# XRCC4 AND XLF IN END BRIDGING REPAIR OF DNA DOUBLE-STRAND BREAK REPAIR

# CHARACTERIZATION OF THE END BRIDGING COMPLEX OF NON-HOMOLOGOUS END JOINING REPAIR OF DNA DOUBLE STRAND BREAKS

CHRISTOPHER M. BROWN, B.SC (HON)

A Thesis

Submitted to the School of Graduate

Studies in Partial Fulfillment of the

Requirements for the Degree

Doctor of Philosophy

**Descriptive Note**

DOCTOR OF PHILOSOPHY (2018) MCMASTER UNIVERSITY
(Biochemistry and Biomedical Sciences) Hamilton, Ontario

Title: Characterization of the end bridging complex of Non-Homologous end joining repair of DNA double strand breaks

Author: Christopher M. Brown, B.Sc. (McMaster University)

Supervisor: Dr. Murray S. Junop

Number of pages: XIX, 164

**Abstract**

DNA double strand breaks represent the single most dangerous type of damage that can afflict the genome. Given the severity of such a lesion, higher eukaryotes possess two distinct pathways to repair such damages.

The work presented here focuses on the role of different protein complexes formed during within Non-Homologous End Joining (NHEJ). Specifically, how the C-terminal tails of XRCC4 and XLF regulate higher-order complex formation of end bridging filaments prior to terminal ligation and release of the intact DNA following repair. The crystal structure of full length XRCC4 was solved to 3.43Å and confirmed that the C-terminal tails or XRCC4 mediate tetramerization but are not required for end bridging of DNA ends. A cluster of residues that stabilized the XRCC4 multi helix bundle were mutated and determined to result in an XRCC4Mutational analysis and SEC-MALS further revealed that this 4-helix bundle stabilizes tetramers, interestingly tetramerization was found to not be required for bridging of DNA ends.

Additional work aimed at determining the mechanism by which XLF binds DNA and how complex filaments are formed was carried out using a combination of structural and biochemical techniques. Mutational analysis of the yeast XLF homologue, Nej1, revealed that the tails of these proteins bind their DNA substrates through an extended interface that may involve wrapping DNA to further stabilize interaction. Also, it was determined that phosphorylation of key residues within this extended DNA binding domain results in a decreased affinity

for DNA and may play a role in DNA repair pathway choice in *Saccharomyces cerevisiae.*

Transmission electron microscopy showed that when bound to DNA, XLF is capable of forming DNA dependent filaments that are capable of bridging DNA ends in a linear manner. Addition of XRCC4 resulted in extensive remodelling of these filaments. Crystals of the XRCC4/XLF/DNA were optimized to diffract to a resolution of greater than 5Å, however further work will be required to determine the structure of this key NHEJ complex.

Finally, attempts to determine the optimal combination of DNA substrates and NHEJ factors to crystallize the terminal repairosome were carried out and initial hit conditions have been identified.

**Acknowledgements**

First and foremost, I wish to thank my supervisor Dr. Murray Junop for his years of support and optimism which drove me to continue. I cannot thank him enough for his encouragement and suggestion to transfer into the Ph.D. program, a decision I have not regretted. In addition to the constant discourse about our science, the opportunities he provided me to teach and serve in a mentorship role were invaluable and I will always be grateful to him for introducing me to the world of being a teacher.

I would also like to thank my committee members Dr. Cecile Fradin and Dr. Giuseppe Melacini for their years of support and fresh perspectives which were vital to moving this project forward.

Every person who has been a part of the Junop lab has helped me immensely, even if it were something as being a sounding board for another crazy idea or teaching me something new. For this I offer my profound thanks, in particular, I would like to express my profound thanks and gratitude to the undergraduate students I was lucky enough to work with on various projects. Brendon McDowell, Alex Vlahos and Chirayu Chokshi for their work on the XRCC4/XLF project, Teagan Telesnicki, Khalid Hossain and Rob Nunn for their work on yeast repair of double strand breaks and Omar Musleh for getting crystals of a complex I never thought we would see. I would be remiss if I didn't specifically thank Joey Laffradi and Beverlee Buzon, two people who started out as colleagues, but have grown into something more than just friends. I would not have been able to finish this thesis were it not for their help and support for so many years.

To my friends and family beyond the lab, your encouragement and reassurances that I could do anything was a constant source of comfort and a real pick me up when the experiments just weren't working. To my parents, Mark and Diane, thank you for more than words can express.

**Table of contents**

## List of Figures

**List of tables**

**Abbreviations**

Adenosine mono-phosphate (AMP)

Adenosine tri-phosphate (ATP)

Advanced photon source (APS)

Altnerative Non-homologous end joining (A-NHEJ)

Aprataxin (APTX)

Aprataxin and PNPK-like factor (APLF)

Argonne National Laboratory (ANL)

BRCA1 C-terminus domain (BRCT)

C-terminal unstructured region(CUR)

Circular Dichroism spectroscopy

Deoxyribonucleic acid (DNA)

DNA dependent protein kinase Catalytic subunit (DNApk CS)

DNA ligase Iv (LigIV)

DNA ligase 4 (DNL4)

Double strand breaks (DSB)

Ethylenediaminetetraacetic acid (EDTA)

Electrophoretic mobility shift assay (EMSA)

Gray (Gy)

Homologous recombination repair (HRR)

Immobilized metal affinity chromatography (IMAC)

Isopropyl β-D-1-thiogalactopyranoside (IPTG)

Laurodimethylamine Oxide (LDAO)

Ligase-interacting factor 1 (Lif1)

MRE11/Rad50/NBS1 (MRN)

Mre11/Rad50/Xrs2 (MRX)

National Synchrotron Light Source (NSLS)

Non-homologous end-joining protein 1 (Nej1)

Non-Homologous end joining (NHEJ)

Nuclear localization sequence (NLS)

Optical Density 600 nM (OD600)

Paralog of XRCC4 and XLF (PAXX)

Phenylmethylsulfonyl fluoride (PMSF)

Poly-acrylamide gel electrophoresis (PAGE)

Poly-ethylene glycol (Peg)

Polynucleotide Kinase (PNK)

Recombination activating gene (Rag1)

Ribonucleic acid (RNA)

Severe combined immunodeficiency (SCID)

Single strand annealing (SSA)

Sodium dodecyl sulfate (SDS)

Size exclusion chromatography – multi angle light scattering (SecMals)

Spindle assembly abnormal homolog 6 (SAS6)

Transmission Electron Microscopy (TEM)

Variable, Diverse, Joining (VDJ)

X-ray repair cross complementing protein 4 (XRCC4)

XRCC4-like factor (XLF)

XRCC4-like Small (XLS)

# Chapter 1 Introduction

## 1.1 DNA double strand breaks

DNA double strand breaks represent one of the single most dangerous types of DNA damage and affect any living organism. In yeast cells, a single un-repaired break can be lethal to a cell and in higher eukaryotes, double strand breaks can lead to apoptosis or even chromosomal re-arrangements of genetic material (Borek et al., 1966; Cavazzana-Calvo, et al., 1993., Hei et al., 1988., Resnick et al., 1976). While not occurring with the same frequency as other types of DNA damage, their results can be far more deleterious to a cell then all other forms of DNA damage.

Double strand breaks are formed in response to a multitude of factors. If two single strand breaks occur within 10 base pairs of one another, DNA secondary structure is no longer sufficient to maintain the DNA duplex (Michael et al., 2000). Anther cause of damage includes ionizing radiation (X-rays, gamma rays) which creates damage through deposition of energy. A single gray (Gy) of ionizing radiation can induce the formation of up to 40 DSBs (Siddiqui et al., 1987). This energization can induce ionization of bases directly or can induce the formation of free radicals within the aqueous environment of the cell (Ward et al., 1973; Ward et al., 1976). These radicals are then able to further damage the DNA, compounding the scope of the damage incurred (Chance et al., 1979; Richter et al., 1988).

Despite being deleterious to cell viability, double strand breaks can be induced deliberately either through the use of chemotherapeutic agents or through

programmed cellular processes during immune cell development. Several chemotherapeutic agents elicit their damage by generating free-radical damage which disproportionately affects the more rapidly dividing tumor cells. Bleomycin is one such compound, which results in non-ligatable DNA ends containing 5'-phosphates and 3'-phosphoglycolate ends (Giloni et.al., 1981). In addition to these incompatible ends, roughly 10% of the damage caused by Bleomycin results in the formation of double strand breaks (Bennet et al., 1993). In addition to Bleomycin, there are many compounds that function by generating DSBs. Some of the more commonly used treatments include ionizing radiation and cross-linking agents

One of the leading causes of double strand breaks is through replication fork collapse during DNA synthesis. Un-repaired abasic sites, nicked single strands and gaps will result in replication fork collapse thereby creating a double strand break (Kuzminov 1995; Kuzminov, 2001). Virtually all other types of DNA damage, if left un-repaired prior to replication can result in the formation of a double strand break. While not as serious as double strand breaks, these lesions must be repaired in a timely manner or they will invariably lead to the formation of a double strand break.

In addition to exposure to mutagenic compounds, double strand breaks are also caused directly by several endogenous cellular processes. Oxidative metabolism can result in the formation of free radicals whose effects are the same as those formed by exposure to ionizing radiation (Michael et al., 2000, Rich, T.,

et al., 2000, Richter, C. et al., 1988). Furthermore, double strand breaks are also introduced so as to expand the genetic diversity of certain types of cells. Meiosis requires the formation of double strand breaks for gene crossovers and VDJ recombination is dependent on successful repair of breaks in immunoglobulin genes (Keeney et al., 1997; Oettinger et al., 1990; Sun et al., 1989). VDJ recombination is utterly essential for a fully functional adaptive immune system in higher eukaryotes (Moshous et al., 2000; Rooney et al., 2001). Genetic diversity generated by VDJ recombination requires the Non-homologous end joining (NHEJ) pathway to repair the double strand breaks introduced into immunoglobulin genes by the recombination activating gene (Rag1) and Rag2 proteins (McBlane et al., 1995; Schatz et al., 1989). Recombination results in the formation of two different types of DNA ends, signal joints and coding joints (Biederman et al., 1991; Gu et al., 1997; Ramsden et al., 1995). Following VDJ recombination, the resulting coding joints are capped by DNA hairpins that must be opened and ligated to the proximal joints so as to allow the newly randomized sequence to be expressed (Ma et al., 2002). No other repair pathway is capable of resolving these joints. (Early et al., 1980; Fukumura et al., 1998; Hendrickson et al., 1991; Schilling et al., 1980; Taccioli et al., 1994). Deficiencies in any of the enzymes or proteins involved in VDJ recombination, if not embryonic lethal, are associated with a severe combined immunodeficiency (SCID) phenotype (Biderman, K et al., 1998). The combination of exogenous insults and

endogenous, pre-planned breaks necessitates one, if not several different ways by which these lesions are detected, treated and ultimately repaired.

### 1.2 Double strand break response

Despite the many ways and means by which DSB can be formed, their detection and repair is essential. Detection of DSBs leads to a signalling cascade which primes the cell to respond to the damage. There are three potential outcomes to a double strand break: 1: cellular apoptosis when damage is unable to be repaired, 2: accurate and effective repair wherein the lesion is repaired correctly, and the cell can resume normal activities, and 3: erroneous repair resulting in chromosomal re-arrangements and a loss of genomic integrity (figure1.1). Of the three, accurate and effective repair is the normal and optimal outcome.

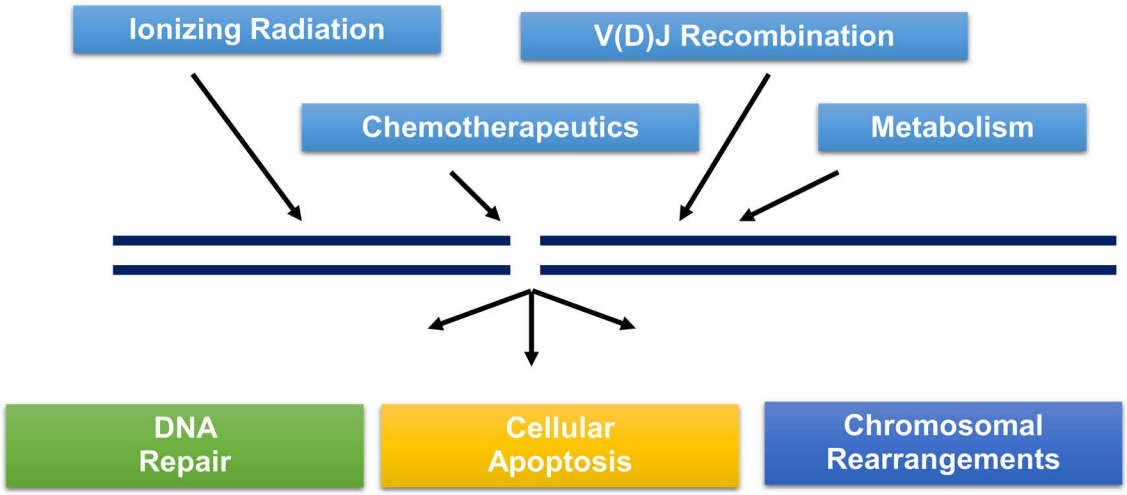



**Figure 1.1 Sources of double strand breaks and potential outcomes**. DNA double strand breaks arise from several different sources, both exogenous and endogenous, but ultimately will results in one of three outcomes: repair, apoptosis or chromosomal rearrangements.

The first step in repair is the detection of DNA damage. Following detection, the break must be repaired. Given the severity of the damage of lesions of this type, higher eukaryotes possess two distinct pathways by which they are repaired. The first is the homologous recombination repair pathway (HRR). This pathway is invoked when the MRE11, Rad50 and NBS1 (MRN) sensor complex arrives at the site of the break and initiates a signalling cascade that results in the repair complex being assembled. (Banin et al., 1998; Canman et al., 1998; Li et al., 2000; Matsuoka et al., 2000). As the name suggests, this pathway requires the presence of a sister chromatid to serve as a homology donor. Accordingly, this pathway is restricted to the late S and G2 phases of the cell cycle (Kadyk, 1992; Resnick, 1976). In addition to the requirement of an available template, this

repair pathway is very slow, often times requiring hours to resolve a single lesion. Despite the opportunity for high-fidelity repair, the limitations and time requirements make HRR the lesser used pathway for repair. The second pathway for DSB repair is the cell-cycle independent Non-homologous end joining. This pathway is rapid, cell cycle independent and capable of directly ligating two strands in the absence of any homology donor. The choice of repair pathway, while largely influenced by cell cycle, becomes committed when one of two events occurs: extensive end resection of the strands at the break site commit homologous recombination repair or binding and engagement of a committed DSB sensor and recruitment of other processing factors commits the cell to Non-homologous end joining. While NHEJ and its initiating proteins are expressed continually throughout the cell cycle, factors which promote end resection are upregulated during those phases of the cell cycle in which HRR is feasible. This provides another mechanism by which pathway choice can be influenced.

Of the two pathways, NHEJ is the dominant and persistent pathway used by higher eukaryotes, with more than 80% of DSBs being repaired by NHEJ (Beucher et al., 2009; Wilson et al., 1982). While the lack of a homologous template means this type of repair may be potentially mutagenic, the rapidity with which it can repair these breaks makes it a better alternative than waiting for a homology donor to be synthesized. A sub-pathway, alternative non-homologous end joining (A-NHEJ) is capable of functioning in the absence of the NHEJ core factors but is dependent on the presence of some sequence homology (between 1

and 4 bp of homology to effect repair) (Bennardo et al., 2008; Yan et al., 2007). Despite similarity in the names of the two pathways, the work presented here will focus on NHEJ and not A-NHEJ.

### 1.3 Non-homologous end joining

In eukaryotes, NHEJ is typically proposed to proceed as a linear pathway following a defined set of steps in a sequential order. More recently, however, due to further clarification of the core NHEJ proteins and identification of new factors, suggest that the repair pathway may be less linear and proceed as more of a dynamic interplay between repair factors. Depending on the complexity of the break, certain factors may be required while others are dispensable. To have effective repair *in vitro*, 7 proteins known as the core NHEJ repair factors are required. These include the KU70/80 heterodimer, the DNA-Dependent protein kinase catalytic subunit, Artemis, XRCC4, XLF and DNA ligaseIV. The NHEJ repair pathway can be broken down into 3 general steps. 1) DNA end detection and initiation of repair, 2) End processing and synthesis and 3) Ligation and strand release. As the work described herein deals exclusively with NHEJ, it will be discussed in detail. (Figure 1.2)

### 1.3.1 End detection and initiation of repair

Repair of double strand breaks by non-homologous end joining is initiated when a DNA break is detected by the Ku70/80 heterodimer (Blier et al., 1993; Griffith et al., 1992, Paillard, S and Strauss, F., 1991).

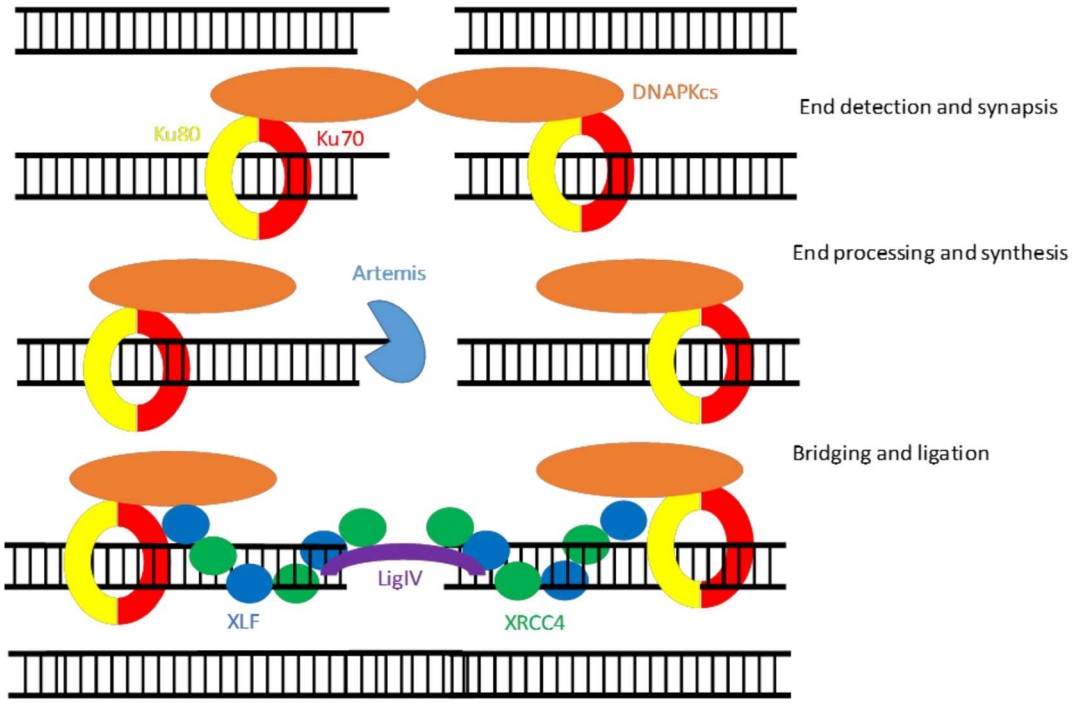



**Figure 1.2 Repair of double strand breaks by Non-homologous end joining repair.** Following the formation of a DNA break, ends are detected and bound by the Ku 70/80 heterodimer (red, yellow) which in turn recruits DNAPKcs (Orange). This holoenzyme mediates initial end synapsis as well as the recruitment of processing factors like the nuclease Artemis (Teal) which renders the ends chemically compatible for ligation. The final stage of repair is the end bridging and alignment step wherein XRCC4 (green) and XLF (blue) align the free DNA ends so that LigIV (purple) can ligate the phospho-diester backbone of the DNA, thereby restoring strand integrity.

Originally discovered in 1981, defects in either of the two Ku proteins result in a severe combined immunodeficient phenotype as well as a severe sensitivity to ionizing radiation (Jackson et al., 1995; Lieber et al., 1988; Mimori et al., 1981. Both proteins have a common domain architecture consisting of an N-terminal α/β domain, a β barrel and a C-terminal α-helical arm. All three domains contribute to the dimerization interface and the resulting ring structure forms a central pore by which the DNA is bound. (Figure 1.3) The structural similarity between the two proteins occurs despite a low level of sequence conservation (roughly ~ 14% identity between monomers). The ring structure formed by Ku70/80 makes it ideal for DNA sequence independent structure specific DNA binding.

The high affinity of Ku for DNA ends as well as its lack of sequence specificity make it the perfect DNA sensor and initiating factor for double strand break repair by NHEJ. The ring structure is large enough to accommodate roughly 20 basepairs of duplex DNA and shows no preference for 5' or 3' overhangs. These features are of vital importance for Ku to serve as an initiator of repair as DSBs occur anywhere in the genome. The interaction between Ku and DNA includes steric interactions between the major and minor grooves of the phospho-diester backbone (figure 1.3) (Walker et al, 2001).

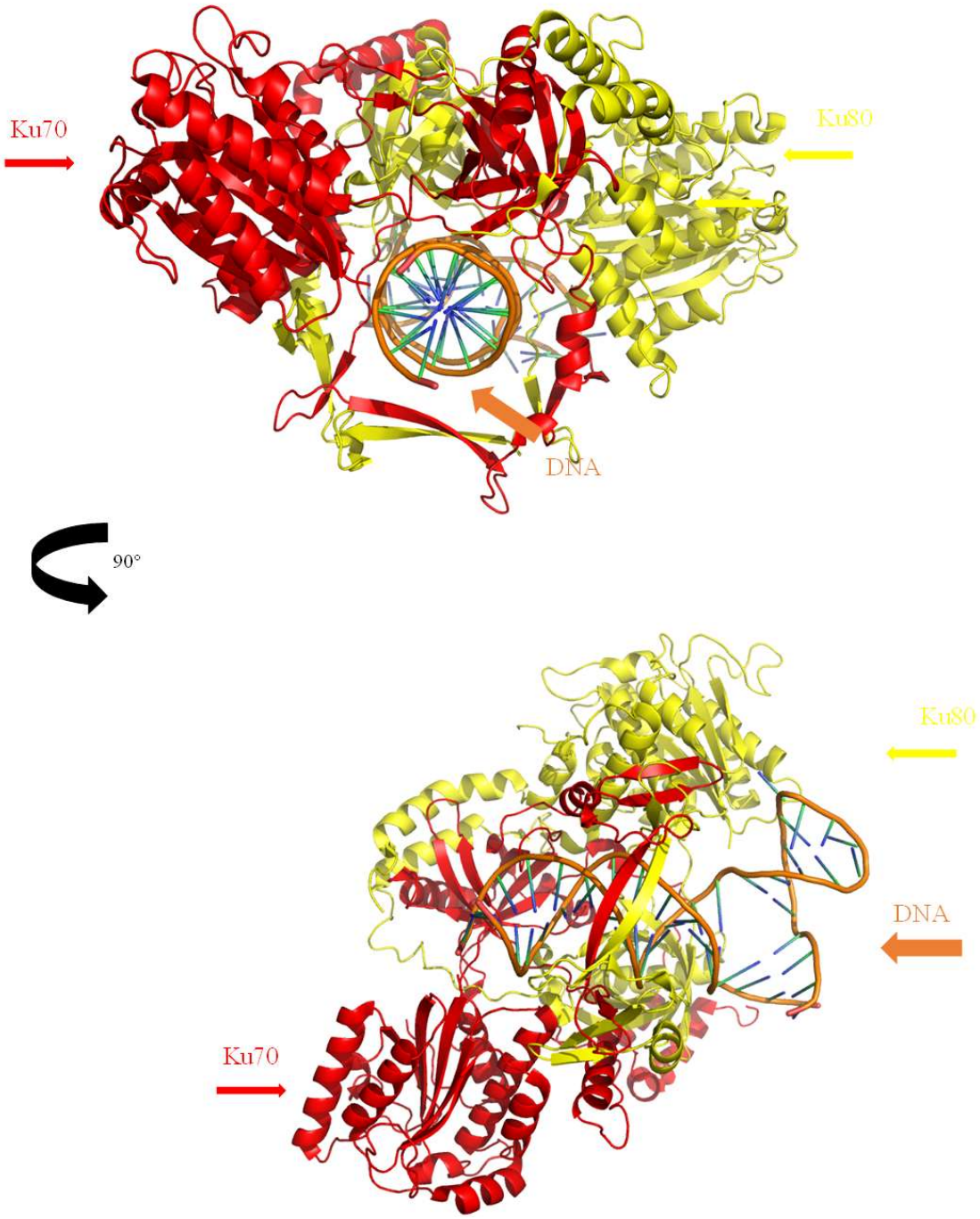



**Figure 1.3 Crystal structure of the Ku70/80 heterodimer bound to DNA.** The ring like structure wraps around and encircles the bound DNA. The interaction requires contributions from both Ku 80 (yellow) and Ku 70 (red). The ring like structure threads along the DNA backbone forming non-specific interactions with the minor groove of the duplex PDB 1jey (Walker et. Al., 2001)

In addition to its vital role in end sensing and initiation of repair, Ku has recently been shown to possess a 5'- deoxyribose-5'-phosphate lyase activity. This lyase activity has been shown to be essential both *in vitro* and *in vivo*, thereby conferring a catalytic activity to the heterodimer in addition to its end sensing ability (Roberts et.Al., 2010). The lyase activity often generates a nick 3' to the base being repaired followed by excision of the base. This activity results in a free DNA end with a 5'phosphate, thereby cleaning up the end in question for further processing and ligation. Beyond DNA binding and lyase activity, Ku is also responsible for the recruitment of nearly all other NHEJ repair proteins to the site of the break (Constantini et al., 2007; Gottlieb et al., 1993; Griffith et al., 1992., Suwa et al., 1994)

Following initial end binding by Ku 70/80, the next step in NHEJ repair is recruitment of the DNA dependant protein kinase catalytic sub-unit (DNAPKcs). This interaction occurs through the C-terminal tail of Ku80 (Gell et al., 1999; West et al., 1998). When bound to the Ku 70/80 heterodimer, this holo-enzyme forms the functional DNA dependent protein kinase. Interestingly, these two proteins do not interact in the absence of DNA and kinase activity is only observed when DNA is present as a co-factor (Downs et.al., 2004; Hartley et al., 1995; Yaneva et al., 1997). Complex formation requires free DNA ends in order for the C-terminal region of Ku80 to interact with DNAPKcs (Radhakrishnan and Miller, 2017). The enzyme is a member of the phosphoinositide-3-kinase serine/threonine kinase. The catalytic subunit is a large 470 kDa protein that is

broken down into three core regions: two arms composed entirely of heat repeats which support the kinase head domain (Sibanda et.Al., 2011). Cryo-electron microscopy studies have shown that like Ku, DNA-PKcs binds and engages both sides of the double strand break, thereby forming a synaptic complex capable of bridging the gap and preventing ends from diffusing away and losing one another within the volume of the nucleus. As with Ku70/80$^{-/-}$, deficiencies in DNAPKcs are associated with severe sensitivity to ionizing radiation and considerable immune defects (Callen et al., 2009).

