic complexes in vitro (Zhao et al., 2019). Nevertheless, evidence from a variety of models show that DNA-PK and its catalytic activity are essential for synapsis (Stinson & Loparo, 2021). XRCC4 appears to be most important component for synapsis since XRCC4 deficiency has a stronger effect than DNA-PKcs deficiency on radiosensitivity (Jiang et al., 2015).

XRCC4 can form tetramers, and associates strongly in its dimeric form with DNA ligase IV (Critchlow et al., 1997; Modesti et al., 2003). XLF (Cernunnos) has a structure similar to XRCC4 with a globular head domain and a coiled-coil C terminus (Y. Li et al., 2008). Filaments of XLF and XRCC4 in complex with each other have been observed in crystallographic and electron microscopy studies (Hammel et al., 2011; Ropars et al., 2011) and it is suggested that a single XLF dimer interacting with 2 XRCC4:LigIV complexes is the minimal component necessary for the transition from long-range to short-range synapsis (Stinson & Loparo, 2021).

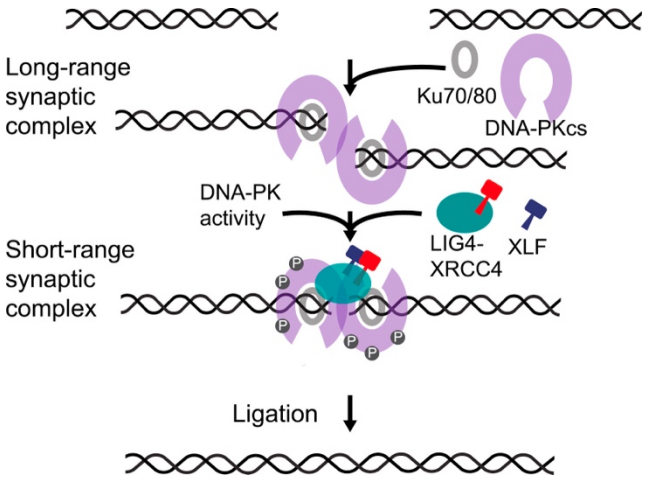


Figure 6 – Progression of end synapsis with DNA-PK. From Graham et al (2016).

Ligase IV (Lig IV) is the enzyme ligating the DNA ends together during NHEJ and its activity is stimulated in the presence of XRCC4 (Grawunder et al., 1998). The addition of XLF to the XRCC4:LigIV complex can even promote the joining of ends that are incompatible for direct ligation through the ligation of at least one of the strands (Tsai et al., 2007). Ligase IV consists of an N-terminal DNA binding domain, a catalytic domain and C-terminal BRCT domains that help in the interaction with Ku and XRCC4 (Costantini et al., 2007; Wu et al., 2009). Recruitment of the ligase appears to be a limiting step in NHEJ, as the preformation of a complex containing Ku and XRCC4-Ligase IV increased the rate of end-joining up to 20-fold in vitro (McElhinny et al., 2000). Aside from its catalytic activity, the physical presence of LigIV seems to be essential for stable synapsis (Cottarel et al., 2013). Consistent with this observation, while the knockout of LigIV is embryonically lethal in mice, a catalytically inactive LigIV leads to alive mice, but with severe growth retardation. In these cases, Ligase 3 appears to apply redundancy to LigIV's catalytic activity (Medina-Suárez et al., 2024).

Recently, a number of factors have been identified as accessory factors in the synaptic complex (Ghosh & Raghavan, 2021). PAXX (Paralog of XRCC4 and XLF) interacts directly with Ku and promotes the stability of the core NHEJ factors on damaged chromatin (Craxton et al., 2015; Ochi et al., 2015). The protein structure is also very similar to that of XLF and XRCC4 – with a N-terminal globular domain and a coiled-coil C terminus that forms homodimers (Xing et al., 2015). PAXX knockout mice develop normally, whereas the simultaneous knockout of PAXX and XLF is embryonically lethal (Balmus et al., 2016). This synthetic lethality points towards a non-essential role for PAXX in NHEJ, and the existence of a functional redundancy between PAXX and XLF.

One of the isoforms of MRI (CYREN) can stimulate NHEJ in vitro through its interaction with Ku (Slavoff et al., 2014). However, it was then shown in cells that MRI regulates cell cycle control of NHEJ pathway choice by inhibiting NHEJ during the S and G2 phases of the cell cycle (Arnoult et al., 2017).

The long non-coding RNA LINP1 is overexpressed in triple-negative breast cancers and regulates the repair of DSBs by its interaction with Ku80 and DNA-PKcs (Zhang et al., 2016). The authors suggest that the RNA acts as a scaffold supporting the interaction between Ku and DNA-PKcs. These recently discovered NHEJ accessory factors are not

always completely essential for successful NHEJ, but their discoveries point towards a high complexity and redundancy among factors in the NHEJ pathway.

End processing during NHEJ

Physiological DNA double strand breaks rarely leave 'clean' DNA breaks that are suitable for direct ligation (Coquerelle et al., 1973; Schipler & Iliakis, 2013). A number of end-processing enzymes are thus associated with NHEJ to 'prepare' the ends for ligation *(Figure 7)*. These enzymes can be broadly classified into damage correction enzymes, nucleases and polymerases.

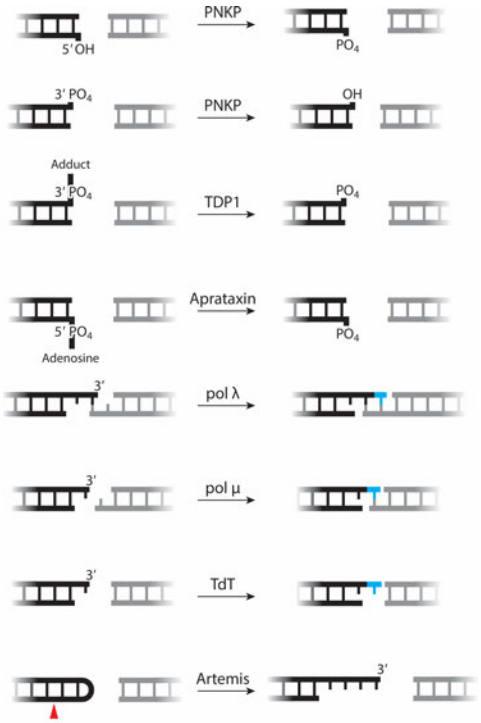


Figure 7 – DNA ends that require processing before ligation and the enzymes responsible for it. From Stinson & Loparo (2021).

PNKP (Polynucleotide kinase 3' phosphatase) is a bifunctional enzyme that phosphorylates 5' OH ends and also removes phosphates from the 3' termini (Chappell et al., 2002; Koch et al., 2004). Abortive DNA ligation can leave an AMP moiety on the 5' end. The enzyme aprataxin hydrolyzes the 5' adenylate to a 5' phosphate fit for ligation (Ahel et al., 2006). Similarly, DNA topoisomerases can leave a covalently cross-linked phosphotyrosine intermediate that is a major block to ligation (Deweese & Osheroff, 2009). TDP1 (Tyrosyl-DNA phosphodiesterase 1) hydrolyzes the 3' phosphotyrosine intermediate generated by Topoisomerase I (S. W. Yang et al., 1996), and TDP2 liberates the 5' phosphotyrosine by-product of Topoisomerase II (Ledesma et al., 2009). TDP1 can also act on the 3'-phosphoglycolate moieties that are commonly generated after irradiation damage (Inamdar et al., 2002). Ku70 has a surprising 5'-dRP/AP lyase activity that excises abasic sites from DNA ends in vitro (Roberts et al., 2010).

Artemis is the most popular nuclease associated with NHEJ due to its role in V(D)J recombination, where it endonucleolytically cleaves the hairpins generated by RAG1/2 (Ma et al., 2002). In complex with DNA-PK, artemis can perform endonucleolytic cleavage

of 5' and 3' overhangs (Niewolik et al., 2006). Independent of DNA-PK, artemis exhibits a mild in vitro 5'-3' exonuclease activity on 5' overhangs and flaps (Ma et al., 2005). Metnase is another recently described endonuclease that is active on ssDNA in flaps/overhangs and contributes to the NHEJ of incompatible ends (Beck et al., 2011).

Terminal deoxyribonucleotide transferase (TdT) and the polymerases λ and μ belonging to the X-family of DNA polymerases are the major polymerases linked to NHEJ. All 3 enzymes contain an evolutionarily conserved Pol X domain at the C terminus and an N terminal DNA-binding BRCT domain. These enzymes are not processive, like those belonging to the B-family polymerases (δ,ε), but are evolved to synthesize short stretches of DNA (Uchiyama et al., 2009). TdT synthesizes from the 3' end in a template-independent fashion and is essential for generating sequence diversity during V(D)J recombination (Gilfillan et al., 1993). Functionally, Pol μ and λ differ in the fact that Pol μ can extend 3' ends that are not base-paired. Unlike TdT, DNA synthesis by Pol μ is still informed by a template sequence (McElhinny et al., 2005). Pol μ is active on a wider variety of DNA ends, whereas Pol λ is more active on a subset of substrates and can be more accurate. Together, the two polymerases enable processing of a wide range of ends (Pryor et al., 2015).

LigIV can effectively act as a 'translesion ligase' during NHEJ, meaning it can tolerate mispairs at the end (Waters et al., 2014). NHEJ also has mechanisms to avoid excessive processing *(Figure 8)*. End-processing has been shown to occur exclusively after formation of the short-range synaptic complex (Stinson et al., 2020). Processing enzymes act hierarchically – end correction enzymes first, followed by the polymerases and nucleases, and ends are joined immediately after compatibility is achieved (Stinson & Loparo, 2021). Though NHEJ has usually been characterized as an 'error-prone' process, this recent evidence suggests that processing is tightly regulated, and that the majority of breaks repaired by NHEJ are done so faithfully (Bétermier et al., 2014; Lin et al., 2013).

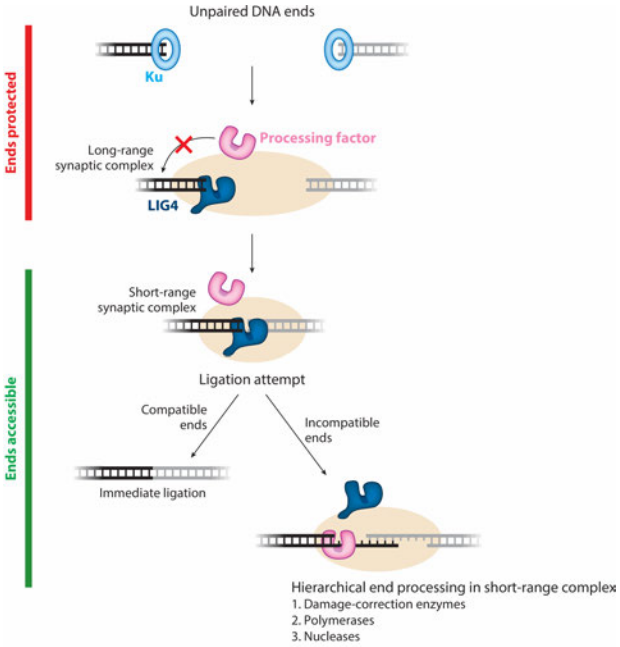


Figure 8 – Model for the regulation of end-processing during NHEJ. From Stinson & Loparo (2021).

NHEJ in health and disease

Being the predominant pathway for repairing DSBs in mammalian cells, the deficiency of NHEJ factors can lead to several adverse health effects (Zhao et al., 2020). NHEJ factor deficiency leads to a loss of junctional diversity during V(D)J recombination (Taccioli et al., 1993). V(D)J recombination is initiated by the binding of RAG1 and RAG2 to specific recombination signal sequences (RSSs) flanking the ends of the V, D and J segments. The RSSs contain a conserved heptamer (CACAGTG) and nonamer (ACAAAAACC) linked by a spacer that is 12 or 23 bp long. The RAG complex nicks the DNA immediately adjacent to the RSS *(Figure 9A)*. This nick is then used as a nucleophile to attack the phosphodiester bond in the complementary strand, leading to the formation of a DSB with hairpins at the coding ends. The NHEJ proteins now take over, with the Artemis endonuclease dependent resolution of the hairpins, followed by the processing of ends by polymerase fill-in or nuclease degradation before ligation (Gellert, 2002; Lieber, 2010). The expression of the proteins RAG1 and RAG2 is tightly regulated, and happens during the maturation of B cells and T cell receptor rearrangement (Kuo & Schlissel, 2009) *(Figure 9B)*.

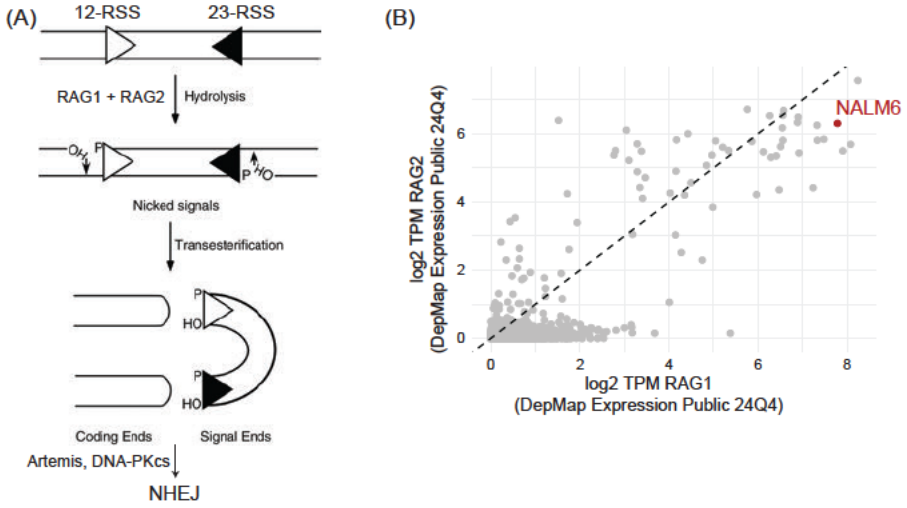


Figure 9 – V(D)J recombination in lymphocytes. (A) RSS cleavage by the RAG proteins to initiate V(D)J recombinaton. Figure modified from Gellert (2002); (B) Expression of RAG1/2 is tightly controlled. TPM (Transcripts per million) counts of RAG1 and RAG2 in 1673 primary cells and cell lines obtained from DepMap. The Pre-B lymphocyte cell line NALM6 used in this study is highlighted.

Mice deficient for Ku experience severe growth retardation and immune defects in the form of severe combined immunodeficiency (SCID) (Gu, Seidl, et al., 1997; Nussenzweig et al., 1996; Zhu et al., 1996). A number of autoimmune diseases produce autoantibodies that react against the Ku heterodimer (Francoeur et al., 1986). In fact, Ku was initially identified when sera from patients with Polymyositis-scleroderma overlap syndrome recognized it as an autoantigen (Mimori et al., 1981).

Mutations or loss of DNA-PKcs (van der Burg et al., 2008; Woodbine et al., 2013), LigIV (Buck, Moshous, et al., 2006; Enders et al., 2006; O'Driscoll et al., 2001; van der Burg et al., 2006), XLF (Buck, Malivert, et al., 2006; G. Li et al., 2008) and Artemis (Bajin et al., 2013; Moshous et al., 2001, 2003; Volk et al., 2015) lead to severe immunodeficiency in humans. This is caused by defective V(D)J recombination which then impairs B and T

lymphocyte development. Curiously, XRCC4 impairment causes defective lymphogenesis and late embryonic lethality in mice (Gao et al., 1998), whereas equivalent symptoms haven't been identified in human patients lacking XRCC4 (Zhao et al., 2020).

NHEJ has a 'mutational signature' in the repair of DSBs with ssDNA overhangs, where it can lead to the formation or disappearance of tandem repeats (Messer & Arndt, 2007). Whereas 5' overhangs can lead to duplication or trimming of the repeats depending on whether synthesis or resection takes place, 3' overhangs generally lead to tandem repeat duplications (Schimmel et al., 2017). The error-prone nature of action of Pol μ makes it a 'mutator' polymerase (Domínguez et al., 2000). In combination with the permissive nature of Ligase IV towards mispairs (Waters et al., 2014), NHEJ can act as a source of genomic variation.

RNA in Double-strand break repair

The view that only proteins act during DNA damage response (DDR) pathways has been challenged in recent times with the emergence of RNA as an important player (Bader et al., 2020; D'Adda di Fagagna, 2014; Hawley et al., 2017; Michelini et al., 2018). miRNAs have been shown to control the expression of the histone H2A.X (Lal et al., 2009), DNA-PKcs (Yan et al., 2010), ATM (Hu et al., 2010), BRCA1 (Moskwa et al., 2011), RAD51 (Y. Wang et al., 2012), PARP1 (Dash et al., 2017), NEIL2 (Kang et al., 2019) and CTIP (L. Yang et al., 2019). These are examples of how RNA influences DDR indirectly by regulating the level of repair factor expression. Lately, we have a confluence of literature pointing towards a more involved role of RNA in DDR *(Figure 10)*.

Wei et al., (2012) is one of the first studies to show that processed RNA from sites of DNA double-strand breaks can have a measurable impact on repair outcome. They were able to sequence small RNA, 21 nt in length, mapping to sites ~600 bp around a DSB in the plant model system Arabidopsis. These small RNA, named DSB-induced small RNAs (diRNAs) were shown to be transcribed by RNA polymerase IV and processed by Dicer-like proteins. They were suggested to be recruited to Ago2 after H2A.X phosphorylation and were shown to support DSB repair through 2 HR subpathways – Single-strand annealing (SSA) and Synthesis-dependent strand annealing (SDSA). Processed small RNA, termed endo-siRNA, were identified around DSBs in a Drosophila cell line system (Michalik et al., 2012). Earlier, a novel class of siRNA named qiRNA were identified upon the induction of DSBs in the rDNA locus of the ascomycete *Neurospora crassa* (Lee et al., 2009).

In human cells, knockdown of the miRNA processing enzymes Dicer and Drosha leads to impaired DDR foci formation (Francia et al., 2012). The formation of γH2A.X foci was unaffected, implying that the RNA synthesis happens downstream of the early stages of the damage response, including break sensing. Using DSBs induced by a site-specific endonuclease, they were able to show that Dicer- and Drosha-dependent RNA (DDRNA) originate from the break site. RNA polymerase II is recruited to break sites via the MRN complex and synthesizes the long non-coding RNAs (dilncRNA), which act as precursors to diRNA/DDRNA (Michelini et al., 2017). Knockdown of the small RNA processing enzymes lead to diminished MDC1 and 53BP1 foci formation (Francia et al., 2016).

DDRNA were essential for the recruitment of chromatin remodelling factors to the break which then improved the accumulation of BRCA1 and Rad51 (Q. Wang & Goldstein, 2016). This leads to the model that small RNA produced at the break and processed by Dicer and Drosha act as amplifiers of DNA damage signalling and enable the recruitment of downstream repair factors (Francia et al., 2016).

Most studies that have identified small RNA produced at the DSB have only done so using reporter constructs. This may not reflect the physiological reality due to the repetitive nature of the sequence and an artificially high level of transcription (Lu et al., 2018). Indeed, in plants, damage-induced RNAs (diRNAs) could not be identified from breaks generated in endogenous loci (Miki et al., 2017). In human cells, when breaks were induced endogenously by expression of the endonuclease I-PpoI, only breaks located in the repetitive 28S rDNA locus led to the production of diRNAs (Bonath et al., 2018). This further demonstrates that the identification of diRNAs/DDRNAs/small RNA around DSBs is not only dependent on the break, but on the presence of repetitive sequences and high transcription around the break. Hence, the model doesn't truly reflect the double-strand break response in the majority non-repetitive regions of the human genome.

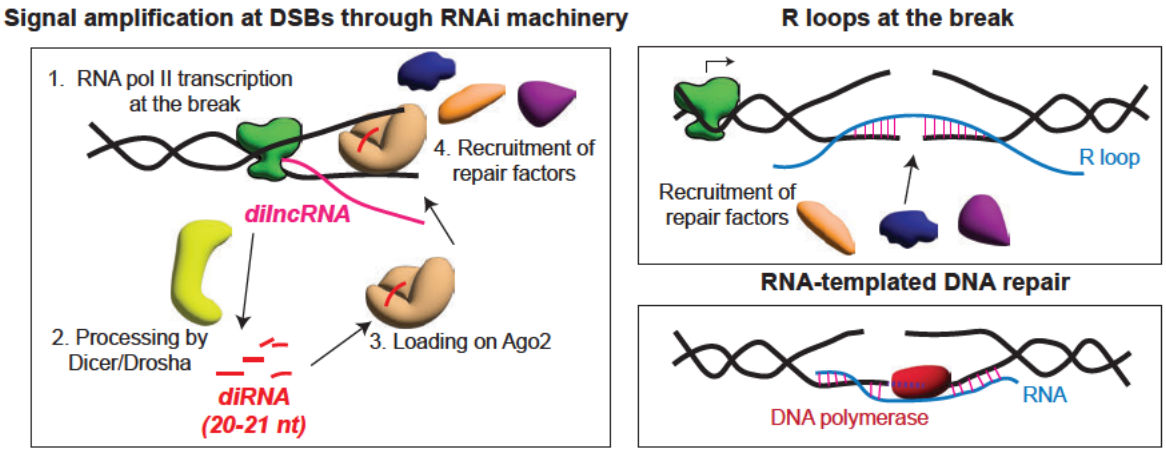


Figure 10 – Suggested mechanisms of RNA involvement in Double-strand break repair

RNA transcription at genes frequently leads to the transient formation of R loops or RNA:DNA hybrids. Indeed, it has been demonstrated that R loops are stabilized around DSBs very early in the double strand break response. Removal of these R loops diminished DSB repair through both HR and NHEJ (Lu et al., 2018). Even post-resection, R loops have been shown to interact with BRCA1 (D'Alessandro et al., 2018).

For double strand breaks in genes with active transcription, RNA can serve as a template to perform DNA synthesis. The ability of DNA polymerases to use RNA as a template during double strand break repair was first shown in the budding yeast *Saccharomyces cerevisiae* (Keskin et al., 2014; Storici et al., 2007) and then later in human cells (Chakraborty et al., 2016; Shen et al., 2011). The use of an RNA oligonucleotide improved the probability of an error-free repair by at least 500-fold (Storici et al., 2007). The pairing of RNA to DNA in this context is shown to be encouraged by the recombinase Rad52 (Keskin et al., 2014; Mazina et al., 2017). RNA with complementarity to both ends of the double strand break participated in NHEJ in an RNaseH2-sensitive manner and even affected the kind and frequency of indels (Jeon et al., 2024). Multiple DNA polymerases

have in vitro reverse transcriptase activity (Chandramouly et al., 2021), and so far, the reverse transcriptase activity of DNA polymerase η has been described to be essential for RNA-templated repair (Chakraborty et al., 2023).

In vitro assays to study NHEJ

The use of in vitro assays has boosted our understanding of various cellular processes like transcription (Yang & Ma, 2016). In vitro assays observing NHEJ were first performed by the incubation of restriction enzyme cut plasmid DNA in Xenopus cell free extracts (Pfeiffer & Vielmetter, 1988). Similar experiments were also then performed with human cell extracts (North et al., 1990). These plasmid-based in vitro NHEJ assays have been crucial in the biochemical characterization of the NHEJ pathway. The requirement of the core NHEJ proteins - Ku heterodimer, DNA-PK, XRCC4 and LigIV towards successful NHEJ was first shown in an in vitro end-joining assay by Baumann & West (1998). The involvement of the DNA kinase activity from human PNKP in processing 5'OH terminal ends for joining was also shown through an in vitro assay (Chappell et al., 2002).

Aside from plasmid-based substrates, genomic DNA from irradiated cells have also been used for in vitro end-joining (Cheong et al., 1998). Though some biochemical knowledge can be gleamed from these assays, these are ultimately limited by the randomness of broken ends generated by radiation, which means the sequences of the junctions cannot be known.

The processing of 'dirty' breaks during NHEJ is of considerable interest (Povirk, 2012). In vitro assays allow customization of the substrate ends and can hence be a valuable addition in our evolving understanding of NHEJ. Substrates with custom 'atypical' ends can be created by radiomimetic drugs like bleomycin or by the ligation of modified oligonucleotides to the ends of a linearized plasmid DNA (Datta et al., 2012).

Results

I. RNA-dependent processing of DNA

RNase treatment of NALM6 cell extracts stimulates in vitro NHEJ

Linearized pBlueScript plasmid DNA incubated with NALM6 cell extracts produce end-joined products visible on an agarose gel *(Figure 11)*. The amount of end-joining was dependent on the amount of extract added to the reaction. Extracts made from LigIV-deficient NALM6 cells were not able to perform any end-joining *(Figure 11A)*, indicating that the joined products are a result of non-homologous end-joining. Inhibition of DNA-PK with the inhibitor NU7441 also produces no end-joining *(Figure 11B)*.

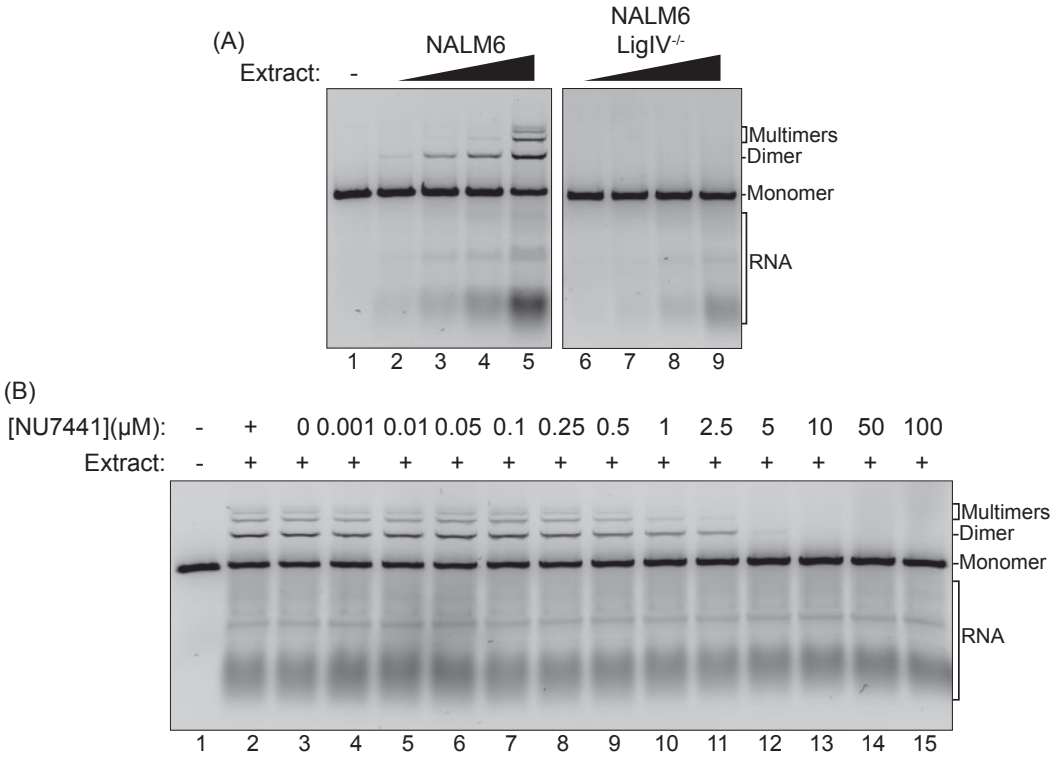


Figure 11 – NALM6 cell extracts perform NHEJ in vitro. In vitro NHEJ is dependent on (A) LigIV and (B) DNA-PK. (A) Lanes 5 and 9 contain 4 µl of the respective extracts; two-fold serial diluted extract in extract buffer were added to lanes 2-4 and 6-8. The substrate is pBlueScript digested with EcoRI (5' `AATT` overhang).

The protein extract preparation protocol also co-purifies some RNA, which is clearly observable on the agarose gels. To degrade the RNA in the extract, the extract was treated with RNase A before addition of the substrate. Surprisingly, the RNase treatment improved the end-joining efficiency of EcoRI-digested plasmid DNA. This RNase-mediated stimulation of end-joining was end-structure dependent. Similar results were observed for other substrates with different sequences in the 5' overhang *(Figure 12A)*. For substrates with 3' overhangs or blunt ends, RNase treatment of the extract either had no effect or a slight repressive effect on end-joining *(Figure 12B)*. This result suggests that RNA has an end-structure dependent inhibitory effect on in vitro end-joining.

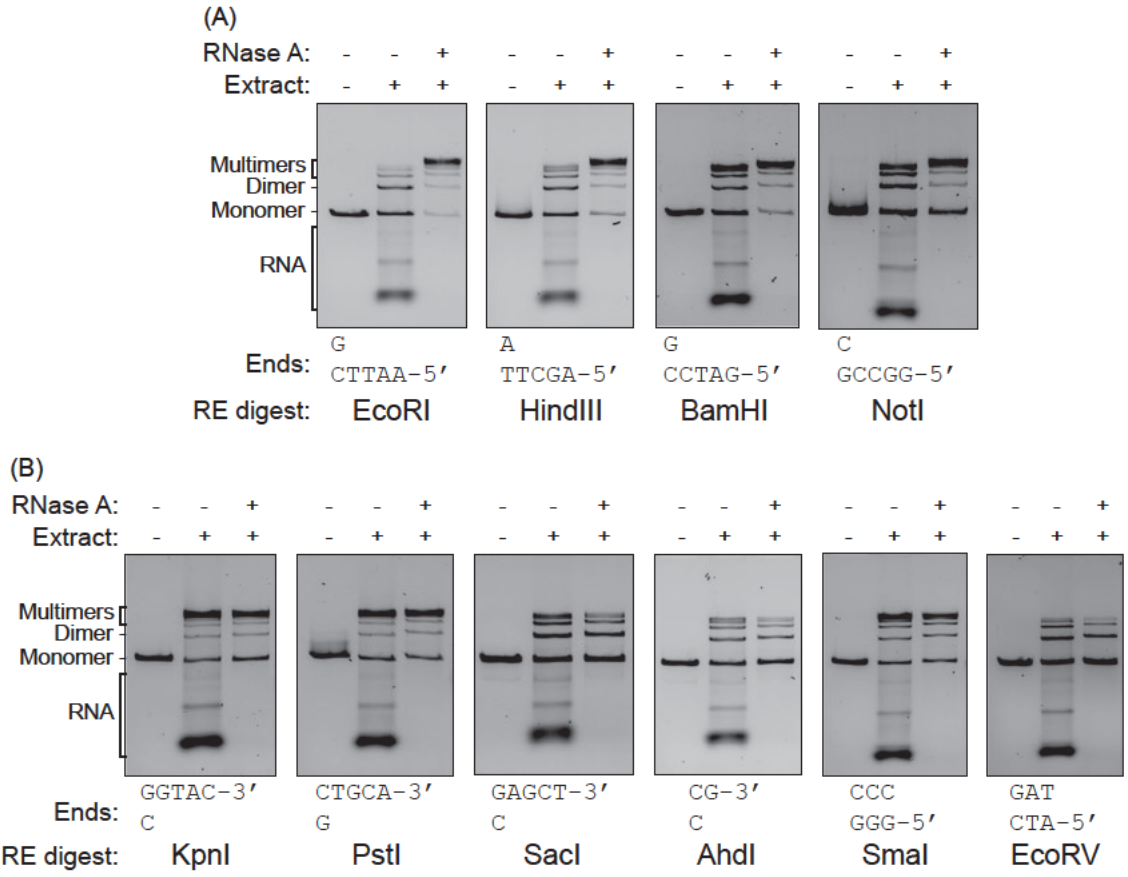


Figure 12 – End-joining is stimulated by RNase A treatment of extract. (A) End-joining of substrates with ends containing 5' overhangs (B) End-joining of substrates with blunt ends or ends with 3' overhangs. All substrates were made by digesting the plasmid pBlueScript. The restriction enzyme used for making the ends is noted under the sequence of the ends.

Addback of RNA doesn't reverse end-joining stimulation

To check if the RNA inhibitory effect is mediated by the general presence of RNA or a specific RNA species, I first checked if the addback of commercially available RNA can reverse the stimulation of end-joining produced by RNase treatment of the extract. After the extract was treated with RNase A, RNasin was added to inhibit further RNase activity. Different amounts of commercial yeast tRNA (Thermo Fischer, #AM7119) or yeast total RNA (Thermo Fischer, #AM7120G) were then added to the end-joining reaction *(Figure 13)*. Neither RNA reversed the RNase stimulation of end-joining, suggesting that the general presence of RNA is not the inhibitory factor for end-joining of 5' overhangs.

Following this, I purified RNA from NALM6 cells using the mirVANA miRNA isolation kit (Thermo Fischer, #AM1560). The total RNA, along with a fraction enriched for small RNAs was extracted. Addback of neither RNA restored end-joining to pre-RNase treatment efficiencies *(Figure 14A)*. There was a slight reduction in end-joining with total RNA, but it required a large amount of RNA, higher than what was already in the extract. Other specific RNA species purified from NALM6 cells – rRNA and tRNA also could not reverse the end-joining stimulation *(Figure 14B)*.

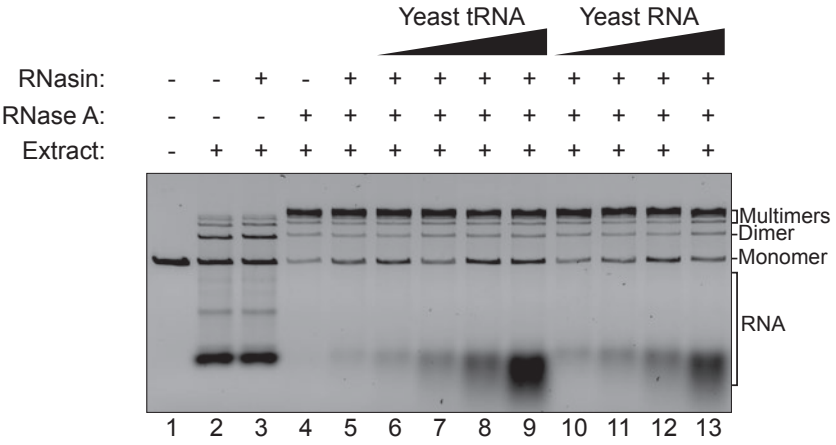


Figure 13 – Addback of general commercially available RNA does not reverse end-joining stimulation by RNase A treatment. DNA substrate is pBlueScript digested with EcoRI (5' AATT overhang). Amounts of RNA added: Lanes 6 & 10 – 0.25 µg; Lanes 7 & 11 – 0.5 µg; Lanes 8 & 12 – 1 µg; Lanes 9 & 13 – 2 µg.

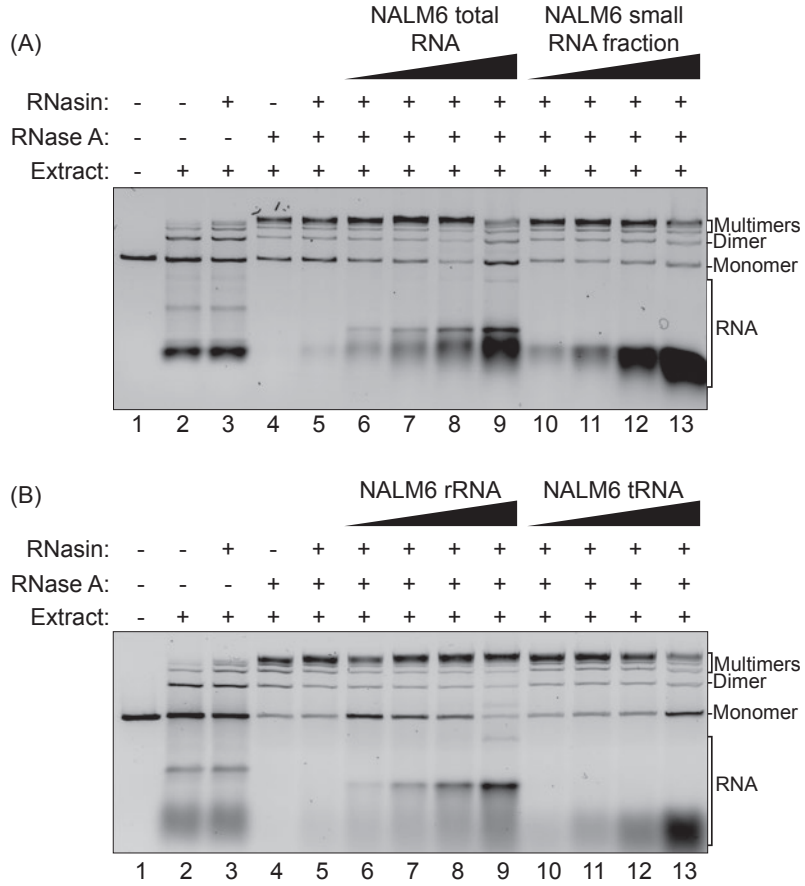


Figure 14 – Addback of NALM6 RNA doesn't fully reverse RNase A mediated end-joining stimulation. (A) Total RNA and small RNA fraction extracted from NALM6 cells with mirVANA miRNA extraction kit (B) NALM6 rRNA and tRNA. DNA substrate is pBlueScript digested with EcoRI (5' AATT overhang). Amounts of RNA added: Lanes 6 & 10 – 0.25 µg; Lanes 7 & 11 – 0.5 µg; Lanes 8 & 12 – 1 µg; Lanes 9 & 13 – 2 µg.

Some models of RNA involvement in DSB repair from recent literature suggest an induction of transcription at DSBs *(Figure 10)*. Dialysis during extract preparation should

deprive the extract of any ribonucleotides. Indeed, almost all end-joining is abolished without the addition of ATP *(Figure 15)*. It is thus reasonable to expect that the other ribonucleotides are also not present at high enough amounts to facilitate de novo RNA synthesis in this system.

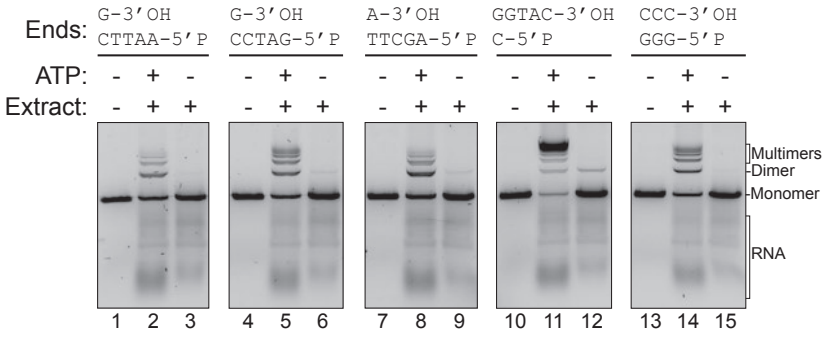


Figure 15 – End-joining reaction without the addition of ATP. End-joining of substrates made from pBlueScript with the described ends was performed for 2 h.

The kinetics of end-joining are influenced by RNase treatment

In order to quantify end-joining and compare across different conditions and substrates, I used end-joining substrates made with the plasmid pKJ025 *(Quantification of end-joining)*. End-joining substrates were created by digesting pKJ025 with various restriction endonucleases.

To study the kinetics of an end-joining reaction, a larger reaction was set up and samples were aspirated at different time points. The end-joining was then stopped with stop buffer (10 mg/ml proteinase K, 100 mM Tris-HCl pH 7.5, 200 mM EDTA and 2.5% SDS), purified and digested with the enzyme AvaII *(Figure 16)*. The intensities of the 3 kb and 1.5 kb fragments were measured by running on a gel and their ratios were calculated *(Figure 17)*.

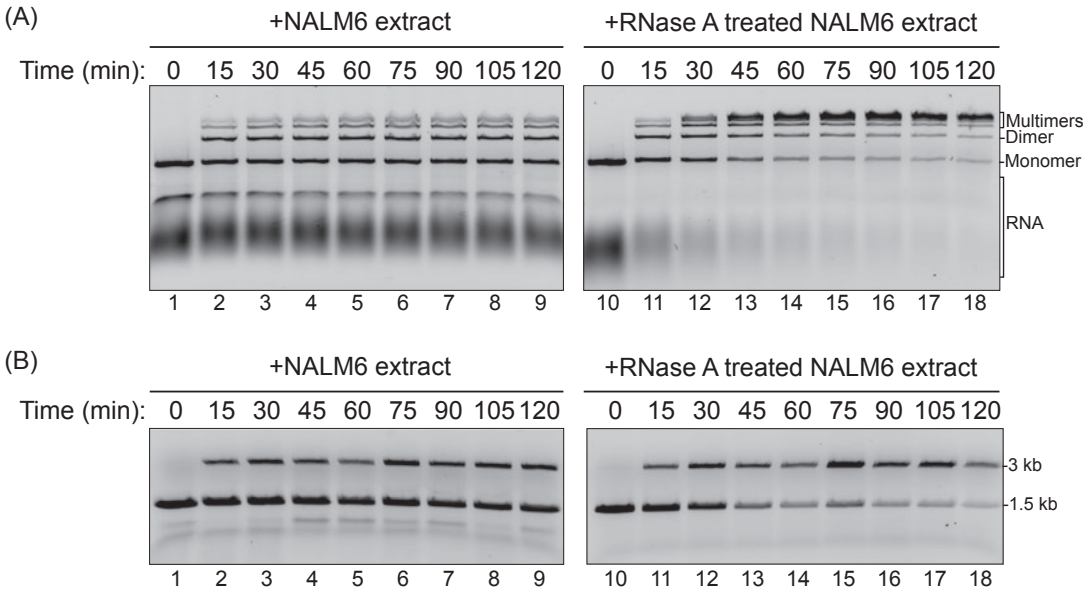


Figure 16 – (A) Representative gel image of a time-resolved end-joining reaction; (B) AvaII digestion of the samples purified from (A). DNA substrate is plasmid pKJ025 digested with EcoRI (5' AATT overhang)

The end-joining of all ends with 5' overhangs followed a similar pattern where end-joining appears to slow down or plateau later in the reaction *(Figure 17A)*. When the extract had been treated with RNase A, no plateauing of end-joining could be observed over 2 hours. A constant level of end-joining is also observed with blunt ended and 3' overhang containing DNA substrates, where the treatment of the extract with RNase A has no effect *(Figure 17B)*. The end-joining kinetics of the 5' overhangs after RNase treatment looks broadly similar to the end-joining of 3' overhangs and blunt ends. This data adds to the conclusion that the extract contains an RNA or an RNA-dependent factor that specifically inhibits end-joining when the DNA ends contain a 5' overhang.

It is also clear that the DNA sequence has a huge effect on the efficiency of end-joining. Consider 2 ends, both with 5' overhangs in *Figure 17A* – one with the overhang sequence of `AATT`, and the second with the sequence `GATC` in the overhang. The initial rate of end-joining and the end-joining after RNase treatment are both lower with the `AATT` overhang. This influence that sequence (possibly both directly at the end and proximal to it) has on end-joining is interesting and is worth exploring further *(III. Sequence biases in end-joining)*.

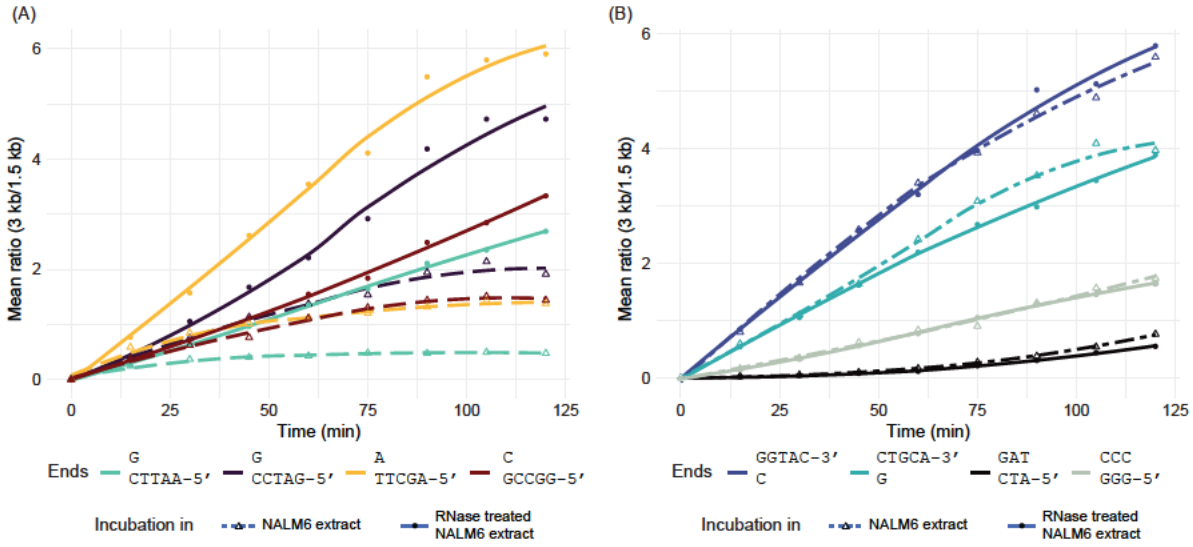


Figure 17 – Kinetics of end-joining with different substrates (A) Ends with 5' overhangs (B) Ends with 3' overhangs and blunt ends. Curve fitting done on R ggplot using the LOESS method.

The increased end-joining of 5' overhangs after RNase treatment is also dependent on the timing of RNase addition relative to the substrate *(Figure 18)*. Typically, the extract was pre-treated with RNase A for 20 minutes before addition of the DNA substrate. Reduction of this pre-treatment period or addition of RNase A after the substrate substantially reduced the end-joining stimulation through RNase A. Addition of the RNase A 30 minutes after the DNA substrate has no effect on the end-joining outcome at all.

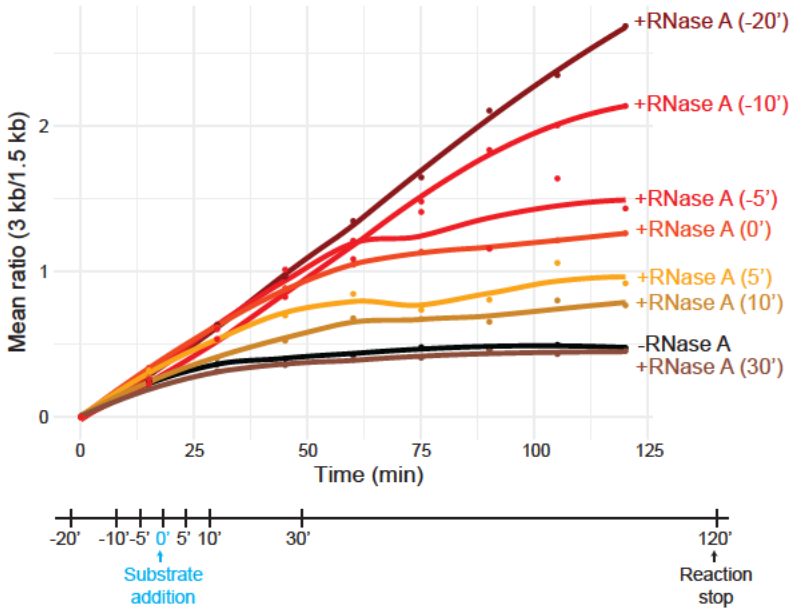


Figure 18 – Effect of the timing of RNase A addition on end-joining stimulation. The value in parentheses denotes the time at which RNase A was added to the extract relative to the substrate. 0' denotes the time of substrate addition. -5' and 5' mean that the RNase A was added to the extract 5 minutes before and after the substrate respectively. DNA substrate is pKJ025 digested with EcoRI (5' AATT overhang). Curve fitting was performed on R ggplot using the LOESS method.

RNase A having little or no effect once the end-joining has already started indicates 'irreversible changes' happening to the DNA very early in the reaction. The RNA present in the extract appears to be responsible for it, either acting alone or as part of a larger protein complex.

Unjoined monomers are not well ligated in further reactions

To understand the 'irreversible changes' happening to DNA ends in the presence of RNA, I decided to inspect the DNA molecules that don't end up as ligated products. A LigIV-deficient NALM6 extract cannot ligate DNA ends together *(Figure 11A)* and thus any DNA incubated in it is an 'unjoined monomer'.

DNA incubated in LigIV-deficient NALM6 extract for different amounts of time (15, 30 and 60 minutes) were purified and incubated in an end-joining reaction with a LigIV-proficient NALM6 extract. If the DNA end contains a 5' overhang, the initial pre-incubation reduces the amount of end-joining observed *(Figure 19A)*. The longer the pre-incubation was, the lower the end-joining is. The stimulation of joining mediated by RNase A treatment is also lost in these samples. If the pre-incubation was performed in extract treated with RNase A, this loss of end-joining is rescued. Such a reduction in end-joining appears to be limited to ends with 5' overhangs, as the pre-incubation has no effect on the further joining of ends with 3' overhangs *(Figure 19B)*.

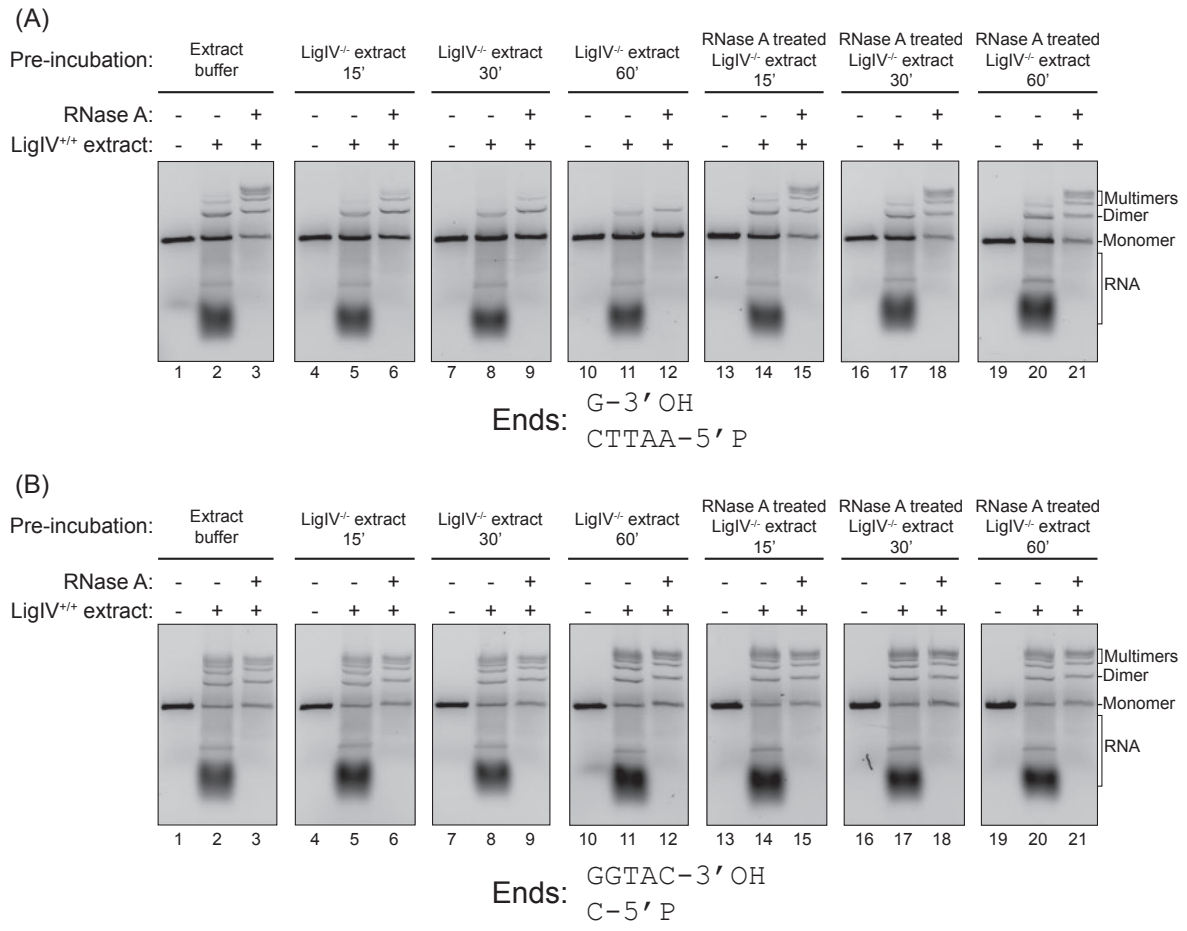


Figure 19 – 5' overhangs pre-incubated in LigIV-deficient NALM6 extract are joined poorly during a further incubation in a LigIV-proficient NALM6 extract. Substrate is pBlueScript digested with EcoRI (A) and KpnI (B).

Similar results can be observed when unjoined monomers are incubated in commercially available ligases too. Both T4 DNA ligase (NEB, #M0202) *(Figure 20)* and Quick ligase (NEB, #M2200) *(Figure 21)* exhibit lower efficiencies when ligating substrates that contain 5' overhangs.

Other substrates with different sequences in the 5' overhang also exhibit the same characteristic after a pre-incubation in LigIV-deficient extract *(Figure 22)*. This indicates an RNA-dependent processing targeted specifically towards ends that have 5' overhangs.

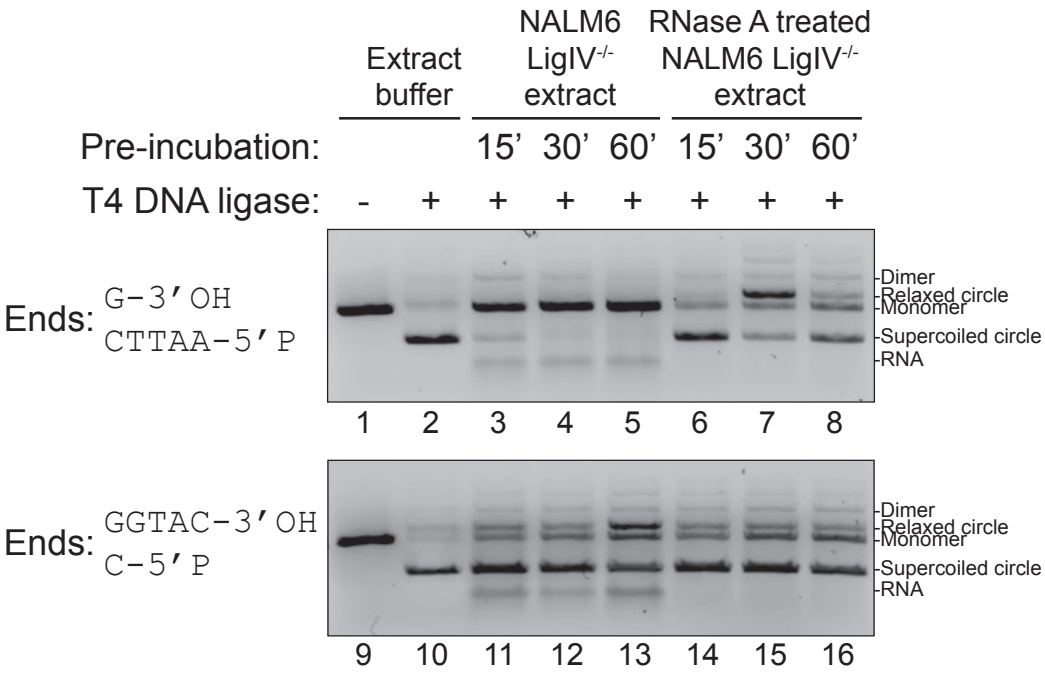


Figure 20 – Incubation of unjoined monomers in T4 DNA ligase. Substrates are pBlueScript digested with EcoRI (above, lanes 1-8) and KpnI (below, lanes 9-16). The ligation reaction was performed in 1x T4 DNA ligase buffer with 200 units of T4 DNA ligase, incubated at room temperature for 10 minutes.

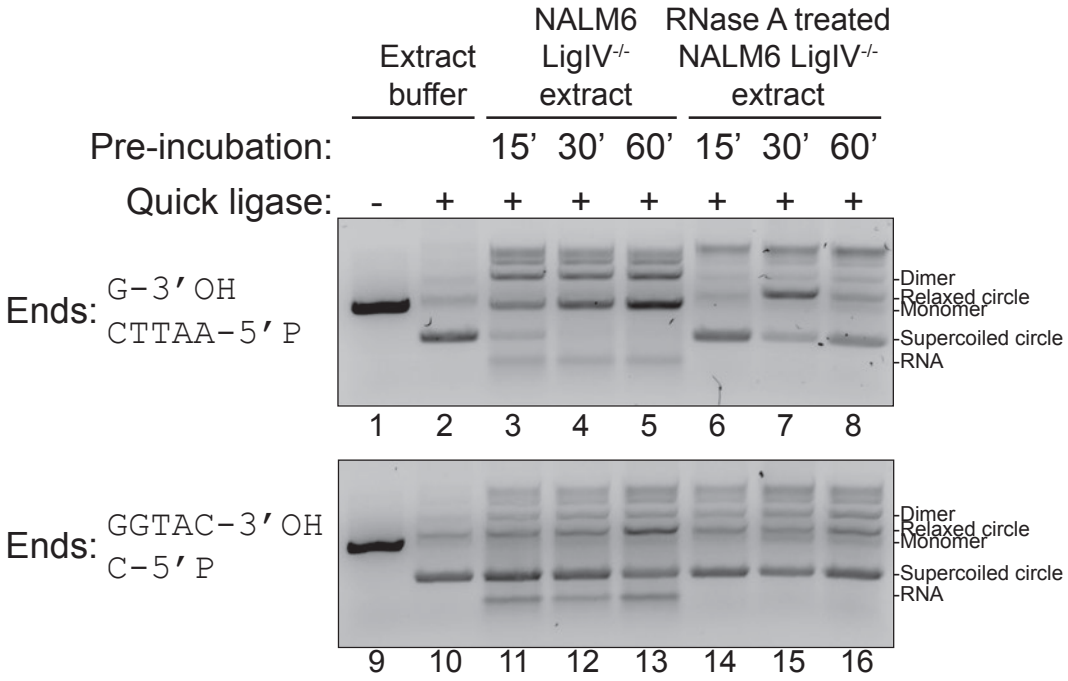


Figure 21 - Incubation of unjoined monomers in Quick ligase. Substrates are pBlueScript digested with EcoRI (above, lanes 1-8) and KpnI (below, lanes 9-16). The ligation reaction was performed in 1x Quick ligase buffer with 1 µl of Quick ligase, incubated at room temperature for 10 minutes.

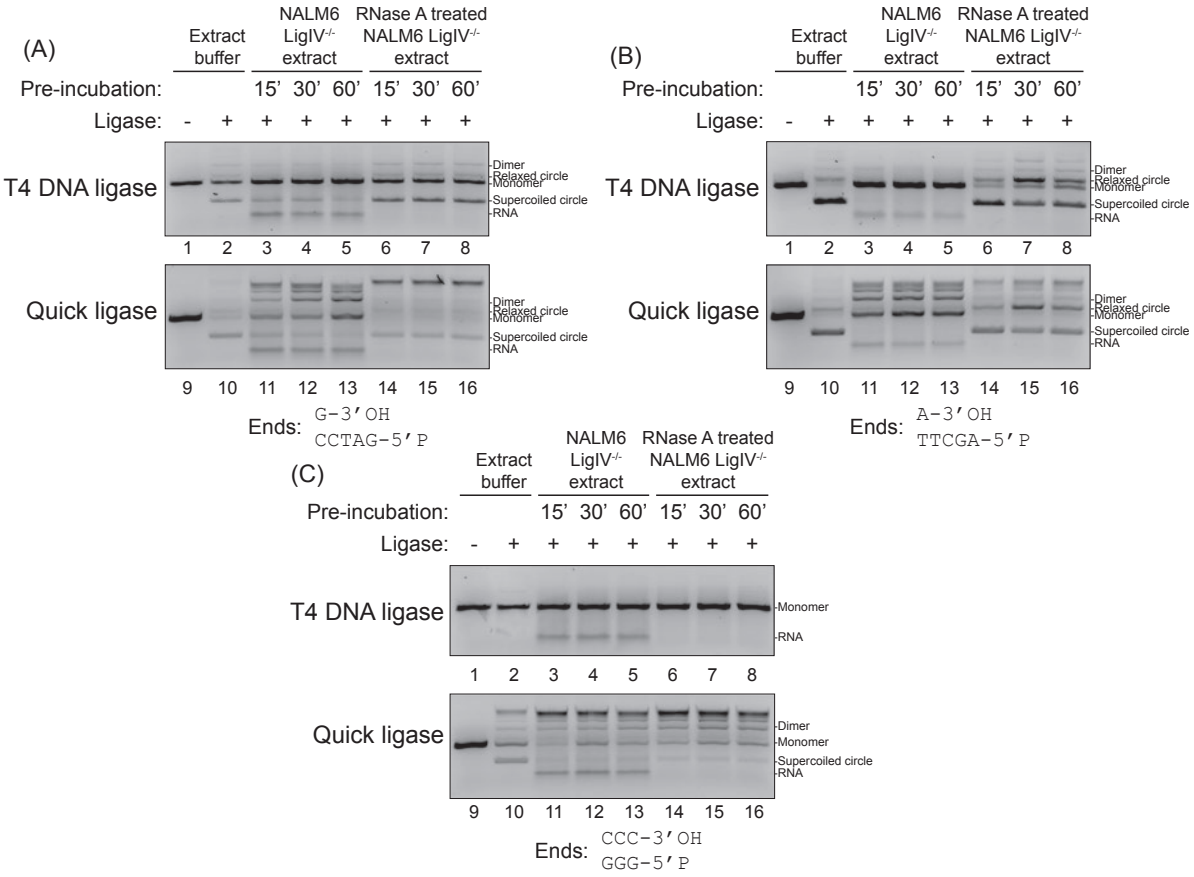


Figure 22 – Incubation of unjoined monomers from various substrates in T4 DNA ligase and Quick ligase. Substrates are pBlueScript digested with BamHI (A), HindIII (B) and SmaI (C). The ligations were performed as described in Figures 20 and 21.

Unjoined monomers show signs of nucleolytic processing

More information about the nature of the unjoined monomer ends can be gleamed by looking at the junction sequences generated by Quick ligase. So, I purified the samples in Figure 21. Fragments containing the junction were then isolated using AvaII digestion and gel extraction. ONT sequencing libraries were prepared and sequenced on a MinION flow cell.

Analysis of the junctions from 5' overhangs reveals deletion of bases at the junction – which implies that the ends are nucleolytically processed in the LigIV-deficient extract *(Figure 23)*. The nucleolytic processing is hindered upon RNase treatment of the extract or if the substrate has a 3' overhang. 4 bp deletions are highly represented in the spectrum of junction sequences. This is possibly an artefact of the assay. A 4 bp deletion leads to the removal of the overhang sequence and creates a blunt end. Such an end is more likely to be ligated than any other length of deletion because of the incompatibility of overhangs other deletion lengths would likely create.

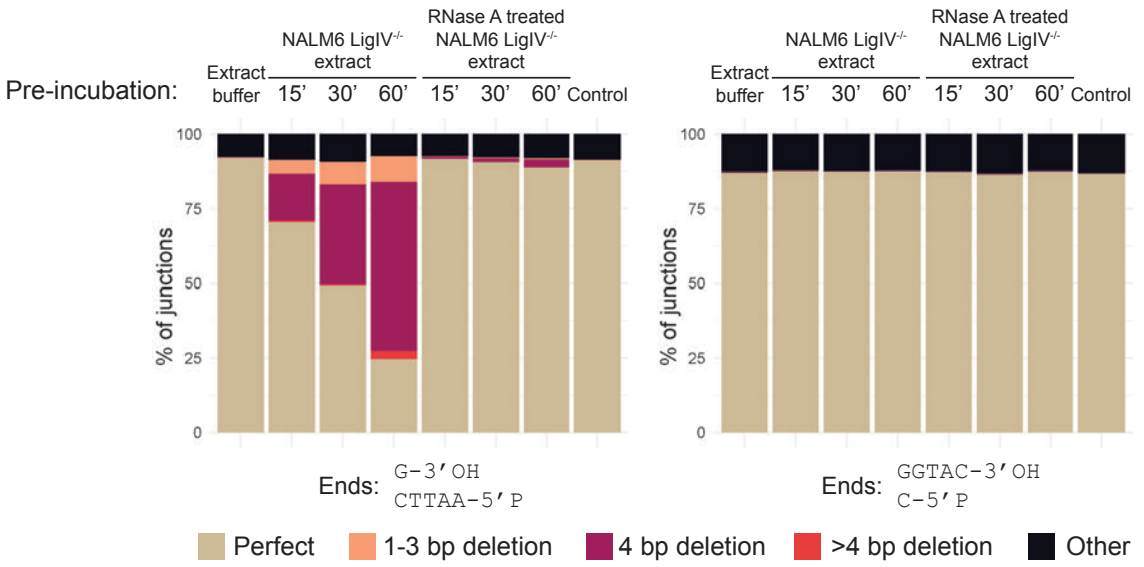


Figure 23 – Analysis of junctions produced by Quick ligase after incubation in LigIV-deficient NALM6 extract. The substrates were made by digesting pKJ025 with EcoRI (left) and KpnI (right). 'Other' sequences are a product of the error-prone nature of ONT long read sequencing. Control is AvaII-digested pKJ025 plasmid DNA. The actual sequences are shown in Figure 24.

Perfect junctions	GCTTGATATC**GAATTC**CTGCAGCCCGG
1-3 bp deletions	GCTTGATATC**GAAT-C**CTGCAGCCCGG GCTTGATATC**G-ATTC**CTGCAGCCCGG
4 bp deletions	GCTTGATATC**G----C**CTGCAGCCCGG
>4 bp deletions	GCCTG----------CAGCCCGG GCTTGATAT-----CTGCAGCCCGG
Other junctions	GCTTGATATC**GAATCC**CTGCAGCCCGG GCTTGATATC**GAAT-C**CTGCAGCCC-G GCTTGATATC**GAATTC**CTGCAGCCCCGG GCTTGATATC**GAATTC**CTGCAGCCC-G GCTTGATATC**GAATTC**CTGCAGCCCGGG GCTTGATATC**GAATTC**CTGCAGCC--G GCTTGATATC**GAATTC**CTGCAGCC-GG GCTTGATATC**GAATTC**CTGCAGCCGGG GCTTGATATC**G----C**CTGCAGCCC-G GCTTGATATC**G----C**CTGCAGCCCGGG GCTTGATATC**G----C**CTGCAGCC-GG GCTTGATATC**GAATTC**CTG-AGCCCGG

Figure 24 – Sequences of the Quick ligase junctions from Figure 23. Only sequences from the EcoRI-digested substrate DNA are shown. 'Other junctions' are sequences that possibly arise due to the error-prone nature of long read sequencing.

In order to make sure this processing is not an artefact of just the LigIV-deficient extract, I performed the same experiment, this time by purifying the unjoined monomers after incubation in a LigIV-proficient NALM6 extract through gel extraction. Similar results as previously were observed *(Figure 25)*. Thus, it can be concluded that DNA ends are likely

nucleolytically processed in the extract irrespective of the status of LigIV. This 'processing' is both RNA and end-structure dependent.

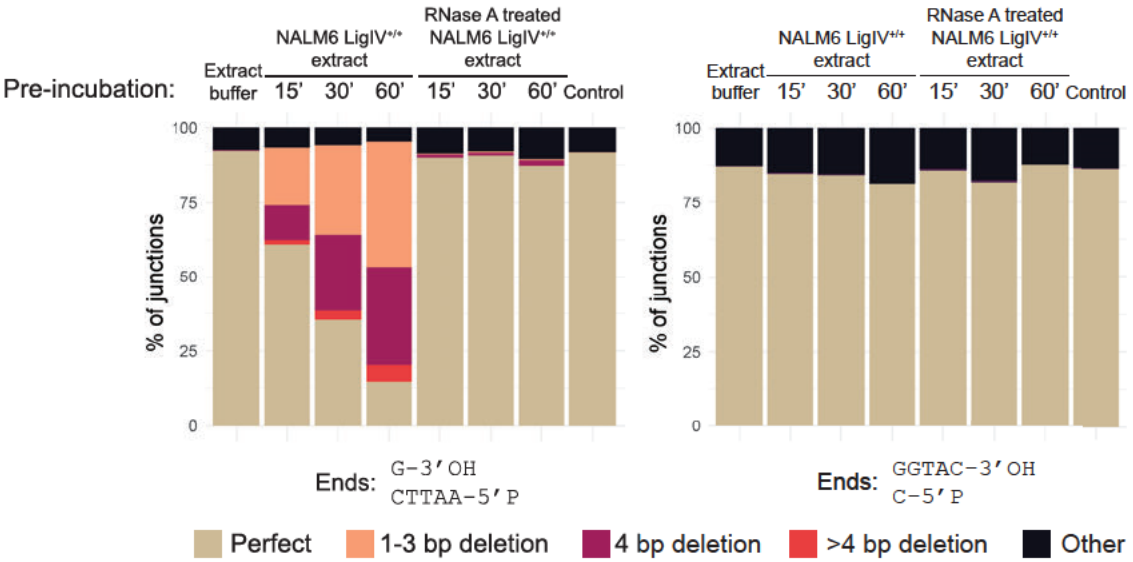


Figure 25 – Quick ligase junctions of unjoined monomers from a LigIV-proficient end-joining reaction. The substrates were made by digesting pKJ025 with EcoRI (left) and KpnI (right). The unjoined monomer fragments were purified by gel extraction and incubated in Quick ligase. The junctions were isolated by AvaII digestion and used in ONT library prep. Control is AvaII-digested pKJ025 plasmid DNA.

5' overhangs are processed by an RNA-dependent nucleolytic factor

Knowing that the 5' overhangs are most likely processed in a nucleolytic manner, I made radiolabelled substrates that can be used to study end-processing in an unbiased manner *(Design of radio-labelled substrates to study 5'-3' processing)*. The substrates were prepared by the ligation of a duplex containing the desired ends to a HindIII-digested pKJ025 plasmid DNA *(Figure 26A)*. The fragments arising from plasmid dimers serve as the internal loading control. After incubation in the LigIV-deficient extract, the DNA was purified and digested with the restriction endonucleases ClaI and SpeI, to get smaller fragments *(Figure 26B)*. Only the fragments from the ClaI digestion will be presented here henceforth since they resolve much better into individual fragments.