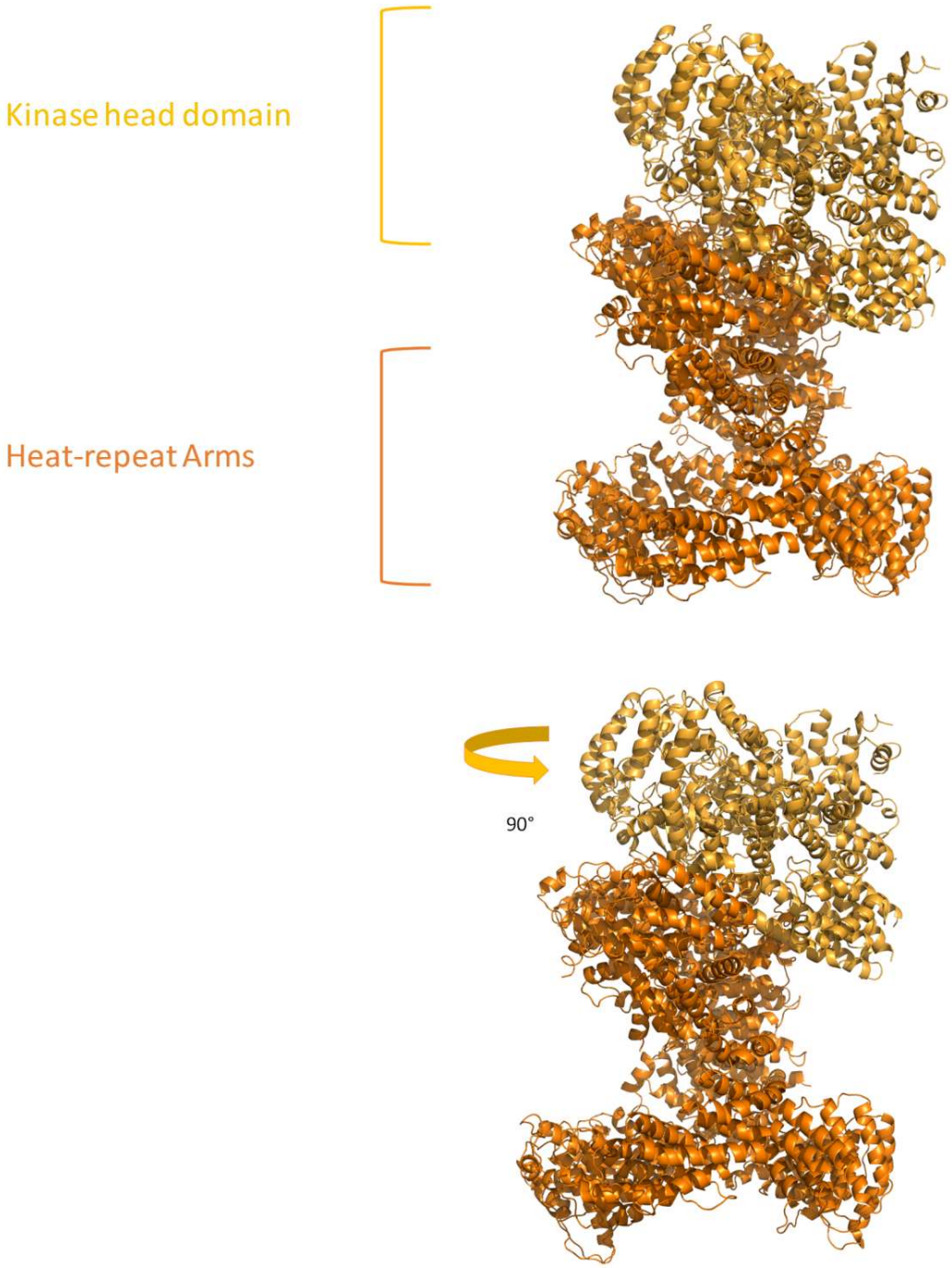



**Figure 1.4: Crystal structure of the DNA dependant protein kinase catalytic subunit.** The catalytic domain resides in the upper head domain (Orange) while heat repeat motifs make up the arms of the protein and mediate DNA binding as well as interaction with the Ku heterodimer (Chirgdaze et. Al., 2017) PDB 5LUQ

In response to DNA damage, DNA-PK is known to autophosphorylate at serines (Ser) 2612 ad 2624 and threonines (Thr) 2609, 2620, 2638 and 2647 (Mahaney et al., 2009). In addition to autophosphoryltion, DNA-PK has been shown to phosphorylate Ku 70/80, XRCC4 and LigIV *in vitro*. Small angle X-ray scattering (SAXS) have revealed that DNA-PK and several of its substrates undergo extensive conformational changes upon phosphorylation (Ding et al., 2003; Hammel et.Al., 2010). These conformational changes suggest that DNA-PK phosphorylation may serve as a regulator of complex formation and function. Furthermore, phosphorylation by DNAPK has been shown to mediate disassembly of XRCC4/XLF filaments and other complexes formed during repair (Andres et al., 20l1). It is likely that once repair has been completed, phosphorylation of the repair complex by DNAPK mediates complex disassembly (Chan et al., 1996; Douglas et al., 2001; Rivera-Calzada et al., 2005). Formation of the full DNAPK holoenzyme marks the transition to the next phase of the repair pathway (processing) wherein the ends are rendered chemically compatible for ligation.

### 1.3.2 End processing and alignment

Once the ends have been aligned by DNA-PK, they must be rendered chemically compatible for ligation. As stated previously, DNA damage that results in a double strand break almost never leaves the 5' phosphate or 3' hydroxyl groups required for ligation. End processing factors include but are not limited to nucleases, polymerases, lyases and hydrolases. The chemistry of groups

on strands at break sites dictates which factors must be recruited. Nucleolytic activity during repair, specifically during VDJ recombination, is carried out by the enzyme Artemis. Artemis is member of the β-CASP family of nucleases and possesses a well characterized 5'- 3' exonuclease activity. Artemis gains further functionality in the form of a structure specific endonuclease when in complex

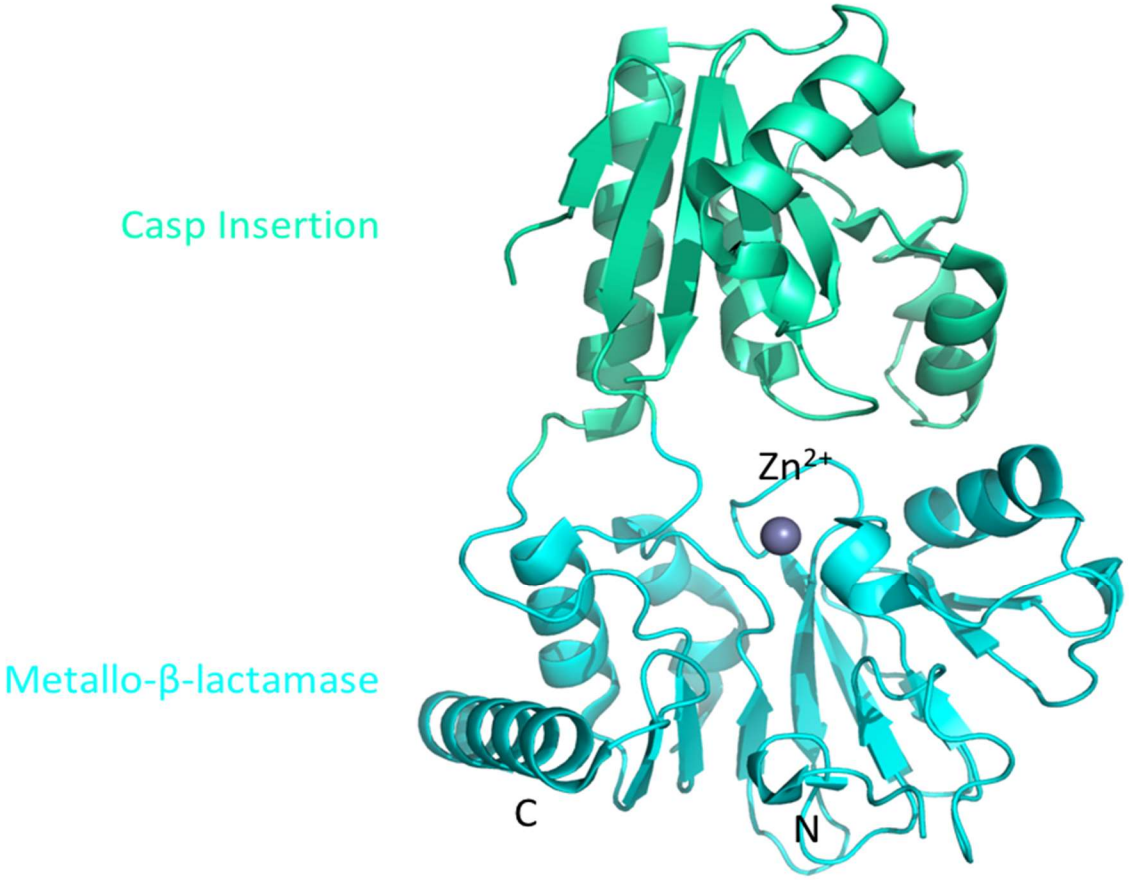



**Figure 1.5 Crystal structure of the β-CASP domain from human SNM1a.** The β-lactamase domain (blue) interrupted by a CASP insertion within a loop region. Catalytic activity requires both domains as well as a catalytic zinc (grey)

with and phosphorylated by DNA-PKcs (Goodzari et al., 2006; Ma et al., 2005; Meek et al., 2007). This endonuclease activity is required for successful VDJ

recombination. Following VDJ recombination, the ends of the coding joints are capped by DNA hairpins. Only DNA-PK activated Artemis is capable of opening these joints. Despite the absence of a crystal structure for Artemis, there do exist several structures of other β-CASP domains and they share a common domain architecture. A Metallo-βlactamase domain is interrupted in the middle of a loop by a CASP domain insertion (Figure 1.5). All β-CASP domains require metal co-factors and Artemis is no exception as zinc ions are required for both types of nuclease activity as well as regulation by DNA-PK to process overhangs and other restrictive structures that impede repair (Callebaut et al., 2002; Carfi et al., 1995., Moshous et al., 2001).

If the ends contain regions of SSDNA, several different polymerases which are capable of filling in gaps within the sequence (Polymerase μ (Polμ) and polymerase λ (Pol λ), members of the Pol X family). Specifically, Pol λ is stimulated by interaction with LigIV (Fan et al., 2004). All members of the Pol X family are single unit polymerases which extend DNA in the 5' to 3' direction. The interaction of polymerases with other NHEJ factors is mediated through BRCA1 C-terminal domains (BRCT). These domains are common within NHEJ factors and mediate several important protein-protein and in some cases, protein-DNA interactions. Once the ends have been processed, the final phase of repair is carried out (DNA ligation).

### 1.3.3 Ligation and strand release

Following end processing, the final ligation complex is assembled at the break. The final ligation complex is formed by three proteins: X-ray repair cross complementing protein 4 (XRCC4), the XRCC4 like factor (XLF) and DNA ligase IV (LigIV) (Ahnesorg et al., 2006; Andres et al., 2007; Giaccia et al., 1990; Junop et al., 1999). The LigIV/XRCC4 complex is recruited to the break site through an interaction between XRCC4 and Ku 70. Once localized at the break, this complex further recruits XLF to form the ternary repair complex. It should be noted that Ku has been shown to recruit XLF and therefore the precise order of repair factor recruitment remains unclear. LigIV is the catalytic ligase which re-seals the phosphor-diester backbone, thereby restoring strand integrity. Defects in any of these three genes (XRCC4, XLF, LigIV) are associated with severe phenotypes. DNA ligaseIV syndrome is observed in patients with mutations in the LigIv gene and are associated with SCID phenotype as well as severe sensitivity to ionizing radiation (Enders et al., 2006; Jiang et al., 2016). Mouse knockouts of either LigIV or XRCC4 are embryonic lethal in mice (Gao et.Al., 1998; Gao et al., 2000; Zha et al., 2007).

DNA LigIV is composed of 5 domains, A DNA binding Domain, nucleotidal transferase domain and OB fold which comprise the catalytic core. Within the C-terminus, there are 2 BRCT domains that mediate protein-protein interactions. The structures of each of these 5 domains have been solved independently. While the individual folds of the catalytic core domains are

conserved relative to other mammalian ligases, LigIV adopts an open conformation rather than the ring like structure that has been previously observed for LigI (figure1.6) (Ochi et al, 2013).

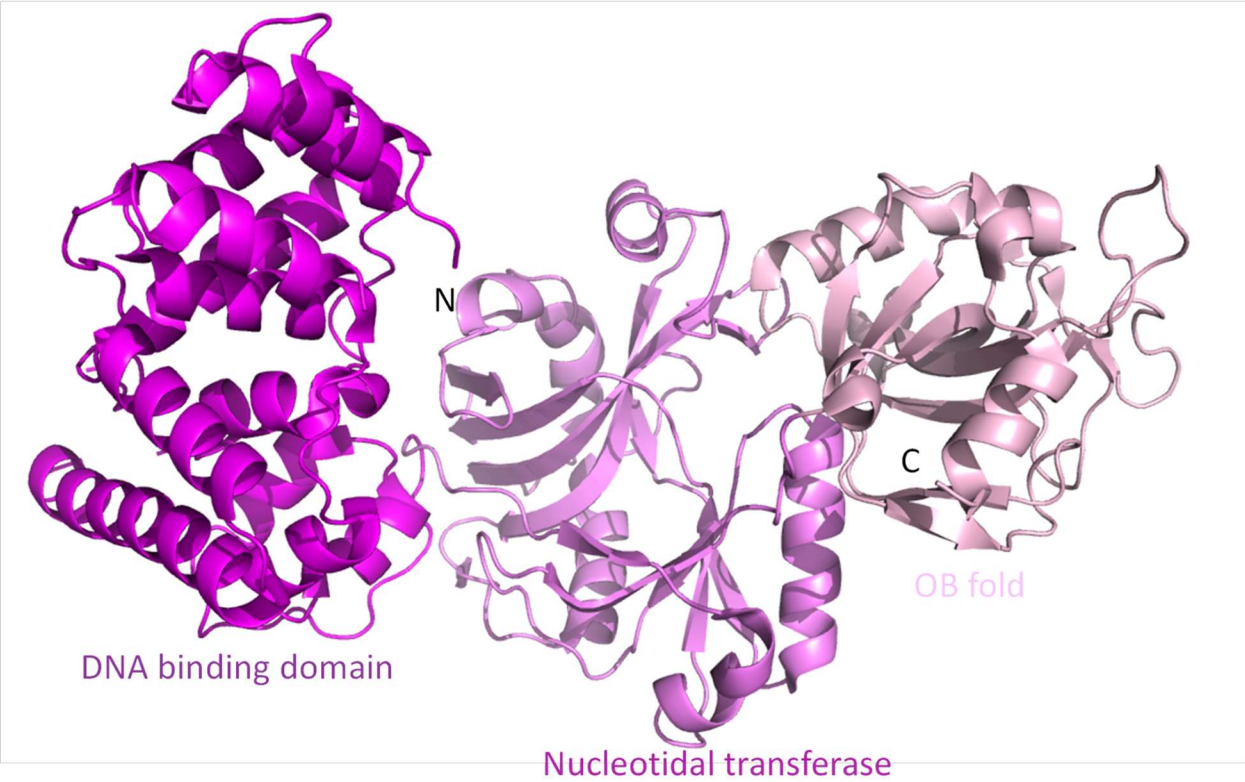



**Figure 1.6 Crystal structure of the DNA Ligase IV catalytic core.** The three domains of catalytic core of LigIV. The DNA binding domain (purple) binds DNA distal to the break site so as to align it for ligation. The nucleotidal transferase domain (violet) contains the catalytic residue, K273 which transfers the AMP moiety to the 5'phosphate at the site of the break. The OB-fold (pink) is associated with oligonucleotide and oligosaccharide binding activities. (Ochi et. Al., 2013) PDB3W50

DNA ligase IV shares a common mechanism with other DNA and RNA ligases: a nucleotide co-factor dependent condensation reaction. An Adenosine mono-phosphate (AMP) moiety from ATP is transferred to the catalytic lysine (K273) resulting in the release of a pyrophosphate molecule and the formation of a phosphor-amide bond. The AMP is then transferred to the 5' phosphate on strands at the break site. Repair is completed when the 3'OH from an opposing DNA strand performs nucelophilic attack the phosphate group, thereby sealing the nick and restoring strand integrity (Teo, S.H. and Jackson, S.P., 2000, Tomkinson, A.E et al., 1991). Interestingly, LigIV is pre-adenylated (unlike other mammalian ligases) and is the only canonical double strand break repair ligase. While LigIV is the catalytic enzyme responsible for sealing the break, its activity is only made possible when it forms a complex with XRCC4 and XLF (two structural homologues that mediate end bridging and stimulation of ligase).

Eukaryotic XRCC4 exists in two isoforms, one of 336 residues the other 334. As the 336 residue isoform is the most commonly expressed, it has been heavily studied. XRCC4 does not possess any known intrinsic enzymatic function but is a scaffolding protein that also stimulates the activity of LigIV. The partial crystal structure of XRCC4 has been solved. XRCC4 exists as a dimer with each monomer adopting an identical fold. The N-terminal head domain (residues 1-110) is formed by a 7-stranded anti parallel beta sandwich. In between strands 4 and 5, there is a helix turn helix motif. The C-terminal tail (residues 111-200) forms an extended alpha helix which, radiates away from the head domain (Junop

et al., 2000; Lieber et al., 1998). XRCC4 plays multiple roles during the repair pathway that are governed by its quaternary structure (Andres et al., 2011; Junop et al., 2000; Modesti et.Al., 2003). XRCC4 forms distinct complexes with LigIV and XLF, and itself (Andres et al., 2011; Hammel et al, 2010. The canonical role of XRCC4 has long been known as its ability to bind and stabilize LigIV in the cell. In cells depleted of XRCC4, no LigIV is detectable. The helical C-terminal tails (160-200) of XRCC4 mediate this interaction by binding with a linker region between the two C-terminal BRCT domains of LigIV forming a highly stable complex (roughly 4.2 nM) (McFadden et. Al., 2012). In addition to LigIV binding, XRCC4 also binds DNA with a preference for longer pieces of DNA (>100 BP) as well as forming homo-tetramers. These interactions were confirmed via gel shift assays and analytical ultracentrifugation (Modesti et. Al., 1999, 2003)

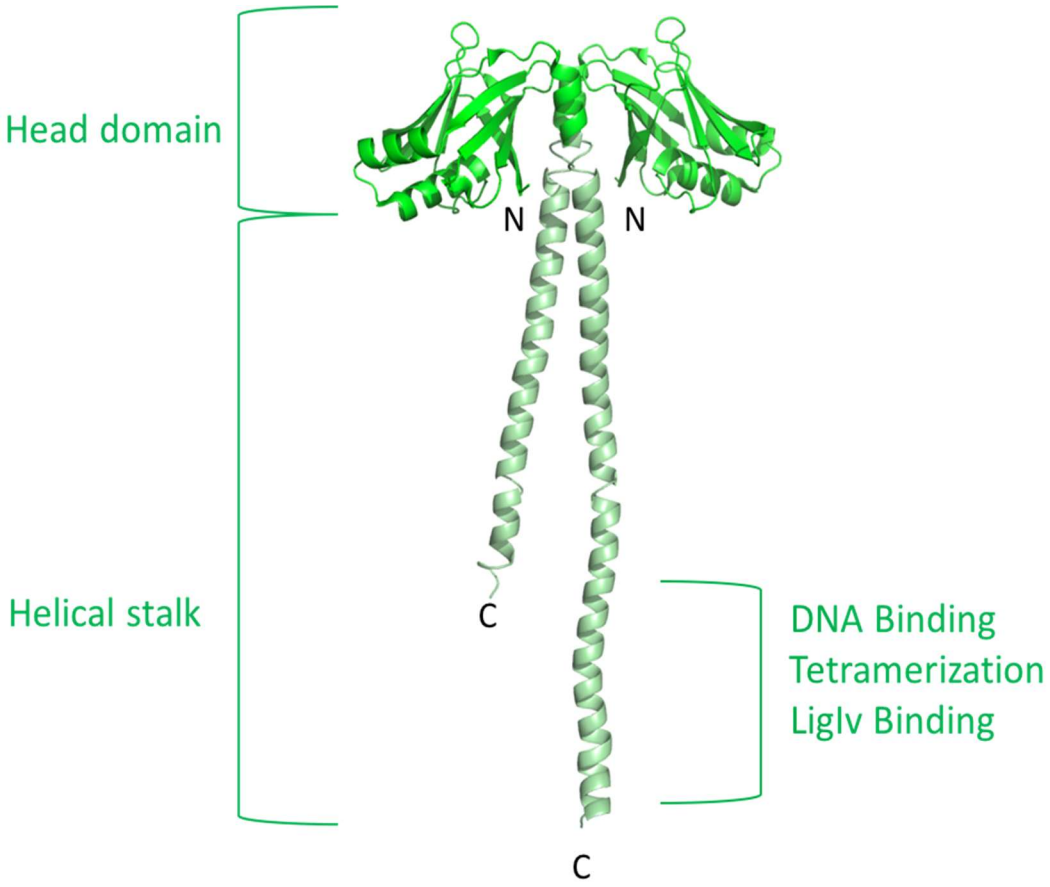



**Figure 1.7 Crystal structure of the XRCC4 dimer.** XRCC4 dimer showing the head domains (green) which mediate interaction with XLF. The XRCC4 head domain is formed by a 7 strand anti-parallel β-sandwich with a helix turn helix motif between inserted strands 4 and 5. The extended helical stalk region mediates DNA binding, LigIV interaction and potentially tetramerization (dark green) Junop et. Al., 2011 PDB 1FU1

.

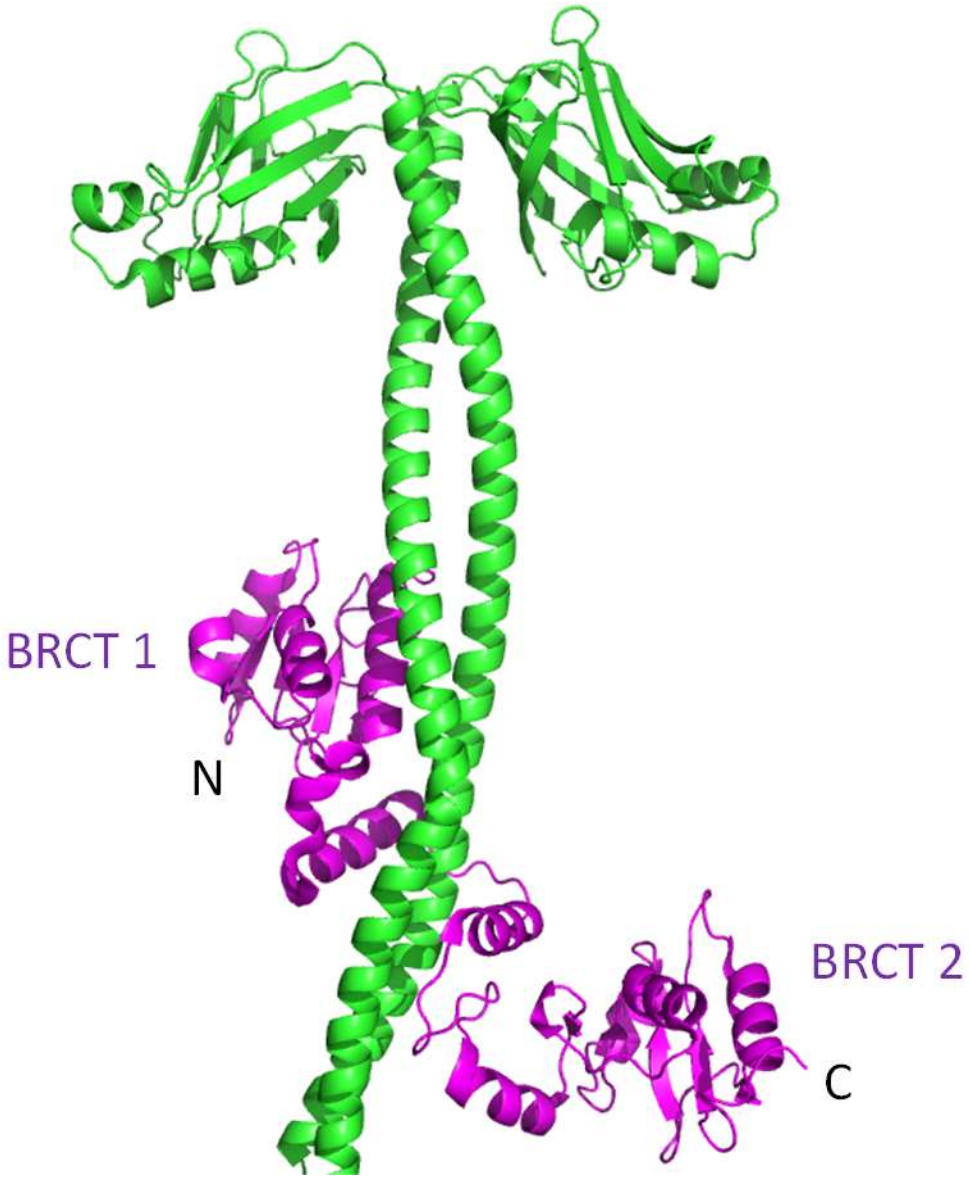



**Figure 1.8 Crystal structure of the tandem BRCT domains of LivIV bound to the coiled coils of XRCC4.** The interaction between XRCC4 and LigIV is mediated by the first BRCT domain and a helix-turn-helix motif between the two BRCT domains of LigIV. The resulting complex is one of the strongest observed with an estimated kd of 2fM. **(**PDB 3II6)

Interestingly, all three functions of XRCC4 map to the same 40 amino acid stretch at the end of the coiled coil (Modesti et al., 2001). The contributions of these individual functions is difficult to ascertain as there is no true separation of function mutant available for this protein to make such studies possible.