The end-processing assay clearly demonstrates the RNA-dependent 5' nucleolytic activity on 5' overhangs *(Figure 27)*. There is a ladder of processed 5' end fragments – only with substrates that contain 5' overhangs. The RNase treatment of the extract also rescues the 5' end processing. The 2 ends with 5' overhangs also show different levels of end processing here. The nucleotide sequence of the 2 substrates is identical, except the 4 nucleotides in the overhang. Based on the data here, it is very likely that the sequence in the overhang carries a large influence on the amount of end-processing, and hence ultimately, the level of end-joining.

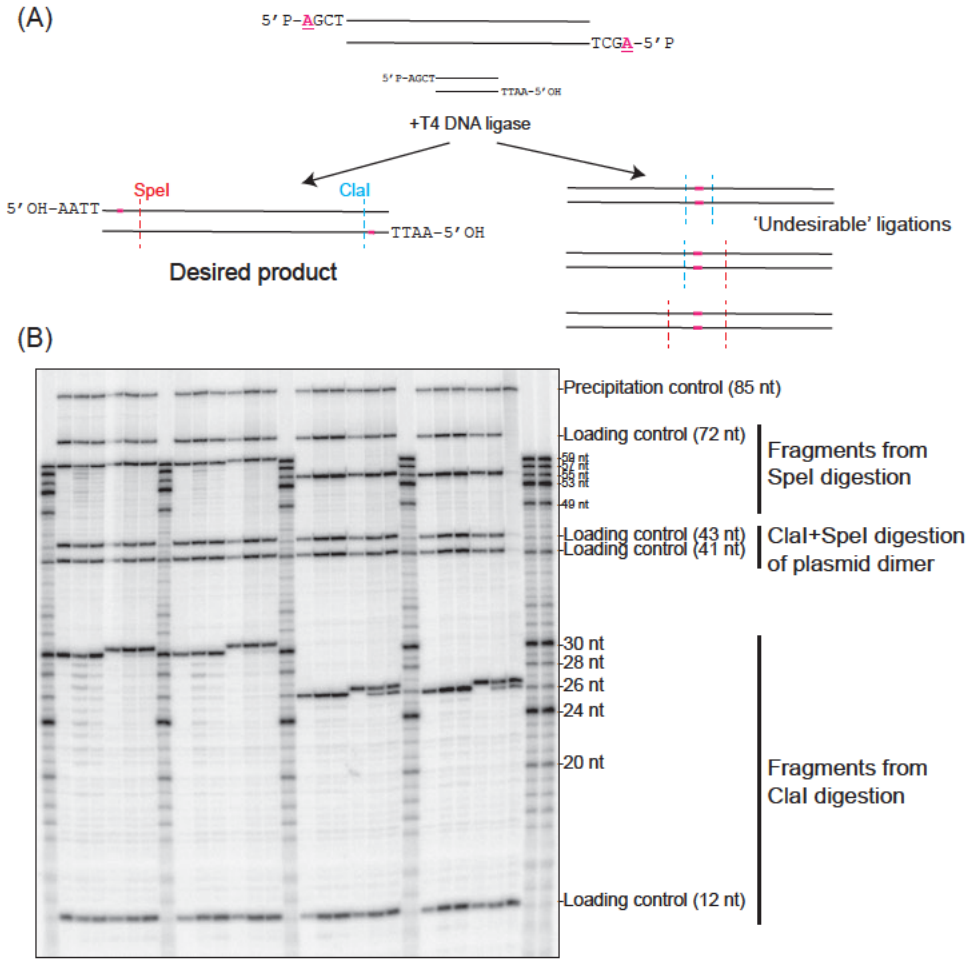


Figure 26 – 5' end processing was examined using radiolabelled substrates. (A) Schematic for the generation of substrates for the end processing assay; (B) Model denaturing polyacrylamide gel showing bands corresponding to 5' end fragments.

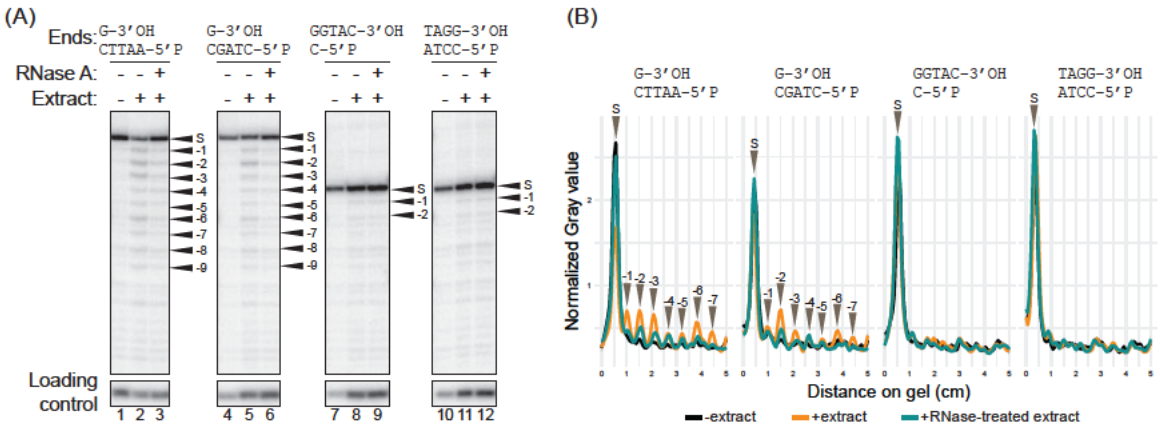


Figure 27 – The 5' ends that are part of a 5' overhang are specifically processed. (A) Radiolabelled substrates with the indicated ends were incubated in extract buffer, NALM6 extract or RNase A treated NALM6 extract for 60 minutes. S denotes the original length of the fragment (B) Quantification of the gel in (A); a line was drawn in each lane on ImageJ and the gray value intensities along the line were measured. The data was normalized to the total intensity in each lane and plotted. Curve fitting was performed on R ggplot with the LOESS method.

The 5’ end processing is proportional to the amount of time that the substrates are incubated in the extract *(Figure 28)*. The use of phosphorothioate bonds for the first 5 bases from the 5’ end to inhibit nuclease activity (Stein et al., 1988) limited the processing *(Figure 29)* – showing that it is indeed an exonucleolytic processing activity. On the other hand, very little/no processing was observed in the 3’-5’ direction *(Figure 30)*. Some processing was detected for 3’ overhangs, but it is neither at the level of the 5’ end processing nor RNA-dependent.

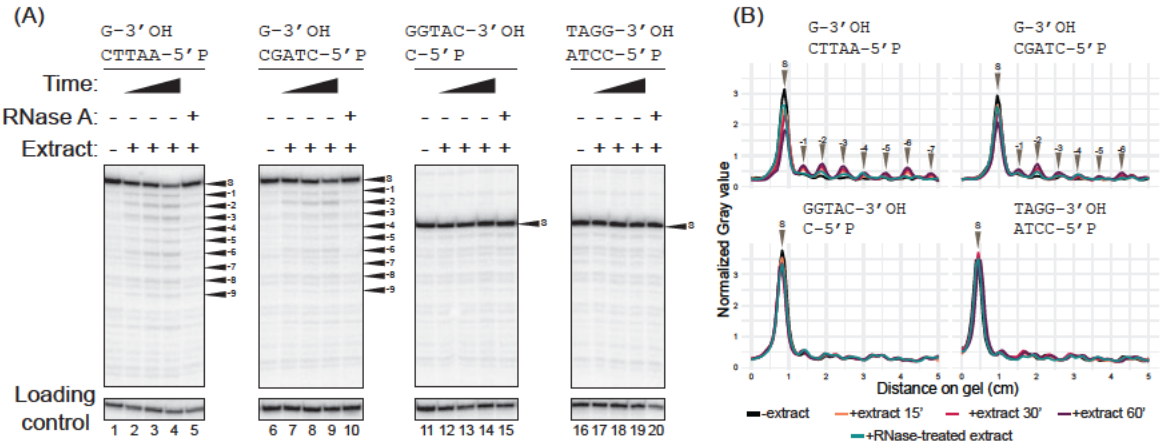


Figure 28 – 5’ end processing resolved by time. (A) The substrates were incubated in LigIV-deficient NALM6 extract for 15, 30 or 60 minutes. S denotes the original length of the fragment. (B) Quantification of the gels in A.

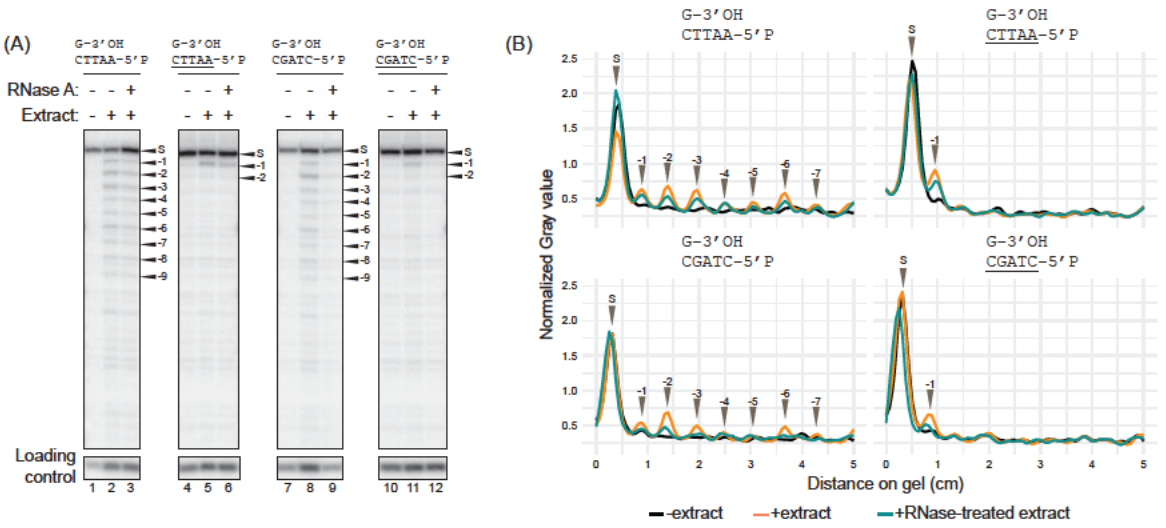


Figure 29 – Phosphorothioate bonds block 5’ end nucleolytic processing.(A) The substrates were incubated in LigIV-deficient NALM6 extract for 60 minutes. S denotes the original length of the fragment. The underlined bases are joined by phosphorothioate bonds. (B) Quantification of the gels in A.

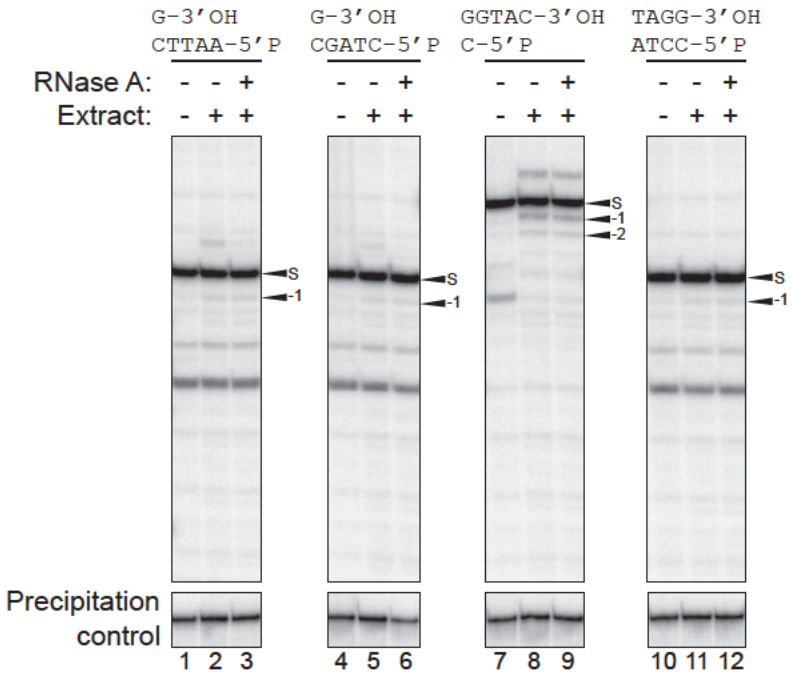


Figure 30 – There is no 3' end processing in the extract. The substrates were incubated in LigIV-deficient NALM6 extract for 60 minutes. S denotes the original length of the fragment.

5' end processing is dependent on the activation of DNA-PK

The observations made on the RNA-dependent 5' end processing so far are interesting – but have limited value in explaining biology. The end processing can be explained as the reason behind the initial result – 5' overhangs are not well joined by NALM6 extracts, unless the extracts are RNase A treated. However, it is yet unclear if the processing occurs during or independent of Non-homologous end-joining.

DNA-PK phosphorylation, which leads to its activation, is one of the major checkpoints of NHEJ and can be inhibited by small molecule inhibitors. NU7441 is a highly selective ATP-competitive inhibitor of DNA-PK (Leahy et al., 2004; Liang et al., 2022). NU7441 was also previously shown to inhibit end-joining in LigIV-proficient NALM6 extracts *(Figure 11B)*.

Inhibiting DNA-PK during the end-processing assay leads to a dramatic reversal of 5' end processing *(Figure 31)*. At a 10 µM concentration, NU7441 completely inhibits end-joining in LigIV-proficient extract *(Figure 11B)*. At the same concentration, end-processing is also almost completely abrogated. This shows that 5' end processing only happens after the activation of DNA-PK during NHEJ.

The removal of ATP from the reaction also has an inhibitory effect on 5' end-processing *(Figure 32)*. Absence of ATP should prevent DNA-PK phosphorylation, among other things. So, on its own, this result may not mean much, but along with NU7441, it supports the claim that the 5' end processing observed in this system happens in the context of NHEJ.

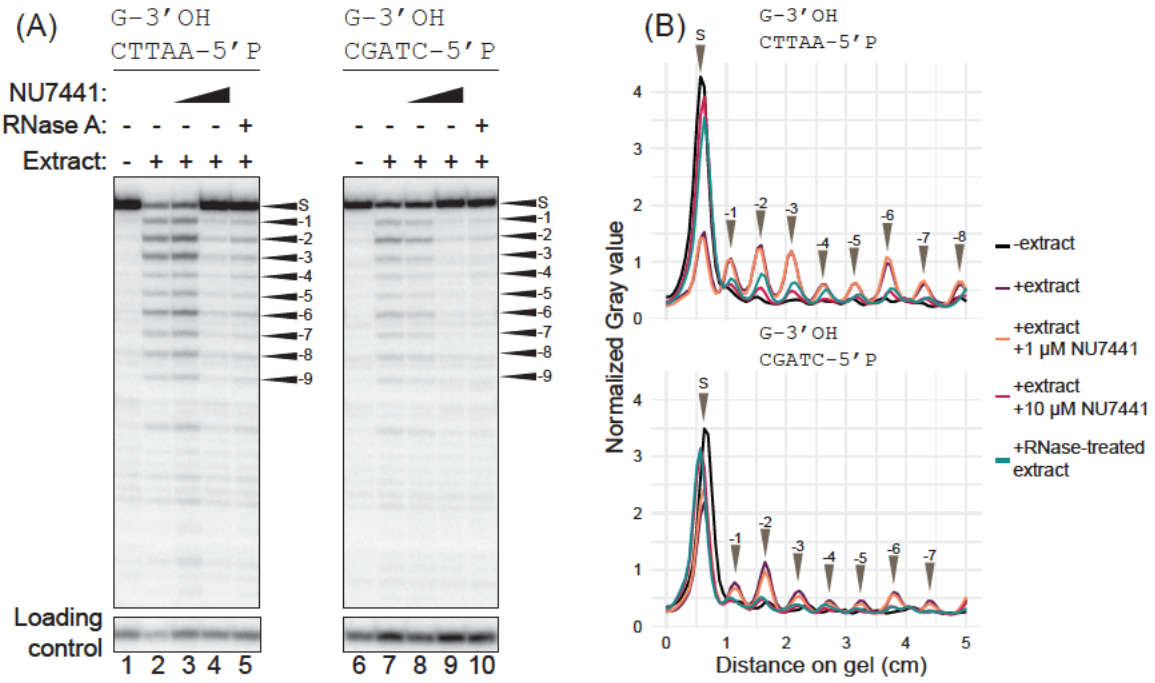


Figure 31 – 5' end processing upon the inhibition of DNA-PK activation by NU7441. (A) The substrates were incubated in LigIV-deficient NALM6 extract for 60 minutes. 1 μM (lanes 3 & 8) or 10 μM (lanes 4 & 9) NU7441 were added to the reaction. In lanes without NU7441, DMSO was added to a concentration of 1%. S denotes the original length of the fragment. (B) Quantification of the gels in A.

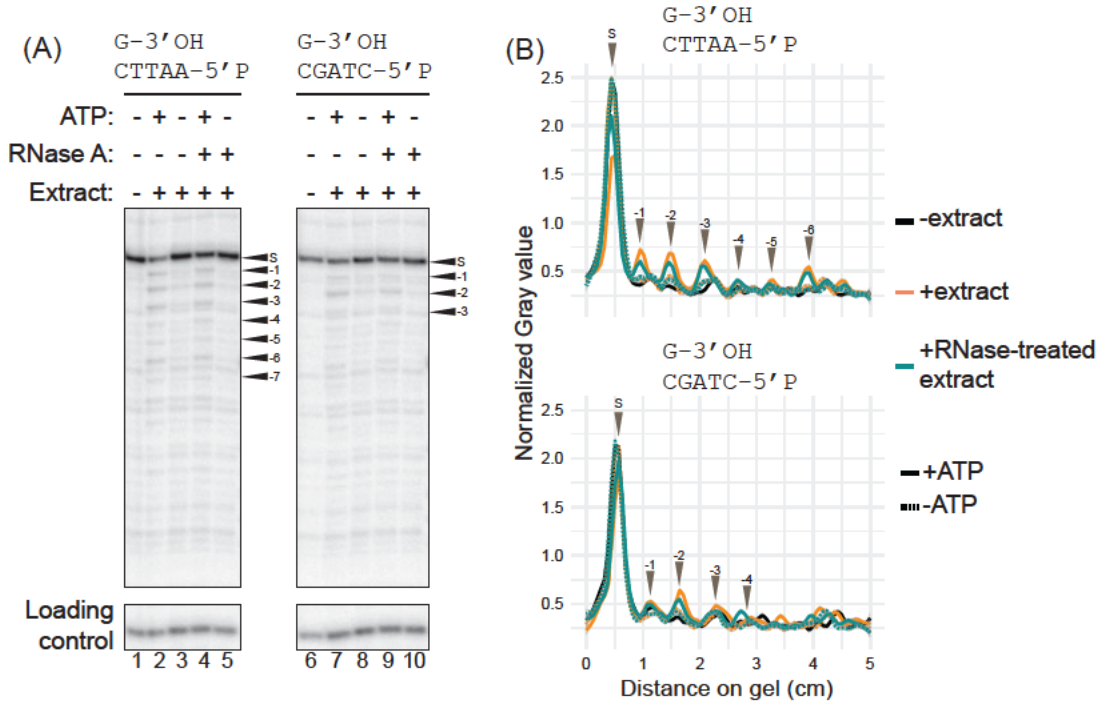


Figure 32 – Effect of ATP on 5' end processing. (A) The substrates were incubated in LigIV-deficient NALM6 extract for 60 minutes. S denotes the original length of the fragment. (B) Quantification of the gels in A.

Aside from end-structure, it appears that the chemistry of the DNA ends is also a determining factor in whether 5' ends are processed *(Figure 33)*. A phosphorylated 5' overhang is necessary for the processing *(Figure 33A)*. When the 5' overhang has been dephosphorylated, the presence of a 3' phosphate on the complementary strand leads to a small amount of processing, but not to the levels observed previously. 5' ends that are part of 3' overhangs remain protected from processing irrespective of their phosphorylation status *(Figure 33B)*.

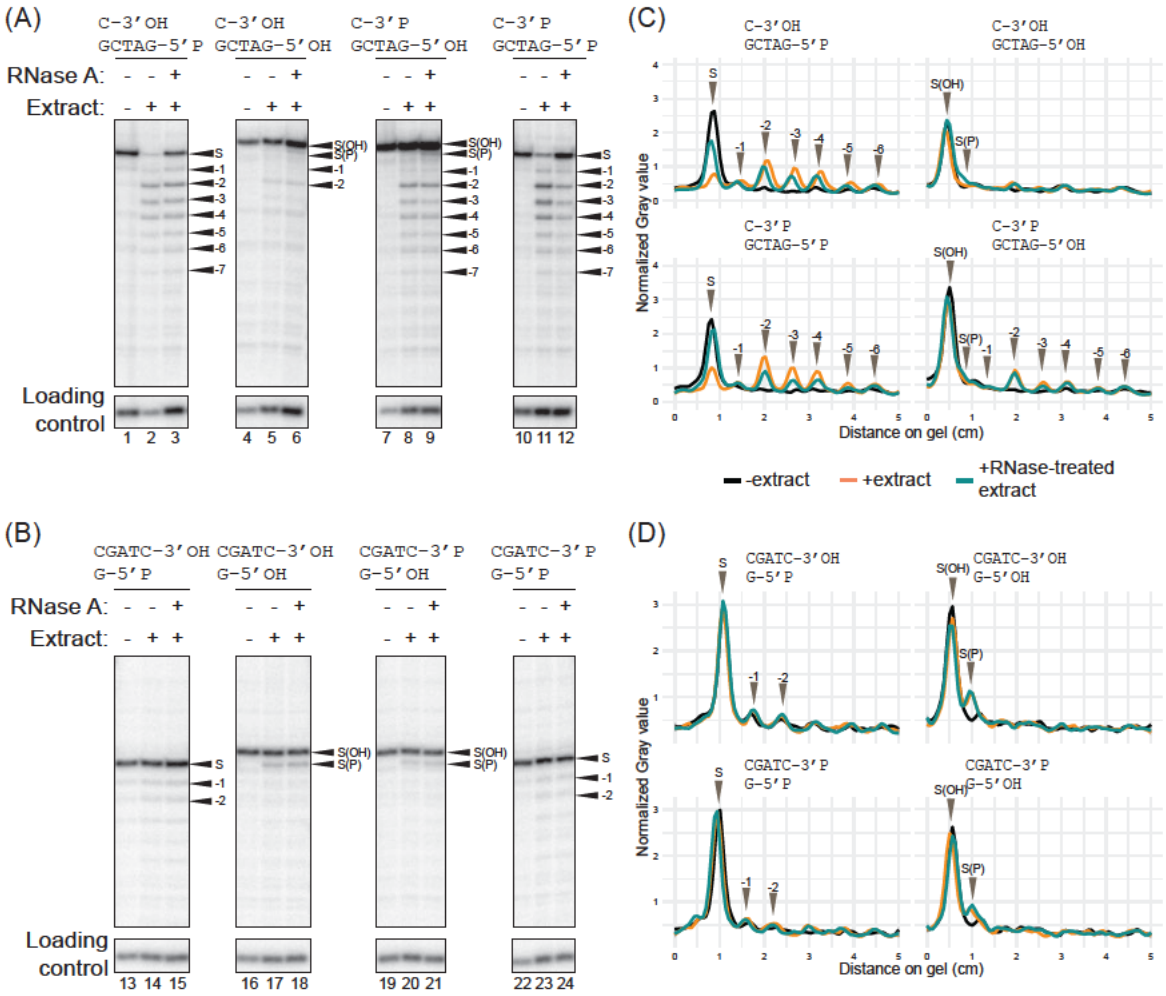


Figure 33 – Effect of DNA end chemistry on 5' end processing. DNA with 'atypical' ends containing 5' overhangs (A) or 3' overhangs (B) were incubated in LigIV-deficient NALM6 extract for 60 minutes. S denotes the original length of the fragment. (C,D) Quantification of the gels in A,B respectively.

This resistance of dephosphorylated 5' ends towards processing also reflects in the end-joining results *(Figure 34)*. Both substrates with 5' phosphates show an increase in end-joining after RNase treatment of the extract. RNase A has no effect on the end-joining for 5'OH+3'OH, whereas there might be a mild improvement for 5'OH+3'P.

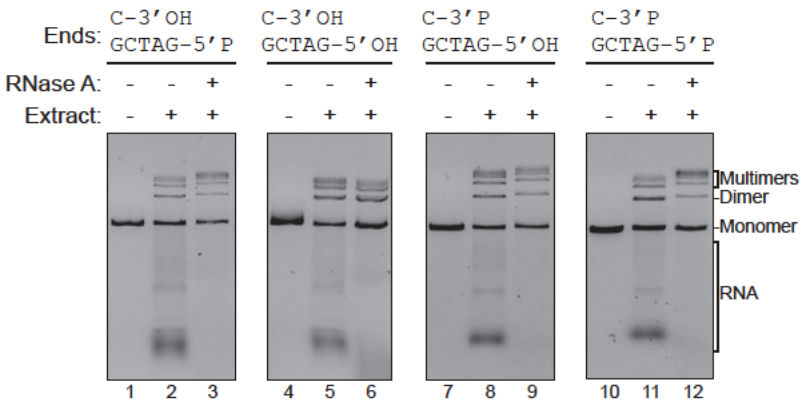


Figure 34 – End-joining of 5' overhang containing substrates with 'atypical' ends. End-joining reactions using substrates with the described ends were performed for 2 hours.

Effect of the 5' end processing on end-joining outcome

The nucleolytic end-processing means that there should be a prevalence of junctions with deletions in an end-joining reaction. To test this hypothesis, I sequenced the junctions from an end-joining reaction with EcoRI-digested pKJ025. The junctions were purified after AvaII digestion and sequenced on a minION flow cell. Sequence analysis showed that the majority of the junctions did not present any signs of processing *(Figure 35)*. Nevertheless, end-joining events with deletions accumulated late into the reaction. RNase treatment of the extract was able to prevent deletions in line with the expectations based on the end-processing assays.

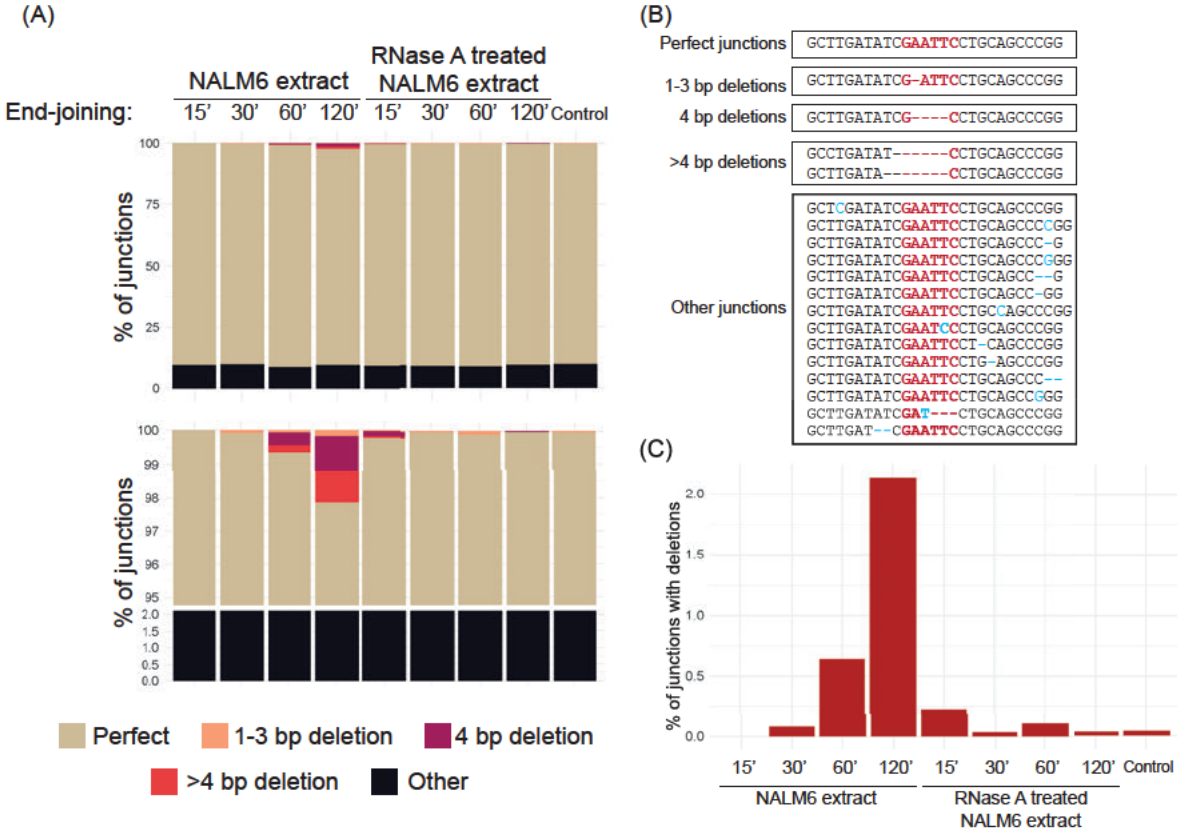


Figure 35 – Analysis of junctions from an end-joining reaction. The substrate is pKJ025 digested with EcoRI (5' AATT overhang). Control is substrate ligated with T4 DNA ligase. (A) (top)

Spectrum of junctions from the end-joining reaction; (bottom) same graph magnified to show junctions with deletions; (B) Sequences of the junctions; (C) Graph showing the percentage of junctions with deletions in each junction.

II. Identification of the processing factor

RNase treatment of the extract causes variance in the DNA-end binding proteome

A proteomics based approach was employed to identify the RNA-dependent processing factor. Biotinylated substrates were made with different end structures and sequences *(Preparation of biotinylated DNA substrates for pulldown)*. These substrates were able to pull down proteins binding to the DNA ends *(Figure 36)*. The use of photocleavable biotin leads to a cleaner pulldown since the proteins that bind non-specifically to the magnetic beads are removed from further analysis. Mass spectrometry identified peptides belonging to core proteins of the non-homologous end joining pathway among the top hits in the photo-released fraction *(Figure 36B)*.

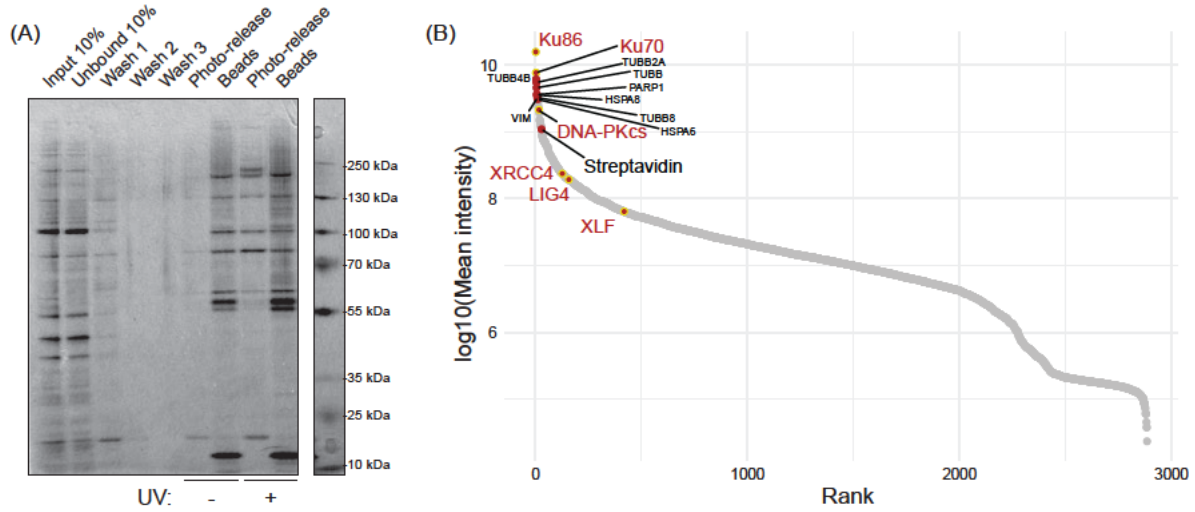


Figure 36 – Pulldown of DNA-end bound proteins. (A) Silver stained gel showing the clean pulldown using photo-cleavable biotin; (B) Rank plot of the MS intensities. Top hits and core proteins of the NHEJ pathway are highlighted.

Principal component analysis (PCA) of the normalized peptide intensities shows a clear difference between the proteins binding to 5' overhangs and others *(Figure 37)*. Whereas the RNase treatment of the extract doesn't make much of a difference in proteins pulled down by 3' overhangs and blunt ends, there is a dramatic shift along PC1 for proteins pulled down by 5' overhangs.

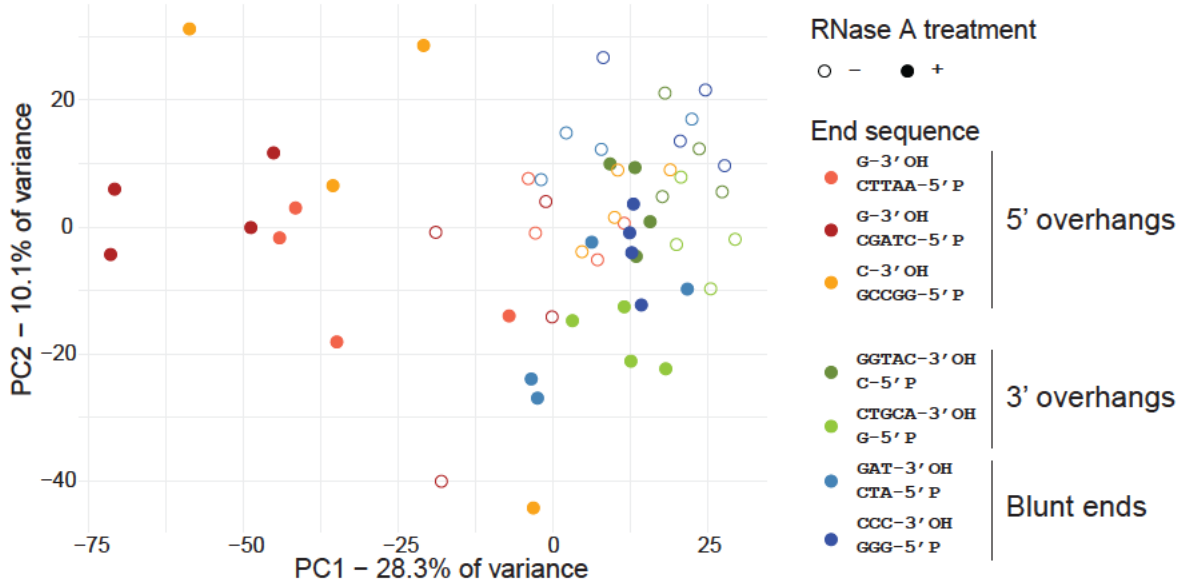


Figure 37 – Principal Component Analysis (PCA) biplot of the proteins pulled down by different ends.

With a log2 Fold Change threshold of 2 along with an adjusted p value (Benjamini-Hochberg correction) under 0.05, 217 proteins were identified to be differentially bound to 5' overhangs after RNase treatment *(Figure 38)*. Out of those, only 16 proteins associated more with the ends after RNase treatment. Among these proteins was an enrichment of aminoacyl tRNA synthetases *(Figure 39)*. Proteins involved in RNA splicing are among the top hits of proteins that bind less efficiently to 5' overhangs after RNase treatment.