XRCC4 is also a substrate for DNAPK, being phosphorylated at several residues, mostly within the C-terminus S53, S193, S260, S302, S313, S318, T 321, S325, S326 (Lieber et al., 1998; Yu et al., 2003). These phosphorylation events, however, have never been shown to be biologically relevant (Yu, Y et al., 2008). The C-terminal tails of XRCC4 (202-336), despite being highly conserved, have yet to have any demonstrated biological relevance. Unlike other members of the XRCC4 family with solved structures, the tails of XRCC4 radiate away from the head domain. All other homologues of XRCC4 including XLF, Spindle assembly atypical 6 (SAS6) and Paralogue of XRCC4/XLF (Paxx) have tails which fold back to interact with their own head domain. Of these three XRCC4 homologues, XLF is the main-most directly implicated in repair by NHEJ.

XLF was identified in 2006 as a novel NHEJ core factor in patients displaying severe combined immunodeficiency and micro-cephaly. Despite very little sequence conservation (<14%), XLF is a structural homologue of XRCC4, possessing the same overall topology (Andres et al., 2007). XLF is a stable homodimer with each subunit

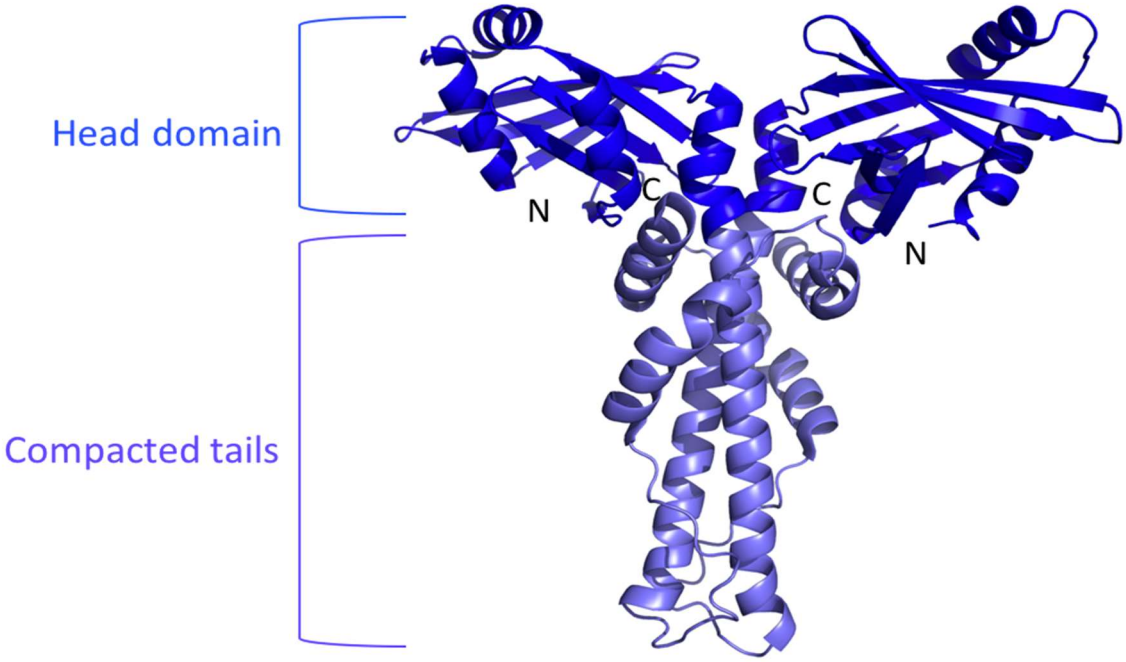



**Figure 1.9 Crystal structure of the XLF dimer.** XLF, like XRCC4, exists as a dimer with each protomer adopting a similar conformation. The head domains (dark blue) adopt a 7 stranded beta sandwich fold with a helix turn helix motif between strands 4 and 5. Unlike XRCC4, however, the C-terminal tails of XLF(light blue) fold back to interact with their own head domains. The DNA binding domains are un-pictured here. **(**Andres et. Al, 2007) PDB 2RWR

formed from a 7-stranded anti parallel β-sandwich with a helix-turn-helix motif inserted between strands 4 and 5 and an extended alpha helical C-terminal tail that extends away from the head domains. Unlike XRCC4, however, the tails of XLF

turn back to interact with their own head domains as opposed to adopting an extended helical conformation. Despite subtle differences in the positioning of the secondary structure elements, XRCC4 and XLF share a high degree of similarity in their overall topology. Both have globular head domains, extended coiled-coil regions and a C-terminal unstructured region (CUR)(Figure 1.10)

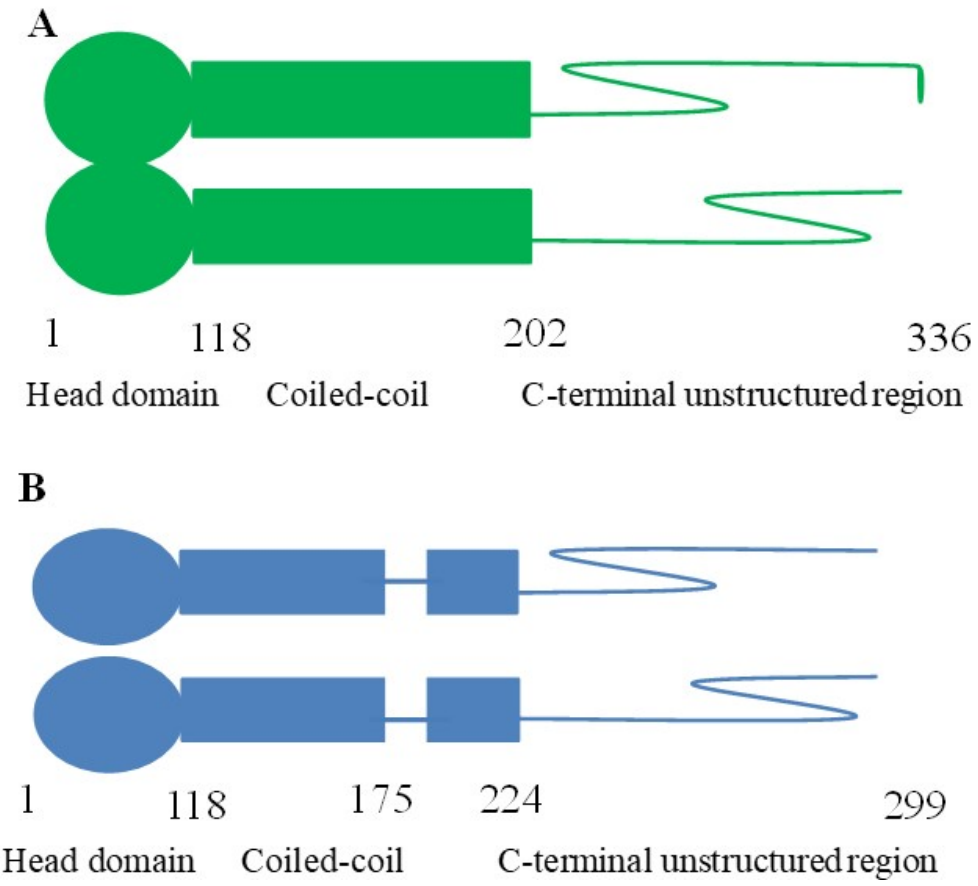



**Figure 1.10 Cartoon representation of the XRCC4 and XLF domain boundaries.** A) Each XRCC4 monomer is composed of a globular head domain (1-117), an extended coiled-coil helix which spans away from the head domain (118-202) and a C-terminal unstructured region which forms the C-terminal tail of each monomer. B) Each XLF monomer is also composed of a globular head domain (1-116), a coiled coil helix (118-224) and CUR (224-299). Unlike the XRCC4 coiled coil, that of XLF is not a uniform, helix but adopts different conformations.

XLF binds DNA with high, non-specific affinity and also shows a preference for longer pieces of DNA. XRCC4 and XLF have been shown to form extended protein-protein filaments *in vivo* which are capable of bridging DNA ends *in vitro* (Andres et al., 2011, Hammel et al., 2010). These filaments are formed by alternating dimers of XRCC4 and XLF at a 30° pitch between dimers. This extended left handed helical conformation creates a large pore in the center of the filament wherein the DNA can be bound. Unlike the tails of XRCC4, the tails of XLF fold back to interact with their own head domains and extend toward the central pore of the filament, wherein the DNA is expected to be likely bound (Andres et al., 2007, Andres et al., 2011). More recently, it has been demonstrated that these XRCC4/XLF filaments form 'sleeves' which coat and thread along the DNA (Brouwer et al, 2016). These sliding sleeves have been further shown to bridge DNA ends prior to ligation, suggesting a mechanism wherein these filaments facilitate end joining by ensuring the two ends are proximal to one another so as to effect ligation.

In addition to forming a complex with XRCC4, XLF has also been shown to stimulate DNA LigIV ligation of blunt or incompatible ends *in vitro* (Andres et al., 2007). This effect has been suggested to reflect the ability of XRCC4/XLF filaments to align DNA ends.

## 1.3 Thesis goals and objectives

The work described in this thesis seeks to elucidate the molecular mechanism of the C-terminal tails of XRCC4 (Lif1) and its structural homologue XLF (Nej1). These tails are predicted to be intrinsically disordered, are key regulatory regions for these proteins and ultimately mediate what higher complex the proteins will form. Chapter 2 focuses on the role of the C-terminal tails of XRCC4 and how they affect which complexes the protein will form. Chapter 3 examines how the C-terminal tails of Lif1 (XLF) bind their DNA substrates. Chapter 4 assesses how the tails of both XRCC4 and XLF are required to mutually stabilize functional end bridging filaments. Chapter 5 provides the first reported crystallization conditions for the complex of DNA ligase IV, XRCC4 and DNA. Chapter 6 discusses the intrinsic details involved in crystallizing proteins with inherently disordered regions as well as multi-component complexes.

All of the proteins studied in this work are central factors in DNA double strand break repair and are consequently of profound medical importance. There exist several types of cancer wherein inhibition of NHEJ is associated with an enhanced sensitivity of the tumor cell to radiation-based treatments (Dolman et al., 2015, Li et al., 2012). Broad screens of compounds targeting protein/protein interactions have been carried out and putative hit molecules have been identified (Mcfadden et al., 2011). Despite the significance of these proteins as drug targets, our understanding of the protein/protein and protein/DNA interactions throughout

NHEJ limits our ability to modulate this pathway. The work presented here represents ongoing work in understanding how complexes in NHEJ form. Ultimately, such knowledge will be helpful to effectively modulate them within the context of diseases. Specifically, work within the following thesis is aimed at understanding the molecular mechanisms of protein-protein and protein-DNA interactions required for repair of DSBs by NHEJ.

# Chapter 2 XRCC4 forms tetramers through a multi-helix bindle

## 2.1Abstract

XRCC4 is a core NHEJ factor which has multiple interacting partners within the repair pathway. In addition to self-association, XRCC4 can form complexes with LigIV, XLF and DNA. The ability of XRCC4 to form these different complexes is regulated through its own oligomeric state. Here, we report the first crystal structure of full length Human XRCC4 to 3.43Å. Based on this new structural information, mutants targeting the tetramerization interface were generated. A single mutant was found to be deficient for tetramerization but still competent for DNA binding, LigIV binding and XLF interaction. Surprisingly, DNA end bridging was not affected by loss of tetramerization. These results delineate the minimum requirements for XRCC4 tetramerization and further raises questions of the biological role of XRCC4 higher order self assembly.

## 2.2 Introduction

XRCC4 is a scaffolding protein that facilitates several different protein/protein and protein/DNA interactions during repair of double strand breaks by the Non-homologous end joining pathway. XRCC4 forms complexes with DNA, LigIV and XLF, and is involved in several different stages of the repair pathway. The first canonical function of XRCC4 was identified as binding to and stabilizing LigIV (Junop et al., 2001., Modesti et al., 2003). Since this discovery, XRCC4 has been shown to assist in retention of XLF at the site of a break and more recently, to bridge DNA ends at the site of the break (Andres et

al., 2011). The nature of how XRCC4/XLF filaments align and bridge DNA ends remains unknown, as does how XRCC4s self association impacts these structures.

Previous structures of XRCC4 suggested two potential tetramerization interfaces, 1) a head to head alignment and 2) a 4-helix bundle which mediates tail dependent tetramers (Figure 2.1). Determining the true tetramerization interface is of profound importance in order to permit further mutational studies aimed at uncovering its mechanism and biological function. If the head to head model is correct, then XRCC4 tetramerization and XLF interaction would be mutually exclusive. Alternatively, if the tail model is correct, then XRCC4 tetramers would be able to bind XLF, but not DNA or DNA LigIV. Another limitation of the existing XRCC4 crystal structures is that they are completely lacking the C-terminal unstructured region of the protein (204-336). While these tails are predicted to be intrinsically disordered, they are highly conserved among eukaryotes. In addition to this high degree of conservation, the tails are known to be DNAPK substrates. Despite extensive efforts. no true separation of function mutants have been discovered. In order to assess the biological function of each function, such mutants must be found.

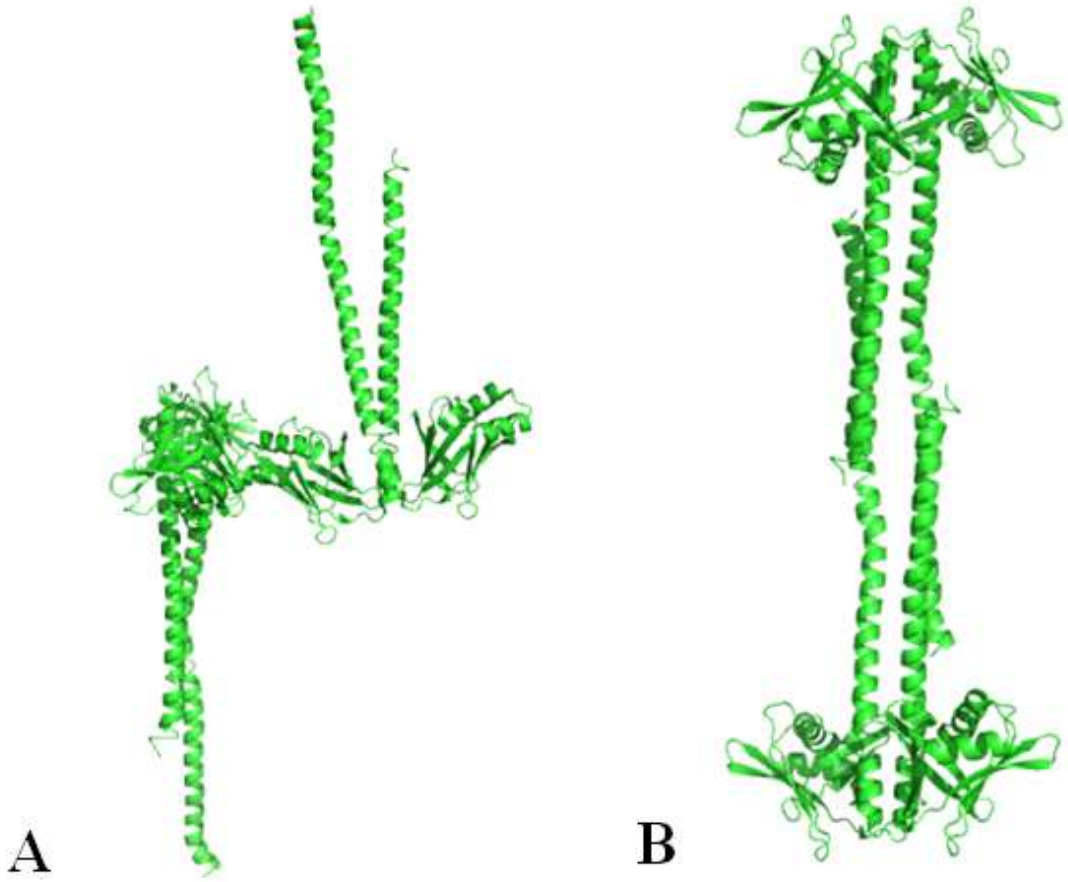



**Figure 2.1 Two potential tetramerization interfaces for XRCC4 self association.** The original crystal structure of XRCC4 suggested two potential tetramerization interfaces within the unit cell. A) A head to head arrangement with the two dimers offset by 90°. B) formation of a multi-helix bundle by the C-terminal tails of the proteins.

There does exist a mutant, XRCC4$^{L184Q/K187D/I191S}$ which is defective for tetramerization. First characterized in 2003, mutations of these residues were found to eliminate XRCC4 tetramerization, as determined by AUC (Modesti et. Al., 2003). It is, however, also defective in DNA binding as well as LigIV binding therefore it is impossible to delineate the loss of which function resulted in the

observed defects of XRCC4.While not a separation of function mutant, it does serve as a perfect negative control for interaction studies of XRCC4.

In order to definitively determine the tetramerization interface, we sought to crystallize full length XRCC4. Further proteolysis studies and sequence conservation also identified several C-terminal truncations of the protein which could potentially serve as the minimal unit capable of stabilizing tetramers and mediating DNA end bridging when in complex with XLF. Here, we report the first crystal structure of full length XRCC4 as well as novel separation of function mutants which allowed for testing of the importance of each potential XRCC4 function.

## 2.3 Materials and methods

### 2.3.1 Growth and expression of native XRCC4

*E. Coli* BL21 DE3 cells were transformed with a plasmid containing the open reading frame for full length human XRCC4 with a C-terminal hexa-histidine affinity tag (PWY 1087). Cells were plated on LB-agar plates supplemented with 100 µg/ml chloramphenicol and incubated overnight at 37° C overnight. A single colony was used to inoculate a 50 ml overnight culture, which was then sub-cultured into 4 litre cultures of Luria Bertani media, also supplemented with 100 µg/ml chloramphenicol. Cells were grown until they reached an optical density at 600 nanometers (OD600) of 0.5. Expression of XRCC4 was induced by the addition of 1 mM isopropyl β-D-1

thiogalactopyranoside (IPTG). Following a three hour induction at 37° C, cells were harvested by centrifugation at 3315 X g for 15 minutes (Avanti J-30I centrifuge with the rotor JLA 9.1000). Cell pellets were recovered and flash frozen and stored at -80°C until further use.

### 2.3.2 Purification of XRCC4

All XRCC4 proteins (wild type and mutants, both point and truncation) were purified in an identical manner. Cell pellets were re-suspended in a buffer containing 20 mM Tris, pH 8.0, 750 mM KCl, 1 mM β-mercaptoethanol (BME) and 3 mM imidazole. Cells were lysed by three passes through a French pressure cell at 20000 psi. Lysate was clarified by centrifugation at 48000 X g for 45 minutes. Clarified lysate was then loaded onto a pre-equilibrated HIStrap-HP nickel affinity column (GE healthcare, United States of America). The protein was eluted via step washes with resuspension buffer containing 15, 30 and 60 mM imidazole. XRCC4 was eluted at 210 mM imidazole. The resulting product was only 50% pure so therefore IMAC was followed by anion exchange chromatography. XRCC4 containing fractions were buffer exchanged into a buffer containing 20 mM TRIS, pH 8.0, 1 mM EDTA, 10 mM DTT, 200 mM KCl and 10% v/v glycerol. This sample was then applied to a 10/100 Mono-Q hp column (GE healthcare, United States of America). XRCC4 was eluted via a linear salt gradient of 200-400 mM KCl over 140 minutes at 0.8 ml/minute. Purity was verified via SDS PAGE.

### 2.3.3 Growth and expression of Seleno-methionine derivatized XRCC4

Seleno-methionine derivatized XRCC4 was grown using an M9-SeMet kit (Shanghai Medcillon). Briefly, plasmid PWY-1087 was transformed into D834 auxotrophic competent cells. These cells were then grown in a minimal media supplemented with Seleno-methionine. Cells were induced with 1mM IPTG at 16°C overnight. The derivatized protein was harvested and purified as described in 2.3.2.

### 2.3.4 Crystallization of full length XRCC4

Purified XRCC4, both native and seleno-methionine derivatized, were crystallized in an identical manner. Since full length XRCC4 contains 7 cysteine residues that readily form non-specific intermolecular disulphide linkages it was necessary to purify XRCC4 under reducing conditions with argon-purged buffers. Crystallization trials were initially carried out using protein concentrations ranging from 2 - 15 mg/mL. Commercial screens (The Classics suite, JCSG I,II,III and IV from Qiagen, Wizards I and II from Jena Bio Sciences) identified several conditions that generated showers of tiny crystals. These crystals could only be obtained when hanging drops were setup in an argon purged anaerobic chamber using reservoir buffer containing cacodylic acid and fresh 10 m*M* 1,4-Dithiothreitol. Varying precipitant and other mother liquor component concentrations failed to control nucleation. Drops contained either hundreds of small crystals or none at all. To limit nucleation and improve crystal quality micro-seeding was required. This was accomplished by crushing micro-crystals

(Bio Plas Pestle; Thomas Scientific) in reservoir buffer. Serial dilutions of the resulting seed stock were mixed in a 1:1 ratio with fresh XRCC4. Seed dilutions ranging from 10^-5 to 10^-7 produced large single crystals after roughly 48 hours (Figure 2.2). The resulting crystals were single and sufficiently large (250 x 100 x 50 $\mu m^3$) to collect diffraction data. Many different cryo-protectant conditions were explored; however, a cryo protecting solution of the mother liquor supplemented with 25% ethylene glycol and 14% w/v polyethylene glycol 6000 was sufficient to cryo protect crystals during freezing. Unfortunately, the crystals rapidly succumbed to radiation damage due to high concentrations of arsenic within cacodylate buffer used during crystal growth and cryo protection. Switching to MES (2-(N-morpholino)ethane-sulfonic acid) containing cryo protecting buffer and soaking crystals for 1-2 minutes prior to freezing dramatically improved resistance to radiation damage allowing for increased exposure time during data collection.

### 2.3.5 X-ray Data collection and processing

Diffraction data were collected at Beam 17-ID at the Advanced Photon Source of Argonne National Labs. Data were collected in half degree wedges with an exposure time of 0.0877 seconds/image. A total of 420 images, representing 210 degrees of rotation comprise the data set. Images were indexed and integrated using iMOSFLM. Reflections were scaled and merged using the Aimless and Ctruncate modules from CCP4i (Battye et al., .2011, Evans et al., 2013)

### 2.3.6 Structure determination and refinement

Merged reflections were used for molecular replacement in PHENIX. The resulting molecular replacement solution contained only minimal phasing information. Following initial phase determination by molecular replacement, phases were further extended by SAD phasing using the autosol module within Phenix. The hybrid substructure search successfully identified one selenium and eight arsenic sites within the asymmetric unit. Models were refined manually in coot and using the refine module from Phenix until the Rfree and Rwork factors converged.

### 2.3.7 Partial proteolysis of XRCC4

Purified XRCC4 was incubated with increasing concentrations of 4 proteases: Trypsin, Chymotrypsin, Papain and subtilisin (Floppy Choppy kit, Jena biosciences). Digestion was arrested by the addition of equal volumes of 2x SDS page load dye. The contents of these digests were then resolved by SDS page. Following coomassie staining, bands common to all 4 digestion patterns were excised and analyzed via mass spectrometry to identify stable domain boundaries.

### 2.3.8 Generation of XRCC4 mutants

Bacterial codon optimized constructs containing the desired patch mutations were synthesized and sub-cloned into a bacterial expression vector (pACYC) by Genscript. All mutants were expressed and purified in an identical manner to wild type XRCC4. C-terminal truncation mutants were generated using

quick-change mutagenesis. Primers to introduce stop codons were mixed with a template encoding full length XRCC4 (PWY 1087). Following amplification and DPN1 digest of the resulting products, DH5α cells were transformed with the harvested plasmid and plated on LB agar plates supplemented with 100 µg/ml chloramphenicol. Colonies were selected and plasmids were isolated via alkaline lysis miniprep. Incorporation of the mutations were verified by sequencing (Robarts research institute)

### 2.3.9 Circular dichroism spectroscopy

Purified XRCC4 was dialyzed into a buffer containing 100 mM potassium-acetate, 10 mM sodium phosphate and 1 mM tris(2-carboxyethyl)phosphine (TCEP). Spectra from 265 nM to 190nM were collected using a Jasco J-810 spectrophotometer (United States of America). Absorbances were converted to molar elipticity co-efficients. Data were analyzed using CDpro software, specifically, the CONTINLL algorithm to quantify the proportion of secondary structure.