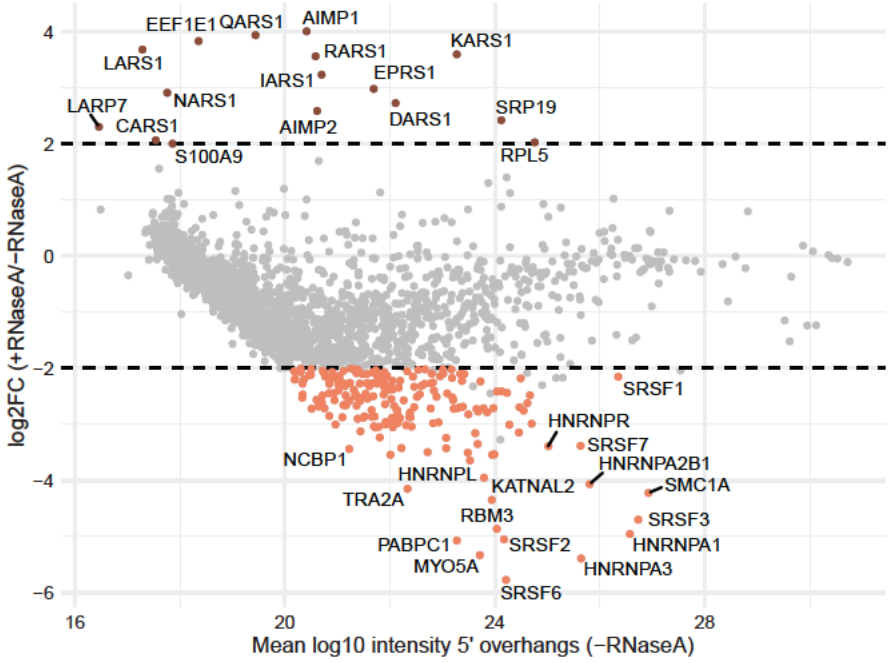


Figure 38 – Proteins differentially binding to 5' overhangs upon RNase A treatment of extract. Proteins with an absolute LFC > 2 and adjusted p value < 0.05 are highlighted.

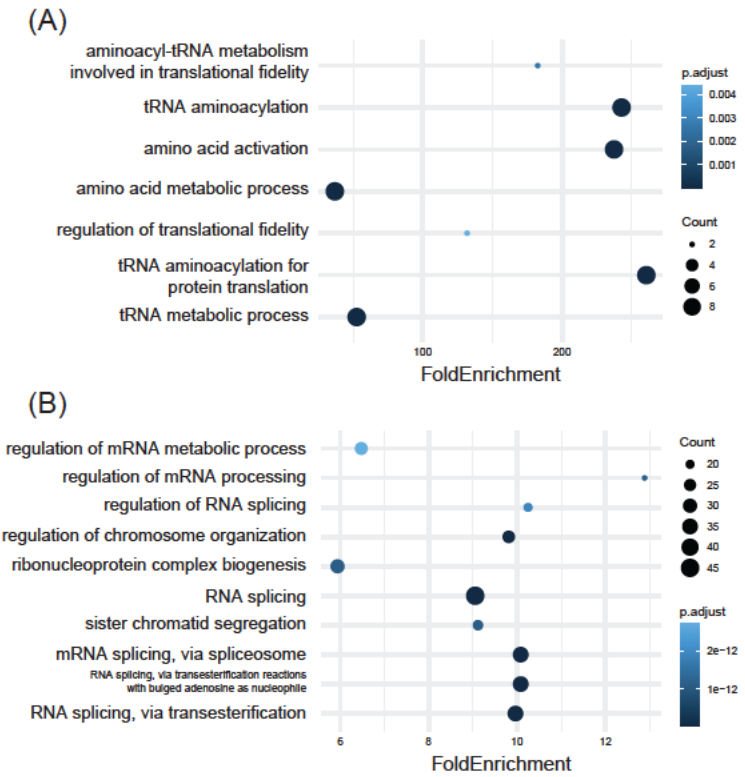


Figure 39 – GO term enrichment of proteins differentially binding to 5' overhangs upon RNase A treatment. (A) Proteins binding after RNase treatment; (B) Proteins released after RNase treatment

Under the same thresholds, only 44 proteins were differentially bound to 3' overhangs and blunt ends *(Figure 40)*. While a similar number of proteins associated with the ends after RNase treatment, there are significantly fewer proteins that bind ends with lower efficiency after RNase treatment.

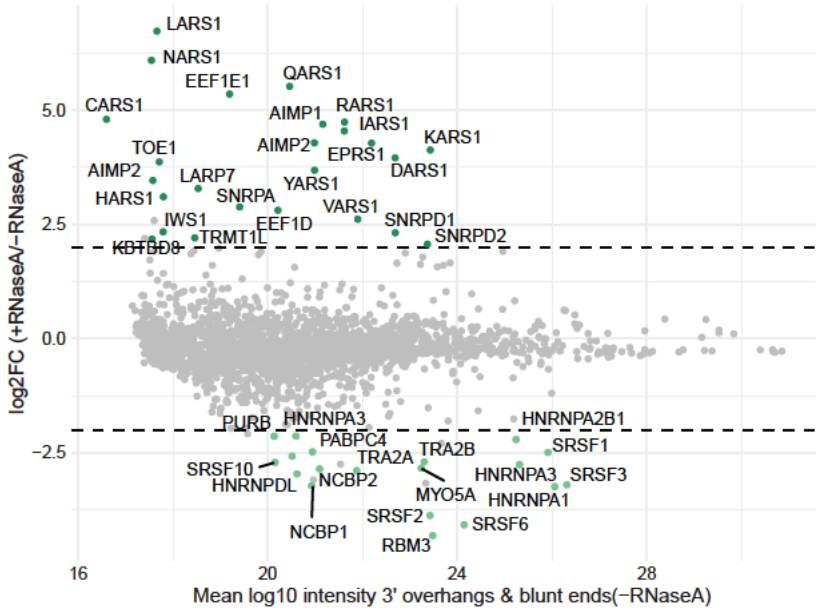


Figure 40 - Proteins differentially binding to 3' overhangs and blunt ends upon RNase A treatment of extract. Proteins with an absolute LFC > 2 and adjusted p value < 0.05 are highlighted.

GO term enrichment analysis also indicates functionally similar proteins among the hits as previously *(Figure 41)*. Aminoacyl tRNA synthetases and proteins involved in RNA splicing are the top GO terms again.

Figure 41 - GO term enrichment of proteins differentially binding to 5' overhangs upon RNase A treatment. (A) Proteins binding after RNase treatment; (B) Proteins released after RNase treatment

The broad similarity in GO terms among the hits implies a common turnover of proteins upon RNase treatment of the extract that is independent of the type/structure of the DNA end. There are 165 proteins that bind differentially only to 5' overhangs, without any change in binding to other substrates upon RNase treatment *(Figure 42)*. Proteins involved in maintaining chromosomal architecture are now among the top hits.

Protein components from 2 well known multi-protein complexes were found enriched among the hits – the pBAF chromatin remodelling complex *(Figure 43)*, and the Cohesin complex *(Figure 44)*. Almost all proteins of both these complexes were found to have lower association with the 5' overhangs after RNase treatment of the extract. For the other substrates, RNase treatment had little or no effect. The presence of all the complex components reduces the possibility of any false positives and increases confidence in the hits.

Chemical allosteric inhibitors for the catalytic subunit of pBAF – BRG (SMARCA4) are commercially available (Papillon et al., 2018). The inhibitor BRM014 was kindly gifted by Dr. Sandra Schick, along with antibodies against various components of the general BAF or pBAF complex.

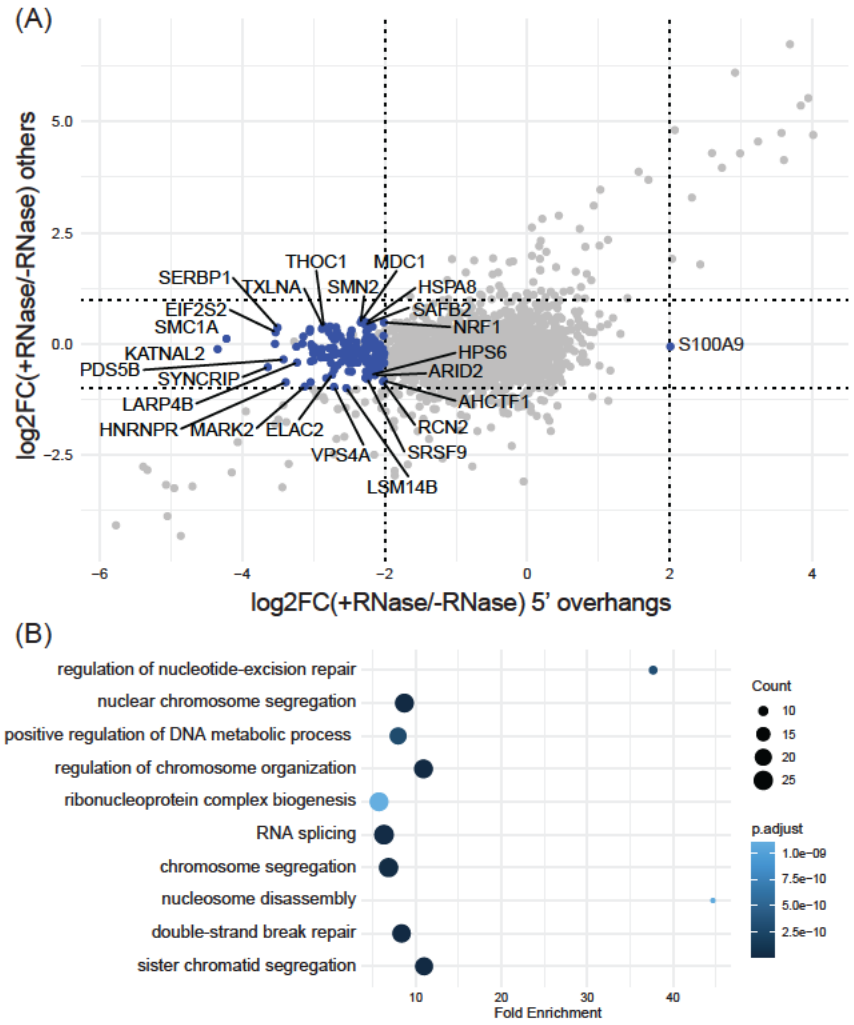


Figure 42 – Hits only in 5' overhangs. (A) Proteins that are hits in 5' overhangs but have an absolute LFC < 1 in the other substrates are highlighted; (B) Enriched GO terms among the 5' overhang-specific hits

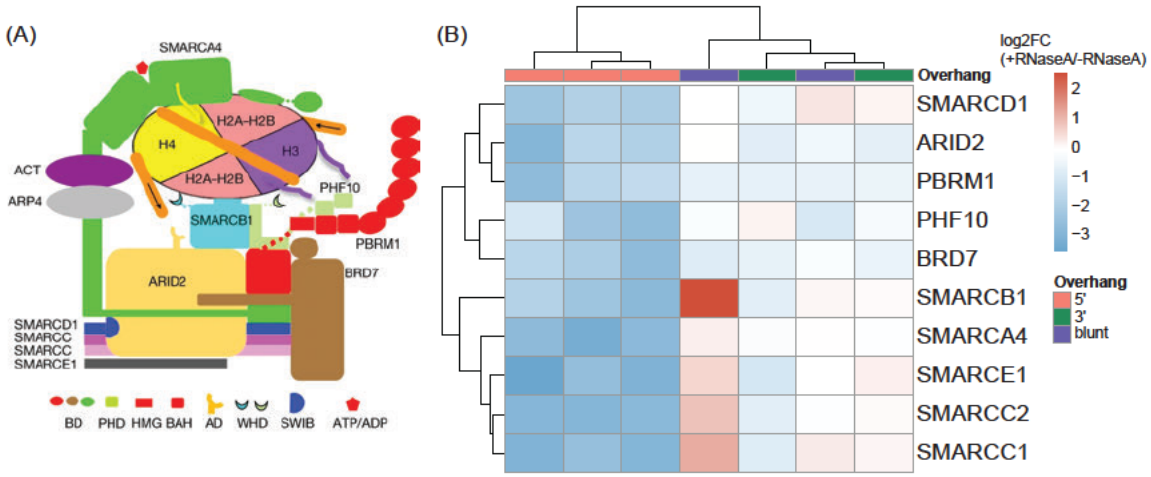


Figure 43 – The pBAF complex was differentially bound to 5' overhangs after RNase treatment. (A) Human pBAF complex, from Yuan et al. (2022); (B) Heatmap of LFC after RNase treatment

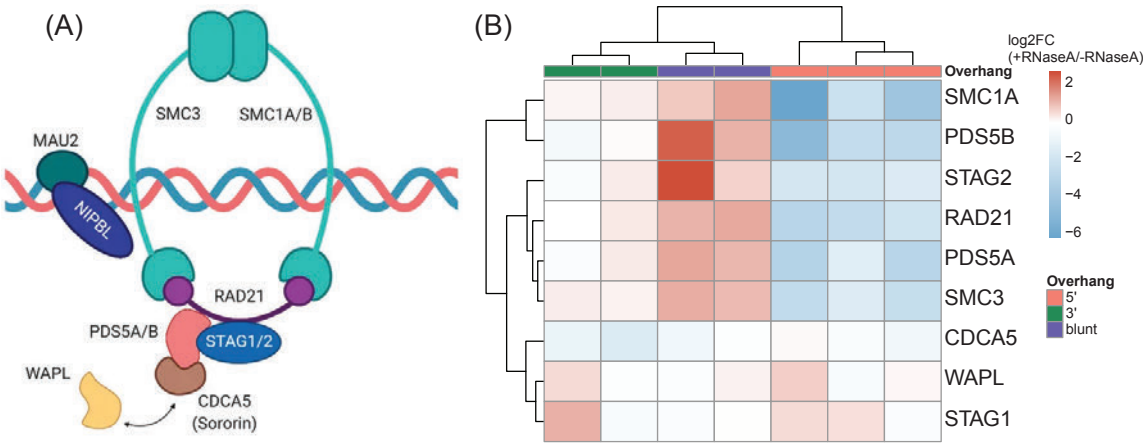


Figure 44 – Components of the cohesin complex were bound less efficiently to 5' overhangs after RNase treatment. (A) Cohesin complex, from Kuru-Schors et al. (2021); (B) Heatmap of LFC after RNase treatment

As a preliminary test, the BRM/BRG1 inhibitor BRM014 was added to the end-joining reaction *(Figure 45A)*. If the pBAF complex was indeed involved in 5' end processing of 5' overhangs, one would expect pBAF inhibition to stimulate end-joining the same way RNase treatment does. In this experiment, addition of the inhibitor had no effect on end-joining. Neither did the incubation of antibodies targeting pBAF components *(Figure 45B)*. These negative results don't necessarily exclude the pBAF complex from performing a functional role in RNA-dependent 5' end-processing since the efficacy of the inhibitors in the assay cannot be verified. More experiments need to be performed to identify the 5' end-processing factor.

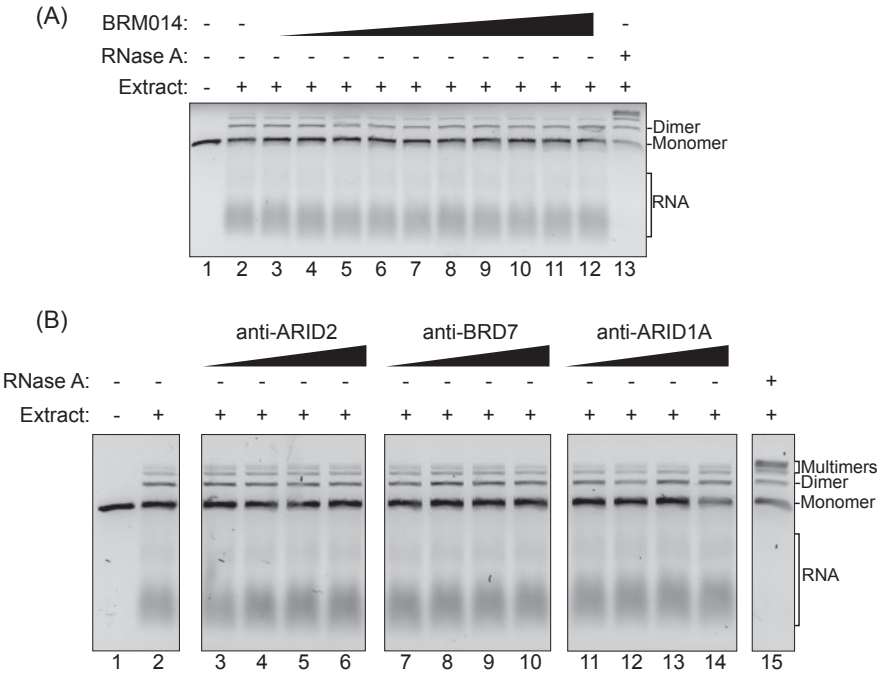


Figure 45 – pBAF inhibition had no effect on end-joining. (A) Inhibitor BRM014 was added at various concentrations from 2 nM to 20 µM in an end-joining reaction.1% DMSO was added to samples where no BRM014 was added. (B) Antibodies against various components of the pBAF complex were added to an end-joining reaction. ARID1A is a non-pBAF control. The substrate is pBlueScript digested with EcoRI (5' `AATT` overhang).

III. Sequence biases in end-joining

Differences in in vitro end-joining efficiencies have been observed depending on the nucleotide sequence of the ends despite similarities in end structures *(Figure 12, Figure 17)*. As a simple example among blunt ended substrates, a SmaI-digested DNA (C:G at the end) is joined much better than an EcoRV-digested DNA (T:A at the end) *(Figure 12)*. Even among substrates with overhangs, the end-joining efficiency varies, dependent on the sequence in the overhangs and possibly even around it *(Figure 17)*. This is of huge biological interest – faster or slower NHEJ in different areas of the genome depending on the sequence context would add to our knowledge on the regulation of double-strand break repair through NHEJ. In order to systematically understand the role sequence plays in NHEJ, substrates with different configurations of random nucleotides at the end or near the end were made *(Generation of random DNA end sequence libraries)*.

Impact of nucleotides at a blunt end

Using the type IIS restriction enzyme MlyI that cleaves outside its recognition sequence (GAGTC), a library of substrates was prepared with a random mix of bases for 5 nucleotides upstream and downstream of the break *(Figure 46A)*. From the individual nucleotide frequencies at each position, it can be observed that A/T is represented more often than G/C, even in the undigested library. The nucleotide frequencies from the junctions made by in vitro NHEJ or T4 DNA ligase were normalized to the undigested control.

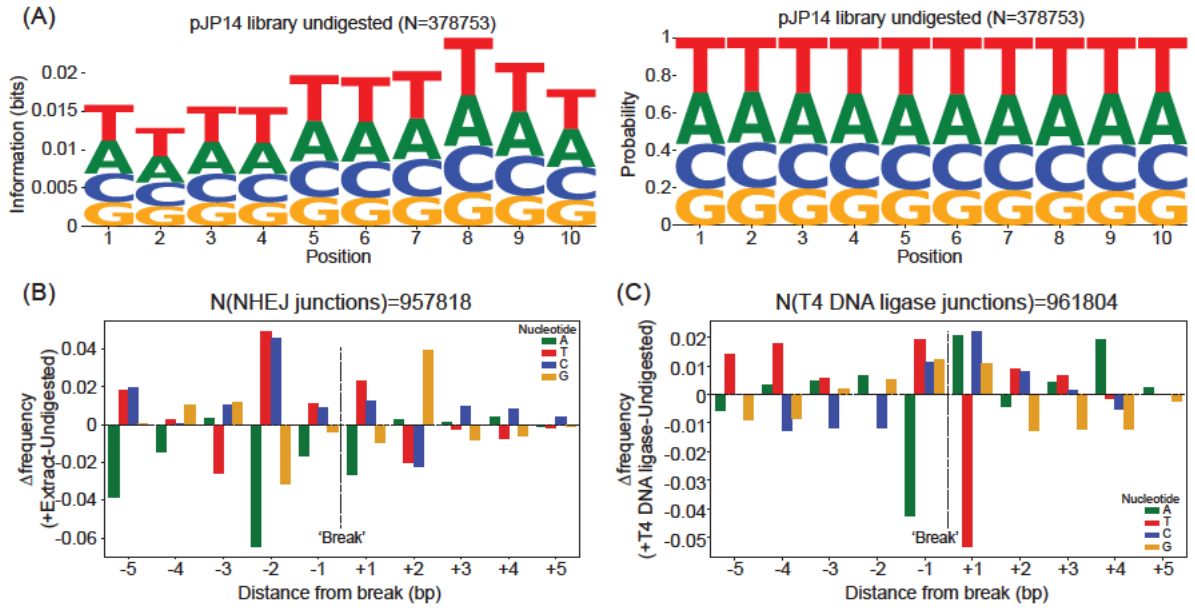


Figure 46 – Nucleotide biases in the end-joining of a blunt-ended substrate. (A) Sequence logo plots of the undigested pJP14 library; change in nucleotide frequency at each position in junctions made by in vitro NHEJ (B) and T4 DNA ligase (C).

Based on the differences in end-joining observed earlier, the initial expectation was for a preference for G/Cs at least at the terminal nucleotide. Surprisingly, no such enrichment of G/Cs was found in the junctions from NHEJ (Distance from the break -1 and +1) *(Figure 46B)*. There is a slight preference for the pyrimidines C & T over the purines A & G one nucleotide away from one of the termini (-2), but that isn't replicated at the corresponding

position on the other side of the break (+2). On the other hand, T4 DNA ligase appears to be less efficient with A:T termini, but the nucleotides at the other positions aren't enriched/depleted much *(Figure 46C)*.

Looking specifically at NHEJ junction sequences that are enriched at a log2 fold change of at least 1 compared to the control, conserved nucleotides at various positions could be identified *(Figure 47A)*. Contrary to initial expectations, 'T's are preferred strongly at almost all positions, but especially at the terminal nucleotides and one nucleotide away from the termini (-2 to +2) *(Figure 47B)*. Relaxing the log2 fold change cut-off to 0.5 still leads to the same conclusions *(Figure 47C)*, but the relatively smaller sample sizes are a caveat.

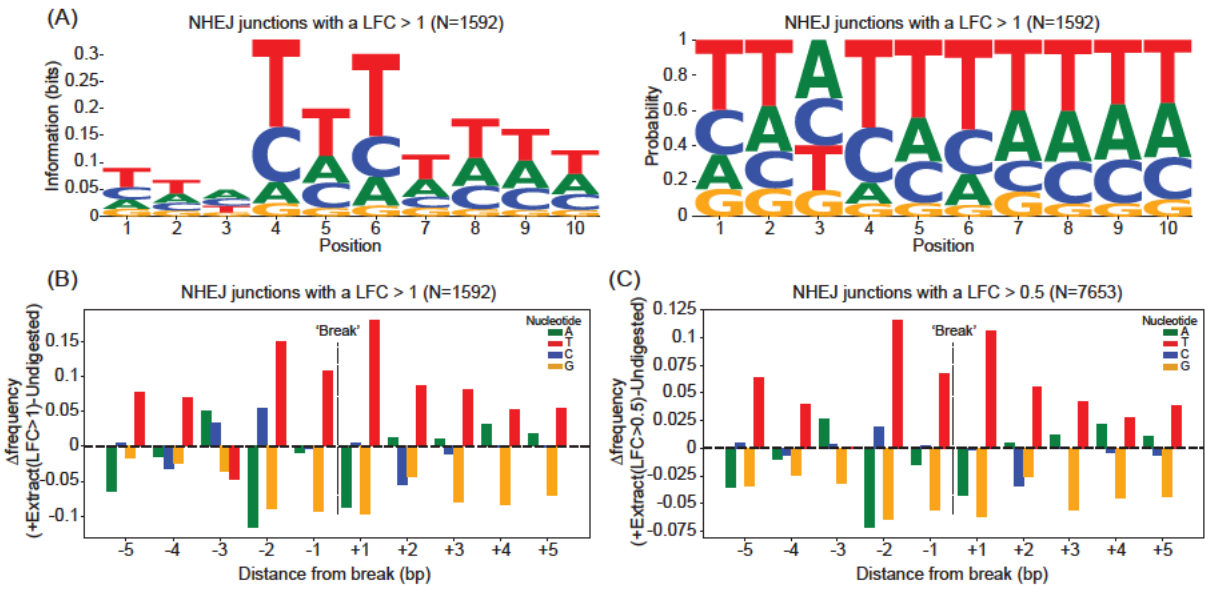


Figure 47 – pJP14 junctions over-represented in the NHEJ junctions. (A) Sequence logo plots of the sequences enriched with a LFC > 1 in NHEJ junctions; Change in nucleotide frequencies among NHEJ junctions enriched at LFC > 1 (B) and LFC > 0.5 (C).

Impact of nucleotides upstream of a blunt end

The libraries pJP17 and pJP18 allow us to specifically look at the influence of nucleotides flanking a defined blunt end (CCC/GGG) *(Figure 48A-B, Figure 49A-B)*. No significant biases for upstream nucleotides in NHEJ could be identified, up to 13 nucleotides away from the break *(Figure 48C, Figure 49C)*. Nevertheless, in both libraries, the nucleotides farthest from the break have the greatest biases.

On the other hand, in the T4 DNA ligase junctions, there was a slight preference for T/A at -4 and +4 bp from the break respectively in both libraries *(Figure 48D, Figure 49D)*. Unlike the NHEJ junctions, the nucleotide frequencies here don't deviate much from the control further upstream from the break. The profiles also tend to be symmetrical either side of the break.

Based on these observations, ligation efficiency by T4 DNA ligase is slightly dependent on the nucleotides closest to the break, but nucleotides > 10 nt away from the termini have almost no impact on it. On the other hand, sequence dependence in NHEJ appears

to be more complex, where nucleotides further away from the break can have a big impact on the efficiency with which an end gets ligated.

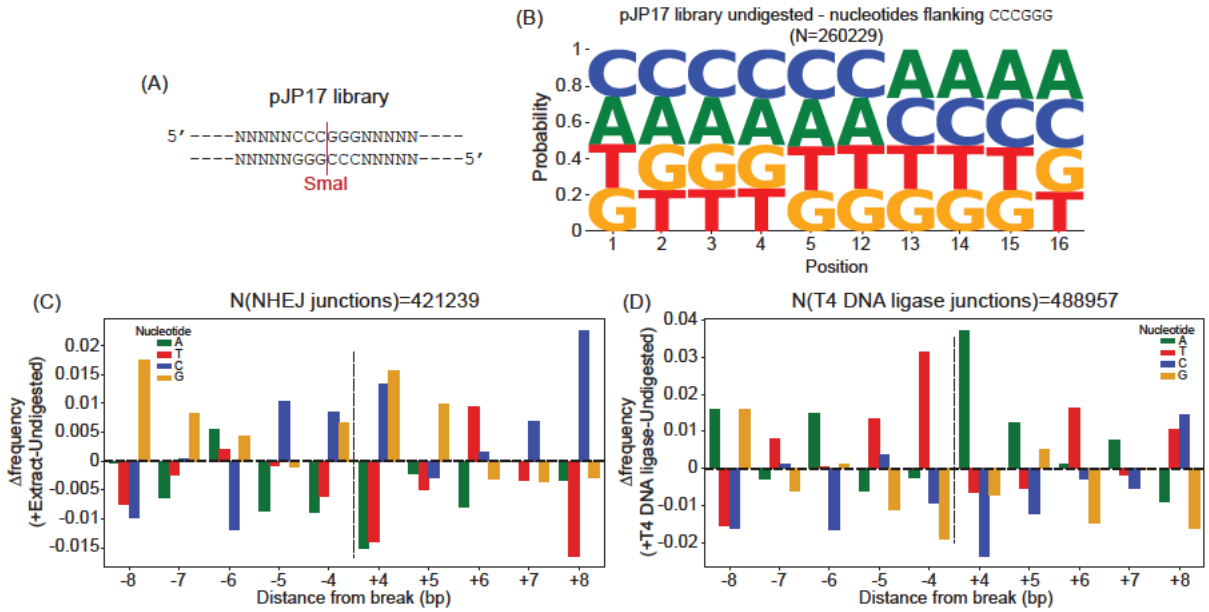


Figure 48 – NHEJ sequence bias upstream from a defined blunt end. (A) Sequence of the pJP17 library; (B) Sequence logo plot of the undigested pJP17 library, only randomized nucleotides shown; change in nucleotide frequency at each position in NHEJ junctions (C) and T4 DNA ligase junctions (D) compared to the undigested control.

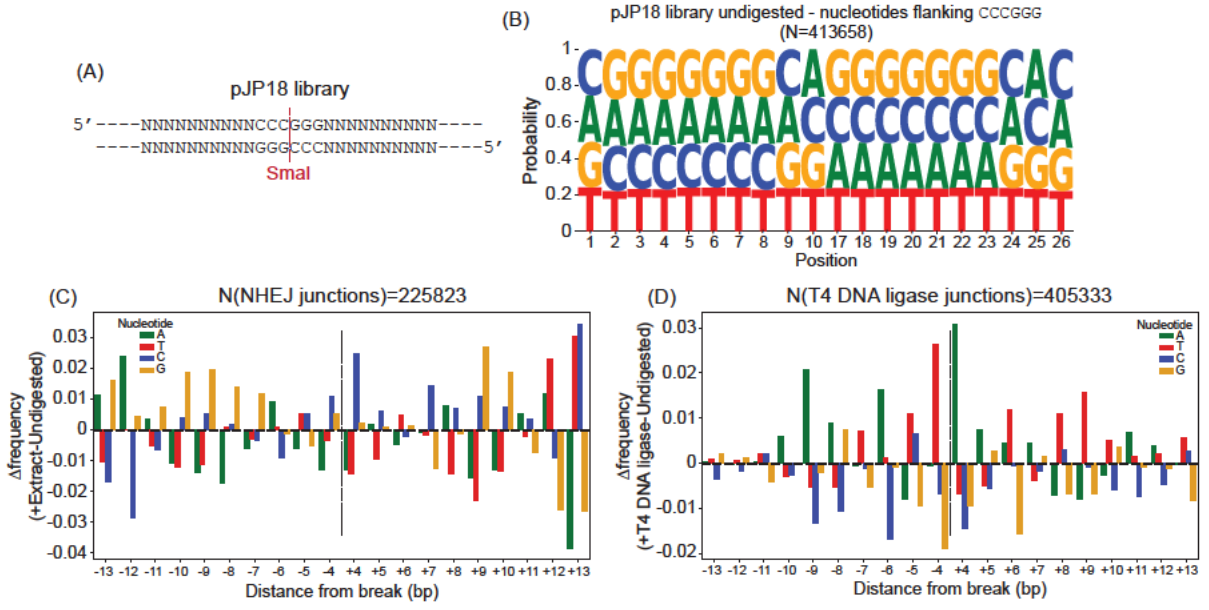


Figure 49 – NHEJ sequence bias upstream from a defined blunt end. (A) Sequence of the pJP18 library; (B) Sequence logo plot of the undigested pJP18 library, only randomized nucleotides shown; change in nucleotide frequency at each position in NHEJ junctions (C) and T4 DNA ligase junctions (D) compared to the undigested control.

Impact of nucleotides upstream of a 5' overhang

The libraries pJP15 and pJP16 were designed to study the impact on NHEJ of nucleotides upstream of a defined 5' overhang *(Figure 50A, Figure 51A)*. Digestion with the enzyme XhoI creates a 5' AGCT overhang. Similar to nucleotides upstream of a blunt end (pJP17

and pJP18), nucleotide biases could be observed at positions far away from the end (>10 bp) *(Figure 50C, Figure 51C)*. But overall, the change in nucleotide frequencies were quite low, especially compared to previous results. For the junctions made by T4 DNA ligase, the change in nucleotide frequencies are under 1% *(Figure 50D, Figure 51D)*, suggesting that T4 DNA ligase is efficient in an upstream sequencebindependent manner while ligating ends with overhangs.

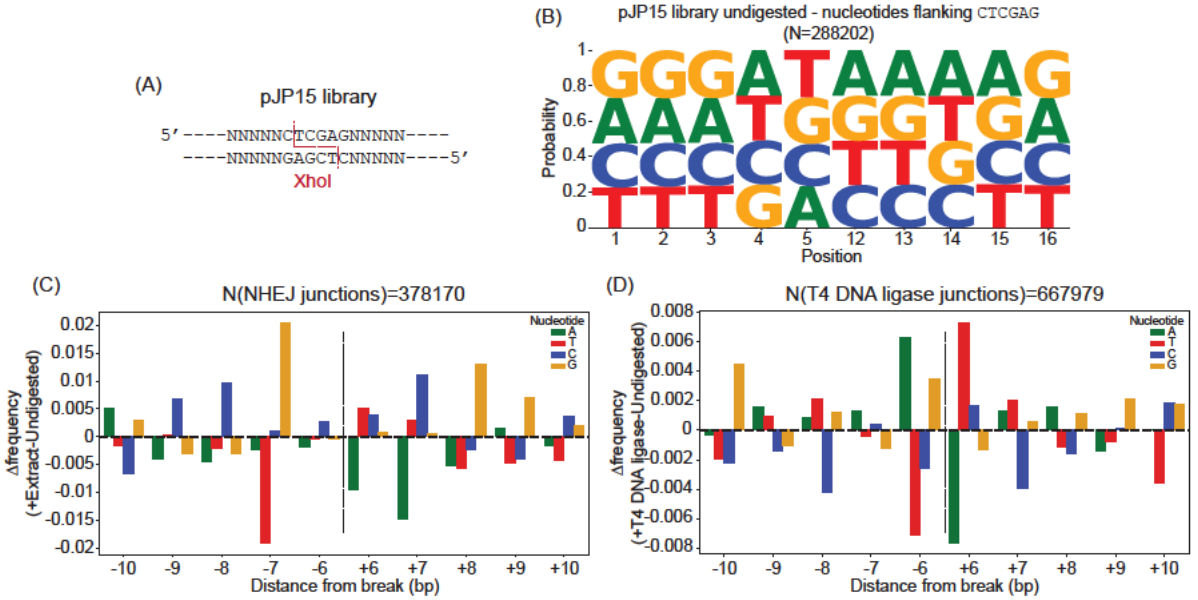


Figure 50 – NHEJ sequence bias upstream from a defined 5' overhang. (A) Sequence of the pJP15 library; (B) Sequence logo plot of the undigested pJP15 library; change in nucleotide frequency at the designated positions in NHEJ junctions (C) and T4 DNA ligase junctions (D) compared to the control.