### 2.3.10 Size exclusion chromatography with multi angle light scattering

Wild type XRCC4 and all mutants were analyzed in an identical manner. Samples were heavily dialyzed into a buffer containing 20 mM Hepes (p.H 8.5), 150 mM potassium chloride, 1 mM EDTA and 0.5 mM TCEP. 100 µl of XRCC4 at varying concentrations were injected onto a SEC 650 size exclusion column (Biorad, USA) at a flow rate of 1 ml/min. Multi angle light scattering data was

collected using a Dawn Heleos multi angle detector (WYATT, United States of America). Light scattering data were analyzed using Astra 6.1 (United States of America).

### 2.3.11 DNA end bridging assay

End bridging was carried out as described in Andres et al., 2011. Briefly, streptavadin labelled 1000 bp DNA strands were attached to biotynylated magnetic beads and incubated with XRCC4, XLF and a 500 bp DS DNA substrate. Bridging activity was read out by the recovery of the 500 bp fragment after magnetic removal of the beads. Biotynlated beads were passivated by washing with binding buffer (20 mM HEPES (pH 8), 75 mM KCl, 0.5 mM EDTA, 1 mM DTT, 5 % glycerol, 400 µg/ml acetylated BSA) and resuspended in an equal volume of buffer. Proteins (XRCC4 and XLF) were added to a final concentration of 2µM along with 200 ng of 500 bp dsDNA. Beads were removed magnetically, de-proteinated by incubation with 1 ng of proteinase K for 20 minutes and the resulting supernatants were visualized on a 1% agarose TAE gel.

## 2.4Results

### 2.4.1 The C-terminal tails of XRCC4 do not fold back like those of XLF

While the structure of XRCC4 has been previously determined, it lacked information regarding the C-terminal region of the protein. We therefore crystallized full length XRCC4 in order to determine if the C-terminal region of XRCC4 folds back similar to XLF or adopts a different conformation. Following optimization of full length XRCC4 crystals, multiple data sets were collected (Fig

2.2). Despite high mosaicity within the lattice, the data sets were of sufficient quality to solve the structure (Figure 2.3, Table 2.1). In order to ensure that the structure of the tails was free of model bias, a minimal search model was used for molecular replacement. Only the head domain of a single XRCC4 monomer was used. While this yielded good phase information for the head domain, the electron density of the tails was poorly defined in the initial electron density maps. Phases were further extended through anomalous dispersion signals obtained from two heavy atom sources: Selenium incorporated into methionine residues and arsenification of the cysteine side chains within the protein. A hybrid substructure search revealed 1 selenium and 8 arsenic atoms within the asymmetric unit (Figure 2.4 A). These heavy atoms allowed for unbiased phase extension and the solution of a near full length XRCC4 structure (Fig 2.4B). Unbiased experimental phasing was necessary as one of the largest questions we sought to answer about XRCC4 is what the true tetramerization interface is and at low resolution it was very important to have confidence in the experimental electron density map that is only available with experimental phasing.

| | |
|---|---|
| Diffraction Source | B-17-ID |
| Wavelength (Å) | 0.9794 |
| Detector | Dectris 6M |
| Crystal-detector distance (mm) | 250 |
| Roation per image (°) | 0.5 |
| Rotation range (°) | 210 |
| Exposure time | 0.12 |
| Space group | C 1 2 1 |
| a,b,c (Å) | 164.0, 74.3, 88.4 |
| α, β, γ (°) | 90, 104.22, 90 |
| Mosaicity (°) | 1.6 |
| Total no. of reflections | 51315 |
| Unique reflections | 13973 |
| Completeness (%) | 99.7 (100) |
| Redundency | 3.7 (3.5) |
| I/σ(I) | 7.9 (2.0) |
| Rmerge (%) | 6.9 (22.5) |

**Table 2.1 Data collection and refinement statistics for the full length XRCC4 structure**

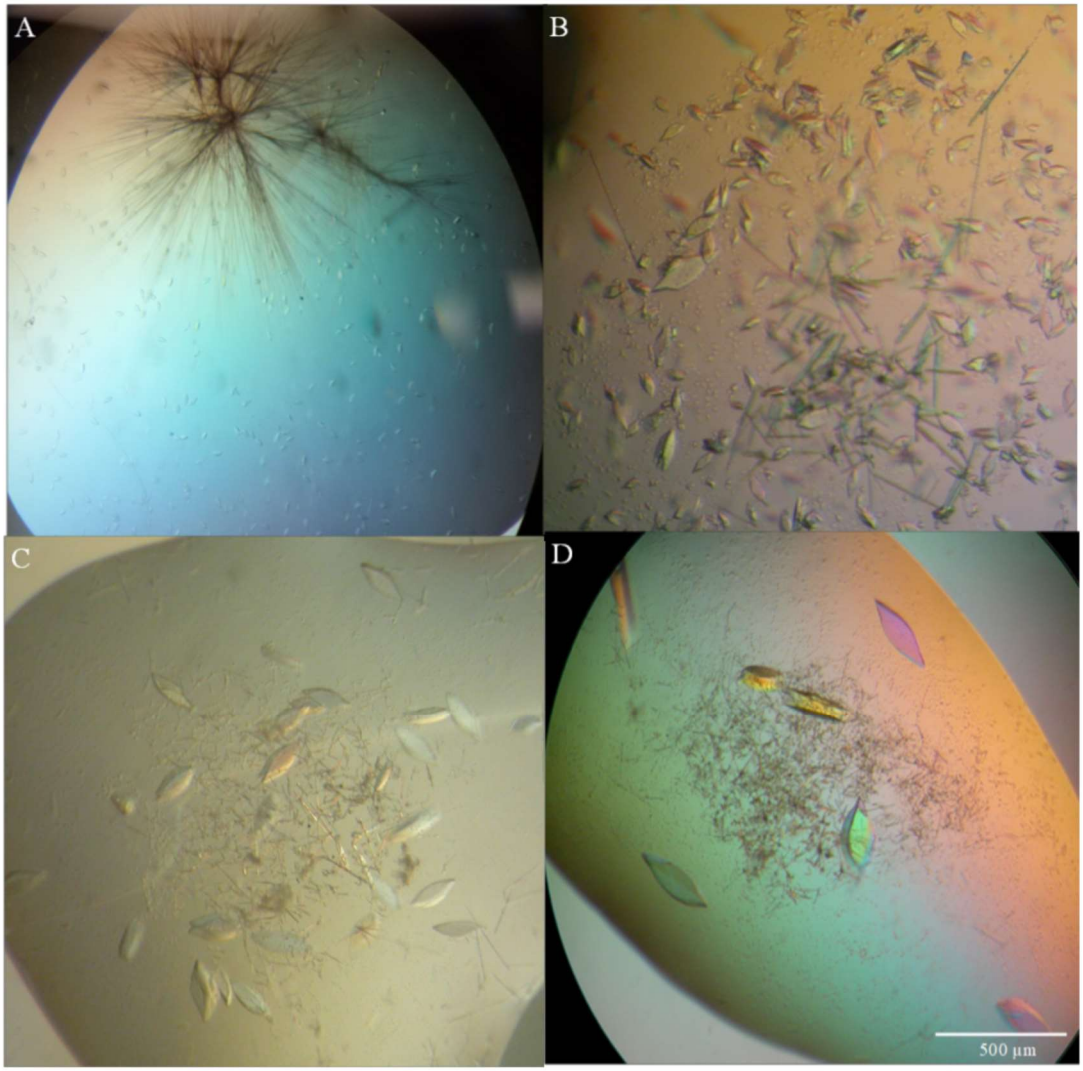



**Figure 2.2 Systematic optimization of full length XRCC4 crystals.** A) Initial XRCC4 crystallization condition obtained via broad screening B) Crystals improved in quality when equilibrated over well solution of the initial mother liquor supplemented with NH2SO4. C) Effect of setting crystal drops in an oxygen purged anaerobic chamber on crystal quality. D) Serial dilutions of harvested microcrystals were mixed with the crystallization condition in a 1:1 ratio and used for serial seeding.

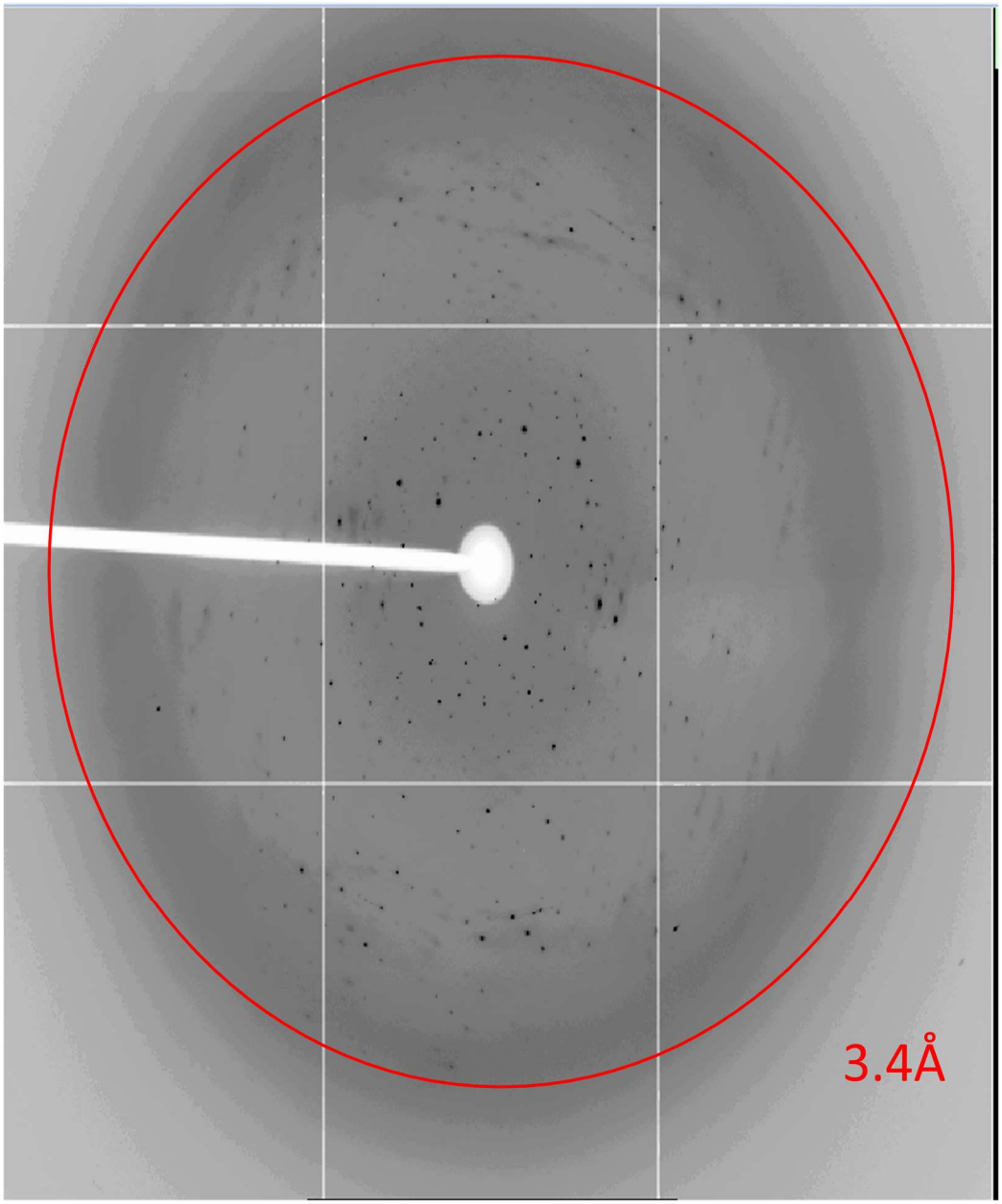



**Figure 2.3 Diffraction pattern of full length XRCC4 crystal showing mosaicity and anisoptropy.** Representative diffraction pattern of diffraction data collected at Advanced photon source (APS)-Argonne beam 17-ID. The irregularity of the spot pattern is indicative of the high degree of mosaicity within the crystal lattice.

In addition to showing that the head to head model of self-association was

not the true tetramerization interface, the resulting density showed several new secondary structure elements of XRCC4 (Figure 2.5 A). Specifically, the tails were shown to extend across the length of the structure and interact asymmetrically with the opposing head domains (Figure 2.5 B). One helix radiates across the entirety of the 4 helix bundle and folds around the cleft of the opposing head domains. The other tail interacts with the underside of the opposing head domain before folding back and forming a much larger multi-helix bundle. From the structure of full length XRCC4, it is clear that the XRCC4 tetramer is stabilized by both extensive contacts within the multi helix bundle as well as newly observed tail to head contacts. As these phases were experimentally determined, it would appear that XRCC4 forms tetramers in an elongated helical stalk manner, and not the head to head model that has been previously proposed (Hammel et.Al., 2010). Due to the low resolution, several gaps were present in the main chain the chain and side chain assignment for the extreme C-terminal portion of the chains was not possible. Chain A comprises the first 267 amino acids while chain B extends to residue 301. While sufficient resolution to assign side chain identities would have been excellent, the structure was still of sufficient quality to determine the overall topology of the quaternary structure of a head to tail tetramer with all 4 tails interacting with the head domains of the opposing dimer.

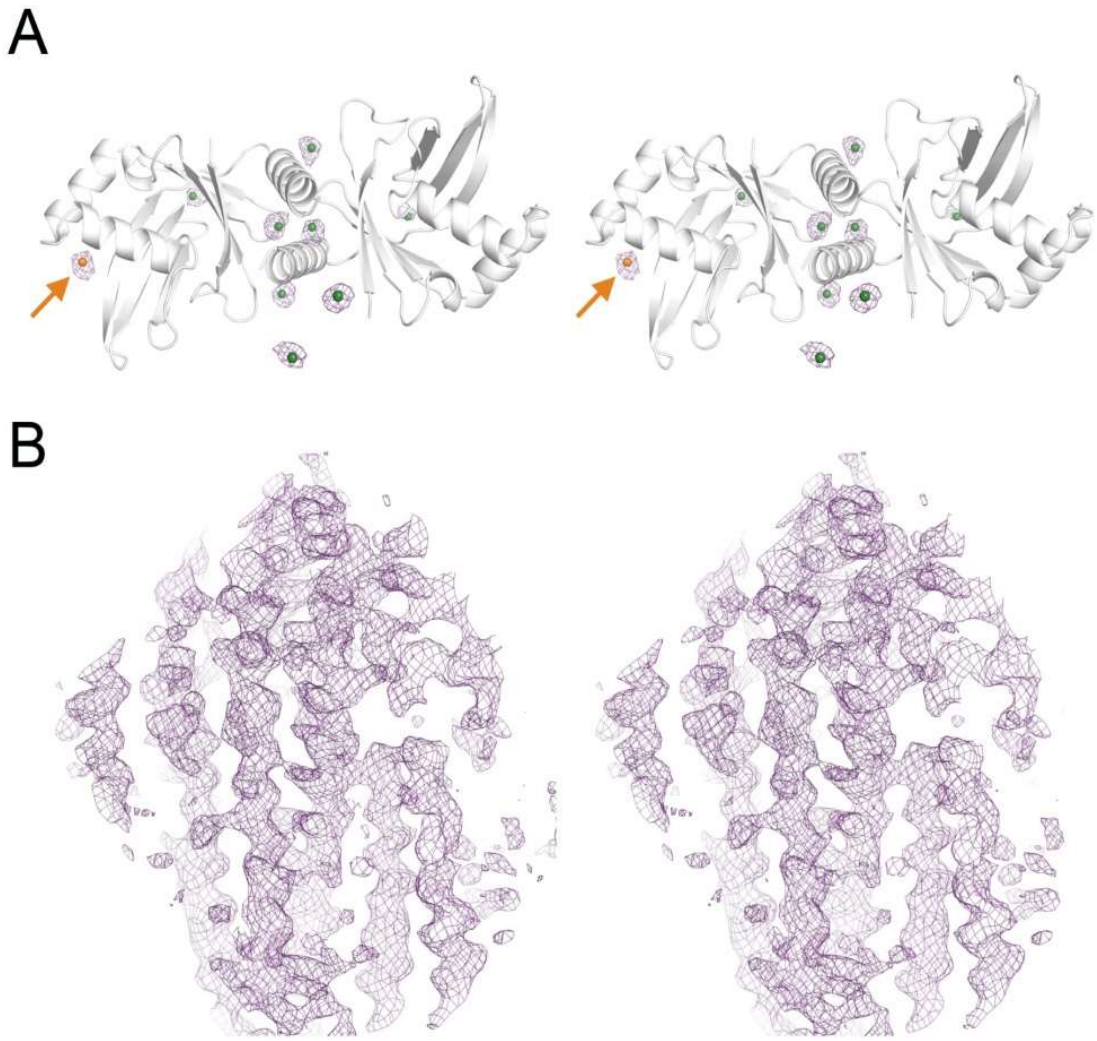
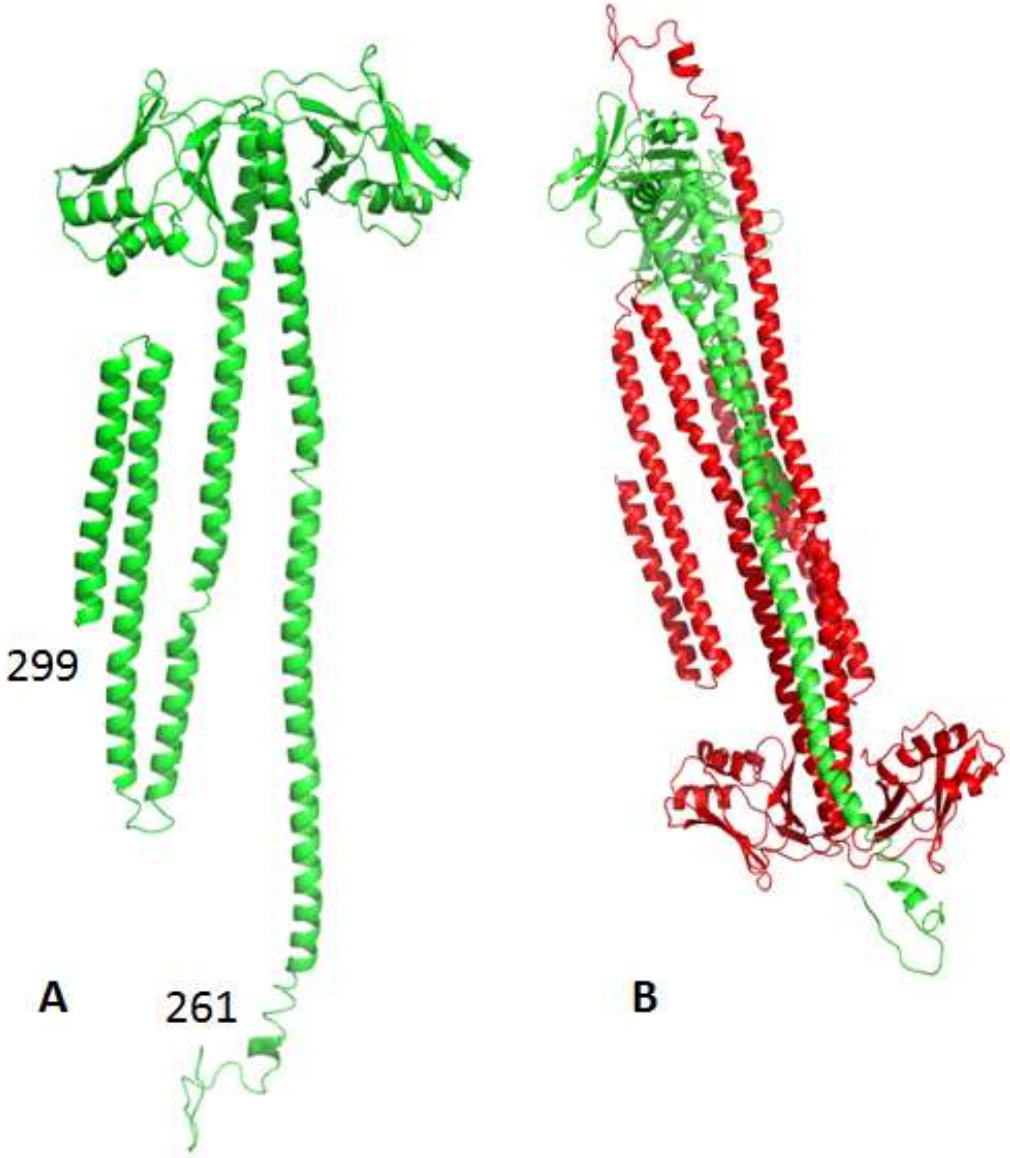



**Figure 2.4 Stereo image of heavy atom sites and resulting density modified maps.** A) Stereo image of the head domains of XRCC4 with the Selenium site (orange) and Arsenic sites (green). B) Density modified maps showing new secondary structure elements within the C-terminus of the XRCC4 tails.




***Fi*gure 2.5 Crystal structure of a more C-terminally extended XRCC4 Dimer (A) and tetramer (B).** The structure of full length XRCC4 showing several new secondary structure elements. The trajectories of the tails are asymmetric. B) The XRCC4 tetramer structure showing the interaction of one tail with the cleft between opposing head domains and the other tail forming a multi helix bundle.

From the current structure of full length XRCC4, it would appear that XRCC4 self association does allow for simultaneous XRCC4 tetramerization and XLF binding, providing a potential mechanism wherein the XRCC4 tetramers stabilize formation of multifilament bundles.

### 2.4.2 XRCC4 forms tetramers through the formation of a multi-helix bundle and asymmetric head/tail interactions with opposing head domains.

Based on details observed in the full length XRCC4 structure as well as sequence conservation, two highly conserved regions were targeted for mutagenesis. Patch 1 corresponds to a region on the underside and face of the XRCC4 head domain of protomer A while Patch 2 spans a cluster of conserved residues in the middle of the multi-helix bundle. Patch 1 (Arg3, Glu25 and Val33) creates a cleft on the head domain which interacts with the tail of the opposing dimer (Figure 2. 6)

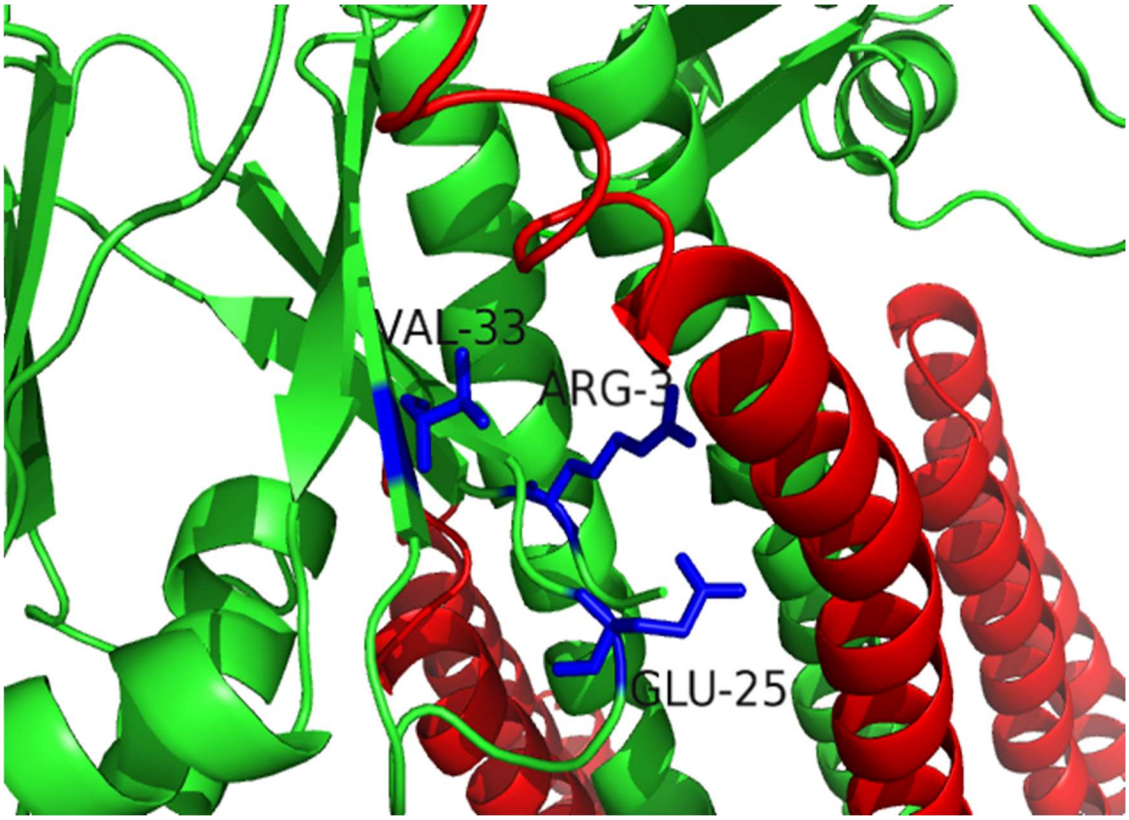



**Figure 2.6 Patch 1 creates a cleft which interacts with the opposing C-terminal tail.** Conserved residues R3, E25 and V33 create a cleft (blue) on the head domain of chain A (green) into which an extended alpha helix from an opposing dimer is bound (red).

The second patch consists of 4 highly conserved residues in the middle of the coiled-coil (Gln145, Lys146, Glu 147 and Arg150) (Figure 2.7). All residues in both patches were mutated to alanines and the resulting proteins were used for further biochemical analysis to probe complex formation and DNA end bridging.

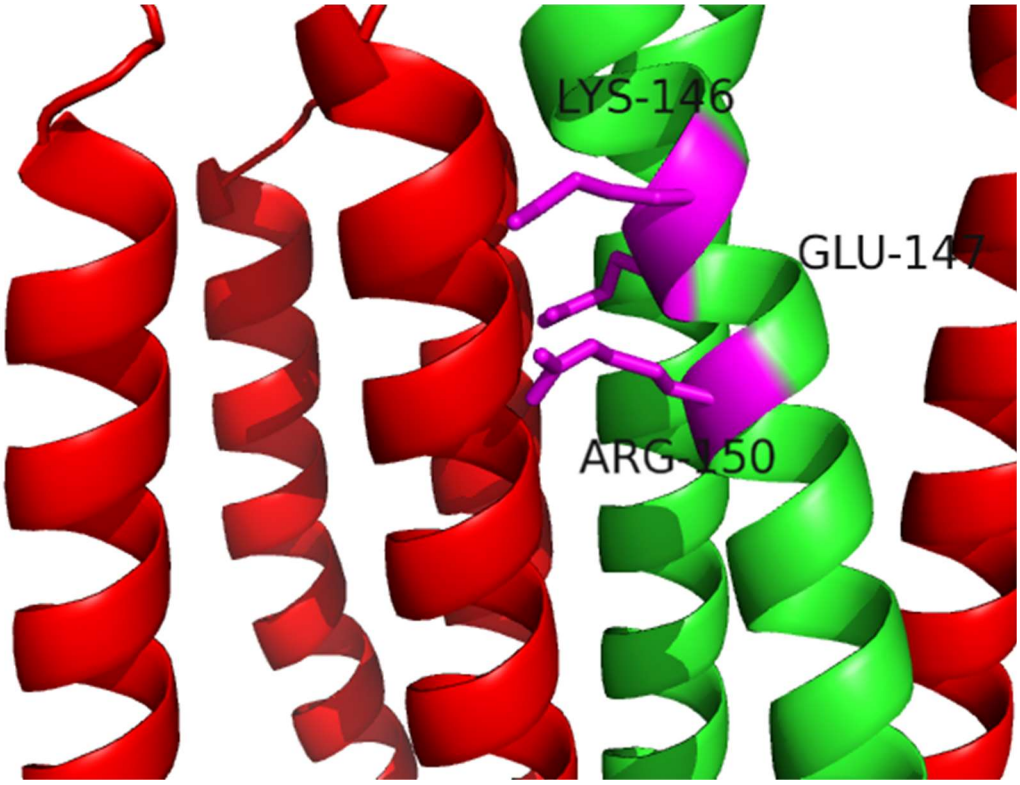



**Figure 2.7 Patch 2 stabilizes a multi-helix bundle.** Residues K146, E147 and R150 form a continuous charged face on each of the alpha helical tails of XRCC4 (green) which interact with the tails of the opposing dimer (Red). This interaction occurs between all 4 of the tails and their opposing partner, stabilizing a multi-helix bundle.

Folding of the mutants were verified using CD spectroscopy. While patch 2 did have a slightly higher percentage of alpha helical content, the proteins were indeed folded. With the proteins being verified as being folded properly, we next sought to characterize their fitness in XRCC4 tetramerization as well as if there were any other defects in the known functions of XRCC4.

Owing to the extended stokes radius of XRCC4, analytical ultracentrifugation (AUC) has proven inconclusive in the past to determine the mass of the XRCC4 complex formed. Accordingly, a different approach was adopted. Size exclusion chromatography with multi angle light scattering (SEC MALS) was selected as it can unambiguously determine the molecular mass of a molecule independent of its conformation. Using SEC MALS, wild type XRCC4 was found to occur in an equilibrium of tetramers and dimers while

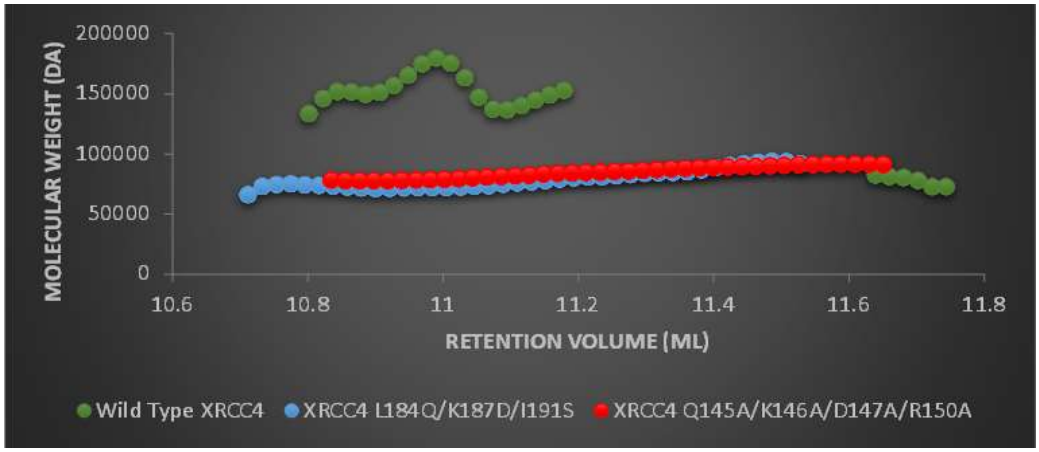



**Figure 2.8 Plot of XRCC4 molecular weights as a function of elution volume as read out by SECMALS.** WT XRCC4 (green) shows a transition from tetramer to dimer over the course of the elution. The previously characterized tetramerization mutant (red) exists only as a dimer in solution. The newly identified mutant (blue) also exists solely as a dimer in solution.

XRCC4$^{L184Q/K187D/I191S}$ existed only as a dimer in solution. Mutants of patch I had the opposite effect in that they favoured tetramers rather than dimers. Patch II mutants, however, were exclusively dimeric in solution (Figure 2.8). This finding was observed with mutant XRCC4 obtained purified from three separate purifications existing exclusively as a dimer in solution. Even when concentrated

to 525 μM, the protein did not form tetramers in solution. While defective for tetramer formation, these mutants were still capable of binding DNA, the BRCT domains of LigIV and XLF at wild type levels (Figure 2.9 A,B and C respectively). Patch II being a true dimer represents the first definitive separation of function mutant of XRCC4. This is consistent

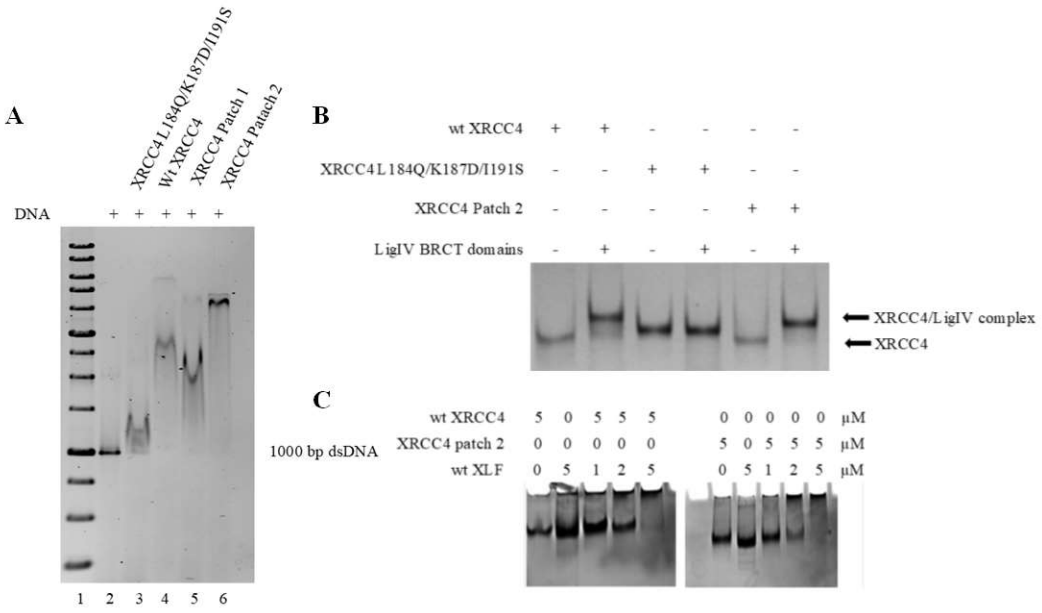



**Figure 2.9 The tetramerization defective mutant of XRCC4, patch 2, retains all other XRCC4 functions.** A) DNA binding was confirmed by gel shift assays which clearly show that the mutant (lane 6) binds as well as, if not better than WT (lane 4) Lane 1 represents a molecular weight ladder of dsDNA. B) Ability to interact with LigIV was confirmed through protein/protein EMSA. C) XLF binding was also assayed trough EMSA analysis. The tetramer defective mutant retains all 3previously known XRCC4 functions.

with previously observed analytical ultra-centrifugation data in which mutants that targeted the potential head to head interface did not have a profound effect on XRCC4 tetramerization (Lee, K.Y.W., 2013). In order to further verify that the

observed 4 helix bundle was the minimum requirement for tetramerization, C-terminal truncations of the protein were generated. The domain boundaries were determined following limited proteolysis of XRCC4 with 4 different proteases (figure 2.10).

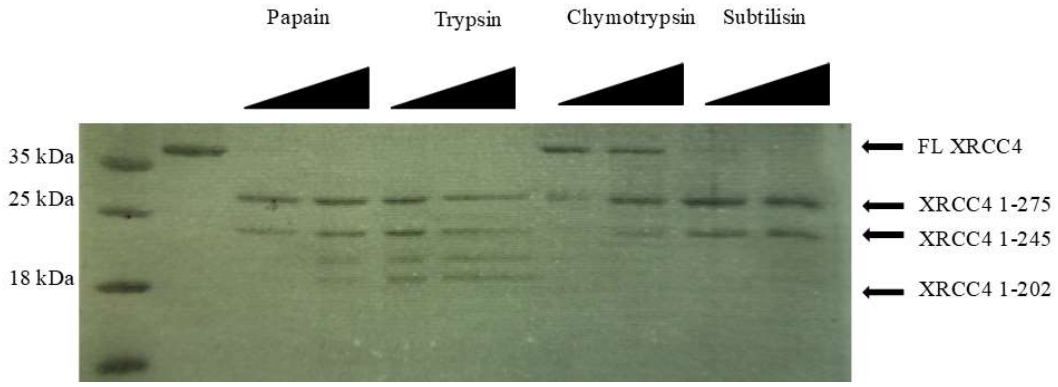



**Figure 2.10 Limited proteolysis of XRCC4 revealed stable domain boundaries within the C-terminal tails of the protein.** SDS page analysis of digested fractions showed several common bands, resistant to all 4 proteases used. Bands were excised and analyzed via Mass Spectrometry to determine fragment composition and subsequently stable, C-terminal domain boundaries of XRCC4.

Following Mass spectrometry analysis, three C-terminal truncations of the protein were generated based on their common observance and resistance to all 4 proteases. C-terminal truncations at residues 275, 245 and 202 were generated and assayed for their ability to tetramerize.

Finally, XRCC4$^{157}$ which completely lacks the multi helix bundle was assayed for its ability to form tetramers by SEC MALS. Even at concentrations considerably higher than the kD of the dimer-tetramer equilibrium, no tetramers

were detected. As this truncation of XRCC4 still retains the head domains, but does not tetramerize, it is further evidence that the head to head mechanism of interaction is not the true tetramer interface.

### 2.4.3 XRCC4 tetramerization is dispensable for DNA end bridging but the multi-helix bundle is absolutely required

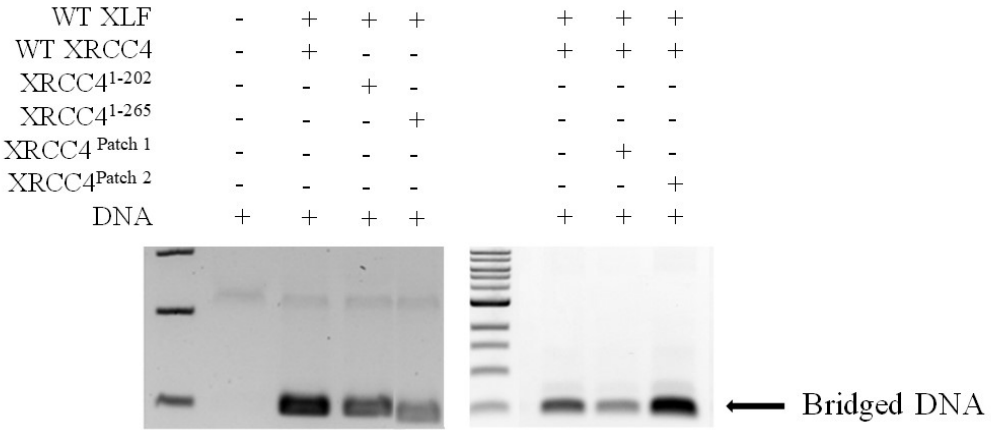



**Figure 2.11 Bridging of DNA ends by XRCC4-XLF complexes.** C-terminal truncations and mutants of XRCC4 (2μM final concentration) were incubated with XLF (2μM final concentration) and analyzed for ability to bridge DNA. ends *in vitro*. Successful bridging was read out by recovery of the 500 bp fragment following de-proteinating of DNA fragments. Lanes 1 and 6 are DNA molecular weight markers.

The filament disrupting mutant as well as C-terminal truncations of XRCC4 were assayed for their ability to bridge DNA ends *in vitro.* These studies revealed that the patch II mutant is capable of bridging DNA ends. In addition to this, C-terminal truncations of XRCC4 (Δ202 and Δ267) which are capable of

forming the multi-helix bundle are capable of end bridging when in complex with XLF (Figure 2.11). This suggests that XRCC4s contribution to end bridging filaments is through a head to head interaction with XLF that does not require any additional protein for the CUR of XRCC4.

## 2.5 Discussion

### 2.5.1 The multi-helix bundle of XRCC4 is sufficient to stabilize tetramers

Tetramerization of XRCC4 was previously reported; however, it remained unclear if the tetramer was stabilized by interaction between head to head or a multi-helix bundle at the end of the helical stalk domains of each monomer. Previous mutational studies based on the truncated crystal structure of XRCC4 were unable to delineate the true tetramerization interface. Interestingly, C-terminal truncations of XRCC4 that include the multi-helix bundle are still capable of bridging DNA ends when in complex with XLF. The complete structure of full length XRCC4 further suggested that all four helical stalks are involved in stabilizing the helix bundle. While the interface was asymmetric with respect to the localization of the C-terminal tails (202-336), all 4 helices from the coiled-coil region were found to contribute in a radial manner to the formation of the multi-helix bundle. Residues involved in the interaction are conserved and their mutation to alanine residues irrevocably eliminated the ability of XRCC4 to tetramerize. Conversely, mutations made in the newly discovered head-tail interacting regions had no effect on the ability of XRCC4 to form a tetramer in solution.

While it is possible that the C-terminal tails do stabilize a head to head model of tetramerization through the formation of alternative protein/protein interactions, the fact that no tetramers were observed by SEC-MALS of XRCC4$^{1-157}$ supports the helical stalk model of tetramerization, not the head to head model.

### 2.5.2 XRCC4 tetramerization is not essential for DNA end bridging

Mutations of the highly conserved patch 2 region (Gln145, Lys146, Glu 147 and Arg150) within the extended coiled coil abolished XRCC4s ability to form tetramers yet this mutant was still capable of bridging DNA ends, however only when forming a complex with wild type XLF. Unlike all previously published XRCC4 mutants, this patch 2 mutant is competent for all other XRCC4 functions (DNA binding, LigIV binding and XLF binding as well as simple filament formation). Previous models have suggested that the formation of XRCC4-XLF multi-filament bundles (functional end bridging filaments) requires the C-terminal tails of XRCC4 to mediate tetramerization, thereby stabilizing the bundles. As the tetramer defective mutant was able to bridge as well as, if not better than, wild type XRCC4, it is clear that this model is incorrect. This is in spite of the fact that these tails are highly conserved, even in residues which are not known kinase substrate residues. Ultimately, this implicates the C-terminal tails of XRCC4 in some other aspect of the repair pathway that has yet to be determined.

The largest question this work raises is that of the biological relevance of XRCC4 tetramerization. Tetramerization was first demonstrated *in vitro* and has

been confirmed to occur in a concentration dependent manner (Modesti et al., 2003). To date, however, there has been no evidence of this being a de-facto necessary function for XRCC4 *in vivo*. Multiple models built around XRCC4 tetramerization have been proposed, specifically wherein the tails stabilize inter-filament bundles through the formation of tetramers. As this behavior is dispensable for bridging, previous models are no longer accurate and must be revised.

Is XRCC4s tetramerization even biologically relevant? In order to answer that question, *in vivo* testing of the tetramer defective mutant must be carried out. Such studies are now capable owing to the newly discovered mutant that represents a true separation of function mutant.

## 2.6 Conclusion

In this work, we solved the crystal structure of full length XRCC4. The new structure revealed several new structural elements and provided insight into construction of mutations which disrupt XRCC4 tetramerization. Mutational analysis and light scattering demonstrated that XRCC4 forms tetramers through the formation of a multi-helix bundle within the coiled coil region.. The interaction is asymmetric between the two protomers which form the dimeric unit. This tetramerization is not required for DNA binding, LigIV interaction or XLF binding. Furthermore, this mechanism of self association is not required for DNA end bridging when in complex with XLF as a tetramer defective mutant is still capable of performing this function.

While the structure was not determined to high enough resolution to assign identities to sidechains, the overall topology map was sufficient. In order to ascertain whether or not tetramerization of XRCC4 is biologically relevant, *in vivo* complementation studies and VDJ recombination assays of the tetramerization defective mutant will be required.

This work reports the first true separation of function mutant observed in XRCC4 and allows for the testing of individual functions of the protein and was only made possible by finally obtaining crystals of the full length XRCC4 complex.

## 2.7 Appendix

| Mutant | Forward | Reverse |
|---|---|---|
| $XRCC4^{1-203}$ | GCATAATAAATTATTAAATGCAGCTCAATAACGAGAAAAGGACATC | GATGTCCTTTTCTCGTTATTGAGCTGCATTTAATAATTTATTATGC |
| $XRCC4^{1-245}$ | GATGAGGAAAGTGAAAACCAATAAGATCTCTCTGGGTTGGCTTC | GAAGCCAACCCAGAGAGATCTTATTGGTTTTCACTTTCCTCATC |
| $XRCC4^{1-275}$ | GAAAAAGGAGACAGCGATAGCAAAGATATCTTGGGAC | GTCCCAAGATATCTTTGCTATCGCTGTCTCCTTTTTC |

**Table 2.2 List of primers used for the generation of XRCC4 C-terminal truncations**

# Chapter 3 Nej1 binds DNA through an extended C-terminal region and its regulation plays a role in repair pathway choice in *Saccharomyces cerevisiae*

## 3.1 Abstract

Despite being essential for DSB repair, the molecular mechanism by which members of the XRCC4 family bind their DNA substrates remains unknown. Low resolution structural techniques and biochemical data have given some information about regions necessary for mediating interaction and preference for larger DNA substrates with cooperative binding being observed. In order to further investigate the mechanism by which these proteins bind DNA non-specifically, mutational, biochemical and mass-spectroscopy analysis were carried out. While specific DNA binding residues could not be definitively identified, proteolysis and biochemical characterization determined that Nej1 binds DNA through an extended interface in the extreme C-terminus of Nej1 (yXLF). Additional work demonstrated that phosphomimetic substitution of Dun1 substrate residues results in a decrease in affinity for DNA and an increase in DSB repair by ESDSA.

## 3.2 Introduction

Similar to all other eukaryotes, *Saccharomyces cerevisiae* are vulnerable to double strand break formation. A single unrepaired break can be lethal to a haploid yeast cell. In yeast, however, the pathway preference is reversed relative to mammals. Instead of NHEJ, *Saccharomyces cerevisiae* primarily employs homologous recombination repair. Despite this difference in pathway preference, several core NHEJs repair factors are found in both eukaryotes and *Saccharomyces cerevisae* (LigIV, XRCC4, XLF and Ku).

Interestingly, there is no definitive DNAPKcs homologue in yeast, representing the largest difference between yeast and higher eukaryotes for NHEJ. Beyond this, the two pathways are very well conserved. Initial DSB end detection is mediated by the yKu70/80 complex which subsequently recruits the Mre11/Rad50/Xrs2 (MRX) complex to the break.

Nej1 is the yeast XLF homologue and shares many similar features, specifically a preference for longer pieces of DNA (>100 bp) and interaction with Lif1, the yeast homologue of XRCC4 (Cavero et al., 2007., Sulek et al., 2007). Furthermore, it has recently been shown that Nej1 is capable of promoting re-adenylation of DNL4 (the yeast homologue of mammalian LigIV), potentially allowing for multiple ligation events to be carried out by the same molecule (Chen, X. and Tomkinson, A.E. 2011). One key difference exists between the Nej1/Lif1 and XLF/XRCC4 complexes. While XLF and XRCC4 are predominantly stabilized by interaction via head to head associations of their N-terminal domains, it is the C-terminus of Nej1 that stabilizes interaction with the head domain of Lif1, as determined by yeast two hybrid analysis (Desphande and Wilson, 2007).

Despite some differences in pathway selection and repair components, yeast represent a good model organism for studying double strand break repair. Yeast can be easily manipulated to facilitate study of key components without compromising the integrity of repair. This ease of control makes *Saccharomyces cerivisiae* an excellent system to study DSB repair in eukaryotes.

Previous work has demonstrated that the DNA binding of Nej1 is localized to the extreme C-terminus of Nej1, specifically residues 268-342 (Sulek et al., 2007). This region contains several clusters of basic residues which were proposed to mediate DNA binding. Mutational analysis of all three clusters (K283A/K285A/K287A, K291A/K293A/R295A and K332A/R333A/K334A), however, showed no detectable changes in DNA binding. While there was an *in vivo* DSB repair defect in the K332A/R333A/K334A mutant, this was likely disruption of a putative nuclear localization sequence (NLS) within Nej1. As mutations of these positively charged residues had no effect on DNA binding, sequence analysis and homology were used to generate new potential DNA binding mutants within the Nej1 C-terminus.

Multiple conserved residues targeting both structural (i.e. proline) and potential DNA binding residues were mutated in an attempt to define the specific DNA binding residues within the extreme C-terminus of Nej1. These studies were further supplemented by hydrogen-deuterium exchange mass spectroscopy analysis of the complex formed by DNA and Nej1. Additionally, regulatory residues within the C-terminal tails were targeted in order to assay the role of phosphorylation within the C-terminal tails.

## 3.3 Materials and methods

### 3.3.1 Growth and expression of full length Nej1

Plasmids containing Hexa-histidine tagged full length Nej1 were transformed into *e.Coli* BL21 DE3 cells via the heat shock method. Cells were plated on agar plates supplemented with 25 µg/ml ampicillin. A single colony was used to inoculate an overnight culture of LB media, also supplemented with 25 µg/ml ampicillin. This overnight culture was then sub-cultured into 4 X 1 litre cultures of LB media. Once the cell cultures reached an OD600 of 0.5, expression was induced by addition of 1mM IPTG. Expression was carried out by a further 3 hours of shaking at 37°C at 225 RPM. Cells were harvested by centrifugation at 3315 X g for 15 minutes (Avanti J-30I centrifuge with the rotor JLA 9.1000). Cell pellets were recovered and flash frozen and stored at -80°C until further use.

### 3.3.2 Purification of full length Nej1

Cell pellets were re-suspended in a nickel binding buffer containing 20 mM Tris, pH 8.5, 1 mM BME, 750 mM KCl and 10% v/v glycerol. Protease inhibitors (benzamidine, Phenylmethylsulfonyl fluoride (PMSF), pepstatin A, aprotinin, and leupeptin) were added for a final concentration of 1 mM. Cells were lysed by three passes through a French pressure cell at 20000 psi. Lysate was clarified by centrifugation at 48000 X g for 45 minutes. Clarified lysate was then loaded onto a pre-equilibrated HIStrap-HP nickel affinity column (GE healthcare, United States of America). The column was then washed with binding buffer supplemented with 30, 45 and 60 mM imidazole before the protein was

eluted with a 210 mM imidazole wash. Following IMAC, Nej1 containing fractions were dialyzed into a heparin binding buffer consisting of 20 mM Hepes, pH 8.0, 100 mM KCl, 0.5 mM EDTA and 2.5 mM DTT. Nej1 was then applied to a pre-equilibrated heparin HPtrap column (GE healthcare). Following a wash of 10 CV with binding buffer, Nej1 was eluted in a linear gradient from 100 to 400 mM KCl over 120 minutes. Purified Nej1 was then buffer exchanged into a buffer containing 20 mM Tris (pH 7.0), 1 mM EDTA, 150 mM KCl and 5 mM DTT. Fractions containing Nej1 were concentrated by centrifugation in a nano-sep centricon, flash frozen with liquid N2 and stored at -80°C until further use.

### 3.3.3 Mutagenesis of full length Nej1

Primers for 12 different mutants were ordered from biobasic Inc. Primers were designed for overlap extension mutagenesis. Primers were re-suspended to final concentration of 125 ng/µl and mixed with 10 ng of template (pMJ4108), 22 µl of ddH2O and 25 µl of 2X iPFU master mix (Froggabio, Canada) and were amplified using the following parameters: 1) initial denaturation at 95°C for 5 minutes, 2) melting at 95°C for 60 seconds, 3) annealing at 62°C for 60 seconds, 4) extension at 72°C for 6 minutes. Steps 2-4 were repeated 20 times followed by a final extension at 72°C for 20 minutes. Following mutagenesis, the resulting product was treated with fast digest DpnI from Fisher scientific for 15 minutes at 37°C. DH5α chemically competent cells were transformed with the digested product. Colonies were selected, and plasmids extracted via alkaline lysis mini

prep as outlined in the High-speed plasmid mini-kit (GeneAid). All mutations were verified by sequencing (Robarts research institute).

### 3.3.3 Growth and expression of Nej1 C-terminus truncations

Nej1$^{268\text{-}342}$ was expressed in the same manner as wild type full length Nej1.