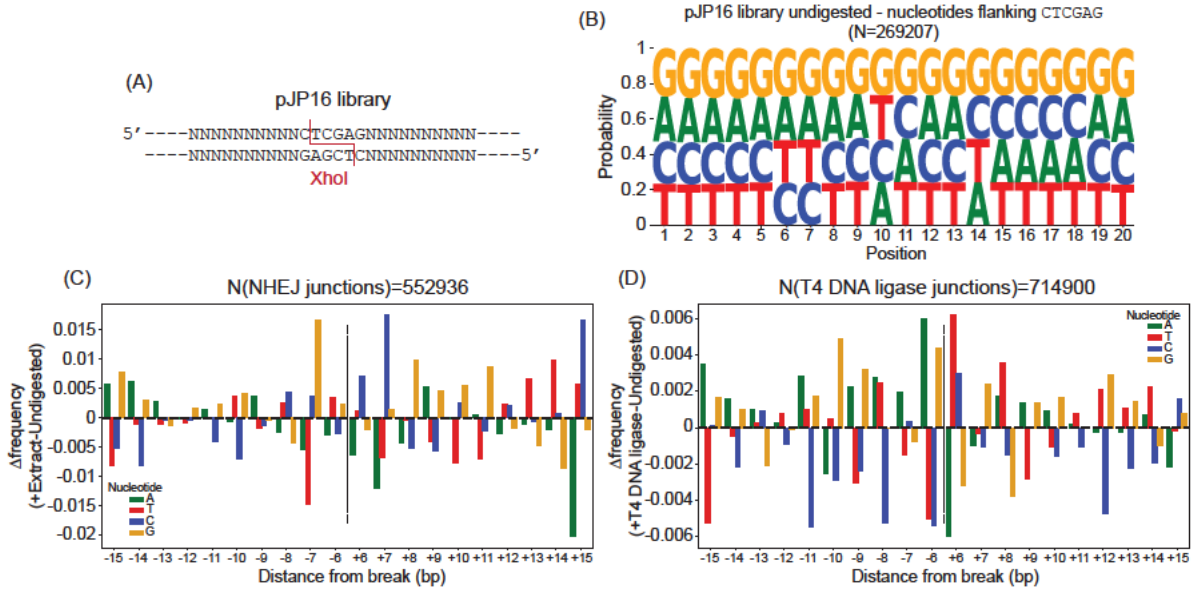


Figure 51 – NHEJ sequence bias upstream from a defined 5' overhang. . (A) Sequence of the pJP16 library; (B) Sequence logo plot of the undigested pJP16 library; change in nucleotide frequency at the designated positions in NHEJ junctions (C) and T4 DNA ligase junctions (D) compared to the control.

Discussion

RNA-dependent 5' end processing during NHEJ

The initial observation made in this study that RNA is potentially involved in NHEJ is an exciting and novel one that builds on some contemporary literature *(Figure 10)*. Neither published model of RNA in DSB repair can fully explain the characteristics of RNA involvement in end-joining in this study. The in vitro nature of the assay takes the emphasis away from the damage signalling pathway and puts the focus on the biochemical and catalytical properties of end-joining. The existence of de novo transcribed RNA in this assay is quite improbable. The pBluescript plasmid, which the substrates for end-joining in this study are based on, contains 3 promoters – lac, T3 and T7. These promoters are bacteria- and bacteriophage- specific promoters that lack the features necessary for eukaryotic RNA polymerase II binding and initiation. Spontaneous promoter-free in vitro transcription from the ends has been described, but only from a 3' overhang. Ends with 5' overhangs or blunt ends initiate very little transcription (Schenborn & Mierendorf, 1985). The ribonucleotides necessary for RNA synthesis are also not expected to be available in the extract *(Figure 15)*.

The RNA-dependent end-processing factor

Addback of RNA extracted from NALM6 cells was unable to reverse the stimulation of end-joining by RNase A *(Figure 14)*. I hypothesize that a specific RNA affects in vitro end-joining in the form of a ribonucleoprotein (RNP) complex. The addback of RNA does not reconstitute the complex possibly because it is a hierarchical process that doesn't happen under the reaction conditions at a very high efficiency. For example, the autonomous in vitro reconstitution of the ribosomal subunits has been a challenge that was only recently solved in *E. coli* (Kosaka et al., 2025). The eukaryotic RNase P holoenzyme was also a problematic RNP for in vitro reconstitution for a long time (Perederina et al., 2018). Hence, it is plausible for a novel RNP that is destabilized by RNase treatment to not reconstitute upon the re-introduction of RNA.

I took a proteomics-based approach to identifying the processing factor. That has provided mixed results – I am able to clearly observe a change in the end-binding proteome to 5' overhangs after RNase treatment of the extracts *(Figure 38)*. Components of multi-protein complexes have been enriched among the hits *(Figure 43, Figure 44)*. Both the pBAF complex and cohesin complex have been described as taking part in the DSB response (Bauerschmidt et al., 2010; Fedkenheuer et al., 2025; Gelot et al., 2016; Kakarougkas et al., 2014, 2015; Kim et al., 2002; Schar, 2004; Thomas-Claudepierre et al., 2013). On the other hand, devising a secondary screen to further filter candidate proteins has been a challenge. The large number of hits was unexpected. In hindsight, fractionating the extract and determining the fraction containing the processing activity would have excluded the majority of non-functional proteins from the list of hits.

Effect of end-structure on processing

Aside from RNA, DNA end-structure is the other major limiting factor for end-processing. Only 5' overhangs were able to get processed, and the 5' ends in 3' overhangs and blunt ends were safe from processing. Similar results have been observed previously for other

mechanisms – in Chappell et al., (2002), the phosphorylation of 5' OH termini by human PNKP was more efficient at protruding termini (5' overhangs) than at recessed termini (3' overhangs). Aside from the relatively reduced accessibility in a 3' overhang or a blunt end, the 5' terminal nucleotides in these overhangs are also paired to a complementary strand. A combination of these factors could explain why blunt ends and 3' overhangs don't allow processing. It also appears that the sequence in the overhang affects the level of processing. Only 2 different sequences were tested in this study, but a significant difference in the amount of processing could be observed.

A dephosphorylated 5' terminus (5' OH) inhibits processing *(Figure 33)*. This suggests that the processing factor could require the presence of a terminal phosphate for its exonucleolytic activity. The bacteriophage λ exonuclease is one such enzyme that exhibits a strong preference for a phosphate moiety at the 5' end (Subramanian et al., 2003). It is not yet clear if this is the case here too, since the presence of a 3' phosphate on the complementary strand restores a small amount of processing.

Looking at the structure of DNA-PK and its interactions with the DNA might provide an alternative explanation for why the processing is so dependent on DNA end structure and end chemistry. The DNA End Blocking helix (DEB) is one of the three regions of DNA-PK that interacts with DNA ends, along with the 4 short DNA-binding patches and the Helix-hairpin-Helix (HhH) motif in the N-terminus. The DEB is a disordered region, which upon interaction with the DNA transitions to a helix with the DNA ends 'split' around it (Liang & Blundell, 2023). The DEB also contains the residues for 'ABCDE' autophosphorylation that is critical for DNA-PK autoactivation. In a study still under peer review, Meek et al. (2023) advance that in cases where DNA ends are not able to split around the DEB helix and stabilize it (like hairpins), the recruitment of XLF during the short-range synaptic complex stage is impaired. This leads to a transient ABCDE autophosphorylation that promotes nucleolytic end-processing. This model of end-structure dependent DNA-PK activation could be an example for the end-structure and end-chemistry dependent nucleolytic processing observed in this study.

Effect of the end-processing on end-joining

The RNA-dependent 5' end processing described here is not only prevalent in a LigIV-deficient extract. When the DNA is incubated in a LigIV-proficient extract, we can observe that the substrates left unjoined in the reaction have been nucleolytically processed *(Figure 25)*. Even in substrates that have been joined, there is an increased prevalence of deletions at later junctions. The RNase treatment of the extract eliminates deletions at the junction *(Figure 35)*. Taken together, these data point toward a competition between end-processing and end-joining as sub-pathways following DNA-PK activation. Processed ends can be ligated, so the pathways perhaps aren't mutually exclusive. Still, DNA ends that have been processed have a lower likelihood of finishing up in a junction because of the incompatibility of the end sequences. It can be convinced that the accumulation of incompatible ends for direct joining is the reason for the reduced end-joining efficiency and its plateauing with 5' overhangs. Other published in vitro end-joining systems have shown comparable efficiencies between the joining of compatible and incompatible ends (Budman & Chu, 2005). This does not appear to be the case in the system used here with NALM6 extracts.

End processing and end synapsis

The use of Ligase IV-deficient extracts enables the decoupling of end processing from ligation and thus helps with the independent characterization of end-processing. Aside from the ligation step, Lig IV performs functions that are independent of its catalytic activity. The presence of Lig IV is essential for DNA-PKcs autophosphorylation after DSB induction in cells, though it does not appear to be the case for in vitro experiments with extracts and linear dsDNA >250 bp (Cottarel et al., 2013). Lig IV also stabilizes the interaction between DNA-PK and other components of the ligation complex, XRCC4 and XLF (Jayaram et al., 2008; Wu et al., 2007). It can thus be expected that the short range synaptic complex formation is impaired with the LigIV-deficient extract and the 5' end processing occurs when the ends are still in long-range synapsis. This is not just an artifact due to the absence of LigIV. Even with the LigIV-proficient extract, the ends of the unjoined monomers appear to be processed *(Figure 25)*, possibly by the same mechanism during long range synapsis. This is in direct contradiction to published literature showing that end processing during NHEJ only occurs after short-range synaptic complex formation (Stinson et al., 2020). The compatibility of the sequences for direct joining is checked during short-range synapsis, so the observed 5' end processing occurs independent of the need for end-processing.

We know that NHEJ repairs breaks generated with compatible ends at almost 100% accuracy in cells (Lin et al., 2013). End-processing during NHEJ is also very tightly regulated to prevent superfluous processing *(Figure 8)*. Based on all the data discussed so far, I hypothesize that the 5' end processing observed in this study is separate from typical end-processing during NHEJ and is independent of end compatibility. It could serve as a mechanism to generate genomic diversity through NHEJ.

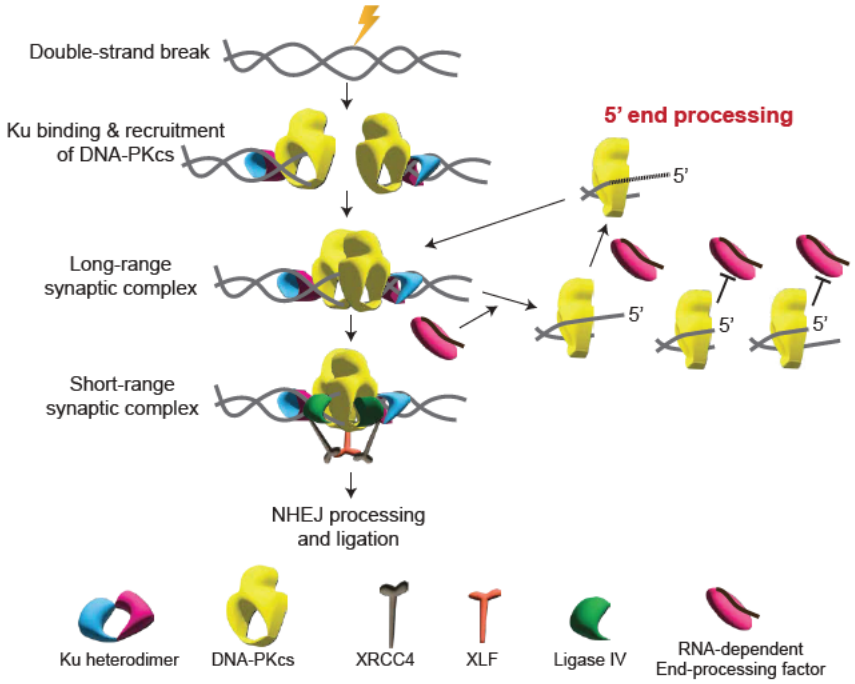


Figure 52 – Model for 5' end processing during NHEJ

NALM6 cells, as precursor B cells isolated from a patient with ALL (Han et al., 1979) can perform V(D)J recombination. The phenomenon observed in this study differs from V(D)J

recombination in a few important ways though. V(D)J recombination is initiated by the recognition and nicking of specific RSS sequences by RAG1/2. The nicking of the DNA and the subsequent generation of hairpins prime the involvement of the NHEJ processing enzymes artemis and TdT. Though RAG1 and RAG2 are expressed by these cells *(Figure 9B)*, their binding sites (RSS) are missing in the pBluescript-based substrate DNA. Besides, during V(D)J recombination, the generation of the hairpin and its processing by Artemis intrinsically gives rise to variation. The RNA-dependent processing described in this study, on the other hand, happens in substrates that don't require any further processing and appears to only happen in a subset of available ends for joining.

NHEJ sequence bias

The efficiency of in vitro end-joining varies among substrates, even of the same end structure *(Figure 12)*. This could be caused by the DNA sequence in the overhang, or even the sequence upstream of the overhang. NHEJ is already regulated in cells by chromatin context (Scully et al., 2019). DSBs in transcriptionally active loci induced by the endonuclease AsiSI were less efficient at recruiting XRCC4, even in G1 (Aymard et al., 2014). To study if the sequence itself has any impact on repair kinetics by NHEJ independent of chromatin, I made substrates with randomized nucleotides at or proximal to a break. In vitro end-joining by the extract doesn't seem to overall prefer any terminal nucleotides in a blunt end *(Figure 46)*. Only when looking at a small subset of sequences that were joined at a very high efficiency was a preference for thymidines observed *(Figure 47)*.

This observation opposes conclusions from an earlier experiment that shows SmaI digested DNA (CCC/GGG) is joined much more efficiently than EcoRV digested DNA (GAT/ATC), thus suggesting a complex relationship between NHEJ repair efficiency and DNA sequence. Indeed, there was a mild selection for certain nucleotides further upstream from the break (> 10 bp), both for blunt ends and those with 5' overhangs. Among the ends that were tested in this study, there is a subtle preference emerging for purines (adenine and guanine) over pyrimidines (cytosine and thymine) on the strand containing the 3' terminus. But this is also not always replicated.

Nucleotides that are not at the terminus playing a role in the efficiency of NHEJ makes perfect mechanistic sense when considering how Ku and DNA-PKcs binds to DNA ends. Ku translocates inwards by about half a helical turn (~5 bp) from the terminus upon the formation of the DNA-PK holoenzyme (Yoo & Dynan, 1999). A recent study solving the structure of an activated DNA-PK complex in high resolution shows that protein residues of the DNA-PK complex are in contact with the DNA/backbone up to 28 bp upstream from the termini *(Figure 53A)* (X. Chen et al., 2021). It is not yet certain if all those interactions are essential for stable DNA-PK binding and activation. The results here suggest that indeed, the identity of the nucleotides in these positions could influence NHEJ repair efficiency, possibly by modulating the stability of DNA-PK interactions with DNA.

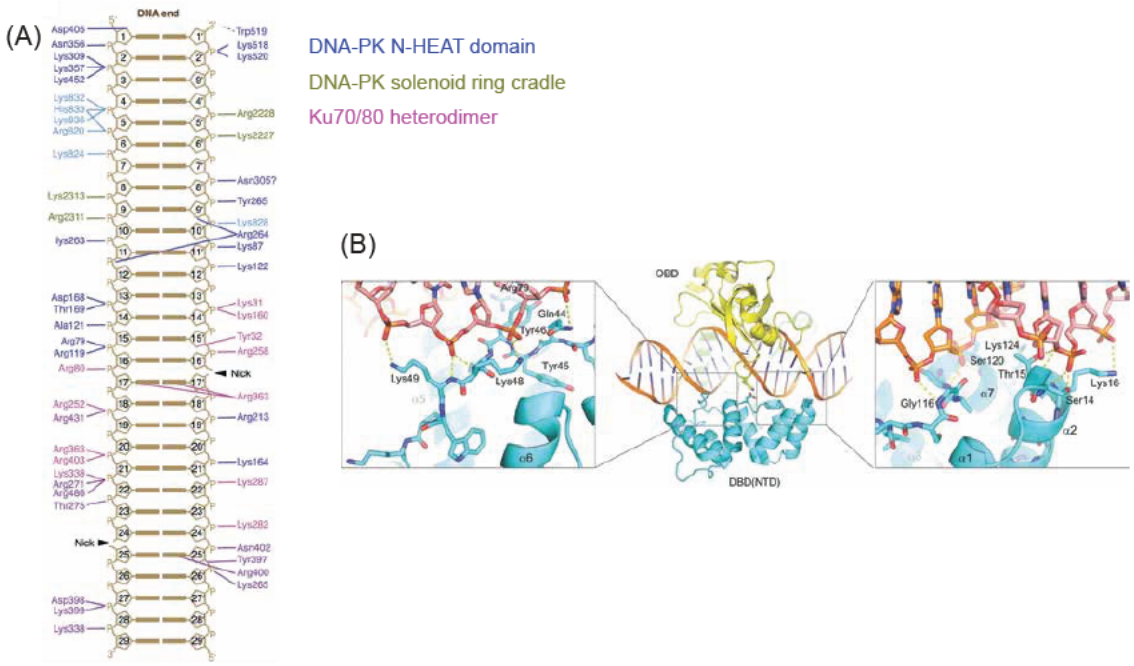


Figure 53 – DNA contact points in (A) activated DNA-PK complex and (B) T4 DNA ligase DNA-binding domain. From X. Chen et al. (2021) and Shi et al. (2018).

The data also provides some insights into the mechanism of ligation by T4 DNA ligase. The DNA-binding domain of T4 DNA ligase only interacts with nucleotides very close to the DNA end/nick *(Figure 53B)* (Shi et al., 2018) unlike the DNA-PK complex. And accordingly, the change in nucleotide frequencies away from the termini was very little, indicating minimal influence for those nucleotides on ligation efficiency. Nevertheless, the data shows that, with blunt ends, T4 DNA ligase does not prefer ligating A:T ends specifically. Interestingly, T:A ligation does not decrease in frequency. Based on this result, one would expect the sequence in an overhang to also affect efficiency of T4 DNA ligase, but it is not simple to analyse and quantify in high-throughput.

Future Experiments

All the end-processing described in this thesis was shown on substrates with 5' overhangs that were 4 nt long. Substrates with overhang lengths starting from 1 nt should be tested in order to fully understand the structural prerequisites for processing. Substrates with random sequences in the overhang can be designed to study the effect of overhang length on processing independent of the sequence.

To further validate the requirement of DNA-PK activation for 5' end processing, the end processing assay can be performed with extracts that have been immunodepleted for DNA-PKcs or with extracts made from cells that are DNA-PKcs deficient.

Several sites within DNA-PK get phosphorylated during NHEJ, leading to varying NHEJ outcomes. The 2 major phosphorylation clusters in DNA-PK, the ABCDE cluster and the PQR cluster can be blocked by site-directed mutagenesis to Alanines. Using these phosphorylation deficient DNA-PK mutants, the exact phosphorylation events on DNA-PK leading to the RNA-dependent 5' end processing can be described.

The LigIV-deficient NALM6 extract needs to be validated by addition of purified Ligase IV or a catalytically dead mutant of Ligase IV (K273A) (Goff et al., 2022).

To further strengthen the hypothesis that RNA-dependent 5' processing happens when the ends are in long range synaptic complex, extracts defective in transitioning to a short-range synaptic complex can be made by the knockout/depletion of XRCC4.

Cells with knockouts of known nucleases (Artemis, Exo1, Wrn and so on) should be generated to test their involvement in the described end-processing.

The extract can be fractionated to purify the 5' end processing activity. This would help in the identification of the processing factor, as the previous pulldown followed by mass spectrometry gave rise to too many candidate proteins to screen. Ammonium sulphate precipitation can be used to fractionate proteins based on solubility, and it has the added advantage of conserving protein structure (Wingfield, 1998). Once the processing factor is identified, cell lines deficient for the protein(s)/RNA can be used to study the impact on NHEJ in cells.

All the experiments in this study were performed on a plasmid-based substrate and in vitro. Even within in vitro end-joining experiments, end-joining efficiencies can vary significantly when comparing ligation of a plasmid-based substrate to irradiated genomic DNA (Cheong et al., 1998). The authors of the study speculate that this could be different sets of activities that play a complementary role in the joining of double strand breaks with NHEJ. Keeping this mind, the results observed here still need to be placed in a cellular context. The use of CRISPR-Cas9 and other inducible endonucleases provide us the ability to create double strand breaks at specific genomic loci whose repair can be followed. This can also be used to further understand and validate the role sequence context plays in NHEJ.

Methods

Cell culture

NALM6 and NALM6 LigIV$^{-/-}$ cells were grown in suspension in RPMI medium (Fischer Scientific, #21875158) supplemented with 10% Fetal Bovine Serum (ATCC, #30-2020). The cells were maintained at 37 ^{0}C with 5% CO_2 and 20% O_2. The cells were split once they reached a density of around 10^6 cells/ml, down to $2*10^5$ cells/ml. To measure cell density, cells were stained with an equal volume of Trypan blue (Fisher Scientific, #11538886) and counted on a hemocytometer under an inverted microscope.

Extract preparation

NALM6 cells were harvested at a density of 10^6 cells/ml from two 3 l spinner flasks (Wheaton, #356887) by centrifugation at 400 g for 10 minutes with brake set to 2 on an Eppendorf 5810R centrifuge. The cell pellet was washed in PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4), followed by a wash in hypotonic lysis buffer (10 mM tris-HCl pH 8, 1 mM EDTA, 5 mM DTT) supplemented with 0.2 M sucrose. Lysis was performed in 2 pellet volumes of hypotonic lysis buffer for 20 minutes at 4 ^{0}C, followed by dounce homogenization (18 strokes with the tight 'A' pestle, Wheaton). PMSF (100 µM) (Merck, #P7626), Aprotinin (10 µg/ml) (Merck, #A6279), Pepstatin A (Merck, #P5318), Chymostatin (Merck, #C7268) and Leupeptin hydrochloride (Merck, #L9783) (1 µg/ml each) were added to the noted final concentrations and incubated on ice for 20 min. 0.5 total volumes of high salt buffer (50 mM tris-HCl pH 7.5, 1 M KCl, 2 mM EDTA, 2 mM DTT) was added and the extract was centrifuged at 37000 rpm for 3 hours at 4 ^{0}C on the Optima XE-100 ultracentrifuge with SW41 rotor buckets. Around 7 ml from the middle part of the supernatant was recovered and dialysed against E buffer (20 mM tris-HCl pH 8, 0.1 M KOAc, 20% glycerol, 0.5 mM EDTA, 1 mM DTT) for 3 hours at 4 ^{0}C. The extract was then clarified by centrifugation at 3900 g for 15 minutes at 4 ^{0}C, snap frozen and stored at -80 ^{0}C.

End-joining assay

End-joining reactions were performed at a volume of 20 µl in 1x End-joining buffer (50 mM Triethanolamine pH 8.5, 60 mM Potassium acetate, 2 mM Magnesium acetate, 1 mM DTT, 0.1 mg/ml BSA and 1 mM ATP) with 4 µl NALM6 extract (around 40-50 µg) and 22 fmol of DNA substrate (40 ng for a ~3 kb long substrate). The extract was first incubated at 37 ^{0}C for 20 minutes before addition of the DNA substrate and ATP. RNase A (Thermo Fischer, #EN0531) was added to a concentration of 1 ng/µl if necessary. End-joining was typically performed for 2 hours at 37 ^{0}C. The reactions were stopped by adding 4 µl of stop buffer (10 mg/ml proteinase K, 100 mM Tris-HCl pH 7.5, 200 mM EDTA and 2.5% SDS) and incubating at 42 ^{0}C for 20 minutes. The end-joining was analyzed by running the samples on 0.6% agarose gels in TBE buffer. The gels were pre-stained with GelRed (Biotium, #41003) and imaged on the BioRad ChemiDoc imaging system.

Extraction of tRNA and rRNA

The extraction protocol was adapted and modified from Avcilar-Kucukgoze et al (2020). NALM6 cells were pelleted by centrifugation at 250 g for 10 minutes. For every 10^7 cells, 1 ml TRIzol reagent (Thermo Fischer, #15596026) was used to resuspend the pellet and

incubated at room temperature for 5 minutes. 0.2 ml Phenol-Chloroform-Isoamyl alcohol (25:24:1, equilibriated with 50 mM NaOAc pH 5.2) was added per ml of TRIzol and mixed vigorously by inverting. After a 3-minute incubation at room temperature, centrifugation was performed at 10000 g for 20 minutes at 4 ^{0}C. 1 µl glycogen (5 mg/ml, Thermo Fischer #AM9510) was added to the aqueous phase along with NaCl to a concentration of 0.2 M. 1 volume of isopropanol was then added and the sample was incubated at -20 ^{0}C for 30 minutes. Nucleic acids were pelleted by centrifugation at 10000 g for 20 minutes at 4 ^{0}C, followed by a wash with cold 70% ethanol.

The pellet was then suspended in 10 ml of cold 1 M NaCl and centrifuged again at 10000 g for 20 minutes at 4 ^{0}C. The supernatant was collected in a fresh tube and the pellet was once again suspended in 5 ml of cold 1 M NaCl and centrifuged. The pellet contains an RNA fraction enriched for rRNA – it was washed with cold 70% ethanol, air-dried and suspended in nuclease-free water. The supernatants from the 2 centrifugations were combined and further processed to purify tRNA.

RNA was precipitated by adding 30 ml cold ethanol to the combined supernatant, incubated at -20 ^{0}C for 30 minutes and centrifuging at 10000 g for 10 minutes at 4 ^{0}C. The pellet was washed with cold 70% ethanol, air-dried and dissolved into 6 ml of 0.3 M NaOAc (pH 5). 3.4 ml isopropanol was added, incubated for 10 minutes at room temperature, then centrifuged at 10000 g for 10 minutes to collect the supernatant. 2.3 ml isopropanol was added to the supernatant and incubated at -20 ^{0}C for 30 minutes. The tRNA was then pelleted by centrifugation at 10000 g for 20 minutes at 4 ^{0}C, washed with cold 70% ethanol, air-dried and suspended in nuclease-free water.

Quantification of end-joining

End-joining could be quantified in experiments where restriction endonuclease-digested pKJ025 was used as substrate. pKJ025 was made by Kristi Jensen by removing one of the two AvaII digestion sites from pBlueScript. The remaining AvaII site is directly across from the multiple cloning site (MCS). To quantify end-joining, the DNA from the reaction was purified with 0.6 volumes of AMPure XP beads (Beckman Coulter, #A63880) and eluted in nuclease-free water. AvaII digestion was performed with 10 units of enzyme in 1x NEB cutsmart buffer for 2 hours at 37 ^{0}C. Unjoined monomers should be digested down to 1.5 kb, whereas the junctions should make a 3 kb band. DNA fragments were separated by electrophoresis in 0.6% agarose gels in TBE buffer at 80 V for 1 hour and imaged on the BioRad ChemiDoc imaging system. The lanes and bands were manually defined on ImageLab and signal intensities were measured. The ratio of signal intensities between the 2 bands was used as a relative quantification of end-joining *(Figure 54)*.

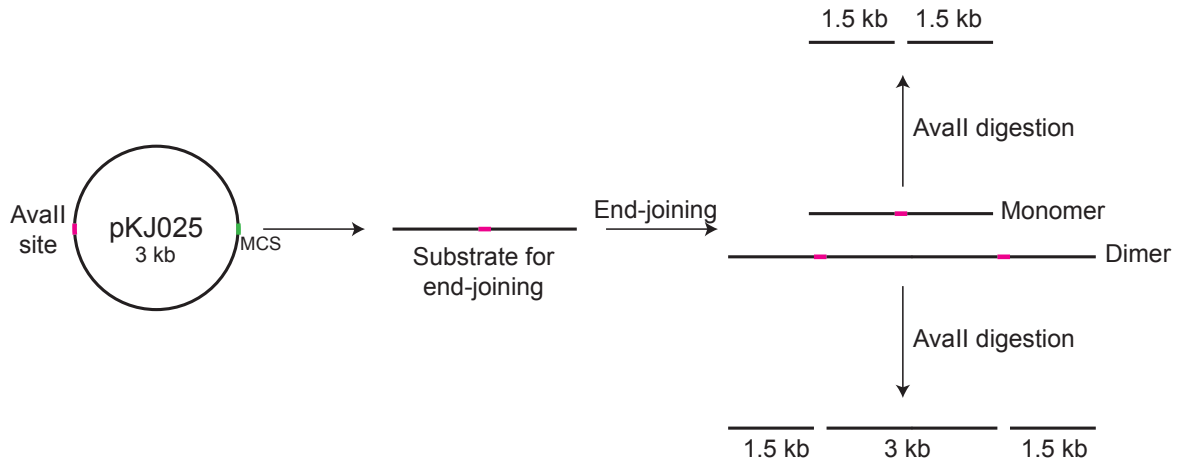


Figure 54 – Quantification of end-joining

Purification of unjoined monomers

DNA substrate molecules that haven't been ligated during an end-joining reaction are designated here as 'unjoined monomers'. Typically, DNA was incubated in extracts made from LigIV-deficient NALM6 cells in an end-joining reaction (see End-joining assay). The DNA was then purified with 0.6 volumes of AMPure XP beads (Beckman Coulter, #A63880). The DNA-bead mixture was incubated for 10 minutes with light vortexing every 2 minutes. The beads were pelleted on a magnetic rack, the supernatant was discarded and the beads were washed twice with 80% ethanol. The beads were then spun down on a mini-centrifuge and returned to the magnetic rack to remove any excess ethanol. The beads were suspended in nuclease-free water and incubated at 37 ^{0}C in a ThermoMixer (Eppendorf, #5382000015) with shaking at 1500-2000 rpm. The beads were pelleted again on a magnetic rack and the eluate containing the DNA was aspirated.

In cases where unjoined monomers were purified after incubation in LigIV-proficient extracts, the end-joining reaction was run on a 0.6% agarose gel. The band corresponding to the monomer was excised and the DNA was purified using the QIAquick Gel Extraction kit (Qiagen, #28704).

Preparation of ONT sequencing libraries

ONT sequencing libraries were prepared by following ONT's protocol for the Native barcoding kit 24 V14 (ONT, SQK-NBD114.24). At least 10 fmol of each DNA sample was made up to 24 µl with nuclease-free water. 1.75 µl each of the NEBNext FFPE DNA repair buffer and the Ultra II End-prep reaction buffer, 1 µl of the NEBNext FFPE DNA repair mix (NEB, #M6630) and 1.5 µl of Ultra II End-prep enzyme mix (NEB, #E7546) were added and incubated at 20 ^{0}C for 5 minutes, followed by 65 ^{0}C for 5 minutes. The DNA was purified with 0.6 volumes of AMPure XP beads each and eluted in 8.5 µl of nuclease-free water. 1.5 µl of a unique native barcode (NB01-NB24) was ligated to each sample with a 1 hour incubation at room temperature with 10 µl of Blunt/TA ligase master mix (NEB, #M0367). The ligation reaction was stopped by adding 2 µl of 0.5 M EDTA. The samples from the various barcodes were combined in a single tube and purified with 0.6 volumes of AMPure XP beads. The elution was performed with 32.5 µl nuclease-free water. 2.5 µl of native adapter was ligated to the pooled, barcoded DNA with 5 µl of Quick T4 DNA ligase (NEB, #E6056) at room temperature for 30 min. The library was purified finally with 0.6

volumes of AMPureXP beads, washed twice in 125 µl Short Fragment Buffer, and eluted in 10 µl of elution buffer. The concentration of the library was measured with Qubit dsDNA High Sensitivity kit (Thermo Fischer, #Q33230) and the average length of the library was verified by running it on TapeStation.

Sequencing analysis of Quick ligase junctions from unjoined monomers

ONT sequencing libraries were loaded on MinION flow cells following ONT protocols. The sequencing run was typically set for 72 h. After an initial basecall with Dorado 0.7.1, the read IDs belonging to the different barcodes were retrieved. The pod5 files were then subset for the different barcodes, and duplex basecalling was performed with the Dorado Super-high accuracy model. The sequences obtained from duplex reads were then filtered for a mean quality score of at least 20 and aligned to a reference junction with minimap2 using the map-ont feature for long reads. The alignment file was filtered and sorted for primary alignments and the junction sequences were retrieved using samtools ampliconclip. The junction sequences from the different samples were combined in a table and filtered for a minimum total occurrence of 10 across all samples. The junctions were then manually annotated and their percentages in different conditions were plotted using R ggplot.

Design of radio-labelled substrates to study 5'-3' processing

An end-labelling reaction was set up with ~13 pmol of HindIII-digested and dephosphorylated pKJ025, 25 pmol of γ-P^{32}-ATP, 100 pmol of 'cold' ATP and 20 units of T4 Polynucleotide Kinase (NEB, #M0201) in 1x T4 PNK buffer at a total reaction volume of 50 µl. The reaction was incubated at 37 ^{0}C for 1 h, followed by heat inactivation of the enzyme at 65 ^{0}C for 20 minutes. The DNA was then purified with 0.6 volumes of AMPure XP beads and eluted in nuclease-free water. The DNA concentration was inferred from the scintillation count based on the following formula:

$$pmol/\mu l\ of\ DNA = \frac{1.9 * 10^{-7} * < scintillation\ count\ in\ cpm/\mu l >}{< P^{32}\ decay\ factor >}$$

Duplexes that can be ligated to the end-labelled DNA were created by annealing oligos *(Table 1)*. A 10-fold molar excess of duplexes were used in a ligation reaction with 400 units of T4 DNA ligase (NEB, #M0202). The ligation was performed at room temperature for 15 minutes, followed by heat inactivation of the enzyme at 65 ^{0}C for 20 minutes. Aside from ligating duplexes onto the HindIII-digested DNA, the reaction would also create dimers from the HindIII-digested DNA ligating to each other *(Figure 26A)*. This artefact can be cleaned up by gel extraction of the desired DNA fragments, but it was not possible in this instance. Instead, these plasmid dimers would serve as an internal loading control in the experiments. The DNA was purified with 0.6 volumes of AMPure beads and eluted in nuclease-free water. The concentration was calculated based on the scintillation count as above.

To get substrates with 5' phosphorylated ends, the DNA was subjected to a futher PNK treatment with 10 units of T4 Polynucleotide kinase and 10 nmol ATP at 37 ^{0}C for 1 h. The DNA was purified and the concentration was measured as previously.

Table 1 – Sequence of duplexes ligated to P^{32}-labelled HindIII-digested DNA to study 5' end processing

Duplex sequence	Substrate
5'P–**AGCT**GTTGACATATGAGCATAGG CAACTGTATACTCGTATCC**TTAA**–5'OH	5' AATT overhang
5'P–**AGCT**GTTGACATATGAGCATAGG CAACTGTATACTCGTATCC**GATC**–5'OH	5' CTAG overhang
5'P–**AGCT**GTTGACATATGAGCATAGG**GTAC** CAACTGTATACTCGTATCC–5'OH	3' GTAC overhang
5'P–**AGCT**GTTGACATATGAGCATAGG CAACTGTATACTCGTATCC–5'OH	Blunt end
5'P–**AGCT**GTTGACATATGAGCATAGG CAACTGTATACTCGTATC**C****TTAA**–5'OH	Nuclease-resistant 5' AATT overhang; underlined bases are joined by phosphorothioate bonds
5'P–**AGCT**GTTGACATATGAGCATAGG CAACTGTATACTCGTATC**C****GATC**–5'OH	Nuclease-resistant 5' CTAG overhang; underlined bases are joined by phosphorothioate bonds
5'P–**AGCT**AAGCATAGGCCGCCTC TTCGTATCCGGCGGAG**CTAG**–5'OH	5' GATC overhang
5'P–**AGCT**AAGCATAGGCCGCCTC**GATC** TTCGTATCCGGCGGAG–5'OH	3' GATC overhang
5'P–**AGCT**AAGCATAGGCCGCCTC–3'P TTCGTATCCGGCGGAG**CTAG**–5'OH	5' GATC overhang, substrate has a 5'OH and 3'P
5'P–**AGCT**AAGCATAGGCCGCCTC**GATC**–3'P TTCGTATCCGGCGGAG–5'OH	3' GATC overhang, substrate has a 5'OH and 3'P
5'P–**AGCT**AAGCATAGGCCGCCTC–3'P TTCGTATCCGGCGGAG**CTAG**–5'P	5' GATC overhang, substrate has a 5'P and 3'P
5'P–**AGCT**AAGCATAGGCCGCCTC**GATC**–3'P TTCGTATCCGGCGGAG–5'P	3' GATC overhang, substrate has a 5'P and 3'P

Design of radio-labelled substrates to study 3'-5' processing

To study 3'-5' end processing, radiolabelled substrates were made similarly as above, but this time the duplex was end-labelled with γ-P^{32}-ATP. 50 pmol of the oligo that would contain the 3' terminus after ligation was labelled with 20 units of T4 Polynucleotide kinase in a reaction with 25 pmol of γ-P^{32}-ATP and 225 pmol of 'cold' ATP. The reaction was incubated at 37 ^{0}C for 1 h, followed by heat inactivation of the enzyme at 65 ^{0}C for 20 minutes. The labelled oligos were precipitated overnight at 4 ^{0}C with 1 volume of 0.6 µM NaOAc, 1 µl glycogen (Invitrogen, #AM9510) and 4 volumes of ice-cold ethanol. The samples were then centrifuged at full speed for 1 hour at 4 ^{0}C, followed by washing the pellet with 70% ethanol. The air-dried pellets were dissolved in nuclease-free water and the concentrations were inferred through scintillation count:

$$pmol/\mu l\ of\ oligo = \frac{7.6 * 10^{-7} * < scintillation\ count\ in\ cpm/\mu l >}{< P^{32}\ decay\ factor >}$$

10 pmol of labelled oligo was annealed into a duplex with an equimolar amount of its complementary oligo and the duplex was used directly in a ligation reaction with 1 pmol of HindIII-digested pKJ025. The ligation was performed with 400 units of T4 DNA ligase at room temperature for 15 min. The ligase was heat inactivated at 65 ^{0}C for 20 minutes, and the DNA was purified with 0.6 volumes of AmpureXP beads. The scintillation counts were again measured with 1 µl of purified DNA and the concentration was derived with the formula –

$$pmol/\mu l\ of\ DNA = \frac{3.8 * 10^{-7} * < scintillation\ count\ in\ cpm/\mu l >}{< P^{32}\ decay\ factor >}$$

The 5' ends of the substrates were phosphorylated as described previously with T4 Polynucleotide kinase.

End-processing assay

The assays to study 5' and 3' end processing were performed similar to the end-joining assay, but with the radiolabelled substrates and with LigIV-deficient NALM6 extracts, at a total volume of 10 µl. After incubation in the desired conditions and amount of time, the reactions were stopped with 2 µl stop buffer (10 mg/ml proteinase K, 100 mM Tris-HCl pH 7.5, 200 mM EDTA and 2.5% SDS) and incubated at 42 ^{0}C for 20 minutes. The DNA was purified using 0.6 volumes of AMPure beads and eluted into 16 µl of nuclease-free water. To be able to resolve the processing intermediates with smaller fragment sizes, the DNA was digested with ClaI (NEB, #R0197) and SpeI-HF (NEB, #R3133) in CutSmart buffer for 2 hours at 37 ^{0}C. The fragments were precipitated overnight at -20 ^{0}C in 1 volume of 5 M NH_4OAc, 0.01 volumes of 1 M $MgCl_2$, 1 µl GlycoBlue (Fisher Scientific, #10301575) and 2.5 volumes of ice-cold ethanol. 5000 cpm of P^{32}-labelled BLoli8483 was also added to act as a precipitation control. The DNA were pelleted by centrifugation at full speed for 15 min, followed by a wash in 70% ethanol. The DNA was dried by placing the tube briefly on a heated block and resuspended in 2x RNA loading dye (NEB, #B0363). Before loading, the samples were heated to 95 ^{0}C for 5 min and snap-cooled in ice-slurry.

A 12% polyacrylamide solution in TBE with 8 M urea was prepared by dissolving 240 g urea in 150 ml of 40% Acrylamide:Bis-acrylamide 19:1 (Fisher Scientific, #10214963) and 50 ml of 10x TBE buffer. The volume was made up to 500 ml with water and the solution was filtered and stored at 4 ^{0}C until needed. 625 µl of 10% APS and 38 µl TEMED were added per 100 ml polyacrylamide solution on the day of making the gel. The gel was poured with a syringe between 2 large glass plates held together with clips. The gel comb was placed, and the gel was allowed to polymerize. Before running samples on it, the gel was warmed up by running in 1x TBE at 80 W for 1 hour. Samples were then loaded onto the gel and run at 80 W for 1 hour and 45 minutes. The gel was removed carefully with a GelBond PAG film (VWR, #LONZ54746) and placed on top of a Whatman paper. This was placed in the gel dryer at 65 ^{0}C for 1 h with the vacuum turned on. The dried gel was then exposed to a phosphor film overnight and the film was imaged using the Typhoon scanner on an IP stage.

Annealing oligos into duplexes

Equimolar amounts of oligos were mixed together in 1x annealing buffer (20 mM tris-acetate (pH 7.5), 50 mM NaCl and 2 mM $MgCl_2$). The mixture was heated up to 99 ^{0}C and cooled down to 20 ^{0}C at the rate of 1 ^{0}C/min on a ThermoMixer.

Preparation of biotinylated DNA substrates for pulldown

2 kb long biotinylated DNA substrates were prepared by PCR. pUC19 plasmid DNA was amplified using a 5'-photocleavable biotin-tagged primer (BLoli8629) and a reverse primer incorporating a restriction enzyme recognition sequence (BLoli8325, BLoli8767-8772). PCRs were set up with 1 ng of pUC19 DNA, 0.5 µM primers, 0.3 µM of each dNTP (NEB, #N0447) and 1 unit of Q5 High-Fidelity DNA polymerase (NEB, #M0491). The reactions were carried out with the following program on a thermocycler – initial denaturation of 98 ^{0}C for 2 minutes, 35 cycles of 98 ^{0}C for 15 seconds, 65 ^{0}C for 15 seconds, 72 ^{0}C for 1 minute and a final extension of 10 minutes. The DNA was purified by gel extraction and restriction digestions with the corresponding enzymes were performed overnight in the recommended buffers. The digested, 5' biotin tagged DNA was then purified using 0.6 volumes of AMPure XP beads and eluted in nuclease-free water.

Pulldown of proteins bound to DNA ends

730 fmol (450 ng) of the biotinylated DNA were bound to 40 µl of Dynabeads MyOne Streptavidin C1 beads (Thermo Fischer, #65001) and resuspended in water. The DNA-bead slurry was incubated with 2 µl of NALM6 extract in 1x End-joining buffer (50 mM Triethanolamine pH 8.5, 60 mM Potassium acetate, 2 mM Magnesium acetate, 1 mM DTT, 0.1 mg/ml BSA and 1 mM ATP) at a total volume of 100 µl. Incubation was performed overnight at 4 ^{0}C on a Hula mixer. The beads were pelleted by placing the tubes on a magnetic rack and washed thrice with the wash/elution buffer (50 mM Triethanolamine pH 8.5, 60 mM Potassium acetate, 2 mM Magnesium acetate, 1 mM DTT, 10% glycerol and 150 mM Sodium chloride). The beads were then suspended in 100 µl of the wash/elution buffer and exposed to 4500 mJ/cm^2 of 320 nm UV light on a transilluminator. This releases the DNA and bound proteins from the beads which can then be collected by pelleting the beads on a magnetic rack. The eluate was incubated with 1 unit of benzonase (Merck, #E1014) at 4 ^{0}C for 2 hours before sample submission for mass spectrometry in quadruplicates.

Analysis of Mass Spectrometry data

MS data analysis was performed on MaxQuant version 2.1.3.0. The peptides were mapped against the UniProt *Homo sapiens* protein database as reference. The intensities were normalized according to the classic LFQ normalization method on MaxQuant. GO term enrichment analysis was performed using the R library clusterProfiler (G. Yu et al., 2012).

Generation of random DNA end sequence libraries

The plasmid pKJ026 generated by Kristi Jensen was used as the vector backbone. BLoli8701, BLoli9286, BLoli9287, BLoli9288 and BLoli9289 were annealed to BLoli8702 and BLoli8703 to form inserts for cloning. A 10-fold molar excess of the annealed duplexes were ligated to 50 ng of digested pKJ026 *(Figure 55)*. For libraries pJP14-16,

pKJ026 was digested with the enzymes EcoRI-HF and KpnI-HF. For libraries pJP17-18, pKJ026 was digested with NotI-HF and KpnI-HF. The ligation was performed with 400 units of T4 DNA ligase (NEB, #M0202) overnight at 16 ^{0}C. The ligase was inactivated by incubating at 65 ^{0}C for 20 minutes. The protocol for library transformation was adapted from Mateyko & de Boer, 2024. 2 µl of the ligation reaction was transformed into OneShot TOP10 Electrocomp *E. coli* (Thermo Fischer, #C404052) by electroporation. The electric pulse was performed on the Bio-Rad MicroPulser Electroporator using a cuvette with a 0.1 cm gap according to the setting "EC1". Immediately after the electric pulse, 1 ml LB medium was added to the cells, and they were grown out at 37 ^{0}C for 1 h with shaking. The complexity of the library was verified by plating 1 µl of the library on LB plates with 100 µg/ml carbenicillin. The desired complexity of the libraries is at least 1 million (4^10 ~ 1.05*10^6).

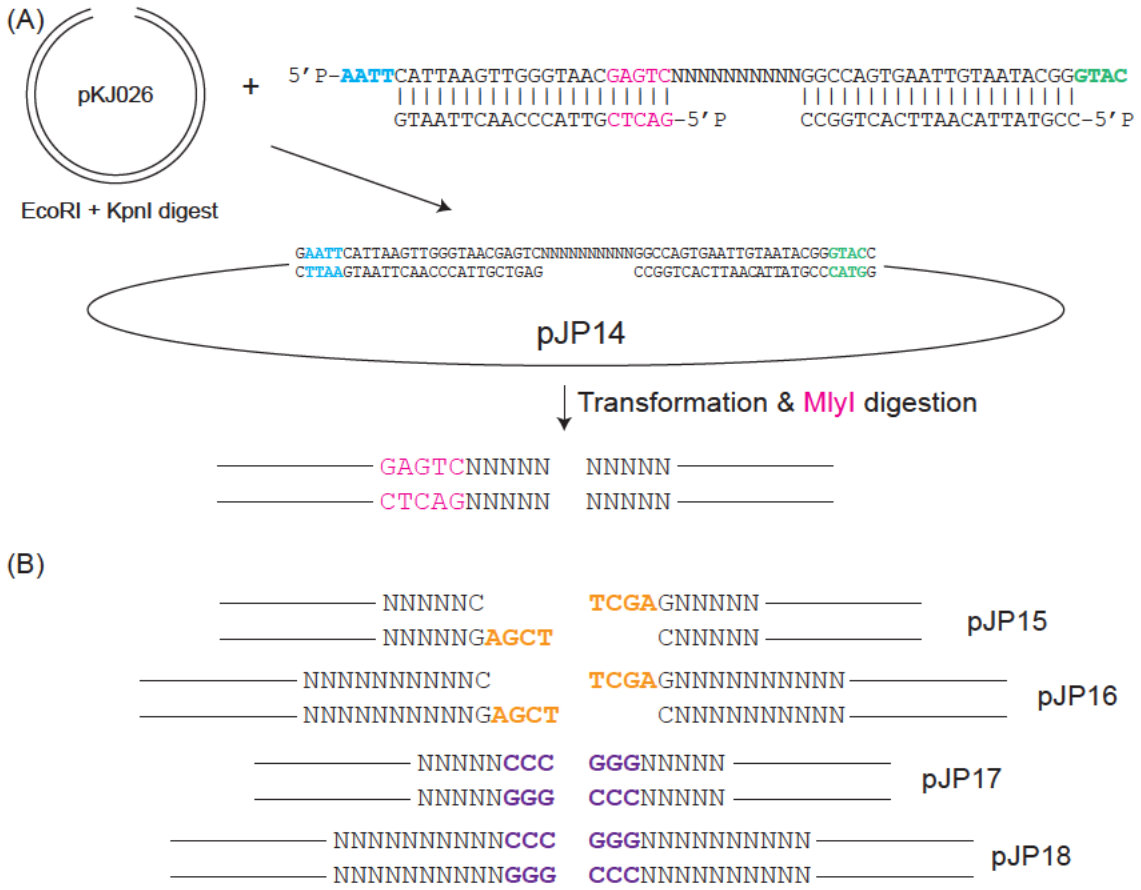


Figure 55 – Generation of random DNA end libraries. (A) Schematic for pJP14 (B) Ends from libraries pJP15-18.

After verifying the complexity, the transformed cells were grown at 37 ^{0}C with shaking for 3 hours. 1/10 of the culture was diluted and grown overnight for plasmid extraction in the presence of 100 µg/ml ampicillin. Plasmid isolation was performed with the PureLink HiPure Plasmid Maxiprep kit (Invitrogen, #K210007). DNA substrates for end-joining were then prepared by restriction digestion with MlyI (NEB, #R0610) (pJP14), XhoI (NEB, #R0146) (pJP15 & pJP16) or SmaI (NEB, #R0141) (pJP17 & pJP18).

End-joining reactions were set up as previously (End-joining assay). The substrates were also incubated in 1 µl of Quick ligase (NEB, #M2200) in Quick Ligase reaction buffer for

10 minutes. The products were purified with 0.6 volumes of AmpureXP beads and digested with AvaII (NEB, #R0153) in rCutsmart buffer for 2 h at 37 ^{0}C. The 2.5 kb fragments containing the junctions were purified by gel extraction using the QIAquick gel extraction kit (Qiagen, #28704) with MinElute columns (Qiagen, #28004).

The illumina sequencing libraries were prepared with a nested PCR reaction. 0.05 ng of the purified junctions or undigested library were first tagged with 6 nucleotide UMIs through 2 cycles of amplification with Q5 High-Fidelity DNA polymerase (NEB, #M0491). An equimolar mix of primers that can contain a further 0/2/3/5/8 Ns were used as the forward primer (BLoli9449-9453 for pJP14-16; BLoli9418-9422 for pJP17-18). The amplicons were then purified with 1 volume of AmpureXP beads and used as template for amplification in a second PCR using XLeap UDI primers with a specific barcode for each sample. Amplification was performed with Q5 High-Fidelity DNA polymerase - initial denaturation of 98 ^{0}C for 2 minutes, 20 cycles of 98 ^{0}C for 15 seconds, 72 ^{0}C for 30 seconds and a final extension of 5 minutes. 5 ng of each library was pooled together and sequenced on a NextSeq P2 flow cell with 64 bp R1 and 58 bp R2 paired-end reads.

Sequencing analysis of random nucleotide libraries

The reads from each sample were first binned into UMI pairs based on the first 6 nucleotides in read 1 and read 2. For each UMI pair, a consensus sequence was called – the most abundant sequence in a UMI pair should make up at least 90% of the sequences in it. The resulting sequences were searched for the string `ATTCATTAAGTTGGGTAACGAGTC` (libraries pJP14-16) or `GCCCATTAAGTTGGGTAACGAGTC` (libraries pJP17 & 18) to identify the start of the randomized nucleotides. The sequences were then extracted and the frequencies of each nucleotide at every position was calculated. The entire analysis was performed using the programming language python.

Oligonucleotides used in the study

Oligo name	Sequence	Comments
Preparation of biotinylated DNA substrates for protein pulldown		
BLoli8325	ATGC**GAATTC**CTCGCTCACTGACTCGCTG	Amplification of pUC19; EcoRI recognition site in overhang highlighted
BLoli8629	/5PCBIO/GCCTCGTGATACGCCTATT	Amplification of pUC19, photocleavable biotin conjugated to the 5' end
BLoli8767	GCAT**GGATCC**CTCGCTCACTGACTCGCTG	Amplification of pUC19; BamHI recognition site in overhang highlighted
BLoli8768	GTATCTGA**GCGGCCGC**CTCGCTCACTGACTCGCTG	Amplification of pUC19; NotI recognition site in overhang highlighted
BLoli8769	GCATCT**GATATC**CTCGCTCACTGACTCGCTG	Amplification of pUC19; EcoRV recognition site in overhang highlighted

BLoli8770	GTAT**CCCGGG**CTCGCTCACTGACTCGCTG	Amplification of pUC19; SmaI recognition site in overhang highlighted
BLoli8771	GCAT**GGTACC**CTCGCTCACTGACTCGCTG	Amplification of pUC19; KpnI recognition site in overhang highlighted
BLoli8772	GCAT**CTGCAG**CTCGCTCACTGACTCGCTG	Amplification of pUC19; PstI recognition site in overhang highlighted
End-processing experiments		
BLoli8483	TGGAAATAAGCGTGCATACTATGCAGTAG CGCGTGGTCGAAATACTGGTATATATAGT ACATGGGATGAGGCGTCGGATCAG	Precipitation control (82 nt)
BLoli9105	/5Phos/**AGCT**AAGCATAGGCCGCCTC	AGCT overhang for ligation to HindIII-digested pKJ025
BLoli9107	/5Phos/**AGCT**AAGCATAGGCCGCCTCGA TC	AGCT overhang for ligation to HindIII-digested pKJ025; 3' GATC overhang
BLoli9109	/5Phos/**AGCT**AAGCATAGGCCGCCTC/3 Phos/	Same as BLoli9105, with a 3' phosphate
BLoli9111	/5Phos/**AGCT**AAGCATAGGCCGCCTCGA TC/3Phos/	Same as BLoli9107, with a 3' phosphate
BLoli9117	GAGGCGGCCTATGCTT	Annealed with BLoli9107 to form a 3' GATC overhang
BLoli9118	GATCGAGGCGGCCTATGCTT	Annealed with BLoli9105 to form a 5' GATC overhang
BLoli9205	/5Phos/**AGCT**GTTGACATATGAGCATAG G	AGCT overhang for ligation to HindIII-digested pKJ025
Bloli9282	**AATT**CCTATGCTCATATGTCAAC	Annealed with BLoli9205 to create a duplex with 5' AATT overhang
Bloli9283	**A*A*T*T***C*CTATGCTCATATGTCAAC	Annealed with BLoli9205 to create a duplex with 5' AATT overhang (phosphorothioate)
Bloli9343	**GATC**CCTATGCTCATATGTCAAC	Annealed with BLoli9205 to create a duplex with 5' GATC overhang
Bloli9344	**G*A*T*C***C*CTATGCTCATATGTCAAC	Annealed with BLoli9205 to create a duplex with 5' GATC overhang (phosphorothioate)
Bloli9345	/5Phos/**AGCT**GTTGACATATGAGCATAG G**GTAC**	AGCT overhang for ligation to HindIII-digested pKJ025; 3' GTAC overhang
Bloli9347	CCTATGCTCATATGTCAAC	Annealed with BLoli9345 to create 3' GATC overhang

Bloli9348	AATTCCTATGCTCATATGTCAACAGCTTAT	Ladder for 5' processing sequencing gel (30 nt)
Bloli9349	TTCCTATGCTCATATGTCAACAGCTTAT	Ladder for 5' processing sequencing gel (28 nt)
Bloli9350	CCTATGCTCATATGTCAACAGCTTAT	Ladder for 5' processing sequencing gel (26 nt)
Bloli9351	TATGCTCATATGTCAACAGCTTAT	Ladder for 5' processing sequencing gel (24 nt)
Bloli9352	AATTCCTATGCTCATATGTCAACAGCTTGATATCGAATTCCTGCAGCCCGGGGGATCCA	Ladder for 5' processing sequencing gel (59 nt)
Bloli9353	TTCCTATGCTCATATGTCAACAGCTTGATATCGAATTCCTGCAGCCCGGGGGATCCA	Ladder for 5' processing sequencing gel (57 nt)
Bloli9354	CCTATGCTCATATGTCAACAGCTTGATATCGAATTCCTGCAGCCCGGGGGATCCA	Ladder for 5' processing sequencing gel (55 nt)
Bloli9355	TATGCTCATATGTCAACAGCTTGATATCGAATTCCTGCAGCCCGGGGGATCCA	Ladder for 5' processing sequencing gel (53 nt)
Bloli9356	CTCATATGTCAACAGCTTGATATCGAATTCCTGCAGCCCGGGGGATCCA	Ladder for 5' processing sequencing gel (49 nt)
Bloli9357	CTCATATGTCAACAGCTTAT	Ladder for 5' processing sequencing gel (20 nt)
Bloli9408	CGATAAGCTGTTGACATATGAGCATAGG	Ladder for 3' processing sequencing gel (28 nt)
Bloli9409	CGATAAGCTGTTGACATATGAGCATA	Ladder for 3' processing sequencing gel (26 nt)
Bloli9410	CGATAAGCTGTTGACATATGAGCA	Ladder for 3' processing sequencing gel (24 nt)
Bloli9411	CGATAAGCTGTTGACATATGAG	Ladder for 3' processing sequencing gel (22 nt)
Bloli9412	CGATAAGCTGTTGACATA	Ladder for 3' processing sequencing gel (18 nt)
Bloli9413	CTAGTGGATCCCCCGGGCTGCAGGAATTCGATATCAAGCTGTTGACATATGAGCATAGG	Ladder for 3' processing sequencing gel (59 nt)
Bloli9414	CTAGTGGATCCCCCGGGCTGCAGGAATTCGATATCAAGCTGTTGACATATGAGCATA	Ladder for 3' processing sequencing gel (57 nt)
Bloli9415	CTAGTGGATCCCCCGGGCTGCAGGAATTCGATATCAAGCTGTTGACATATGAGCA	Ladder for 3' processing sequencing gel (55 nt)
Bloli9416	CTAGTGGATCCCCCGGGCTGCAGGAATTCGATATCAAGCTGTTGACATATGAG	Ladder for 3' processing sequencing gel (53 nt)
Bloli9417	CTAGTGGATCCCCCGGGCTGCAGGAATTCGATATCAAGCTGTTGACATA	Ladder for 3' processing sequencing gel (49 nt)
NHEJ sequence bias study		
BLoli8701	/5PHOS/**AATT**CATTAAGTTGGGTAAC**GAGTC**NNNNNNNNNNGGCCAGTGAATTGTAATACGG**GTAC**	Top oligo for pJP14 library; Contains EcoRI and KpnI overhangs for ligation; MlyI restriction site in red

BLoli8702	/5PHOS/GACTCGTTACCCAACTTAATG	Oligos complementary to BLoli8701/9286/9287/9288/9289
BLoli8703	/5PHOS/CCGTATTACAATTCACTGGCC	
BLoli9286	/5PHOS/**AATT**CATTAAGTTGGGTAACGAGTCNNNNN**CTCGAG**NNNNNGGCCAGTGAATTGTAATACGG**GTAC**	Top oligo for pJP15 library; Contains EcoRI and KpnI overhangs for ligation; XhoI restriction site in red
BLoli9287	/5PHOS/**AATT**CATTAAGTTGGGTAACGAGTCNNNNNNNNNN**CTCGAG**NNNNNNNNNNGGCCAGTGAATTGTAATACGG**GTAC**	Top oligo for pJP16 library; Contains EcoRI and KpnI overhangs for ligation; XhoI restriction site in red
BLoli9288	/5PHOS/**GGCC**CATTAAGTTGGGTAACGAGTCNNNNN**CCCGGG**NNNNNGGCCAGTGAATTGTAATACGG**GTAC**	Top oligo for pJP17 library; Contains NotI and KpnI overhangs for ligation; SmaI restriction site in red
BLoli9289	/5PHOS/**GGCC**CATTAAGTTGGGTAACGAGTCNNNNNNNNNN**CCCGGG**NNNNNNNNNNGGCCAGTGAATTGTAATACGG**GTAC**	Top oligo for pJP18 library; Contains NotI and KpnI overhangs for ligation; SmaI restriction site in red
BLoli9418	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNGCCCATTAAGTTGGGTAAC	Forward primer - PCR1 for illumina library prep pJP17-18. Overhang anneals to PCR2 primers; contains 6 nt UMI
BLoli9419	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNGCCCATTAAGTTGGGTAAC	Forward primer - PCR1 for illumina library prep pJP17-18. Overhang anneals to PCR2 primers; contains 6 nt UMI and an additional 2 Ns to offset sequence homogeneity
BLoli9420	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNNGCCCATTAAGTTGGGTAAC	Forward primer - PCR1 for illumina library prep pJP17-18. Overhang anneals to PCR2 primers; contains 6 nt UMI and an additional 3 Ns to offset sequence homogeneity
BLoli9421	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNNNGCCCATTAAGTTGGGTAAC	Forward primer - PCR1 for illumina library prep pJP17-18. Overhang anneals to PCR2 primers; contains 6 nt UMI and an additional 5 Ns to offset sequence homogeneity
BLoli9422	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNNNNNGCCCATTAAGTTGGGTAAC	Forward primer - PCR1 for illumina library prep pJP17-18. Overhang anneals to PCR2 primers; contains 6 nt UMI and an additional 8 Ns to offset sequence homogeneity

BLoli9423	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNGTTATTGTCTCATGAGCGG	Reverse primer - PCR1 for illumina library prep. Overhang anneals to PCR2 primers; contains 6 nt UMI
BLoli9449	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNATTCATTAAGTTGGGTAAC	Forward primer - PCR1 for illumina library prep pJP14-16. Overhang anneals to PCR2 primers; contains 6 nt UMI
BLoli9450	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNATTCATTAAGTTGGGTAAC	Forward primer - PCR1 for illumina library prep pJP14-16. Overhang anneals to PCR2 primers; contains 6 nt UMI and an additional 2 Ns to offset sequence homogeneity
BLoli9451	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNNATTCATTAAGTTGGGTAAC	Forward primer - PCR1 for illumina library prep pJP14-16. Overhang anneals to PCR2 primers; contains 6 nt UMI and an additional 3 Ns to offset sequence homogeneity
BLoli9452	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNNNATTCATTAAGTTGGGTAAC	Forward primer - PCR1 for illumina library prep pJP14-16. Overhang anneals to PCR2 primers; contains 6 nt UMI and an additional 5 Ns to offset sequence homogeneity
BLoli9453	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNNNNNNNATTCATTAAGTTGGGTAAC	Forward primer - PCR1 for illumina library prep pJP14-16. Overhang anneals to PCR2 primers; contains 6 nt UMI and an additional 8 Ns to offset sequence homogeneity
XLeap_UDI_xxx_P5	AATGATACGGCGACCACCGAGATCTACAC**NNNNNNNN**ACACTCTTTCCCTACACGACGCTCTTCCGATCT	Forward primer – PCR2 for illumina library prep. 8 nt sample barcode underlined
XLeap_UDI_xxx_P7	CAAGCAGAAGACGGCATACGAGAT**NNNNNNNN**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	Reverse primer – PCR2 for illumina library prep. 8 nt sample barcode underlined

Antibodies used in the study

Antibodies against ARID2 (Santa Cruz, #sc-166117), BRD7 (Santa Cruz, #sc-376180) and ARID1A (Abcam, #ab182560) were a gift from Dr. Sandra Schick, IMB Mainz.

Acknowledgements

This journey to a doctorate has been a long and challenging one. I extend my thanks to every single individual who has mentored me over my career for helping shape the scientist that I am today.

Huge thanks to for giving me the opportunity to work in his wonderful lab and for being a mentor. I appreciate the encouragement to bring my own ideas, while also supporting me through my failures. You have had a massive influence on my outlook as a scientist. Special thanks to and for being part of my TAC and helping me with some useful suggestions and perspectives.

Big thanks to my colleagues and friends at the lab for welcoming me. Almost every single member has contributed to my project by being the person to share their expertise, or listen to my ideas and help polish them. Special thanks to the Bioinformatics team for encouraging my curiosity with bioinformatics analysis. You guys must have answered so many dumb questions!

I should also recognize the work done by the core facilities at IMB and BioZentrum I in facilitating my research. Particularly, and from IMB Proteomics CF for assisting with MS experiments and analysis; and from the IMB Genomics CF.

I feel eternally grateful to my parents and for encouraging me to follow my dreams and helping me along the way to make it a reality. I am a product of your efforts and hopes, this achievement is vindication for your struggles.

Last but not the least, to my lovely wife , thank you for being a cheerful part of my life and providing me perspective on my bad days. You have shown me how to enjoy my life outside the lab. You have allowed me to go on 20-minute monologues about NHEJ or some cool result I got even when all you wanted to do was go to sleep. You contribution here is massive and cannot be understated.

References

Ahel, I., Rass, U., El-Khamisy, S. F., Katyal, S., Clements, P. M., McKinnon, P. J., Caldecott, K. W., & West, S. C. (2006). The neurodegenerative disease protein aprataxin resolves abortive DNA ligation intermediates. *Nature*, *443*(7112), 713–716. https://doi.org/10.1038/nature05164

Arnoult, N., Correia, A., Ma, J., Merlo, A., Garcia-Gomez, S., Maric, M., Tognetti, M., Benner, C. W., Boulton, S. J., Saghatelian, A., & Karlseder, J. (2017). Regulation of DNA repair pathway choice in S and G2 phases by the NHEJ inhibitor CYREN. *Nature*, *549*(7673), 548–552. https://doi.org/10.1038/nature24023

Avcilar-Kucukgoze, I., Gamper, H., Hou, Y. M., & Kashina, A. (2020). Purification and Use of tRNA for Enzymatic Post-translational Addition of Amino Acids to Proteins. *STAR Protocols*, *1*(3), 100207. https://doi.org/10.1016/j.xpro.2020.100207

Aymard, F., Bugler, B., Schmidt, C. K., Guillou, E., Caron, P., Briois, S., Iacovoni, J. S., Daburon, V., Miller, K. M., Jackson, S. P., & Legube, G. (2014). Transcriptionally active chromatin recruits homologous recombination at DNA double-strand breaks. *Nature Structural & Molecular Biology*, *21*(4), 366–374. https://doi.org/10.1038/nsmb.2796

Bader, A. S., Hawley, B. R., Wilczynska, A., & Bushell, M. (2020). The roles of RNA in DNA double-strand break repair. *British Journal of Cancer*, *122*(5), 613–623. https://doi.org/10.1038/s41416-019-0624-1

Bajin, İ. Y., Ayvaz, D. Ç., Ünal, Ş., Özgür, T. T., Çetin, M., Gümrük, F., Tezcan, İ., de Villartay, J.-P., & Sanal, Ö. (2013). Atypical combined immunodeficiency due to Artemis defect: A case presenting as hyperimmunoglobulin M syndrome and with LGLL. *Molecular Immunology*, *56*(4), 354–357. https://doi.org/10.1016/j.molimm.2013.05.004

Bakkenist, C. J., & Kastan, M. B. (2003). DNA damage activates ATM through intermolecular autophosphorylation and dimer dissociation. *Nature*, *421*(6922), 499–506. https://doi.org/10.1038/nature01368

Balmus, G., Barros, A. C., Wijnhoven, P. W. G., Lescale, C., Hasse, H. L., Boroviak, K., le Sage, C., Doe, B., Speak, A. O., Galli, A., Jacobsen, M., Deriano, L., Adams, D. J., Blackford, A. N., & Jackson, S. P. (2016). Synthetic lethality between PAXX and XLF in mammalian development. *Genes & Development*, *30*(19), 2152–2157. https://doi.org/10.1101/gad.290510.116

Bauerschmidt, C., Arrichiello, C., Burdak-Rothkamm, S., Woodcock, M., Hill, M. A., Stevens, D. L., & Rothkamm, K. (2010). Cohesin promotes the repair of ionizing radiation-induced DNA double-strand breaks in replicated chromatin. *Nucleic Acids Research*, *38*(2), 477–487. https://doi.org/10.1093/nar/gkp976

Baumann, P., & West, S. C. (1998). DNA end-joining catalyzed by human cell-free extracts. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(24), 14066–14070. https://doi.org/10.1073/pnas.95.24.14066

Beck, B. D., Lee, S.-S., Williamson, E., Hromas, R. A., & Lee, S.-H. (2011). Biochemical Characterization of Metnase's Endonuclease Activity and Its Role in NHEJ Repair. *Biochemistry*, *50*(20), 4360–4370. https://doi.org/10.1021/bi200333k

Bétermier, M., Bertrand, P., & Lopez, B. S. (2014). Is Non-Homologous End-Joining Really an Inherently Error-Prone Process? *PLoS Genetics*, *10*(1), e1004086. https://doi.org/10.1371/journal.pgen.1004086

Beucher, A., Birraux, J., Tchouandong, L., Barton, O., Shibata, A., Conrad, S., Goodarzi, A. A., Krempler, A., Jeggo, P. A., & Löbrich, M. (2009). ATM and Artemis promote homologous recombination of radiation-induced DNA double-strand breaks in G2. *The EMBO Journal*, *28*(21), 3413–3427. https://doi.org/10.1038/emboj.2009.276

Blier, P. R., Griffith, A. J., Craft, J., & Hardin, J. A. (1993). Binding of Ku protein to DNA. Measurement of affinity for ends and demonstration of binding to nicks. *Journal of Biological Chemistry*, *268*(10), 7594–7601. https://doi.org/10.1016/S0021-9258(18)53216-6

Block, W. D., Yu, Y., Merkle, D., Gifford, J. L., Ding, Q., Meek, K., & Lees-Miller, S. P. (2004). Autophosphorylation-dependent remodeling of the DNA-dependent protein kinase catalytic subunit regulates ligation of DNA ends. *Nucleic Acids Research*, *32*(14), 4351–4357. https://doi.org/10.1093/nar/gkh761

Bonath, F., Domingo-Prim, J., Tarbier, M., Friedländer, M. R., & Visa, N. (2018). Next-generation sequencing reveals two populations of damage-induced small RNAs at endogenous DNA double-strand breaks. *Nucleic Acids Research*, *46*(22), 11869–11882. https://doi.org/10.1093/nar/gky1107

Buck, D., Malivert, L., de Chasseval, R., Barraud, A., Fondanèche, M.-C., Sanal, O., Plebani, A., Stéphan, J.-L., Hufnagel, M., le Deist, F., Fischer, A., Durandy, A., de Villartay, J.-P., & Revy, P. (2006). Cernunnos, a Novel Nonhomologous End-Joining Factor, Is Mutated in Human Immunodeficiency with Microcephaly. *Cell*, *124*(2), 287–299. https://doi.org/10.1016/j.cell.2005.12.030

Buck, D., Moshous, D., de Chasseval, R., Ma, Y., le Deist, F., Cavazzana-Calvo, M., Fischer, A., Casanova, J., Lieber, M. R., & de Villartay, J. (2006). Severe combined immunodeficiency and microcephaly in siblings with hypomorphic mutations in DNA ligase IV. *European Journal of Immunology*, *36*(1), 224–235. https://doi.org/10.1002/eji.200535401

Budman, J., & Chu, G. (2005). Processing of DNA for nonhomologous end-joining by cell-free extract. *The EMBO Journal*, *24*(4), 849–860. https://doi.org/10.1038/sj.emboj.7600563

Burma, S., Chen, B. P., Murphy, M., Kurimasa, A., & Chen, D. J. (2001). ATM Phosphorylates Histone H2AX in Response to DNA Double-strand Breaks. *Journal of Biological Chemistry*, *276*(45), 42462–42467. https://doi.org/10.1074/jbc.C100466200