### 3.3.4 Mutagenesis of the Nej1 C-terminus

The Nej1 C-terminus was mutated in the same manner as wild type with the only difference being that the template plasmid encoded Nej1$^{268\text{-}342}$ (pMJ4510), and not full length Nej1.

### 3.3.5 Purification of the Nej1 C-terminus

Cell pellets were re-suspended in a nickel binding buffer containing 20 mM Tris, pH 8.5, 1 mM BME, 1.5 M KCl, 3 mM imidazole and 10% V/V glycerol. Cells were lysed by three passes through a French pressure cell at 20000 psi. Lysate was clarified by centrifugation at 48000 X g for 45 minutes. Clarified lysate was then loaded onto a pre-equilibrated HIStrap-HP nickel affinity column (GE healthcare, United States of America). The column was then washed with binding buffer supplemented with 30, 45 and 60 mM imidazole before the protein was eluted with a 210mM imidazole wash.

Following initial nickel purification, cells were buffer exchanged into a Tobacco etch virus protease (TEV) digestion buffer containing 20 mM Tris, pH

8.0, 3 mM BME, and 500 mM KCl. The high salt concentration is necessary so as to partially remove contaminating genomic DNA.

Nej1 was then mixed with the in a ratio of 1 mg of TEV for every 15 mgs of Nej1 and digested for 2 hours at room temperature. Following digestion, the reaction contents were re-exchanged back into the initial nickel binding buffer and re-applied to the His-trap column so as to separate the Nej1 from TEV and the hexa-histidine tag. The purified protein was collected in the flowthrough. Purity was verified by SDS page analysis. The resulting protein was then dialyzed into a buffer containing 20 mM TRIS, pH 8.0, 1 mM EDTA, 10 mM DTT, 200 mM KCl and 10% v/v glycerol.

A260/280 ratios were recorded, and the samples were found to be free of contaminating genomic DNA that had been previously observed. DNA free Nej1 was then flash frozen in liquid N2 and frozen until further use.

### 3.3.6 Generation of DNA substrates

DNA substrates were generated using two separate methods: 1) annealing of complimentary synthetic olignucleotides (15- 50 BP in length) and 2) Polymerase chain reaction for large substrates (200-2620 BP). For smaller substrates, page purified oligonucleotides were ordered from Biobasic Inc. The lyophylized strands were re-suspended in Tris-EDTA (TE) buffer to a concentration of 600 μM. Strands were mixed in a 1:1 ratio and annealed heating

to 95°C followed by cooling at a rate of 1° per minute over 75 minutes. Annealing was verified by native page analysis.

Longer DNA substrates were generated by mixing appropriate primers (125 ng forward and reverse) and template (10 ng) with 25 µl of 2X iPFU master mix (Froggabio, Canada) and amplified using the following thermocycler parameters: 1) initial denaturation at 95°C for 5 minutes, 2) melting at 95°C for 60 seconds, 3) annealing at 62°C for 60 seconds, 4) extension at 72°C for 1 minute per KB length of substrate. Steps 2-4 were repeated 27 times followed by a final extension at 72°C for 5 minutes. Resulting products were purified using a fast-q cleanup kit (Qiagen).

### 3.3.7 DNA binding assays and electrophoretic mobility shift assays

Purified Nej1 (full length and the C-terminus) was dialyzed overnight into a reaction buffer composed of 50 mM Tris, pH 8.0, 150 mM KCl, 0.5 mM EDTA and 10% v/v glycerol. A fixed amount of DNA (100 ng) resuspended in TE buffer was incubated with increasing concentrations of Nej1 (0.5-20 µM) for 30 minutes at room temperature. Samples were then loaded onto a 6% TBE native page gel and electrophoresed at 150 volts for 100 minutes. Gels were stained by soaking in 1XTBE buffer supplemented with 50 µg/ml. Gels were visualized using a Gel-Doc EZ imager (Biorad). Defects in binding were determined by a lower affinity for DNA at identical protein concentrations.

### 3.3.8 Unfolding and isolation of DNA free Nej1 C-terminus

Purified Nej1$^{268-322}$ bearing a C-terminal hexa-histidine tag was re-applied to a HIStrap Nickel affinity column (GE, United States of America). The column was then washed with 5 column volumes of nickel binding buffer (20 mM Tris, pH 8.5, 1 mM BME, 1.5 M KCl, 3 mM imidazole and 10% V/V glycerol) supplemented with 3 M Gaunidine hydrocholoride. The protein was left to rest in this unfolding buffer before being washed with 20 column volumes of guanidine free binding buffer. The unfolded protein was eluted with a binding buffer containing 210 mM imidazole. Nucleic acid content was measured by a direct A260 reading as this fragment contains no residues which absorb at 280 nm.

## 3.4 Results

### 3.4.1 Mutation of Dun1 substrate residues result in a decrease in Nej1 affinity for DNA

Previous work identified Serines 297 and 298 as phosphorylation sites of the cell cycle checkpoint kinase Dun1 (Ahnersong, P. and Jasckon, S.P., 2006). Phosphorylation of these residues results in the recruitment of the SRS helicase to the site of the break. In order to assay the impact of phosphorylation of these residues on DNA binding, alanine and phosphomimetic substitutions were made at both residues. Alanine substitutions at either or both residue had no effect on DNA binding. In contrast, glutamic acid substitutions at both residues had a marked impact on DNA binding (Figure 3.1).

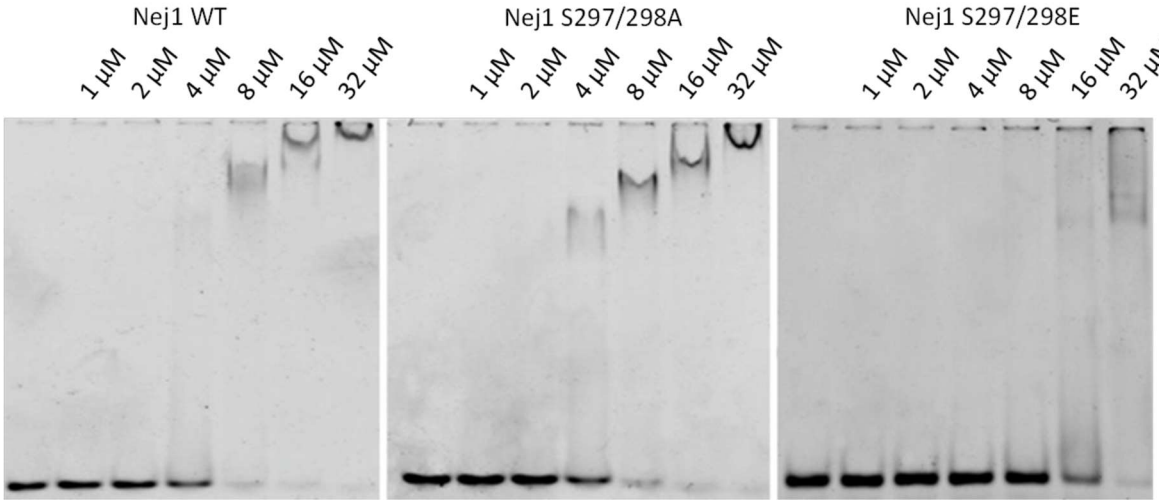
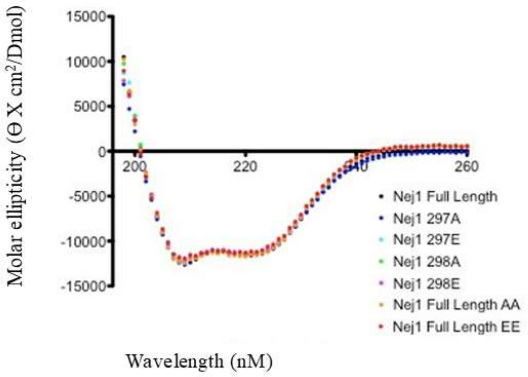



**Figure 3.1 Phosphomimetic substitutions of Dun1 phosphorylation sites reduces Nej1s affinity for DNA.** Full length Nej1, Nej1AA and Nej1EE were all incubated with a 50 bp ds DNA substrate. Alanine substitutions at the Dun1 substrate residues had no impact on DNA binding. Phosphomimetic mutations at these sites resulted in a marked decrease in affinity for the 50 bp DNA substrate

Single glutamic acid substitutions at either position had no observable impact on DNA binding, suggesting that a double phosphorylation event that is required to reduce Nej1 DNA binding. CD spectroscopy was carried out for all mutants being assayed. All mutants were shown to be properly folded relative to the wild type control (Figure 3.2).

While these mutations did result in a decrease in DNA binding affinity, DNA binding was not completely abolished. In order to determine the specific mechanism by which Nej1 binds DNA, a more in-depth analysis of the DNA binding domain was undertaken.




| Nej1 Mutant | % Alpha helical | % Beta strand | % Turns | % Disordered coil |
|---|---|---|---|---|
| FL | 38.9 | 13.0 | 19.2 | 28.9 |
| S297A | 38.2 | 13.1 | 19.4 | 29.3 |
| S297E | 38.9 | 13.0 | 18.7 | 29.5 |
| S298A | 38.9 | 13.1 | 18.3 | 29.7 |
| S298E | 38.2 | 13.0 | 18.7 | 30.1 |
| AA | 38.5 | 12.8 | 18.4 | 30.3 |
| EE | 36.3 | 14.3 | 19.5 | 29.9 |

**Figure 3.2 Nej1 Phosphomimetic mutants retain their secondary structure composition.** All individual and combinatorial mutants of Dun1 substrate residues were analysed via CD spectroscopy. All are properly folded indicating that any defects in DNA binding are not due to simple global misfolding.

### 3.4.2 The C-terminus of Nej1 binds DNA through an extended interface

As previous mutational analysis targeting positively charged regions within the C-terminus of Nej-1 had no impact on DNA binding, other potential DNA binding interacting motifs were targeted. The C-terminal fragment (residues 268-342) was analyzed for any putative structure-based DNA binding motifs. While no consensus motif (e.g. AT hook) was identified, several clusters of alternating structural (i.e. proline) and basic amino acids were identified. Six mutations of residues which may contribute to structure-based DNA binding were created, P278G, G282P, P286G, P292G, P315G and P329G. A further 3 residues were mutated based on sequence conservation and positive charge (R281M, K317M and K319M). In addition, the previously characterized regulatory mutant, S297E/S298E and a further highly conserved residue, E312L were mutated. All mutations were made in an expression vector containing only the DNA binding

domain (resides268-342 of Nej1). All 12 constructs were expressed and purified to homogeneity (Figure 3.3). Each of the individual mutations had no effect on

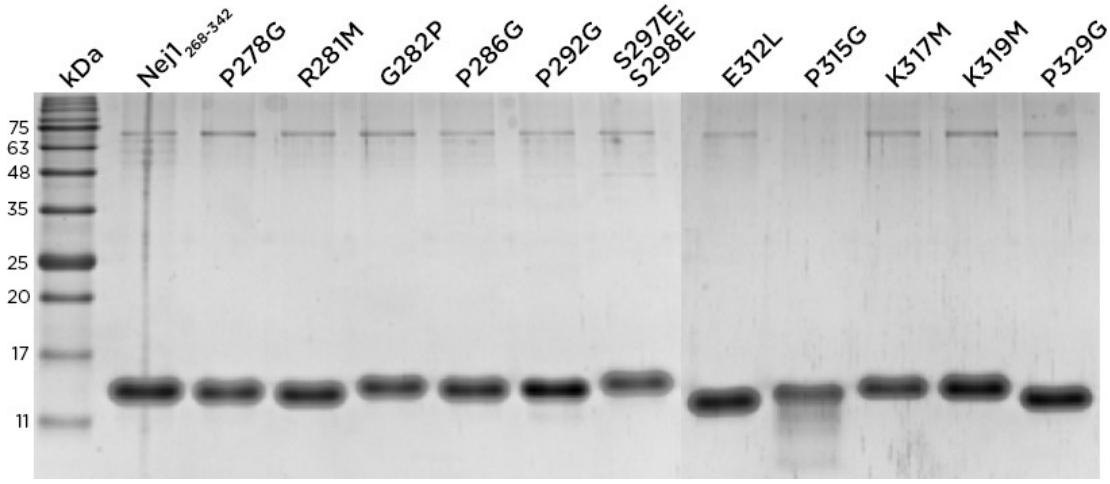



**Figure 3.3 SDS page gel of all 12 Nej1 DNA binding domain mutants.** All 12 mutants were purified to homogeneity using a two-step purification process. With the exception of P315G, no mutants showed obvious signs of heterogeneity or proteolysis.

DNA binding of the Nej1 C-terminus to a 1000 bp substrate (Figure 3.4).

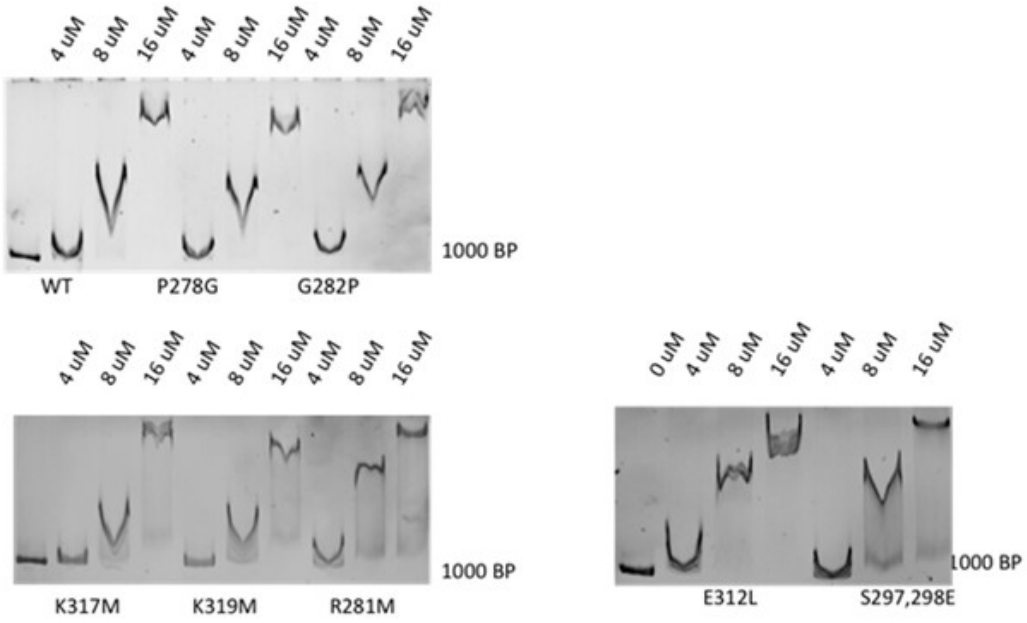



**Figure 3.4 EMSA analysis of individual DNA binding point mutants in the DNA binding domain (268-342).** Each of the individual DNA binding mutants were incubated with a 1000 BP dsDNA substrate. Following electrophoresis and staining, none of the individual mutants were observed to have a marked impact on DNA binding.

While most of the biochemical characterization of the DNA binding of Nej1 was carried out using a C-terminal fragment of Nej1, the affinity of the DNA binding domain on its own for DNA is considerably higher than that of full length Nej1. A previously identified regulatory mutant, S297E/S298E was demonstrated to possess a lower affinity (~5fold decrease in affinity) relative to wild type protein. When this same mutation was made in the DNA binding domain construct, there was no observable defect in binding (Figure 3.4 C). As it was

possible that this higher affinity was masking a DNA binding defect that would be observed within the full-length protein, the mutational analysis repeated using full length Nej1. As with XRCC4, patches of residues were mutated together as opposed to individual residues. Two regions were selected based on sequence

conservation as well as their prevalence in other non-specific DNA binding motifs. The two regions were P278G/G282P/P286G/P292G (patch 1) and P315G/E316L/K317A/K319A (Patch 2). The mutations were introduced into

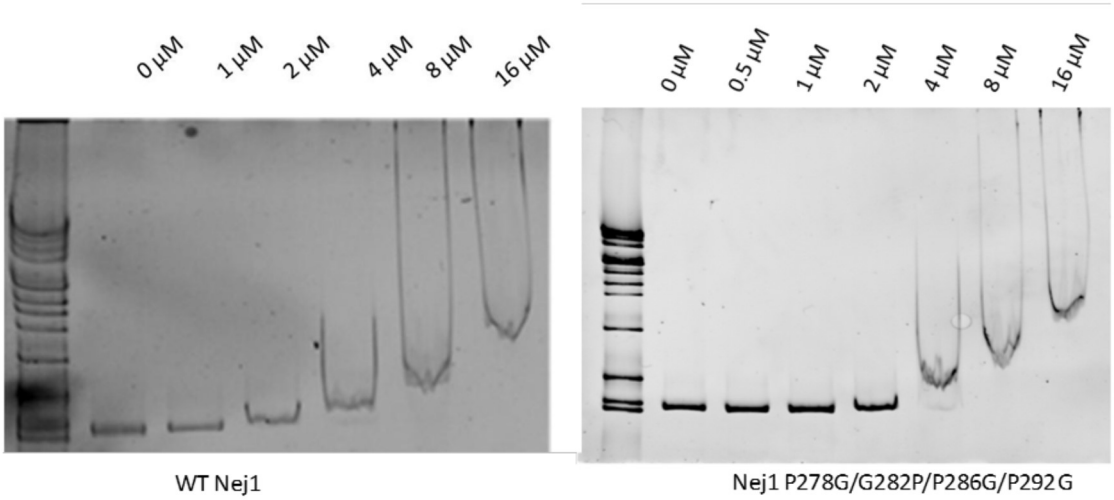



**Figure 3.5 EMSA analysis of multiple mutations of potential DNA binding residues in the Nej1 DNA binding domain.** Mutations of entire clusters of potential DNA binding residues also had no effect on DNA binding within the C-terminus of Nej1. Increasing concentrations of Nej1 were incubated with a 1000 bp blunt DNA substrate.

both the full-length protein as well as the C-terminal DNA binding domain. As with every previous other mutation made to this region, mutating the residues in bulk had no effect on DNA binding to a 1000 base pair substrate (Figure 3.5) or shorter.

### 3.4.3 DNA binding induces a conformational change in Nej1

As the mutational analysis of the C-terminus of Nej1 was unable to identify a residue or residues within the C-terminus responsible for mediating DNA binding, we subsequently began exploring other avenues through which the DNA binding motif of Nej1 could be determined. Hydrogen-deuterium exchange (HDX) mass spectroscopy was carried out. Despite considerable effort, HDX was unable to identify the specific DNA binding region within the C-terminus as there was no observable change in deuterium uptake.

In order to determine if the C-terminal tail of Nej1 required the entire DNA binding domain or a more C-terminally truncated fragment would suffice, new constructs for the DNA binding domain were generated. While the C-terminus is conserved, the last 20 residues (with the exception of the NLS) show a lower degree of conservation. Based on these data, a new construct of Nej1$^{286\text{-}322}$ was generated.

Interestingly, Nej1$^{268\text{-}322}$ had such high affinity for genomic DNA that purification required the presence of 2M KCl in order to yield a sample free of DNA. Protein purified without high salt washes, while bound by genomic DNA,

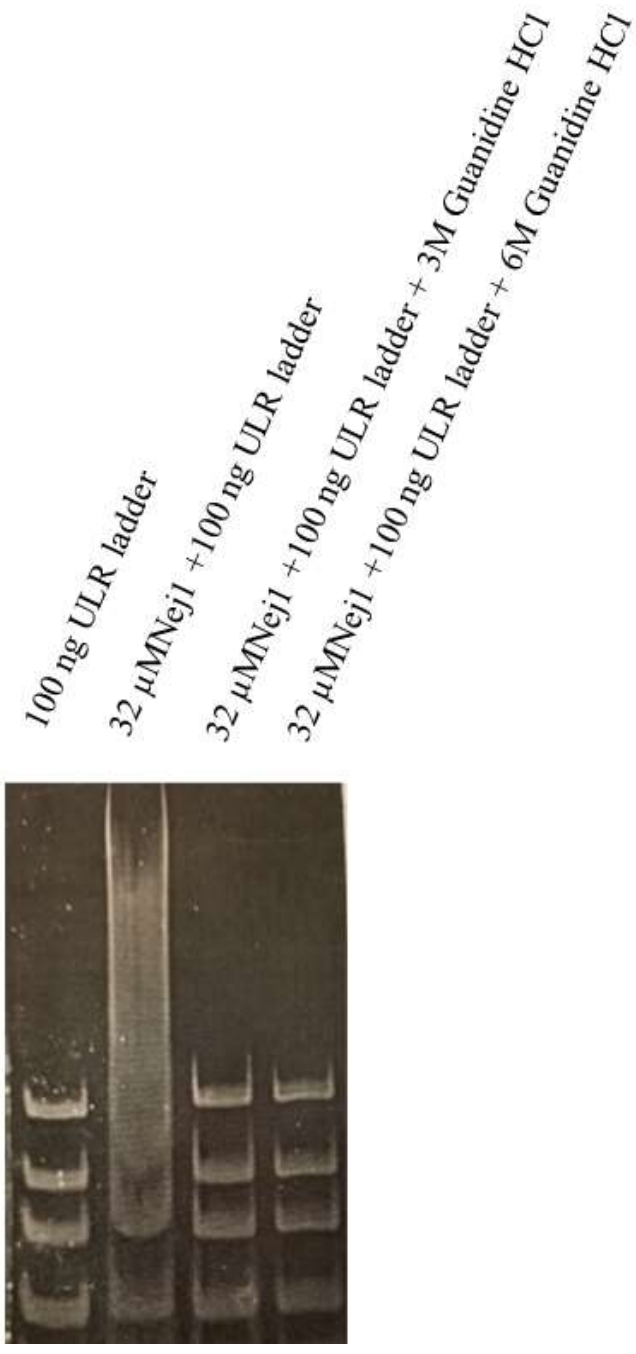



**Figure 3.6 Unfolded Nej1 C-terminal tails are unable to bind DNA.** Following unfolding by incubation with guanidine HCl, DNA free tails of Nej1 were unable to re-bind DNA substrates of any length. This strongly suggests that once a necessary structural motif is removed, the DNA binding ability of Nej1 is abolished.

was still capable of binding additional DNA. In order to assess the DNA binding

capabilities of this fragment, it was necessary to remove all DNA contamination. This was carried out in two ways, 1) a high salt wash during the purification of the fragment or 2) treatment of the protein with guanidine HCl.

Most surprisingly, Nej1$^{268-322}$, upon treatment with either salt or guanidine, was completely incapable of re-binding DNA despite all potential DNA binding residues being present (Figure 3.6). This strongly suggests that the interaction between the Nej1 C-terminus and DNA is not simply due to electrostatic interactions but requires a defined structure. Once the structural motif is abolished through denaturation, the ability to bind DNA is also diminished.

## 3.5 Discussion

### 3.5.1 Dun1 phosphorylation of Nej1 may affect repair pathway choice through reduction of affinity for DNA

Phosphorylation of Nej1 by the kinase Dun1 results in DSBs in yeast being repaired by a synthesis dependent strand annealing mechanism. This is due, in part, to recruitment of the SRS helicase to the site of the break (Carter et al., 2009). As our results show that phosphomimetic substitutions of the Dun1 substrate residues resulted in a marked decrease in DNA binding by Nej1, it is possible that the decrease in affinity for DNA following phosphorylation is what mediates a transition to DSB repair by a single stranded annealing like mechanism. Individual mutations of these residues had no discernable impact on DNA binding. An important consideration is that phosphorylation is a post

translational modification that cannot be completely matched by mutational analysis. CD-spectroscopy confirmed that the phosphomimetic mutants were properly folded and therefore the DNA binding defect is due to a true decrease in affinity and not a global misfolding effect.

### 3.5.2 Nej1 binds DNA through an extended C-terminal interface

Previous yeast two hybrid work has shown that it is the tails of Nej1that mediate binding to the head domain of Lif1 (Desphande et al.,2007). Work reported here further demonstrates by deletional analysis of the 268-342 region that the C-terminal tail of Nej1 also functions as the chief DNA binding region of the protein.

Prior studies have shown that mutations of highly conserved, basic clusters within the C-terminus of Nej1 have no impact on DNA binding (Sulek et al., 2007). While charge based interactions are a very common mechanism for protein/DNA interactions, they are by no means the sole mechanism of interaction. Structure based DNA binding motifs are also common mechanisms by which proteins bind DNA non-specifically. Motifs resembling putative AT-Hook motifs (PRG repeats and palindromes within the primary sequence) were targeted for mutational analysis that may contribute to non-specific, charge based DNA interactions. (R281, R304, K317, K319)

That several different series of mutations, both individual and combinatorial substitutions were unable to disrupt DNA binding suggests that the

C-terminus of Nej1 uses an extended interface to bind DNA. The only mutations capable of decreasing DNA binding were phosphomimetic substitutions of the Dun1 substrate residues. Furthermore, the fact that de-natured Nej1 268-342 does not retain DNA binding after being refolded strongly suggests that the entirety of the tail is required for DNA binding and that it is a structural interaction that is required to mediate this very strong interaction. The only way to completely abolish DNA binding of Nej1 is through deletion of the entire DNA binding domain.

### 3.5.3 The DNA binding domain of Nej1 forms a stable complex with DNA

When expressed as a free peptide, the DNA binding domain of Nej1 binds DNA with an extremely high affinity. Initial purifications were contaminated with genomic DNA from the bacterial cells from which the protein was isolated during purification.

Regardless of the means by which the DNA was removed from the C-terminal fragment (guanidine treatment or extremely high salt), the fragment was unable to re-bind DNA.