Callen, E., Zong, D., Wu, W., Wong, N., Stanlie, A., Ishikawa, M., Pavani, R., Dumitrache, L. C., Byrum, A. K., Mendez-Dorantes, C., Martinez, P., Canela, A., Maman, Y., Day, A., Kruhlak, M. J., Blasco, M. A., Stark, J. M., Mosammaparast, N., McKinnon, P. J., & Nussenzweig, A. (2020). 53BP1 Enforces Distinct Pre- and Post-resection Blocks on Homologous Recombination. *Molecular Cell*, *77*(1), 26-38.e7. https://doi.org/10.1016/j.molcel.2019.09.024

Carter, T., Vančurová, I., Sun, I., Lou, W., & DeLeon, S. (1990). A DNA-Activated Protein Kinase from HeLa Cell Nuclei. *Molecular and Cellular Biology*, *10*(12), 6460–6471. https://doi.org/10.1128/mcb.10.12.6460-6471.1990

Cary, R. B., Peterson, S. R., Wang, J., Bear, D. G., Bradbury, E. M., & Chen, D. J. (1997). DNA looping by Ku and the DNA-dependent protein kinase. *Proceedings of the National Academy of Sciences*, *94*(9), 4267–4272. https://doi.org/10.1073/pnas.94.9.4267

Chakraborty, A., Tapryal, N., Islam, A., Sarker, A. H., Manohar, K., Mitra, J., Hegde, M. L., & Hazra, T. (2023). Human DNA polymerase η promotes RNA-templated error-free repair of DNA double-strand breaks. *Journal of Biological Chemistry*, *299*(3). https://doi.org/10.1016/j.jbc.2023.102991

Chakraborty, A., Tapryal, N., Venkova, T., Horikoshi, N., Pandita, R. K., Sarker, A. H., Sarkar, P. S., Pandita, T. K., & Hazra, T. K. (2016). Classical non-homologous end-joining pathway utilizes nascent RNA for error-free double-strand break repair of transcribed genes. *Nature Communications*, *7*. https://doi.org/10.1038/ncomms13049

Chan, D. W., Chen, B. P.-C., Prithivirajsingh, S., Kurimasa, A., Story, M. D., Qin, J., & Chen, D. J. (2002). Autophosphorylation of the DNA-dependent protein kinase catalytic subunit is required for rejoining of DNA double-strand breaks. *Genes & Development*, *16*(18), 2333–2338. https://doi.org/10.1101/gad.1015202

Chan, D. W., & Lees-Miller, S. P. (1996). The DNA-dependent Protein Kinase Is Inactivated by Autophosphorylation of the Catalytic Subunit. *Journal of Biological Chemistry*, *271*(15), 8936–8941. https://doi.org/10.1074/jbc.271.15.8936

Chandramouly, G., Zhao, J., McDevitt, S., Rusanov, T., Hoang, T., Borisonnik, N., Treddinick, T., Lopezcolorado, F. W., Kent, T., Siddique, L. A., Mallon, J., Huhn, J., Shoda, Z., Kashkina, E., Brambati, A., Stark, J. M., Chen, X. S., & Pomerantz, R. T. (2021). Polθ reverse transcribes RNA and promotes RNA-templated DNA repair. *Science Advances*, *7*(24), 1–12. https://doi.org/10.1126/sciadv.abf1771

Chang, H. H. Y., Pannunzio, N. R., Adachi, N., & Lieber, M. R. (2017). Non-homologous DNA end joining and alternative pathways to double-strand break repair. In *Nature Reviews Molecular Cell Biology* (Vol. 18, Issue 8, pp. 495–506). Nature Publishing Group. https://doi.org/10.1038/nrm.2017.48

Chapman, J. R., Sossick, A. J., Boulton, S. J., & Jackson, S. P. (2012). BRCA1-associated exclusion of 53BP1 from DNA damage sites underlies temporal control of DNA repair. *Journal of Cell Science*, *125*(15), 3529–3534. https://doi.org/10.1242/jcs.105353

Chappell, C., Hanakahi, L. A., Karimi-Busheri, F., Weinfeld, M., & West, S. C. (2002). Involvement of human polynucleotide kinase in double-strand break repair by non-homologous end joining. *The EMBO Journal*, *21*(11), 2827–2832. https://doi.org/10.1093/emboj/21.11.2827

Chen, B. P. C., Chan, D. W., Kobayashi, J., Burma, S., Asaithamby, A., Morotomi-Yano, K., Botvinick, E., Qin, J., & Chen, D. J. (2005). Cell Cycle Dependence of DNA-dependent Protein Kinase Phosphorylation in Response to DNA Double Strand Breaks. *Journal of Biological Chemistry*, *280*(15), 14709–14715. https://doi.org/10.1074/jbc.M408827200

Chen, B. P. C., Uematsu, N., Kobayashi, J., Lerenthal, Y., Krempler, A., Yajima, H., Löbrich, M., Shiloh, Y., & Chen, D. J. (2007). Ataxia Telangiectasia Mutated (ATM) Is Essential for DNA-PKcs Phosphorylations at the Thr-2609 Cluster upon DNA Double Strand Break. *Journal of Biological Chemistry*, *282*(9), 6582–6587. https://doi.org/10.1074/jbc.M611605200

Chen, S., Lee, L., Naila, T., Fishbain, S., Wang, A., Tomkinson, A. E., Lees-Miller, S. P., & He, Y. (2021). Structural basis of long-range to short-range synaptic transition in NHEJ. *Nature*, *593*(7858), 294–298. https://doi.org/10.1038/s41586-021-03458-7

Chen, X., Xu, X., Chen, Y., Cheung, J. C., Wang, H., Jiang, J., de Val, N., Fox, T., Gellert, M., & Yang, W. (2021). Structure of an activated DNA-PK and its implications for NHEJ. *Molecular Cell*, *81*(4), 801-810.e3. https://doi.org/10.1016/j.molcel.2020.12.015

Chen, Z., & Tyler, J. K. (2022). The Chromatin Landscape Channels DNA Double-Strand Breaks to Distinct Repair Pathways. *Frontiers in Cell and Developmental Biology*, *10*. https://doi.org/10.3389/fcell.2022.909696

Cheong, N., Perrault, A. R., & Iliakis, G. (1998). In vitro rejoining of DNA double strand breaks: a comparison of genomic-DNA with plasmid-DNA-based assays. *International Journal of Radiation Biology*, *73*(5), 481–493. https://doi.org/10.1080/095530098142022

Chiu, C.-F., Lin, T.-Y., & Chou, W.-G. (2001). Direct transfer of Ku between DNA molecules with nonhomologous ends. *Mutation Research/DNA Repair*, *486*(3), 185–194. https://doi.org/10.1016/S0921-8777(01)00080-5

Coquerelle, T., Bopp, A., Kessler, B., & Hagen, U. (1973). Strand Breaks and 5' End-groups in DNA of Irradiated Thymocytes. *International Journal of Radiation Biology and Related Studies in Physics, Chemistry and Medicine*, *24*(4), 397–404. https://doi.org/10.1080/09553007314551251

Costantini, S., Woodbine, L., Andreoli, L., Jeggo, P. A., & Vindigni, A. (2007). Interaction of the Ku heterodimer with the DNA ligase IV/Xrcc4 complex and its regulation by DNA-PK. *DNA Repair*, *6*(6), 712–722. https://doi.org/10.1016/j.dnarep.2006.12.007

Cottarel, J., Frit, P., Bombarde, O., Salles, B., Négrel, A., Bernard, S., Jeggo, P. A., Lieber, M. R., Modesti, M., & Calsou, P. (2013). A noncatalytic function of the ligation complex during nonhomologous end joining. *Journal of Cell Biology*, *200*(2), 173–186. https://doi.org/10.1083/jcb.201203128

Craxton, A., Somers, J., Munnur, D., Jukes-Jones, R., Cain, K., & Malewicz, M. (2015). XLS (c9orf142) is a new component of mammalian DNA double-stranded break repair. *Cell Death & Differentiation*, *22*(6), 890–897. https://doi.org/10.1038/cdd.2015.22

Cristini, A., Ricci, G., Britton, S., Salimbeni, S., Huang, S. N., Marinello, J., Calsou, P., Pommier, Y., Favre, G., Capranico, G., Gromak, N., & Sordet, O. (2019). Dual Processing of R-Loops and Topoisomerase I Induces Transcription-Dependent DNA Double-Strand Breaks. *Cell Reports*, *28*(12), 3167-3181.e6. https://doi.org/10.1016/j.celrep.2019.08.041

Critchlow, S. E., Bowater, R. P., & Jackson, S. P. (1997). Mammalian DNA double-strand break repair protein XRCC4 interacts with DNA ligase IV. *Current Biology : CB*, *7*(8), 588–598. https://doi.org/10.1016/s0960-9822(06)00258-2

Cui, X., Yu, Y., Gupta, S., Cho, Y.-M., Lees-Miller, S. P., & Meek, K. (2005). Autophosphorylation of DNA-dependent protein kinase regulates DNA end processing and may also alter double-strand break repair pathway choice. *Molecular and Cellular Biology*, *25*(24), 10842–10852. https://doi.org/10.1128/MCB.25.24.10842-10852.2005

D'Adda di Fagagna, F. (2014). A direct role for small non-coding RNAs in DNA damage response. In *Trends in Cell Biology* (Vol. 24, Issue 3, pp. 171–178). https://doi.org/10.1016/j.tcb.2013.09.008

D'Alessandro, G., Whelan, D. R., Howard, S. M., Vitelli, V., Renaudin, X., Adamowicz, M., Iannelli, F., Jones-Weinert, C. W., Lee, M. Y., Matti, V., Lee, W. T. C., Morten, M. J.,

Venkitaraman, A. R., Cejka, P., Rothenberg, E., & d'Adda di Fagagna, F. (2018). BRCA2 controls DNA:RNA hybrid level at DSBs by mediating RNase H2 recruitment. *Nature Communications*, *9*(1). https://doi.org/10.1038/s41467-018-07799-2

D'Amours, D., & Jackson, S. P. (2002). The MRE11 complex: at the crossroads of DNA repair and checkpoint signalling. *Nature Reviews Molecular Cell Biology*, *3*(5), 317–327. https://doi.org/10.1038/nrm805

Dash, S., Balasubramaniam, M., Rana, T., Godino, A., Peck, E. G., Goodwin, J. S., Villalta, F., Calipari, E. S., Nestler, E. J., Dash, C., & Pandhare, J. (2017). Poly (ADP-Ribose) Polymerase-1 (PARP-1) Induction by Cocaine Is Post-Transcriptionally Regulated by miR-125b. *Eneuro*, *4*(4), ENEURO.0089-17.2017. https://doi.org/10.1523/ENEURO.0089-17.2017

Datta, K., Purkayastha, S., Neumann, R. D., & Winters, T. A. (2012). *An In Vitro DNA Double-Strand Break Repair Assay Based on End-Joining of Defined Duplex Oligonucleotides* (pp. 485–500). https://doi.org/10.1007/978-1-61779-998-3_33

de Vries, E., van Driel, W., Bergsma, W. G., Arnberg, A. C., & van der Vliet, P. C. (1989). HeLa nuclear protein recognizing DNA termini and translocating on DNA forming a regular DNA-multimeric protein complex. *Journal of Molecular Biology*, *208*(1), 65–78. https://doi.org/10.1016/0022-2836(89)90088-0

DeFazio, L. G., Stansel, R. M., Griffith, J. D., & Chu, G. (2002). Synapsis of DNA ends by DNA-dependent protein kinase. *The EMBO Journal*, *21*(12), 3192–3200. https://doi.org/10.1093/emboj/cdf299

Dereli, I., Telychko, V., Papanikos, F., Raveendran, K., Xu, J., Boekhout, M., Stanzione, M., Neuditschko, B., Imjeti, N. S., Selezneva, E., Tuncay, H., Demir, S., Giannattasio, T., Gentzel, M., Bondarieva, A., Stevense, M., Barchi, M., Schnittger, A., Weir, J. R., … Tóth, A. (2024). Seeding the meiotic DNA break machinery and initiating recombination on chromosome axes. *Nature Communications*, *15*(1), 2941. https://doi.org/10.1038/s41467-024-47020-1

Deweese, J. E., & Osheroff, N. (2009). The DNA cleavage reaction of topoisomerase II: wolf in sheep's clothing. *Nucleic Acids Research*, *37*(3), 738–748. https://doi.org/10.1093/nar/gkn937

Ding, Q., Reddy, Y. V. R., Wang, W., Woods, T., Douglas, P., Ramsden, D. A., Lees-Miller, S. P., & Meek, K. (2003). Autophosphorylation of the Catalytic Subunit of the DNA-Dependent Protein Kinase Is Required for Efficient End Processing during DNA Double-Strand Break Repair. *Molecular and Cellular Biology*, *23*(16), 5836–5848. https://doi.org/10.1128/MCB.23.16.5836-5848.2003

Dobbs, T. A., Tainer, J. A., & Lees-Miller, S. P. (2010). A structural model for regulation of NHEJ by DNA-PKcs autophosphorylation. *DNA Repair*, *9*(12), 1307–1314. https://doi.org/10.1016/j.dnarep.2010.09.019

Domínguez, O., Ruiz, J. F., Laín de Lera, T., García-Díaz, M., González, M. A., Kirchhoff, T., Martínez-A, C., Bernad, A., & Blanco, L. (2000). DNA polymerase mu (Pol mu), homologous to TdT, could act as a DNA mutator in eukaryotic cells. *The EMBO Journal*, *19*(7), 1731–1742. https://doi.org/10.1093/emboj/19.7.1731

Douglas, P., Gupta, S., Morrice, N., Meek, K., & Lees-Miller, S. P. (2005). DNA-PK-dependent phosphorylation of Ku70/80 is not required for non-homologous end joining. *DNA Repair*, *4*(9), 1006–1018. https://doi.org/10.1016/j.dnarep.2005.05.003

Douglas, P., Sapkota, G. P., Morrice, N., Yu, Y., Goodarzi, A. A., Merkle, D., Meek, K., Alessi, D. R., & Lees-Miller, S. P. (2002). Identification of in vitro and in vivo phosphorylation sites in the catalytic subunit of the DNA-dependent protein kinase. *The Biochemical Journal*, *368*(Pt 1), 243–251. https://doi.org/10.1042/BJ20020973

Downs, J. A., & Jackson, S. P. (2004). A means to a DNA end: the many roles of Ku. *Nature Reviews Molecular Cell Biology*, *5*(5), 367–378. https://doi.org/10.1038/nrm1367

Enders, A., Fisch, P., Schwarz, K., Duffner, U., Pannicke, U., Nikolopoulos, E., Peters, A., Orlowska-Volk, M., Schindler, D., Friedrich, W., Selle, B., Niemeyer, C., & Ehl, S. (2006). A Severe Form of Human Combined Immunodeficiency Due to Mutations in DNA Ligase IV. *The Journal of Immunology*, *176*(8), 5060–5068. https://doi.org/10.4049/jimmunol.176.8.5060

Fedkenheuer, M., Shang, Y., Jung, S., Fedkenheuer, K., Park, S., Mazza, D., Sebastian, R., Nagashima, H., Zong, D., Tan, H., Jaiswal, S. K., Fu, H., Cruz, A., Vartak, S. V., Wisniewski, J., Sartorelli, V., O'Shea, J. J., Elnitski, L., Nussenzweig, A., … Casellas, R. (2025). A dual role of Cohesin in DNA DSB repair. *Nature Communications*, *16*(1), 843. https://doi.org/10.1038/s41467-025-56086-4

Francia, S., Cabrini, M., Matti, V., Oldani, A., & d'Adda di Fagagna, F. (2016). DICER, DROSHA and DNA damage response RNAs are necessary for the secondary recruitment of DNA damage response factors. *Journal of Cell Science*, *129*(7), 1468–1476. https://doi.org/10.1242/jcs.182188

Francia, S., Michelini, F., Saxena, A., Tang, D., De Hoon, M., Anelli, V., Mione, M., Carninci, P., & D'adda Di Fagagna, F. (2012). Site-specific DICER and DROSHA RNA products control the DNA-damage response. *Nature*, *488*(7410), 231–235. https://doi.org/10.1038/nature11179

Francoeur, A. M., Peebles, C. L., Gompper, P. T., & Tan, E. M. (1986). Identification of Ki (Ku, p70/p80) autoantigens and analysis of anti-Ki autoantibody reactivity. *Journal of Immunology (Baltimore, Md. : 1950)*, *136*(5), 1648–1653.

Friesner, J. D., Liu, B., Culligan, K., & Britt, A. B. (2005). Ionizing Radiation–dependent γ-H2AX Focus Formation Requires Ataxia Telangiectasia Mutated and Ataxia Telangiectasia Mutated and Rad3-related. *Molecular Biology of the Cell*, *16*(5), 2566–2576. https://doi.org/10.1091/mbc.e04-10-0890

Gao, Y., Sun, Y., Frank, K. M., Dikkes, P., Fujiwara, Y., Seidl, K. J., Sekiguchi, J. M., Rathbun, G. A., Swat, W., Wang, J., Bronson, R. T., Malynn, B. A., Bryans, M., Zhu, C., Chaudhuri, J., Davidson, L., Ferrini, R., Stamato, T., Orkin, S. H., … Alt, F. W. (1998). A Critical Role for DNA End-Joining Proteins in Both Lymphogenesis and Neurogenesis. *Cell*, *95*(7), 891–902. https://doi.org/10.1016/S0092-8674(00)81714-6

Gauss, G., & Lieber, M. R. (1996). Mechanistic Constraints on Diversity in Human V(D)J Recombination. *Molecular and Cellular Biology*, *16*(1), 258–269. https://doi.org/10.1128/MCB.16.1.258

Gell, D., & Jackson, S. P. (1999). Mapping of protein-protein interactions within the DNA-dependent protein kinase complex. *Nucleic Acids Research*, *27*(17), 3494–3502. https://doi.org/10.1093/nar/27.17.3494

Gellert, M. (2002). V(D)J recombination: RAG proteins, repair factors, and regulation. *Annual Review of Biochemistry*, *71*(D), 101–132. https://doi.org/10.1146/annurev.biochem.71.090501.150203

Gelot, C., Guirouilh-Barbat, J., Le Guen, T., Dardillac, E., Chailleux, C., Canitrot, Y., & Lopez, B. S. (2016). The Cohesin Complex Prevents the End Joining of Distant DNA Double-Strand Ends. *Molecular Cell*, *61*(1), 15–26. https://doi.org/10.1016/j.molcel.2015.11.002

Ghosh, D., & Raghavan, S. C. (2021). Nonhomologous end joining: new accessory factors fine tune the machinery. *Trends in Genetics*, *37*(6), 582–599. https://doi.org/10.1016/j.tig.2021.03.001

Gilfillan, S., Dierich, A., Lemeur, M., Benoist, C., & Mathis, D. (1993). Mice Lacking TdT: Mature Animals with an Immature Lymphocyte Repertoire. *Science*, *261*(5125), 1175–1178. https://doi.org/10.1126/science.8356452

Goff, N. J., Brenière, M., Buehl, C. J., De Melo, A. J., Huskova, H., Ochi, T., Blundell, T. L., Mao, W., Yu, K., Modesti, M., & Meek, K. (2022). Catalytically inactive DNA ligase IV promotes DNA repair in living cells. *Nucleic Acids Research*, *50*(19), 11058–11071. https://doi.org/10.1093/nar/gkac913

Gottlieb, T. M., & Jackson, S. P. (1993). The DNA-dependent protein kinase: Requirement for DNA ends and association with Ku antigen. *Cell*, *72*(1), 131–142. https://doi.org/10.1016/0092-8674(93)90057-W

Graham, T. G. W., Walter, J. C., & Loparo, J. J. (2016). Two-Stage Synapsis of DNA Ends during Non-homologous End Joining. *Molecular Cell*, *61*(6), 850–858. https://doi.org/10.1016/j.molcel.2016.02.010

Grawunder, U., Zimmer, D., Kulesza, P., & Lieber, M. R. (1998). Requirement for an Interaction of XRCC4 with DNA Ligase IV for Wild-type V(D)J Recombination and DNA Double-strand Break Repairin Vivo. *Journal of Biological Chemistry*, *273*(38), 24708–24714. https://doi.org/10.1074/jbc.273.38.24708

Gu, Y., Jin, S., Gao, Y., Weaver, D. T., & Alt, F. W. (1997). Ku70-deficient embryonic stem cells have increased ionizing radiosensitivity, defective DNA end-binding activity, and inability to support V(D)J recombination. *Proceedings of the National Academy of Sciences*, *94*(15), 8076–8081. https://doi.org/10.1073/pnas.94.15.8076

Gu, Y., Seidl, K. J., Rathbun, G. A., Zhu, C., Manis, J. P., van der Stoep, N., Davidson, L., Cheng, H.-L., Sekiguchi, J. M., Frank, K., Stanhope-Baker, P., Schlissel, M. S., Roth, D. B., & Alt, F. W. (1997). Growth Retardation and Leaky SCID Phenotype of Ku70-Deficient Mice. *Immunity*, *7*(5), 653–665. https://doi.org/10.1016/S1074-7613(00)80386-6

Haahr, P., Hoffmann, S., Tollenaere, M. A. X., Ho, T., Toledo, L. I., Mann, M., Bekker-Jensen, S., Räschle, M., & Mailand, N. (2016). Activation of the ATR kinase by the RPA-binding protein ETAA1. *Nature Cell Biology*, *18*(11), 1196–1207. https://doi.org/10.1038/ncb3422

Hammel, M., Rey, M., Yu, Y., Mani, R. S., Classen, S., Liu, M., Pique, M. E., Fang, S., Mahaney, B. L., Weinfeld, M., Schriemer, D. C., Lees-Miller, S. P., & Tainer, J. A. (2011). XRCC4 Protein Interactions with XRCC4-like Factor (XLF) Create an Extended Grooved Scaffold for DNA Ligation and Double Strand Break Repair. *Journal of Biological Chemistry*, *286*(37), 32638–32650. https://doi.org/10.1074/jbc.M111.272641

Han, T., Dadey, B., & Minowada, J. (1979). Cultured human leukemic non-T/non-B lymphoblasts and their stimulating capacity in “one-way”mixed lymphocyte reaction. Suggestive evidence for early T-cell or B-cell precursors. *Cancer*, *44*(1), 136–140. https://doi.org/10.1002/1097-0142(197907)44:1<136::AID-CNCR2820440124>3.0.CO;2-I

Hartley, K. O., Gell, D., Smith, G. C. M., Zhang, H., Divecha, N., Connelly, M. A., Admon, A., Lees-Miller, S. P., Anderson, C. W., & Jackson, S. P. (1995). DNA-dependent protein kinase catalytic subunit: A relative of phosphatidylinositol 3-kinase and the ataxia telangiectasia gene product. *Cell*, *82*(5), 849–856. https://doi.org/10.1016/0092-8674(95)90482-4

Hawley, B. R., Lu, W. T., Wilczynska, A., & Bushell, M. (2017). The emerging role of RNAs in DNA damage repair. *Cell Death and Differentiation*, *24*(4), 580–587. https://doi.org/10.1038/cdd.2017.16

Hu, H., Du, L., Nagabayashi, G., Seeger, R. C., & Gatti, R. A. (2010). ATM is down-regulated by N-Myc–regulated microRNA-421. *Proceedings of the National Academy of Sciences*, *107*(4), 1506–1511. https://doi.org/10.1073/pnas.0907763107

Huen, M. S. Y., Grant, R., Manke, I., Minn, K., Yu, X., Yaffe, M. B., & Chen, J. (2007). RNF8 Transduces the DNA-Damage Signal via Histone Ubiquitylation and Checkpoint Protein Assembly. *Cell*, *131*(5), 901–914. https://doi.org/10.1016/j.cell.2007.09.041

Inamdar, K. V., Pouliot, J. J., Zhou, T., Lees-Miller, S. P., Rasouli-Nia, A., & Povirk, L. F. (2002). Conversion of Phosphoglycolate to Phosphate Termini on 3' Overhangs of DNA Double Strand Breaks by the Human Tyrosyl-DNA Phosphodiesterase hTdp1. *Journal of Biological Chemistry*, *277*(30), 27162–27168. https://doi.org/10.1074/jbc.M204688200

Jackson, S. P., MacDonald, J. J., Lees-Miller, S., & Tjian, R. (1990). GC box binding induces phosphorylation of Sp1 by a DNA-dependent protein kinase. *Cell*, *63*(1), 155–165. https://doi.org/10.1016/0092-8674(90)90296-Q

Jacquet, K., Fradet-Turcotte, A., Avvakumov, N., Lambert, J.-P., Roques, C., Pandita, R. K., Paquet, E., Herst, P., Gingras, A.-C., Pandita, T. K., Legube, G., Doyon, Y., Durocher, D., & Côté, J. (2016). The TIP60 Complex Regulates Bivalent Chromatin Recognition by 53BP1 through Direct H4K20me Binding and H2AK15 Acetylation. *Molecular Cell*, *62*(3), 409–421. https://doi.org/10.1016/j.molcel.2016.03.031

Jayaram, S., Ketner, G., Adachi, N., & Hanakahi, L. A. (2008). Loss of DNA ligase IV prevents recognition of DNA by double-strand break repair proteins XRCC4 and XLF. *Nucleic Acids Research*, *36*(18), 5773–5786. https://doi.org/10.1093/nar/gkn552

Jeggo, P. A., & Löbrich, M. (2007). DNA double-strand breaks: their cellular and clinical impact? *Oncogene*, *26*(56), 7717–7719. https://doi.org/10.1038/sj.onc.1210868

Jeon, Y., Lu, Y., Ferrari, M. M., Channagiri, T., Xu, P., Meers, C., Zhang, Y., Balachander, S., Park, V. S., Marsili, S., Pursell, Z. F., Jonoska, N., & Storici, F. (2024). RNA-mediated double-strand break repair by end-joining mechanisms. *Nature Communications* , *15*(1). https://doi.org/10.1038/s41467-024-51457-9

Jiang, W., Crowe, J. L., Liu, X., Nakajima, S., Wang, Y., Li, C., Lee, B. J., Dubois, R. L., Liu, C., Yu, X., Lan, L., & Zha, S. (2015). Differential Phosphorylation of DNA-PKcs Regulates the Interplay between End-Processing and End-Ligation during

Nonhomologous End-Joining. *Molecular Cell*, *58*(1), 172–185. https://doi.org/10.1016/j.molcel.2015.02.024
Kakarougkas, A., Downs, J. A., & Jeggo, P. A. (2015). The PBAF chromatin remodeling complex represses transcription and promotes rapid repair at DNA double-strand breaks. *Molecular & Cellular Oncology*, *2*(1), e970072. https://doi.org/10.4161/23723548.2014.970072
Kakarougkas, A., Ismail, A., Chambers, A. L., Riballo, E., Herbert, A. D., Künzel, J., Löbrich, M., Jeggo, P. A., & Downs, J. A. (2014). Requirement for PBAF in Transcriptional Repression and Repair at DNA Breaks in Actively Transcribed Regions of Chromatin. *Molecular Cell*, *55*(5), 723–732. https://doi.org/10.1016/j.molcel.2014.06.028
Kakarougkas, A., & Jeggo, P. A. (2014). DNA DSB repair pathway choice: an orchestrated handover mechanism. *The British Journal of Radiology*, *87*(1035), 20130685. https://doi.org/10.1259/bjr.20130685
Kang, L., Zou, X., Zhang, G., Xiang, J., Wang, Y., Yang, M., Chen, X., Wu, J., & Guan, A. H. (2019). A variant in a microRNA binding site in NEIL2 3'UTR confers susceptibility to age-related cataracts. *The FASEB Journal*, *33*(9), 10469–10476. https://doi.org/10.1096/fj.201802291R
Karanam, K., Kafri, R., Loewer, A., & Lahav, G. (2012). Quantitative Live Cell Imaging Reveals a Gradual Shift between DNA Repair Mechanisms and a Maximal Use of HR in Mid S Phase. *Molecular Cell*, *47*(2), 320–329. https://doi.org/10.1016/j.molcel.2012.05.052
Keskin, H., Shen, Y., Huang, F., Patel, M., Yang, T., Ashley, K., Mazin, A. V., & Storici, F. (2014). Transcript-RNA-templated DNA recombination and repair. *Nature*, *515*(7527), 436–439. https://doi.org/10.1038/nature13682
Khan, F. A., & Ali, S. O. (2017). Physiological Roles of DNA Double-Strand Breaks. *Journal of Nucleic Acids*, *2017*, 1–20. https://doi.org/10.1155/2017/6439169
Kim, J.-S., Krasieva, T. B., LaMorte, V., Taylor, A. M. R., & Yokomori, K. (2002). Specific Recruitment of Human Cohesin to Laser-induced DNA Damage. *Journal of Biological Chemistry*, *277*(47), 45149–45153. https://doi.org/10.1074/jbc.M209123200
Koch, C. A., Agyei, R., Galicia, S., Metalnikov, P., O'Donnell, P., Starostine, A., Weinfeld, M., & Durocher, D. (2004). Xrcc4 physically links DNA end processing by polynucleotide kinase to DNA ligation by DNA ligase IV. *The EMBO Journal*, *23*(19), 3874–3885. https://doi.org/10.1038/sj.emboj.7600375
Kolas, N. K., Chapman, J. R., Nakada, S., Ylanko, J., Chahwan, R., Sweeney, F. D., Panier, S., Mendez, M., Wildenhain, J., Thomson, T. M., Pelletier, L., Jackson, S. P., & Durocher, D. (2007). Orchestration of the DNA-Damage Response by the RNF8 Ubiquitin Ligase. *Science*, *318*(5856), 1637–1640. https://doi.org/10.1126/science.1150034
Kosaka, Y., Miyawaki, Y., Mori, M., Aburaya, S., Nishizawa, C., Chujo, T., Niwa, T., Miyazaki, T., Sugita, T., Fukuyama, M., Taguchi, H., Tomizawa, K., Sugase, K., Ueda, M., & Aoki, W. (2025). Autonomous ribosome biogenesis in vitro. *Nature Communications*, *16*(1), 514. https://doi.org/10.1038/s41467-025-55853-7
Krasner, D. S., Daley, J. M., Sung, P., & Niu, H. (2015). Interplay between Ku and Replication Protein A in the Restriction of Exo1-mediated DNA Break End

Resection. *Journal of Biological Chemistry*, *290*(30), 18806–18816. https://doi.org/10.1074/jbc.M115.660191
Kuo, T. C., & Schlissel, M. S. (2009). Mechanisms controlling expression of the RAG locus during lymphocyte development. *Current Opinion in Immunology*, *21*(2), 173–178. https://doi.org/10.1016/j.coi.2009.03.008
Kuru-Schors, M., Haemmerle, M., & Gutschner, T. (2021). The Cohesin Complex and Its Interplay with Non-Coding RNAs. *Non-Coding RNA*, *7*(4), 67. https://doi.org/10.3390/ncrna7040067
Lal, A., Pan, Y., Navarro, F., Dykxhoorn, D. M., Moreau, L., Meire, E., Bentwich, Z., Lieberman, J., & Chowdhury, D. (2009). miR-24–mediated downregulation of H2AX suppresses DNA repair in terminally differentiated blood cells. *Nature Structural & Molecular Biology*, *16*(5), 492–498. https://doi.org/10.1038/nsmb.1589
Leahy, J. J. J., Golding, B. T., Griffin, R. J., Hardcastle, I. R., Richardson, C., Rigoreau, L., & Smith, G. C. M. (2004). Identification of a highly potent and selective DNA-dependent protein kinase (DNA-PK) inhibitor (NU7441) by screening of chromenone libraries. *Bioorganic & Medicinal Chemistry Letters*, *14*(24), 6083–6087. https://doi.org/10.1016/j.bmcl.2004.09.060
Ledesma, F. C., El Khamisy, S. F., Zuma, M. C., Osborn, K., & Caldecott, K. W. (2009). A human 5'-tyrosyl DNA phosphodiesterase that repairs topoisomerase-mediated DNA damage. *Nature*, *461*(7264), 674–678. https://doi.org/10.1038/nature08444
Lee, H.-C., Chang, S.-S., Choudhary, S., Aalto, A. P., Maiti, M., Bamford, D. H., & Liu, Y. (2009). qiRNA is a new type of small interfering RNA induced by DNA damage. *Nature*, *459*(7244), 274–277. https://doi.org/10.1038/nature08041
Lees-Miller, S. P., Chen, Y.-R., & Anderson, C. W. (1990). Human Cells Contain a DNA-Activated Protein Kinase that Phosphorylates Simian Virus 40 T Antigen, Mouse p53, and the Human Ku Autoantigen. *Molecular and Cellular Biology*, *10*(12), 6472–6481. https://doi.org/10.1128/mcb.10.12.6472-6481.1990
Li, G., Alt, F. W., Cheng, H.-L., Brush, J. W., Goff, P. H., Murphy, M. M., Franco, S., Zhang, Y., & Zha, S. (2008). Lymphocyte-Specific Compensation for XLF/Cernunnos End-Joining Functions in V(D)J Recombination. *Molecular Cell*, *31*(5), 631–640. https://doi.org/10.1016/j.molcel.2008.07.017
Li, Y., Chirgadze, D. Y., Bolanos-Garcia, V. M., Sibanda, B. L., Davies, O. R., Ahnesorg, P., Jackson, S. P., & Blundell, T. L. (2008). Crystal structure of human XLF/Cernunnos reveals unexpected differences from XRCC4 with implications for NHEJ. *The EMBO Journal*, *27*(1), 290–300. https://doi.org/10.1038/sj.emboj.7601942
Liang, S., & Blundell, T. L. (2023). Human DNA-dependent protein kinase activation mechanism. *Nature Structural & Molecular Biology*, *30*(2), 140–147. https://doi.org/10.1038/s41594-022-00881-w
Liang, S., Thomas, S. E., Chaplin, A. K., Hardwick, S. W., Chirgadze, D. Y., & Blundell, T. L. (2022). Structural insights into inhibitor regulation of the DNA repair protein DNA-PKcs. *Nature*, *601*(7894), 643–648. https://doi.org/10.1038/s41586-021-04274-9
Lieber, M. R. (2010). The Mechanism of Double-Strand DNA Break Repair by the Nonhomologous DNA End-Joining Pathway. *Annual Review of Biochemistry*, *79*(1), 181–211. https://doi.org/10.1146/annurev.biochem.052308.093131
Lin, W. Y., Wilson, J. H., & Lin, Y. (2013). Repair of chromosomal double-strand breaks by precise ligation in human cells. *DNA Repair*, *12*(7), 480–487. https://doi.org/10.1016/j.dnarep.2013.04.024

Lu, W. T., Hawley, B. R., Skalka, G. L., Baldock, R. A., Smith, E. M., Bader, A. S., Malewicz, M., Watts, F. Z., Wilczynska, A., & Bushell, M. (2018). Drosha drives the formation of DNA:RNA hybrids around DNA break sites to facilitate DNA repair. *Nature Communications*, *9*(1). https://doi.org/10.1038/s41467-018-02893-x

Ma, Y., Pannicke, U., Schwarz, K., & Lieber, M. R. (2002). Hairpin Opening and Overhang Processing by an Artemis/DNA-Dependent Protein Kinase Complex in Nonhomologous End Joining and V(D)J Recombination. *Cell*, *108*(6), 781–794. https://doi.org/10.1016/S0092-8674(02)00671-2