In addition to the extensive mutagenic analysis of the DNA binding domain, other techniques were employed to investigate the mechanism of DNA binding by Nej1. Although these methods produced less concrete results, some information could be derived. The most surprising of these results was a complete lack of any change in deuterium uptake of the complex plus/minus DNA when

assayed by Hydrogen-deuterium exchange mass spectrometry (HDX). Despite a massive excess of DNA to protein (DNA present in a 10-fold excess of the 3.4 μM kD), no change in uptake was observed in the tails, or in any of the protein for that matter. The most likely reason for this outcome is that the resulting protein-DNA complex was so strong, it was resistant to denaturation and proteolytic digest following HD exchange. If the DNA bound complex would not fly in the mass-spectrometer then there would be no significant changes in the deuteration profile of the protein. This is consistent with other proteolytic and nucleolytic characterization of this complex which have shown the tail region to be strongly resistant to degradation.

### 3.6 Conclusion

Nej1 is a key protein in yNHEJ repair. In this work, we examined its role in DSB repair by further characterizing the mechanism through which it and other XLF homologues bind their DNA substrates. Phosphomimetic mutants of Nej1 were found to result in a marked decrease in DNA binding affinity while alanine substitutions at these positions still showed wild type levels of DNA binding. Since phosphorylation of these sites has been shown to alter repair pathway choice, our result suggests that regulation of Nej1 through phosphorylation occurs through reduction of the DNA binding affinity.

In an attempt to definitively determine the mechanism by which members of the XRCC4 family bind their DNA substrates, we undertook a combination of approaches to determine the specific residues which mediate DNA binding in

Nej1. Twelve individual and two multiple patch mutants were generated that targeted the C-terminal DNA binding domain of Nej1. Surprisingly, every mutant tested showed no measurable impact on DNA binding. Given the number of techniques that have proven unsuccessful at determining the mechanism by which Nej1 (yXLF) binds DNA, the best remaining option may be a gain of direct structural information through co-crystallization or NMR studies of the DNA binding domain in complex with DNA. To this end, numerous combinations of Nej1 tails (268-322, 268-342, 280-322) and varied DNA substrates were used for crystallographic studies but did not yet yield diffracting crystals of the complex. Capturing this complex in-crystallo remains the best way of discerning the mechanism by which XLF, Nej1 and members of this family bind their DNA substrates.

## 3.7 Appendix

| Mutant | Forward | Reverse |
|---|---|---|
| R281M | CCAAATAGCATGGGTAAAATTAAACCAAAGAC | AATTTTACCCATGCTATTTGGTCGAGCCTC |
| G282P | GCCGCCCTAAAATTAAACCAAAGACAGATTTC | TTAATTTTAGGGCGGCTATTTGGTCGAGC |
| P286G | AAATTAAAGGAAAGACAGATTTCAAGCC | CTGTCTTTCCTTTAATTTTACCGCGGCTATTTGG |
| P292G | GATTTCAAGGGAAAATCTAGAGAAAGTAGTAC | CTAGATTTTCCCTTGAAATCTGTCTTTGG |
| S297E, S298E | CTAGAGAAGAAGAAACCAGTTCGCAACTTCGCC | CGAACTGGTTTCTTCTTCTCTAGATTTTGGC |
| R304M | CGCAACTTATGCTGGAGAATTTTTCTG | CTCCAGCATAAGTTGCGAACTGGTACTACT |
| E312L | GAATCTCTGGCTACACCAGAAAAAACAAAATC | GGTGTAGCCAGAGATTCAGAAAAATTCTCC |
| P315G | GGCTACAGGAGAAAAAACAAAATCAAGTTC | GTTTTTTCTCCTGTAGCCTCAGATTCAG |
| K317M | CCAGAAATGACAAAATCAAGTTCTTCTTTGG | GATTTTGTCATTTCTGGTGTAGCCTCAGATTC |
| K319M | GAAAAAACAATGTCAAGTTCTTCTTTGGTCG | GAACTTGACATTGTTTTTTCTGGTGTAGCCTC |
| P329G | GAATACGGACAGAAAAAAAGAAAATTTGG | TTTCTGTCCGTATTCTTCGACCAAAGAAG |

**Table 3.1 List of primers used to generate mutants within the DNA binding domain of Nej1**

# Chapter 4 Multiple protein/protein and protein/DNA interactions are required to form end bridging filaments during NHEJ

## 4.1Abstract

XRCC4 and XLF form simple filaments through interactions within their N-terminal head domains. These simple (single) filaments are insufficient, however, for DNA end bridging and stimulation of ligation (Andres et al., 2011). Formation of functional end bridging filaments is dependent on the C-terminal tails of both proteins (Andres et al., 2011). Tails of XLF mediate primary DNA binding while the tails of XRCC4 are thought to be required for formation of the full end bridging complex. Recently, it was found that that these filaments form sleeves that rapidly slide along the length of duplex DNA. What remains unknown, however, is the functional composition of XRCC4 and XLF within these filaments. Here, we employed several biochemical and structural characterization, to ascertain the mechanism by which these filaments bridge DNA ends. Our results suggest that XLF is capable of forming DNA-dependent filaments in the absence of XRCC4 and that XRCC4 induces a morphological change in these XLF/DNA filaments *in vitro*. Finally, crystallographic have generated the first reported crystals of an XLF/XRCC4/DNA complex that diffract to 5 Å.

## 4.2 Introduction

One of the most critical aspects of DSB repair is the need for DNA ends to remain associated following formation of a DSB. Should the two ends diffuse away from one another, successful end relegation would be difficult. Previously, end synapsis has been thought to be mediated by the Ku/DNA-PKcs complex

(Defazio, L.G et al., 2002). With the more recent discovery that XRCC4 and XLF can form highly stable DNA end bridging filaments, there is growing consensus that end bridging may be largely dependent on XLF/XRCC4 and not Ku/DNAPK. XLF/XRCC4 Filaments have been shown to form sliding sleeves or multi-filament bundles able to slide along duplex DNA and also align DNA ends *In Vitro* (Brouwer et al., 2016). Importantly, these filaments have been shown to form within the cell (Ropars et al., 2011). Unknown, however, are the biological requirements for formation of functional end bridging structures. Unanswered questions include: are both proteins present in equal numbers, do both proteins contribute to DNA binding and perhaps, most importantly, does the DNA wrap around a protein or does the protein coat the DNA? Determining the details of interaction is key to our understanding of the mechanism by which the ends are made compatible and brought together for ligation and strand release.

Here, we employed several structural and biochemical techniques so as to determine how XLF/XRCC4 filaments are formed. Specifically, we looked at individual contributions of each of three factors- XRCC4, XLF and DNA to the formation and functioning of filaments. Based on these analyses, we suggest that the DNA binding activity of XLF is responsible for engaging DNA within the filament and that XRCC4 is likely required for mediating other protein/protein interactions necessary for repair.

## 4.3Materials and methods

### 4.3.1 Purification of XRCC4 and XLF

XRCC4 was purified as previously described in chapter 2. XLF was initially purified using IMAC under conditions identical to XRCC4. Following this, XLF was buffer exchanged into a heparin buffer containing 20 mM Hepes (pH 8.0), 1 mM EDTA, 5 mM DTT and 200 mM KCl. The sample was then applied to a pre-equilibrated Heparin-HP column (GE Healthcare, United States of America). Protein was eluted over a linear salt gradient of 200-500 mM KCl over 120 minutes at a flow rate of 1 ml/min. Purity was verified via SDS page. Fractions containing XLF were pooled, concentrated and stored at -80°C until further use.

### 4.3.2 Negative stain transmission electron microscopy

Continuous carbon grids were purchased from Electron Microscopy Services. Grids were glow discharged for 30 seconds. Following glow discharge, protein-DNA complexes were blotted onto the discharged surface for 30 seconds. Once the samples had dried, a 1% Uranyl acetate stain was applied to the grids for 30 seconds. Excess stain was removed and the grids were cured by exposure to a high temperature heat lamp for 30 seconds. Fixed grids were then visualized using a Jeol 1200EX-Temscan transmission electron microscope at a maximum magnification of 12500x.

### 4.3.3 Analytical size exclusion chromatography

Purified XLF was extensively dialyzed into an SEC buffer containing 150 mM KCl, 20 mM Tris, pH 8.0, 0.5 mM EDTA and 10 mM DTT. XLF (at various concentrations) were injected onto a bio-rad SEC650 gel filtration column (Biorad, United States of America) equilibrated with the same SEC buffer at a rate of 0.5 ml/min. Determination of size was made in reference to a calibration curve of pre-determined molecular weight standards and retention volume

### 4.3.4 Generation of DNA substrates

DNA substrates were generated in a manner identical to those in chapter three. Briefly, smaller substrates were generated by annealing complementary synthetic oligonucleotides. Longer substrates were generated via PCR.

### 4.3.5 DNA electrophoretic mobility shift assays

Products from mixing protein and DNA were resolved using a 0.6% TBE agarose gel initially run at 150 volts for 5 minutes followed by a further 720 minutes at 80 volts. Gels were stained in 0.5X TBE buffer supplemented with 50 µg/µl ETBR. Gels were de-stained by successive washes with ddH2O with. Gels were visualized using a Chemidock gel visualizer (Biorad, United States of America).

### 4.3.6 Two-dimensional gel electrophoresis analysis

XRCC4 and XLF were extensively dialyzed into a reaction buffer containing 20 mM Hepes (p.H. 8.0), 0.5 mM EDTA, 120 mM KCl and 10%

glycerol. Proteins were mixed in a 1:1 ratio either in the presence or absence of 500 ng of linearized pUC19 dsDNA. Reactions were then loaded onto a 6% acrylamide TBE gel and electrophoresed at 200 volts for 120 minutes. Gels were stained with coomasie brilliant blue and bands containing the complex were excised from the gel. Harvested bands were boiled for 5 minutes in 2X SDS page load dye. Excised bands were subsequently loaded into wells of a 9% denaturing acrylamide gel and electrophoresed at 150 volts for 90 minutes. Gels were again stained using coomassie brilliant and the band intensities were quantified using to determine the approximate contribution of each protein to DNA-dependent and independent filament formation (Abramof et al., 2004).

### 4.3.7 Crystallization of the XRCC4/XLF/DNA complex

Purified XRCC4 and XLF were extensively dialyzed into a crystallization buffer of 20 mM TRIS, pH 8.0, 1 mM EDTA, 10 mM DTT, 200 mM KCl and 10% v/v glycerol and mixed in a 1:1 ratio at room temperature and incubated for 30 minutes. Following this, DNA was added in varying concentrations to allow formation of XRCC4/XLF/DNA filaments. The hanging drop vapor diffusion method was used to screen for crystal formation. Drops were set by mixing 1µl of complex with 1µl of mother liquor and incubating over a well containing 1.5 M $NH_2SO_4$. In total, 750 different crystallization conditions were screened. Optimization of the initial hits consisted of additive screens (Hampton International, United States of America) as well as systematic optimization of DNA substrate size and individual component parameters within the

crystallization condition. DNA substrates consisted of a repeating 19 bp complimentary core flanked by varying lengths of self associating ssDNA extensions. Initial crystals of the XLF/XRCC4$^{1\text{-}157}$/DNA complex were with a mother liquor composed of 100 mM MES-NaOH (pH 6.5), 200 mM $MgCl_2$ and 10 % w/v PEG 4000. A second crystal form was obtained when the complex was mixed in a 1:1.2 ratio with a mother liquor composed of 100 mM Hepes (pH 7.5), 200 mM $Li_2SO_4$ and 25% W/V PEG 3350. The first crystal form initially diffracted to higher resolution and the crystals were physically stronger, as a result, this crystal form was selected for optimization. Form one was optimized through the addition of additives and controlled dehydration over varying concentrations of $NH_2SO_4$.

Crystals of the XLF/XRCC4$^{L184Q/K187D/I191S}$/DNA complex were only observed when mixed in a 1:0.95 ratio with a mother liquor composed of 100 mM Sodium Citrate/Citric acid (pH 5.4) and 20% w/v PEG 3350. Crystal growth required controlled micro-seeding of crystal nuclei into fresh drops equilibrated over 2.0 M $NH_2SO_4$. The highest resolution diffracting crystals required the addition of alcohol-based compounds to the reaction mix prior to vapor diffusion.

### 4.3.8 Silver staining of DNA-bound crystals

Putative crystals of the XRCC4/XLF/DNA complex were removed from crystallization drops and washed extensively with crystallization buffer. Crystals were then dissolved in protein storage buffer and run on a 9% denaturing SDS page gel. The gel was fixed in a solution of 50% methanol, 10% acetic acid for 20

minutes at room temperature. Following washes with ddH2O, the gel was activated by soaking in a solution of 200mg/L $Na_2S_2O_3$ for 1 minute and stained in a solution of 2g/L $AgNO_3$ for 20 minutes. Bands were developed until visibility using 37% v/v formaldehyde. Development was arrested by addition of sodium carbonate. Gels were imaged using a GelDOC easy imager (Biorad, United States of America).

### 4.3.9 X-ray diffraction Data collection and processing

X-ray diffraction data were collected at beamline 17-ID of Argonne National Laboratory. Images were collected at quarter degree oscillations at a rate of 12 images per second. A full 360° of data was collected for a total of 1440 images. Data was processed using iMOSFLM (Battye et al., 2011). Molecular replacement was carried out using the Phaser module from Phenix with different combinations of search models consisting of XRCC4 and XLF.

## 4.4 Results

### 4.4.1 Transmission electron microscopy analysis of XLF bound to DNA

It has been previously reported that XLF is capable of bridging DNA ends *in vitro* at high enough concentrations (> 2μM), however, it was unknown if XLF is capable of forming filaments in the absence of XRCC4.When diluted to concentrations appropriate for EM, no large oligomers were observed on the EM grids (Figure 4.1A). At identical protein concentrations in the presence of a 1000 bp dsDNA substrate, large, continuous filaments were observed (Figure 4.1B-D).

These filaments were considerably longer than a single 1000 bp piece of DNA (384 nm) suggesting that XLF must be bridging DNA ends in a linear fashion. Interestingly, no multi-filament bundles or junctions of multiple DNA ends were observed, consistent with the formation of simple XLF only filaments. Indeed, the length of the bundle was dependent on protein concentration, with higher XLF concentration resulting longer filaments. This is consistent with results from gel shift analysis where DNA shift events are concentration dependent. Increasing concentrations of XLF, even as high as 128μM reproducibly resulted in the formation of discrete species. A known DNA binding mutant of XLF (K293A) was also visualized by TEM. At identical protein and DNA concentrations as wild type XLF, no filaments were observed of $XLF^{K293A}$(Figure 4.3B). Furthermore, XLF concentrated to 425 μM displayed no signs of filament formation in the absence of DNA (Figure4.4). Size exclusion chromatography (SEC) showed that XLF still existed exclusively as a dimer in solution, even at high concentrations.

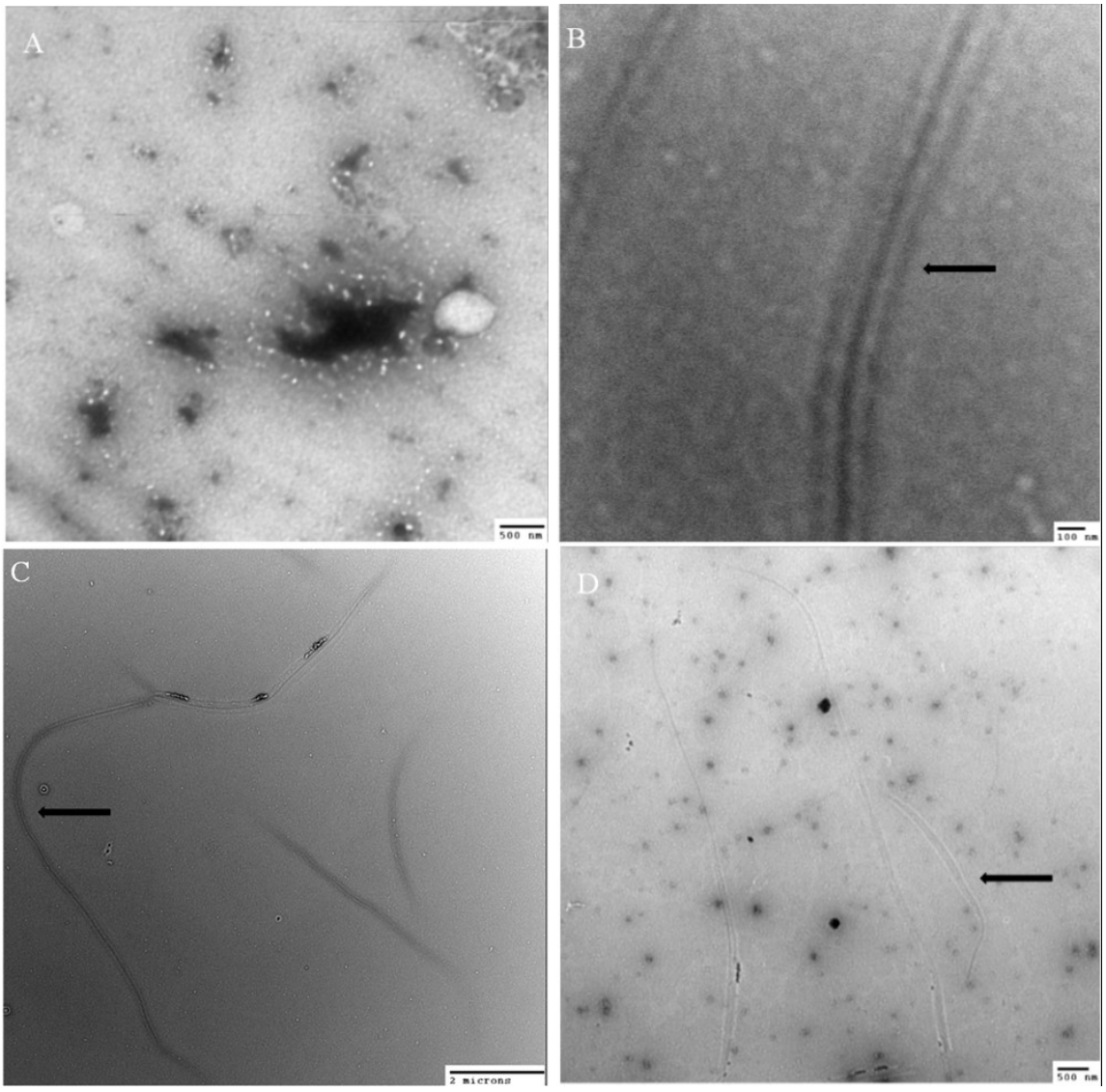



**Figure 4.1 Transmission electron micrographs of XLF/DNA filaments.** A) 22μM XLF blotted onto 400 mesh grids in the absence of DNA B) close up of an individual XLF-DNA filament, 78000 X magnification C) low magnification image showing the long filaments formed by XLF bound to DNA, 25000X magnification D) 22μM XLF pre-incubated with a 1000 bp dsDNA substrate, 78000 X magnification. All grids were stained with 1% uranyl acetate.

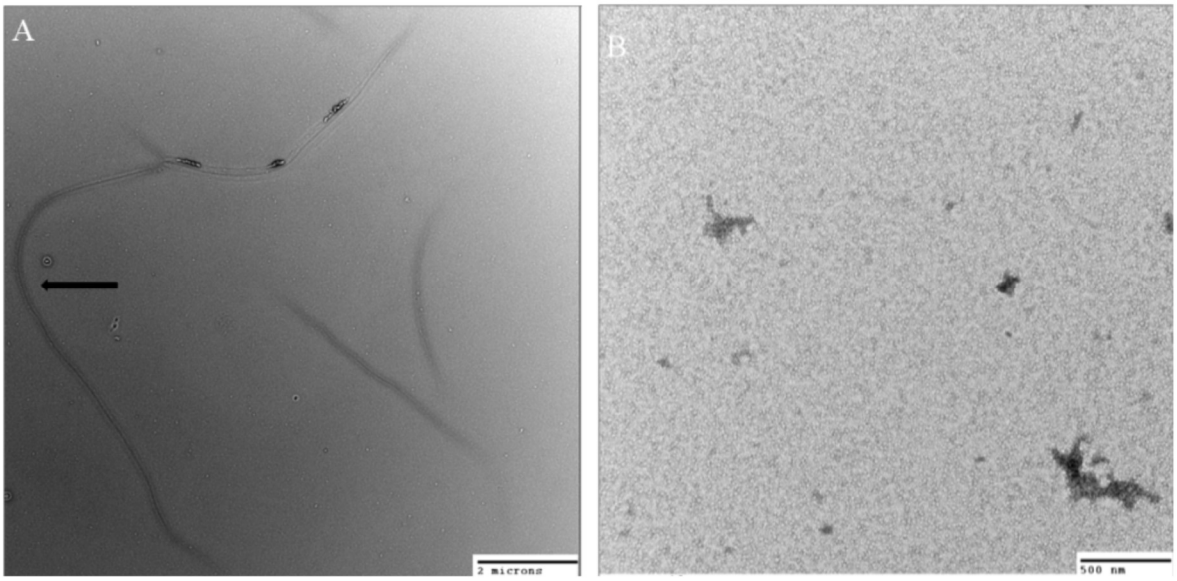



**Figure 4.2 Transmission electron micrographs of WT and DNA binding defective XLF with a 1000 bp dsDNA substrate.** A) 22 μM WT XLF bound to a 1000 bp dsDNA substrate. Clear fibers are seen at lengths far greater than the 384 μm in length expected of a 1000 bp DNA fragment, 78000 X magnification B) 500 μM $XLF^{k293A}$ incubated with a 1000 bp dsDNA substrate. Unlike wt XLF, no filaments are observed indicating that the formation of XLF only filaments is, in fact, DNA dependent 500X magnification.

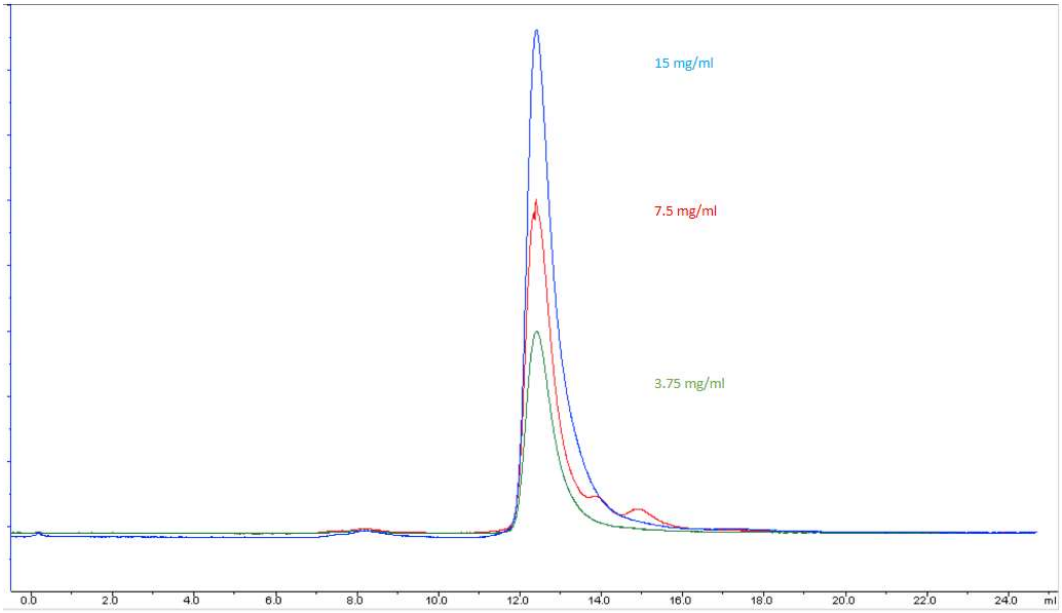



**Figure 4.3 Size exclusion chromatography analysis of XLF oligomerization.** Twofold increasing concentrations of WT XLF in the absence of DNA were analysed by SEC. Based on elution volumes, even the most concentrated XLF samples existed almost exclusively as a dimer in solution.

### 4.4.2 XRCC4 induces a morphological change in XLF/DNA filaments

Following the observation that XLF is capable of forming filaments in a DNA dependent manner, we sought further characterize the role that XRCC4 plays in filament formation and regulation. While the X-ray structure of the simple XLF/XRCC4filament has been solved by several groups, the structures are all lacking the DNA binding domains of both proteins as well as the DNA component of the filament. This represents one of the largest gaps in our structural knowledge of the complex.

A complex of XLF and DNA was formed in the absence and presence of various XRCC4 species were added to the filaments in order to assess what impact XRCC4 has on the pre-formed XLF/DNA filaments.

Addition of XRCC4 to XLF/DNA filaments resulted in distinct morphological changes of XLF filaments. This was observed both by TEM and electrophoretic mobility shift changes. EM micrographs showed a transition from single, long filaments (Figure 4.4 A) into multiple, condensed filaments. These new XRCC4/XLF/DNA species were observed across all XRCC4 populations analyzed, including full length (Figure 4.4B), tetramer defective (figure 4.4 C) and C-terminally truncated (figure 4.4 D).