Ma, Y., Schwarz, K., & Lieber, M. R. (2005). The Artemis:DNA-PKcs endonuclease cleaves DNA loops, flaps, and gaps. *DNA Repair*, *4*(7), 845–851. https://doi.org/10.1016/j.dnarep.2005.04.013

Mailand, N., Bekker-Jensen, S., Faustrup, H., Melander, F., Bartek, J., Lukas, C., & Lukas, J. (2007). RNF8 Ubiquitylates Histones at DNA Double-Strand Breaks and Promotes Assembly of Repair Proteins. *Cell*, *131*(5), 887–900. https://doi.org/10.1016/j.cell.2007.09.040

Mao, Z., Bozzella, M., Seluanov, A., & Gorbunova, V. (2008). Comparison of nonhomologous end joining and homologous recombination in human cells. *DNA Repair*, *7*(10), 1765–1771. https://doi.org/10.1016/j.dnarep.2008.06.018

Mateyko, N., & de Boer, C. G. (2024). Culture Wars: Empirically Determining the Best Approach for Plasmid Library Amplification. *ACS Synthetic Biology*, *13*(8), 2328–2334. https://doi.org/10.1021/acssynbio.4c00377

Mazina, O. M., Keskin, H., Hanamshet, K., Storici, F., & Mazin, A. V. (2017). Rad52 Inverse Strand Exchange Drives RNA-Templated DNA Double-Strand Break Repair. *Molecular Cell*, *67*(1), 19-29.e3. https://doi.org/10.1016/j.molcel.2017.05.019

Medina-Suárez, D., Han, L., O'Reilly, S., Liu, J., Wei, C., Brenière, M., Goff, N. J., Chen, C., Modesti, M., Meek, K., Harrington, B., & Yu, K. (2024). Lig3-dependent rescue of mouse viability and DNA double-strand break repair by catalytically inactive Lig4. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkae1216

Meek, K., Buehl, C., Goff, N., Mikhova, M., Hardwick, S., Blundell, T., Modesti, M., Schmidt, J., & Chaplin, A. (2023). *Unravelling the complexities of DNA-PK activation by structure-based mutagenesis*. https://doi.org/10.21203/rs.3.rs-3627471/v1

Meek, K., Douglas, P., Cui, X., Ding, Q., & Lees-Miller, S. P. (2007). trans Autophosphorylation at DNA-dependent protein kinase's two major autophosphorylation site clusters facilitates end processing but not end joining. *Molecular and Cellular Biology*, *27*(10), 3881–3890. https://doi.org/10.1128/MCB.02366-06

Merkle, D., Douglas, P., Moorhead, G. B. G., Leonenko, Z., Yu, Y., Cramb, D., Bazett-Jones, D. P., & Lees-Miller, S. P. (2002). The DNA-Dependent Protein Kinase Interacts with DNA To Form a Protein−DNA Complex That Is Disrupted by Phosphorylation. *Biochemistry*, *41*(42), 12706–12714. https://doi.org/10.1021/bi0263558

Messer, P. W., & Arndt, P. F. (2007). The Majority of Recent Short DNA Insertions in the Human Genome Are Tandem Duplications. *Molecular Biology and Evolution*, *24*(5), 1190–1197. https://doi.org/10.1093/molbev/msm035

Michalik, K. M., Böttcher, R., & Förstemann, K. (2012). A small RNA response at DNA ends in Drosophila. *Nucleic Acids Research*, *40*(19), 9596–9603. https://doi.org/10.1093/nar/gks711

Michelena, J., Pellegrino, S., Spegg, V., & Altmeyer, M. (2021). Replicated chromatin curtails 53BP1 recruitment in BRCA1-proficient and BRCA1-deficient cells. *Life Science Alliance*, *4*(6), e202101023. https://doi.org/10.26508/lsa.202101023

Michelini, F., Jalihal, A. P., Francia, S., Meers, C., Neeb, Z. T., Rossiello, F., Gioia, U., Aguado, J., Jones-Weinert, C., Luke, B., Biamonti, G., Nowacki, M., Storici, F., Carninci, P., Walter, N. G., & Fagagna, F. D. A. Di. (2018). From "cellular" RNA to "smart" RNA: Multiple Roles of RNA in Genome Stability and beyond. In *Chemical Reviews* (Vol. 118, Issue 8, pp. 4365–4403). American Chemical Society. https://doi.org/10.1021/acs.chemrev.7b00487

Michelini, F., Pitchiaya, S., Vitelli, V., Sharma, S., Gioia, U., Pessina, F., Cabrini, M., Wang, Y., Capozzo, I., Iannelli, F., Matti, V., Francia, S., Shivashankar, G. V., Walter, N. G., & Di Fagagna, F. D. A. (2017). Damage-induced lncRNAs control the DNA damage response through interaction with DDRNAs at individual double-strand breaks. *Nature Cell Biology*, *19*(12), 1400–1411. https://doi.org/10.1038/ncb3643

Miki, D., Zhu, P., Zhang, W., Mao, Y., Feng, Z., Huang, H., Zhang, H., Li, Y., Liu, R., Zhang, H., Qi, Y., & Zhu, J.-K. (2017). Efficient Generation of diRNAs Requires Components in the Posttranscriptional Gene Silencing Pathway. *Scientific Reports*, *7*(1), 301. https://doi.org/10.1038/s41598-017-00374-7

Mimori, T., Akizuki, M., Yamagata, H., Inada, S., Yoshida, S., & Homma, M. (1981). Characterization of a high molecular weight acidic nuclear protein recognized by autoantibodies in sera from patients with polymyositis-scleroderma overlap. *Journal of Clinical Investigation*, *68*(3), 611–620. https://doi.org/10.1172/JCI110295

Mimori, T., & Hardin, J. A. (1986). Mechanism of interaction between Ku protein and DNA. *Journal of Biological Chemistry*, *261*(22), 10375–10379. https://doi.org/10.1016/S0021-9258(18)67534-9

Modesti, M., Junop, M. S., Ghirlando, R., van de Rakt, M., Gellert, M., Yang, W., & Kanaar, R. (2003). Tetramerization and DNA Ligase IV Interaction of the DNA Double-strand Break Repair Protein XRCC4 are Mutually Exclusive. *Journal of Molecular Biology*, *334*(2), 215–228. https://doi.org/10.1016/j.jmb.2003.09.031

Moshous, D., Callebaut, I., de Chasseval, R., Corneo, B., Cavazzana-Calvo, M., Le Deist, F., Tezcan, I., Sanal, O., Bertrand, Y., Philippe, N., Fischer, A., & de Villartay, J.-P. (2001). Artemis, a Novel DNA Double-Strand Break Repair/V(D)J Recombination Protein, Is Mutated in Human Severe Combined Immune Deficiency. *Cell*, *105*(2), 177–186. https://doi.org/10.1016/S0092-8674(01)00309-9

Moshous, D., Pannetier, C., Chasseval, R. de, Deist, F. le, Cavazzana-Calvo, M., Romana, S., Macintyre, E., Canioni, D., Brousse, N., Fischer, A., Casanova, J.-L., & Villartay, J.-P. de. (2003). Partial T and B lymphocyte immunodeficiency and predisposition to lymphoma in patients with hypomorphic mutations in Artemis. *Journal of Clinical Investigation*, *111*(3), 381–387. https://doi.org/10.1172/JCI16774

Moskwa, P., Buffa, F. M., Pan, Y., Panchakshari, R., Gottipati, P., Muschel, R. J., Beech, J., Kulshrestha, R., Abdelmohsen, K., Weinstock, D. M., Gorospe, M., Harris, A. L., Helleday, T., & Chowdhury, D. (2011). miR-182-Mediated Downregulation of BRCA1 Impacts DNA Repair and Sensitivity to PARP Inhibitors. *Molecular Cell*, *41*(2), 210–220. https://doi.org/10.1016/j.molcel.2010.12.005

Nick McElhinny, S. A., Havener, J. M., Garcia-Diaz, M., Juárez, R., Bebenek, K., Kee, B. L., Blanco, L., Kunkel, T. A., & Ramsden, D. A. (2005). A Gradient of Template Dependence Defines Distinct Biological Roles for Family X Polymerases in

Nonhomologous End Joining. *Molecular Cell*, *19*(3), 357–366. https://doi.org/10.1016/j.molcel.2005.06.012

Nick McElhinny, S. A., Snowden, C. M., McCarville, J., & Ramsden, D. A. (2000). Ku Recruits the XRCC4-Ligase IV Complex to DNA Ends. *Molecular and Cellular Biology*, *20*(9), 2996–3003. https://doi.org/10.1128/MCB.20.9.2996-3003.2000

Niewolik, D., Pannicke, U., Lu, H., Ma, Y., Wang, L.-C. V., Kulesza, P., Zandi, E., Lieber, M. R., & Schwarz, K. (2006). DNA-PKcs Dependence of Artemis Endonucleolytic Activity, Differences between Hairpins and 5' or 3' Overhangs. *Journal of Biological Chemistry*, *281*(45), 33900–33909. https://doi.org/10.1074/jbc.M606023200

Noordermeer, S. M., Adam, S., Setiaputra, D., Barazas, M., Pettitt, S. J., Ling, A. K., Olivieri, M., Álvarez-Quilón, A., Moatti, N., Zimmermann, M., Annunziato, S., Krastev, D. B., Song, F., Brandsma, I., Frankum, J., Brough, R., Sherker, A., Landry, S., Szilard, R. K., … Durocher, D. (2018). The shieldin complex mediates 53BP1-dependent DNA repair. *Nature*, *560*(7716), 117–121. https://doi.org/10.1038/s41586-018-0340-7

North, P., Ganesh, A., & Thacker, J. (1990). The rejoining of double-strand breaks in DNA by human cell extracts. *Nucleic Acids Research*, *18*(21), 6205–6210. https://doi.org/10.1093/nar/18.21.6205

Nussenzweig, A., Chen, C., da Costa Soares, V., Sanchez, M., Sokol, K., Nussenzweig, M. C., & Li, G. C. (1996). Requirement for Ku80 in growth and immunoglobulin V(D)J recombination. *Nature*, *382*(6591), 551–555. https://doi.org/10.1038/382551a0

Ochi, T., Blackford, A. N., Coates, J., Jhujh, S., Mehmood, S., Tamura, N., Travers, J., Wu, Q., Draviam, V. M., Robinson, C. V., Blundell, T. L., & Jackson, S. P. (2015). PAXX, a paralog of XRCC4 and XLF, interacts with Ku to promote DNA double-strand break repair. *Science*, *347*(6218), 185–188. https://doi.org/10.1126/science.1261971

O'Driscoll, M., Cerosaletti, K. M., Girard, P.-M., Dai, Y., Stumm, M., Kysela, B., Hirsch, B., Gennery, A., Palmer, S. E., Seidel, J., Gatti, R. A., Varon, R., Oettinger, M. A., Neitzel, H., Jeggo, P. A., & Concannon, P. (2001). DNA Ligase IV Mutations Identified in Patients Exhibiting Developmental Delay and Immunodeficiency. *Molecular Cell*, *8*(6), 1175–1185. https://doi.org/10.1016/S1097-2765(01)00408-7

Pannunzio, N. R., Li, S., Watanabe, G., & Lieber, M. R. (2014). Non-homologous end joining often uses microhomology: Implications for alternative end joining. *DNA Repair*, *17*, 74–80. https://doi.org/10.1016/j.dnarep.2014.02.006

Papillon, J. P. N., Nakajima, K., Adair, C. D., Hempel, J., Jouk, A. O., Karki, R. G., Mathieu, S., Möbitz, H., Ntaganda, R., Smith, T., Visser, M., Hill, S. E., Hurtado, F. K., Chenail, G., Bhang, H.-E. C., Bric, A., Xiang, K., Bushold, G., Gilbert, T., … Jagani, Z. (2018). Discovery of Orally Active Inhibitors of Brahma Homolog (BRM)/SMARCA2 ATPase Activity for the Treatment of Brahma Related Gene 1 (BRG1)/SMARCA4-Mutant Cancers. *Journal of Medicinal Chemistry*, *61*(22), 10155–10172. https://doi.org/10.1021/acs.jmedchem.8b01318

Perederina, A., Berezin, I., & Krasilnikov, A. S. (2018). In vitro reconstitution and analysis of eukaryotic RNase P RNPs. *Nucleic Acids Research*, *46*(13), 6857–6868. https://doi.org/10.1093/nar/gky333

Pfeiffer, P., & Vielmetter, W. (1988). Joining of nonhomologous DNA double strand breaks *in vitro*. *Nucleic Acids Research*, *16*(3), 907–924. https://doi.org/10.1093/nar/16.3.907

Pommier, Y., Nussenzweig, A., Takeda, S., & Austin, C. (2022). Human topoisomerases and their roles in genome stability and organization. *Nature Reviews Molecular Cell Biology*, *23*(6), 407–427. https://doi.org/10.1038/s41580-022-00452-3

Povirk, L. F. (2012). Processing of Damaged DNA Ends for Double-Strand Break Repair in Mammalian Cells. *ISRN Molecular Biology*, *2012*, 1–16. https://doi.org/10.5402/2012/345805

Pryor, J. M., Waters, C. A., Aza, A., Asagoshi, K., Strom, C., Mieczkowski, P. A., Blanco, L., & Ramsden, D. A. (2015). Essential role for polymerase specialization in cellular nonhomologous end joining. *Proceedings of the National Academy of Sciences*, *112*(33). https://doi.org/10.1073/pnas.1505805112

Reddy, Y. V. R., Ding, Q., Lees-Miller, S. P., Meek, K., & Ramsden, D. A. (2004). Non-homologous end joining requires that the DNA-PK complex undergo an autophosphorylation-dependent rearrangement at DNA ends. *The Journal of Biological Chemistry*, *279*(38), 39408–39413. https://doi.org/10.1074/jbc.M406432200

Ribes-Zamora, A., Indiviglio, S. M., Mihalek, I., Williams, C. L., & Bertuch, A. A. (2013). TRF2 Interaction with Ku Heterotetramerization Interface Gives Insight into c-NHEJ Prevention at Human Telomeres. *Cell Reports*, *5*(1), 194–206. https://doi.org/10.1016/j.celrep.2013.08.040

Rinaldi, C., Pizzul, P., Casari, E., Mangiagalli, M., Tisi, R., & Longhese, M. P. (2023). The Ku complex promotes DNA end-bridging and this function is antagonized by Tel1/ATM kinase. *Nucleic Acids Research*, *51*(4), 1783–1802. https://doi.org/10.1093/nar/gkad062

Roberts, S. A., Strande, N., Burkhalter, M. D., Strom, C., Havener, J. M., Hasty, P., & Ramsden, D. A. (2010). Ku is a 5'-dRP/AP lyase that excises nucleotide damage near broken ends. *Nature*, *464*(7292), 1214–1217. https://doi.org/10.1038/nature08926

Ropars, V., Drevet, P., Legrand, P., Baconnais, S., Amram, J., Faure, G., Márquez, J. A., Piétrement, O., Guerois, R., Callebaut, I., Le Cam, E., Revy, P., de Villartay, J.-P., & Charbonnier, J.-B. (2011). Structural characterization of filaments formed by human Xrcc4–Cernunnos/XLF complex involved in nonhomologous DNA end-joining. *Proceedings of the National Academy of Sciences*, *108*(31), 12663–12668. https://doi.org/10.1073/pnas.1100758108

Salunkhe, S., Daley, J. M., Kaur, H., Tomimatsu, N., Xue, C., Raina, V. B., Jasper, A. M., Rogers, C. M., Li, W., Zhou, S., Mojidra, R., Kwon, Y., Fang, Q., Ji, J.-H., Badamchi Shabestari, A., Fitzgerald, O., Dinh, H., Mukherjee, B., Habib, A. A., … Sung, P. (2024). Promotion of DNA end resection by BRCA1–BARD1 in homologous recombination. *Nature*, *634*(8033), 482–491. https://doi.org/10.1038/s41586-024-07910-2

Schar, P. (2004). SMC1 coordinates DNA double-strand break repair pathways. *Nucleic Acids Research*, *32*(13), 3921–3929. https://doi.org/10.1093/nar/gkh716

Schatz, D. G., & Ji, Y. (2011). Recombination centres and the orchestration of V(D)J recombination. *Nature Reviews Immunology*, *11*(4), 251–263. https://doi.org/10.1038/nri2941

Schenborn, E. T., & Mierendorf, R. C. (1985). A novel transcription property of SP6 and T7 RNA polymerases: dependence on template structure. *Nucleic Acids Research*, *13*(17), 6223–6236. https://doi.org/10.1093/nar/13.17.6223

Schimmel, J., Kool, H., van Schendel, R., & Tijsterman, M. (2017). Mutational signatures of non-homologous and polymerase theta-mediated end-joining in embryonic stem cells. *The EMBO Journal*, *36*(24), 3634–3649. https://doi.org/10.15252/embj.201796948

Schipler, A., & Iliakis, G. (2013). DNA double-strand-break complexity levels and their possible contributions to the probability for error-prone processing and repair pathway choice. *Nucleic Acids Research*, *41*(16), 7589–7605. https://doi.org/10.1093/nar/gkt556

Scully, R., Panday, A., Elango, R., & Willis, N. A. (2019). DNA double-strand break repair-pathway choice in somatic mammalian cells. *Nature Reviews Molecular Cell Biology*. https://doi.org/10.1038/s41580-019-0152-0

Shen, Y., Nandi, P., Taylor, M. B., Stuckey, S., Bhadsavle, H. P., Weiss, B., & Storici, F. (2011). RNA-driven genetic changes in bacteria and in human cells. *Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis*, *717*(1–2), 91–98. https://doi.org/10.1016/j.mrfmmm.2011.03.016

Shi, K., Bohl, T. E., Park, J., Zasada, A., Malik, S., Banerjee, S., Tran, V., Li, N., Yin, Z., Kurniawan, F., Orellana, K., & Aihara, H. (2018). T4 DNA ligase structure reveals a prototypical ATP-dependent ligase with a unique mode of sliding clamp interaction. *Nucleic Acids Research*, *46*(19), 10474–10488. https://doi.org/10.1093/nar/gky776

Sibanda, B. L., Chirgadze, D. Y., Ascher, D. B., & Blundell, T. L. (2017). DNA-PKcs structure suggests an allosteric mechanism modulating DNA double-strand break repair. *Science*, *355*(6324), 520–524. https://doi.org/10.1126/science.aak9654

Sibanda, B. L., Chirgadze, D. Y., & Blundell, T. L. (2010). Crystal structure of DNA-PKcs reveals a large open-ring cradle comprised of HEAT repeats. *Nature*, *463*(7277), 118–121. https://doi.org/10.1038/nature08648

Slavoff, S. A., Heo, J., Budnik, B. A., Hanakahi, L. A., & Saghatelian, A. (2014). A Human Short Open Reading Frame (sORF)-encoded Polypeptide That Stimulates DNA End Joining. *Journal of Biological Chemistry*, *289*(16), 10950–10957. https://doi.org/10.1074/jbc.C113.533968

Soubeyrand, S., Pope, L., Pakuts, B., & Haché, R. J. G. (2003). Threonines 2638/2647 in DNA-PK are essential for cellular resistance to ionizing radiation. *Cancer Research*, *63*(6), 1198–1201.

Stavnezer, J., & Schrader, C. E. (2014). IgH chain class switch recombination: mechanism and regulation. *Journal of Immunology (Baltimore, Md. : 1950)*, *193*(11), 5370–5378. https://doi.org/10.4049/jimmunol.1401849

Stein, C. A., Subasinghe, C., Shinozuka, K., & Cohen, J. S. (1988). Physicochemical properties of phospborothioate oligodeoxynucleotides. *Nucleic Acids Research*, *16*(8), 3209–3221. https://doi.org/10.1093/nar/16.8.3209

Stiff, T., O'Driscoll, M., Rief, N., Iwabuchi, K., Löbrich, M., & Jeggo, P. A. (2004). ATM and DNA-PK Function Redundantly to Phosphorylate H2AX after Exposure to Ionizing Radiation. *Cancer Research*, *64*(7), 2390–2396. https://doi.org/10.1158/0008-5472.CAN-03-3207

Stinson, B. M., & Loparo, J. J. (2021). Repair of DNA Double-Strand Breaks by the Nonhomologous End Joining Pathway. *Annual Review of Biochemistry*, *90*(1), 137–164. https://doi.org/10.1146/annurev-biochem-080320-110356

Stinson, B. M., Moreno, A. T., Walter, J. C., & Loparo, J. J. (2020). A Mechanism to Minimize Errors during Non-homologous End Joining. *Molecular Cell*, *77*(5), 1080-1091.e8. https://doi.org/10.1016/j.molcel.2019.11.018

Storici, F., Bebenek, K., Kunkel, T. A., Gordenin, D. A., & Resnick, M. A. (2007). RNA-templated DNA repair. *Nature*, *447*(7142), 338–341. https://doi.org/10.1038/nature05720

Stucki, M., & Jackson, S. P. (2006). γH2AX and MDC1: Anchoring the DNA-damage-response machinery to broken chromosomes. *DNA Repair*, *5*(5), 534–543. https://doi.org/10.1016/j.dnarep.2006.01.012

Subramanian, K., Rutvisuttinunt, W., Scott, W., & Myers, R. S. (2003). The enzymatic basis of processivity in lambda exonuclease. *Nucleic Acids Research*, *31*(6), 1585–1596. https://doi.org/10.1093/nar/gkg266

Taccioli, G. E., Rathbun, G., Oltz, E., Stamato, T., Jeggo, P. A., & Alt, F. W. (1993). Impairment of V(D)J Recombination in Double-Strand Break Repair Mutants. *Science*, *260*(5105), 207–210. https://doi.org/10.1126/science.8469973

Takata, M., Sasaki, M. S., Sonoda, E., Morrison, C., Hashimoto, M., Utsumi, H., Yamaguchi-Iwai, Y., Shinohara, A., & Takeda, S. (1998). Homologous recombination and non-homologous end-joining pathways of DNA double-strand break repair have overlapping roles in the maintenance of chromosomal integrity in vertebrate cells. *The EMBO Journal*, *17*(18), 5497–5508. https://doi.org/10.1093/emboj/17.18.5497

Tang, J., Cho, N. W., Cui, G., Manion, E. M., Shanbhag, N. M., Botuyan, M. V., Mer, G., & Greenberg, R. A. (2013). Acetylation limits 53BP1 association with damaged chromatin to promote homologous recombination. *Nature Structural & Molecular Biology*, *20*(3), 317–325. https://doi.org/10.1038/nsmb.2499

Thomas-Claudepierre, A.-S., Schiavo, E., Heyer, V., Fournier, M., Page, A., Robert, I., & Reina-San-Martin, B. (2013). The cohesin complex regulates immunoglobulin class switch recombination. *Journal of Experimental Medicine*, *210*(12), 2495–2502. https://doi.org/10.1084/jem.20130166

Tsai, C. J., Kim, S. A., & Chu, G. (2007). Cernunnos/XLF promotes the ligation of mismatched and noncohesive DNA ends. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(19), 7851–7856. https://doi.org/10.1073/pnas.0702620104

Tubbs, A., & Nussenzweig, A. (2017). Endogenous DNA Damage as a Source of Genomic Instability in Cancer. *Cell*, *168*(4), 644–656. https://doi.org/10.1016/j.cell.2017.01.002

Uchiyama, Y., Takeuchi, R., Kodera, H., & Sakaguchi, K. (2009). Distribution and roles of X-family DNA polymerases in eukaryotes. *Biochimie*, *91*(2), 165–170. https://doi.org/10.1016/j.biochi.2008.07.005

van den Bosch, M., Bree, R. T., & Lowndes, N. F. (2003). The MRN complex: coordinating and mediating the response to broken chromosomes. *EMBO Reports*, *4*(9), 844–849. https://doi.org/10.1038/sj.embor.embor925

van der Burg, M., IJspeert, H., Verkaik, N. S., Turul, T., Wiegant, W. W., Morotomi-Yano, K., Mari, P.-O., Tezcan, I., Chen, D. J., Zdzienicka, M. Z., van Dongen, J. J. M., & van Gent, D. C. (2008). A DNA-PKcs mutation in a radiosensitive T–B– SCID patient inhibits Artemis activation and nonhomologous end-joining. *Journal of Clinical Investigation*. https://doi.org/10.1172/JCI37141

van der Burg, M., van Veelen, L. R., Verkaik, N. S., Wiegant, W. W., Hartwig, N. G., Barendregt, B. H., Brugmans, L., Raams, A., Jaspers, N. G. J., Zdzienicka, M. Z., van Dongen, J. J. M., & van Gent, D. C. (2006). A new type of radiosensitive T-B-NK+ severe combined immunodeficiency caused by a LIG4 mutation. *The Journal of Clinical Investigation*, *116*(1), 137–145. https://doi.org/10.1172/JCI26121

Vogt, A., & He, Y. (2023). Structure and mechanism in non-homologous end joining. *DNA Repair*, *130*. https://doi.org/10.1016/j.dnarep.2023.103547

Volk, T., Pannicke, U., Reisli, I., Bulashevska, A., Ritter, J., Björkman, A., Schäffer, A. A., Fliegauf, M., Sayar, E. H., Salzer, U., Fisch, P., Pfeifer, D., Di Virgilio, M., Cao, H., Yang, F., Zimmermann, K., Keles, S., Caliskaner, Z., Güner, S., … Grimbacher, B. (2015). *DCLRE1C* (ARTEMIS) mutations causing phenotypes ranging from atypical severe combined immunodeficiency to mere antibody deficiency. *Human Molecular Genetics*, *24*(25), 7361–7372. https://doi.org/10.1093/hmg/ddv437

Wakasugi, M., Sasaki, T., Matsumoto, M., Nagaoka, M., Inoue, K., Inobe, M., Horibata, K., Tanaka, K., & Matsunaga, T. (2014). Nucleotide Excision Repair-dependent DNA Double-strand Break Formation and ATM Signaling Activation in Mammalian Quiescent Cells. *Journal of Biological Chemistry*, *289*(41), 28730–28737. https://doi.org/10.1074/jbc.M114.589747

Walker, A. I., Hunt, T., Jackson, R. J., & Anderson, C. W. (1985). Double-stranded DNA induces the phosphorylation of several proteins including the 90 000 mol. wt. heat-shock protein in animal cell extracts. *The EMBO Journal*, *4*(1), 139–145. https://doi.org/10.1002/j.1460-2075.1985.tb02328.x

Walker, J. R., Corpina, R. A., & Goldberg, J. (2001). Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair. *Nature*, *412*(6847), 607–614. https://doi.org/10.1038/35088000

Wang, Q., & Goldstein, M. (2016). Small RNAs Recruit Chromatin-Modifying Enzymes MMSET and Tip60 to Reconfigure Damaged DNA upon Double-Strand Break and Facilitate Repair. *Cancer Research*, *76*(7), 1904–1915. https://doi.org/10.1158/0008-5472.CAN-15-2334

Wang, Y., Huang, J.-W., Calses, P., Kemp, C. J., & Taniguchi, T. (2012). MiR-96 Downregulates REV1 and RAD51 to Promote Cellular Sensitivity to Cisplatin and PARP Inhibition. *Cancer Research*, *72*(16), 4037–4046. https://doi.org/10.1158/0008-5472.CAN-12-0103

Waters, C. A., Strande, N. T., Pryor, J. M., Strom, C. N., Mieczkowski, P., Burkhalter, M. D., Oh, S., Qaqish, B. F., Moore, D. T., Hendrickson, E. A., & Ramsden, D. A. (2014). The fidelity of the ligation step determines how ends are resolved during nonhomologous end joining. *Nature Communications*, *5*(1), 4286. https://doi.org/10.1038/ncomms5286

Wei, W., Ba, Z., Gao, M., Wu, Y., Ma, Y., Amiard, S., White, C. I., Danielsen, J. M. R., Yang, Y. G., & Qi, Y. (2012). A role for small RNAs in DNA double-strand break repair. *Cell*, *149*(1), 101–112. https://doi.org/10.1016/j.cell.2012.03.002

Wingfield, P. (1998). Protein Precipitation Using Ammonium Sulfate. *Current Protocols in Protein Science*, *13*(1). https://doi.org/10.1002/0471140864.psa03fs13

Woodbine, L., Neal, J. A., Sasi, N.-K., Shimada, M., Deem, K., Coleman, H., Dobyns, W. B., Ogi, T., Meek, K., Davies, E. G., & Jeggo, P. A. (2013). PRKDC mutations in a SCID patient with profound neurological abnormalities. *Journal of Clinical Investigation*, *123*(7), 2969–2980. https://doi.org/10.1172/JCI67349

Wu, P.-Y., Frit, P., Malivert, L., Revy, P., Biard, D., Salles, B., & Calsou, P. (2007). Interplay between Cernunnos-XLF and Nonhomologous End-joining Proteins at DNA Ends in the Cell. *Journal of Biological Chemistry*, *282*(44), 31937–31943. https://doi.org/10.1074/jbc.M704554200

Wu, P.-Y., Frit, P., Meesala, S., Dauvillier, S., Modesti, M., Andres, S. N., Huang, Y., Sekiguchi, J., Calsou, P., Salles, B., & Junop, M. S. (2009). Structural and Functional Interaction between the Human DNA Repair Proteins DNA Ligase IV and XRCC4. *Molecular and Cellular Biology*, *29*(11), 3163–3172. https://doi.org/10.1128/MCB.01895-08

Xing, M., Yang, M., Huo, W., Feng, F., Wei, L., Jiang, W., Ning, S., Yan, Z., Li, W., Wang, Q., Hou, M., Dong, C., Guo, R., Gao, G., Ji, J., Zha, S., Lan, L., Liang, H., & Xu, D. (2015). Interactome analysis identifies a new paralogue of XRCC4 in non-homologous end joining DNA repair pathway. *Nature Communications*, *6*(1), 6233. https://doi.org/10.1038/ncomms7233

Yan, D., Ng, W. L., Zhang, X., Wang, P., Zhang, Z., Mo, Y.-Y., Mao, H., Hao, C., Olson, J. J., Curran, W. J., & Wang, Y. (2010). Targeting DNA-PKcs and ATM with miR-101 Sensitizes Tumors to Radiation. *PLoS ONE*, *5*(7), e11397. https://doi.org/10.1371/journal.pone.0011397

Yang, L., Yang, B., Wang, Y., Liu, T., He, Z., Zhao, H., Xie, L., & Mu, H. (2019). The CTIP-mediated repair of TNF-α-induced DNA double-strand break was impaired by miR-130b in cervical cancer cell. *Cell Biochemistry and Function*, *37*(7), 534–544. https://doi.org/10.1002/cbf.3430

Yang, S. W., Burgin, A. B., Huizenga, B. N., Robertson, C. A., Yao, K. C., & Nash, H. A. (1996). A eukaryotic enzyme that can disjoin dead-end covalent complexes between DNA and type I topoisomerases. *Proceedings of the National Academy of Sciences*, *93*(21), 11534–11539. https://doi.org/10.1073/pnas.93.21.11534

Yang, X., & Ma, C. (2016). In Vitro Transcription Assays and Their Application in Drug Discovery. *Journal of Visualized Experiments : JoVE*, *115*. https://doi.org/10.3791/54256

Yano, K., Morotomi-Yano, K., Wang, S., Uematsu, N., Lee, K., Asaithamby, A., Weterings, E., & Chen, D. J. (2008). Ku recruits XLF to DNA double-strand breaks. *EMBO Reports*, *9*(1), 91–96. https://doi.org/10.1038/sj.embor.7401137

Yoo, S., & Dynan, W. S. (1999). Geometry of a complex formed by double strand break repair proteins at a single DNA end: recruitment of DNA-PKcs induces inward translocation of Ku protein. *Nucleic Acids Research*, *27*(24), 4679–4686. https://doi.org/10.1093/nar/27.24.4679

Yu, G., Wang, L.-G., Han, Y., & He, Q.-Y. (2012). clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS: A Journal of Integrative Biology*, *16*(5), 284–287. https://doi.org/10.1089/omi.2011.0118

Yu, Y., Mahaney, B. L., Yano, K.-I., Ye, R., Fang, S., Douglas, P., Chen, D. J., & Lees-Miller, S. P. (2008). DNA-PK and ATM phosphorylation sites in XLF/Cernunnos are not required for repair of DNA double strand breaks. *DNA Repair*, *7*(10), 1680–1692. https://doi.org/10.1016/j.dnarep.2008.06.015

Yu, Y., Wang, W., Ding, Q., Ye, R., Chen, D., Merkle, D., Schriemer, D., Meek, K., & Lees-Miller, S. P. (2003). DNA-PK phosphorylation sites in XRCC4 are not required for survival after radiation or for V(D)J recombination. *DNA Repair*, *2*(11), 1239–1252. https://doi.org/10.1016/S1568-7864(03)00143-5

Yuan, J., Chen, K., Zhang, W., & Chen, Z. (2022). Structure of human chromatin-remodelling PBAF complex bound to a nucleosome. *Nature*, *605*(7908), 166–171. https://doi.org/10.1038/s41586-022-04658-5

Yue, X., Bai, C., Xie, D., Ma, T., & Zhou, P.-K. (2020). DNA-PKcs: A Multi-Faceted Player in DNA Damage Response. *Frontiers in Genetics*, *11*. https://doi.org/10.3389/fgene.2020.607428

Zhang, Y., He, Q., Hu, Z., Feng, Y., Fan, L., Tang, Z., Yuan, J., Shan, W., Li, C., Hu, X., Tanyi, J. L., Fan, Y., Huang, Q., Montone, K., Dang, C. V, & Zhang, L. (2016). Long noncoding RNA LINP1 regulates repair of DNA double-strand breaks in triple-negative breast cancer. *Nature Structural & Molecular Biology*, *23*(6), 522–530. https://doi.org/10.1038/nsmb.3211

Zhao, B., Rothenberg, E., Ramsden, D. A., & Lieber, M. R. (2020). The molecular basis and disease relevance of non-homologous DNA end joining. In *Nature Reviews Molecular Cell Biology* (Vol. 21, Issue 12, pp. 765–781). Nature Research. https://doi.org/10.1038/s41580-020-00297-8

Zhao, B., Watanabe, G., Morten, M. J., Reid, D. A., Rothenberg, E., & Lieber, M. R. (2019). The essential elements for the noncovalent association of two DNA ends during NHEJ synapsis. *Nature Communications*, *10*(1), 3588. https://doi.org/10.1038/s41467-019-11507-z

Zhao, Y., Thomas, H. D., Batey, M. A., Cowell, I. G., Richardson, C. J., Griffin, R. J., Calvert, A. H., Newell, D. R., Smith, G. C. M., & Curtin, N. J. (2006). Preclinical Evaluation of a Potent Novel DNA-Dependent Protein Kinase Inhibitor NU7441. *Cancer Research*, *66*(10), 5354–5362. https://doi.org/10.1158/0008-5472.CAN-05-4275

Zhu, C., Bogue, M. A., Lim, D.-S., Hasty, P., & Roth, D. B. (1996). Ku86-Deficient Mice Exhibit Severe Combined Immunodeficiency and Defective Processing of V(D)J Recombination Intermediates. *Cell*, *86*(3), 379–389. https://doi.org/10.1016/S0092-8674(00)80111-7

Zou, L., & Elledge, S. J. (2003). Sensing DNA Damage Through ATRIP Recognition of RPA-ssDNA Complexes. *Science*, *300*(5625), 1542–1548. https://doi.org/10.1126/science.1083430