Coupled with microscopy data, results from gel shift assays suggest that XLF is capable of forming filaments that are remodelled by XRCC4. Even at substoichiometric quantities of XRCC4 relative to XLF, all XRCC4 truncations as well as the LigIV/XRCC4 complex resulted in band tightening and an even more discrete band as observed by EMSA (Figure 4.6). This is contrary to what would be expected if XRCC4 and XLF were associating in a 1:1 ratio through their N-terminal head domains. Even as XRCC4 is titrated in, the resulting complexes migrate with an apparent lower molecular weight. If XRCC4 were simply associating with XLF through their head domains, then an increasing molecular weight filament would be expected and not the more discrete bands observed in Figure 4.5

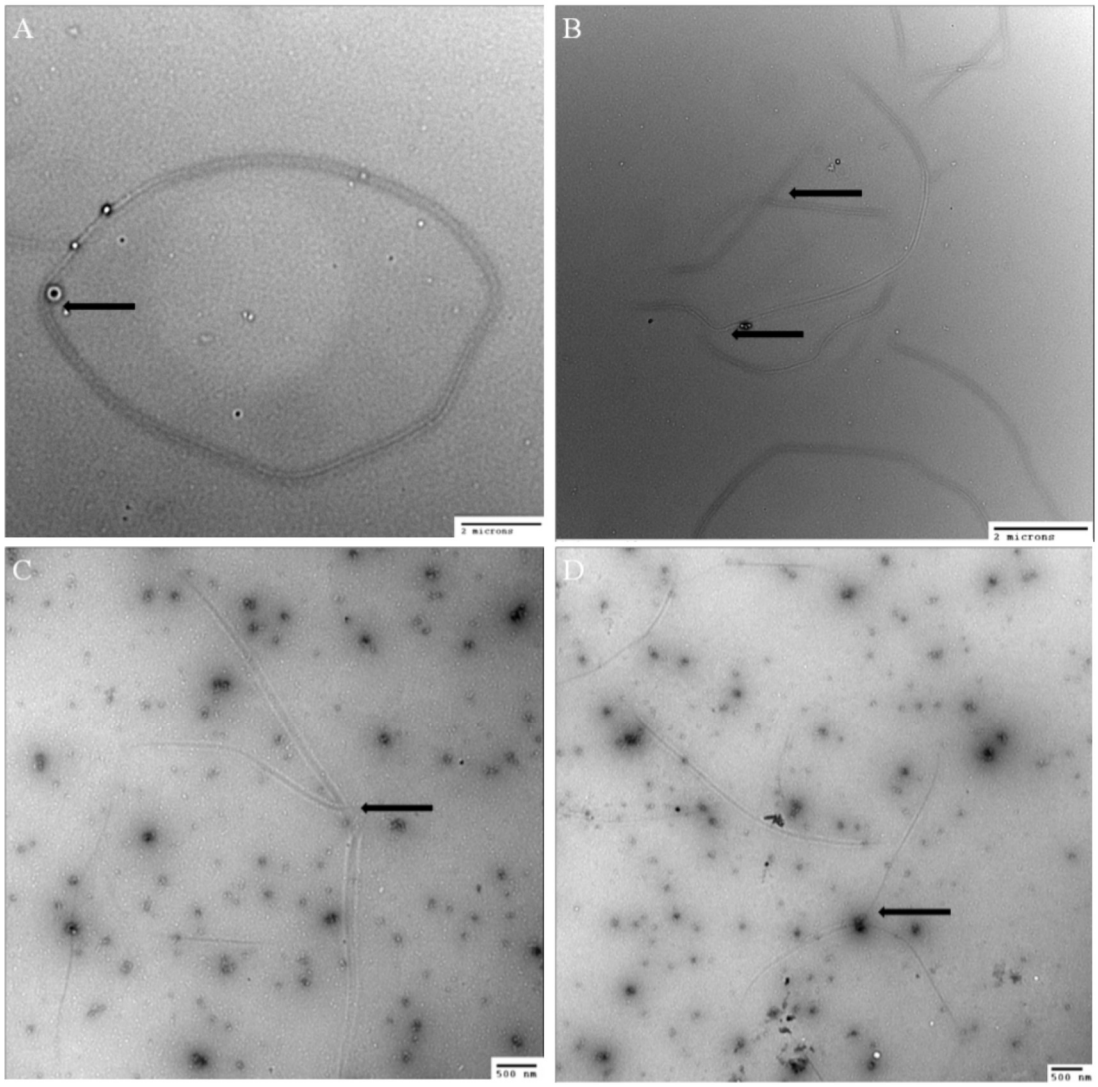



**Figure 4.4 Transmission electron micrographs of XLF/DNA complexes incubated with XRCC4.** A) A single XLF/DNA filament formed in the absence of XRCC4. B) Complex filament bundles formed when full length XRCC4 is added in equimolar concentrations to the XLF/DNA complexes 25000 X magnification. C) Inter-filament junctions formed when XRCC4$^{patch\ 2}$ is used in placed of WT XRCC4. D) Diffuse bundles formed in the presence of a C-terminally truncated XRCC4, 78000 X magnification

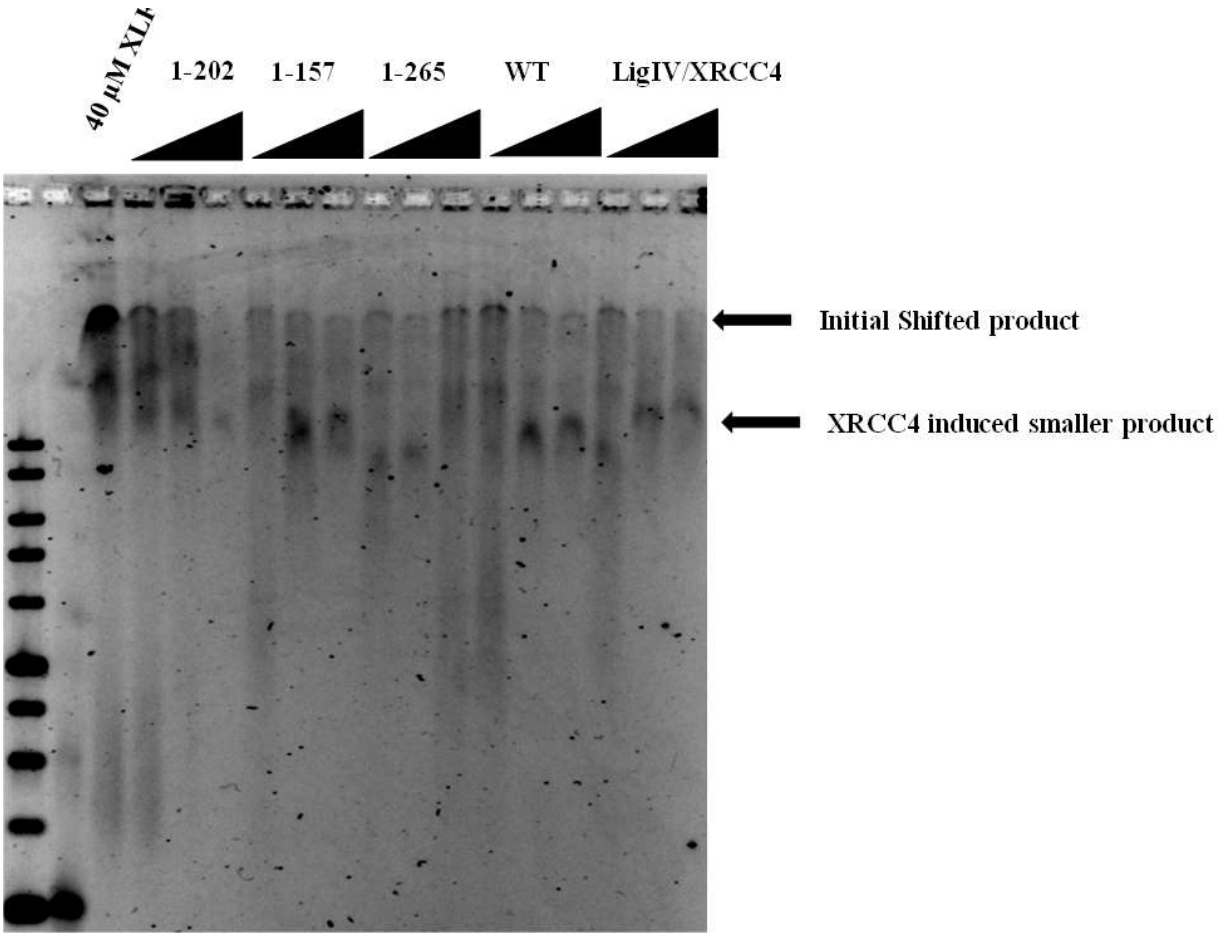



**Figure 4.5 Electrophoretic mobility shift analysis of XLF/DNA complexes incubated with XRCC4.** XLF/DNA filaments undergo a re-modelling event when incubated with XRCC4. Fixed concentrations of XLF(20 μM) and 1000 bp ds DNA were incubated with a titration of 1 to 10 μM XRCC4 and in all cases, the same discrete bands were observed

### 4.4.3 Crystallization of the XRCC4/XLF/DNA complex

It is known that XRCC4 and XLF form DNA end bridging filaments, however, the mechanism by which these filaments bind DNA remains unclear. In order to further characterize this mechanism, we sought to crystallize the final end bridging complex of XRCC4/XLF/DNA. In addition to wt XRCC4, mutants known to be defective in tetramerization were used to favor a complex filament incapable of incorporating into a multi-filament bundle. In this way the pool of

complex would be more likely to form a homogeneous population more amenable to crystallization.

Equally as important as the choice of XRCC4 mutant was the DNA substrate for crystallization. Initial attempts to crystallize the complex involved screening complexes formed with different lengths of blunt ended DNA. Results from these screens showed heavy amounts of phase separation and extremely bi-refringent crystals, which were composed only of DNA. While blunt ended duplex DNA of varying length have been used previously, the minimal length of a DNA bound by an XRCC4/XLF complex has yet to be conclusively determined. Rather than empirically testing arbitrary lengths of DNA substrate for crystallization, a series of self-associating substrates, dubbed the stacking series, were designed. These substrates were based around a common 20 basepair complementary core flanked by single stranded overhangs on the 3' end. Overhang length varied7 to 11 nucleotides and allowed for the overhangs to anneal with a second substrate, thereby creating elongated repeats of the core 20

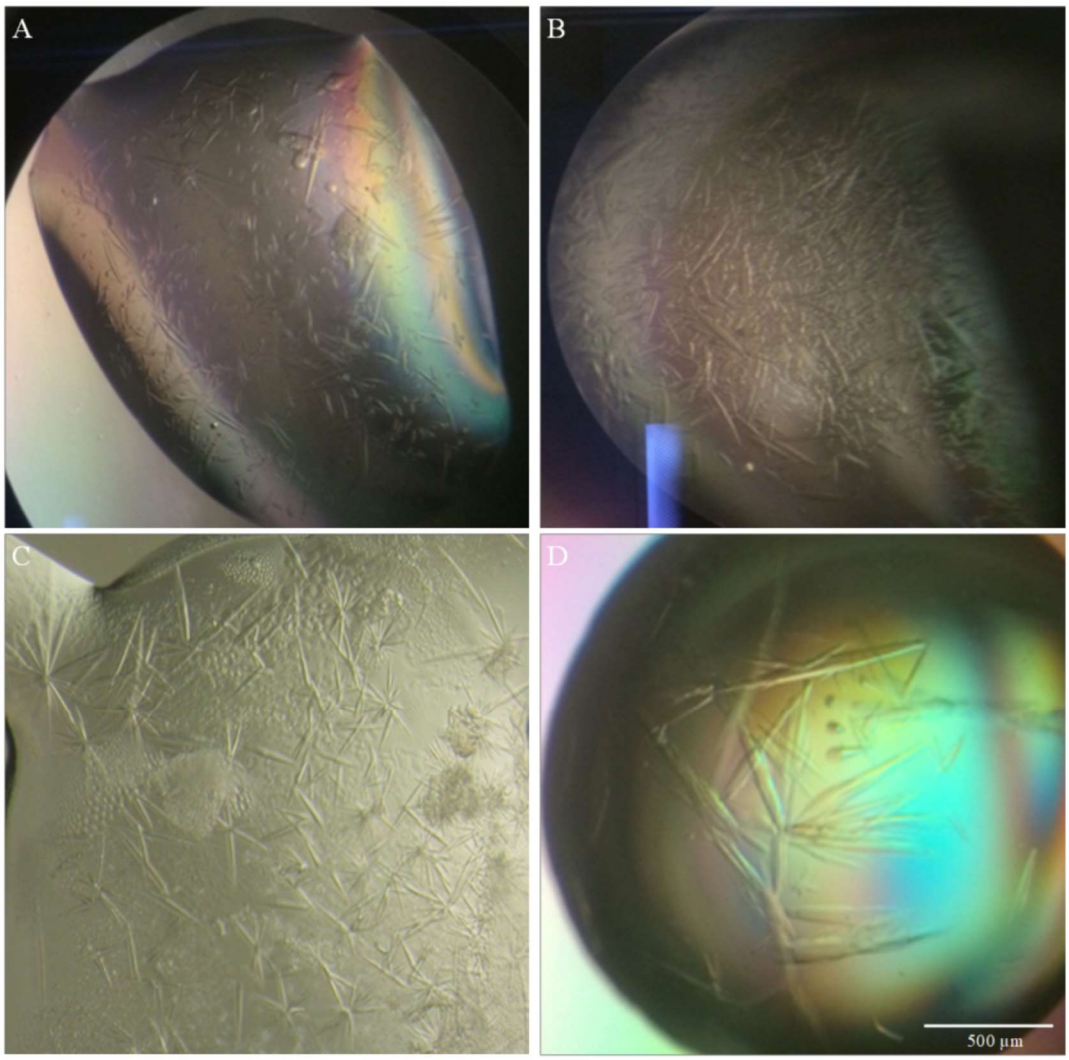



**Figure 4.6 Systematic optimization of XLF/RCC4$^{157}$/DNA crystals.** A) Initial hit identified by broad screening 750 sparse matrix conditions. B) Crystal form 2, also obtained through broad screening. C) Optimization of form one through controlled micro seeding of nuclei D) Crystal growth improved by controlled dehydration at lower temperatures.

basepair unit augmented in length by the length of the overhang (Table 4.2). This substrate design allowed the XRCC4/XLF protein components to 'select' for an ideal length of DNA substrate.

Eight broad screens of 96 conditions were carried out for complexes formed with each of the 5 stacking substrates. All preliminary hits were identified with complexes formed with either the stack 9 or stack 10 DNA as the substrate.

Initial crystals formed when protein/DNA complexes were equilibrated with a buffer containing 100 mM HEPES-NaOH (pH7.5), 200 mM $LiSO_4$ and 25%w/v peg3350. The small, needle like crystals grew after 48 hours at room temperature. Crystals obtained directly from primary screening diffracted very weakly, with only low-resolution spots observed (~20Å). Optimization of crystal growth involved the same micro-seeding strategy that was used to generate the highest resolution crystals of full length XRCC4 as well as controlled dehydration of the drops at lower temperatures (Figure 4.6). While these modifications did improve the diffraction resolution of the crystals, the increase was insufficient to warrant extensive further optimization. Furthermore, the crystals were physically weak and prone to damage during the mounting process and data collection. After limited ability to optimize these hits, a decision was made to pursue a different, more heterogeneous complex. Electrophoretic mobility shift assays were used to evaluate the relative stability and homogeneity of different combinations of XRCC4 mutants in complex with XLF and DNA.

The previously employed negative control (XRCC4L184Q/K187D/I191S) was shown to form a very discrete complex by EMSA when bound to XLF and the stack 9 substrate. Accordingly, this form of XRCC4 was selected for further structural studies of the XRCC4/XLF/DNA complex. Broad screens were again

carried out with each of the stacking substrates. Again, the best crystals were obtained using the stack 9 and 10 substrates, with stack 9 yielding the best quality

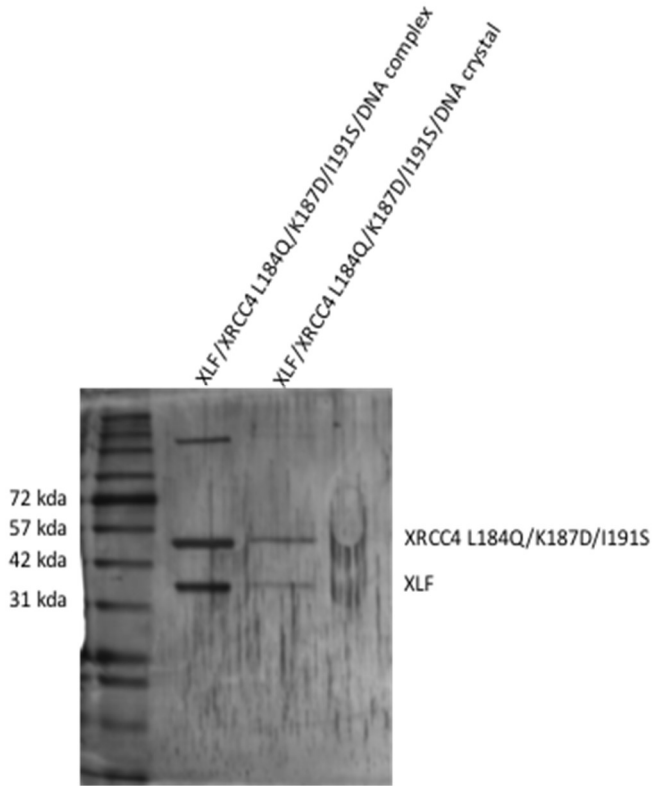



**Figure 4.7 Silver stained SDS page gel of dissolved crystals alongside pre-formed complex.** Washed crystals were dissolved and visualized via SDS page gel. Lane 2 contains pre-formed complex run alongside the dissolved crystals. Lanes show an identical protein composition indicating that the crystals formed do contain both proteins in equivalent amounts. DNA was not observed in either the pre-formed complex or dissolved crystal

crystals. Crystal diffraction greatly improved (resolution rising from 20 Å to 5Å) with the addition of 2,2-butanediol when added to the hanging drop. In order to confirm that the crystals contained both proteins and DNA, crystals were isolated,

washed, dissolved, run on SDS PAGE and silver stained to assess their composition.

Crystals were isolated from the hanging drops and washed extensively to remove any contaminating protein from the drops remaining on the crystal. Crystals were then dissolved in 2x SDS PAGE load dye and boiled for 5 minutes. Following electrophoresis, silver staining of the gel showed that the crystal did in fact contain both proteins in equivalent amounts (figure 4.7). No DNA was observed on the gel, however, the crystals were unable to grow in the absence of DNA substrate strongly suggesting the presence of DNA within the crystal. Once the crystals were confirmed to have the entirety of the complex, more extensive optimization was carried out beyond the addition of additives. Controlled de-hydration, serial seeding and varying the ratio of additives each improved the overall quality of the crystals (figure 4.8). After further pre-screening of the crystals to identify suitable cryo-protection of the condition, multiple crystals were harvested, frozen and were sent for data collection at beam 17-ID, APS Argonne (table 4.2)

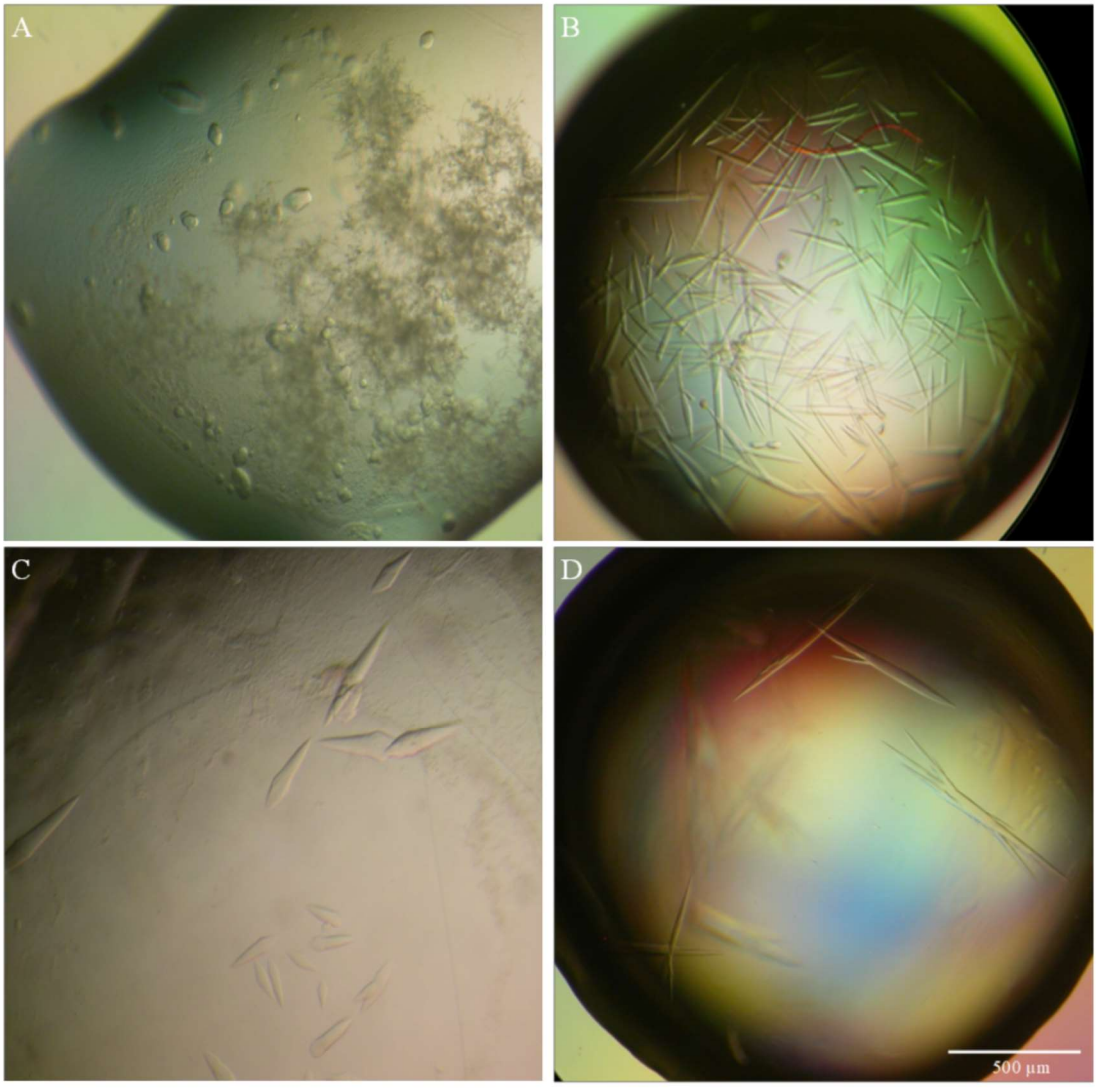



**Figure 4.8: Optimization of XRCC4 $^{L184Q/K187D/I191S}$ /XLF/DNA co-crystal.** A) Initial hit as identified by broad screening. B) Optimization of DNA substrate yielded significantly larger crystals C) Micro seeding and controlled dehydration of the drops produced more discrete crystals D) Best optimized crystals obtained by the addition of 4 µl of 10% w/v 1,2-butanediol while incubated at 22°C

| | |
|---|---|
| Diffraction Soure | B-17-ID |
| Wavelength (Å) | 0.979 |
| Detector | Dectris 6M |
| Crystal-detector distance (mm) | 250 |
| Rotation per image (°) | 0.25 |
| Rotation range (°) | 360 |
| Exposure time | 0.12 |
| Space group | P3 |
| a,b,c (Å) | 165.5, 165.5, 2271.2 |
| α,β,γ (°) | 90.0, 90.0, 120.0 |
| Mosaicity (°) | 1.9 |
| Total no. of reflections | 274940 |
| Unique reflections | 8481 |
| Completeness (%) | 100 (100) |
| Redundency | 32.4 (31.9) |
| I/σ (I) | 5.2 (2.0) |
| Rmerge (%) | 15.8 (45) |

**Table 4.1 Diffraction and processing statistics for XLF/XRCC4/DNA crystals diffracting to 5Å**

Diffraction data were collected to a resolution of 5Å (Table 4.2). According to Xtriage analysis of the data, it was estimated that there are some 1900 protein residues and an indeterminate number of DNA residues were present within the asymmetric crystal unit (Brunger et al., 1998). Despite extensive attempts, no molecular replacement solution could be obtained. This was most likely due to an altered packing of XRCC4/XLF molecules within the asymmetric unit.

## 4.5 Discussion

### 4.5.1 XLF can form DNA dependent filaments in the absence of XRCC4

XLF has been shown to be capable of bridging DNA ends on its own at concentrations greater than 2μM. What has remained unclear, however, is if this bridging is a result of XLF filaments or some other type of structure-based bridging. Results presented here clearly demonstrate that XLF is capable of forming DNA dependent end bridging filaments. Indeed, the length of several species by EM was considerably longer than the 384nm length which would be expected for a 1000 bp linear piece of DNA. This alignment of strands was observed whenever XLF was incubated with DNA and completely absent whenever DNA was not present or a DNA binding mutant of XLF was used. The fact that intermediate species were not observed suggests that all of the XLF present was involved in creating several continuous filaments that extended until all the individual strands of DNA were aligned.

XRCC4 was also examined visually by EM through a similar manner to XLF. In all cases, no filaments were observed even in the presence of DNA. This inability to form filaments on its own is consistent with previous findings that suggest XRCC4 is incapable of bridging DNA ends in the absence of XLF. Our results further suggest that this lack of bridging may be due to an inability to form homo filaments. While these results are among the first to directly link XLF oligomerization to other functions (DNA binding in this case), the protein/protein interactions of XLF still remain poorly characterized.

### 4.5.2 XRCC4 induces a morphological change in XLF/DNA filaments

Interestingly, addition of XRCC4 to pre-formed XLF/DNA filaments was found to alter filament morphology. The changes observed in the mobility of XLF/DNA filaments (EMSA) upon XRCC4 addition, both in topology of filaments formed (by EM) and migration during electrophoresis suggest that XRCC4 is capable of remodelling existing XLF-DNA filaments. As our work demonstrated that tetramerization of XRCC4 is not required for bridging, remodelling of XLF/DNA filaments by XRCC4 is most likely dependent on uncharacterized protein-protein interaction. Previous atomic force microscopy performed on the full-length complex of XRCC4/XLF/DNA contained junctions of filaments and alignment of filament bundles (Andres et al., 2011) Such structures were completely absent in the XLF/DNA micrographs and only appeared when XRCC4 was added to the reaction mix prior to visualization. This implicates XRCC4 in alignment, condensation and re-modelling of XLF/DNA

filaments through a protein/protein interaction. This idea is further supported by the fact that XRCC4$^{1-157}$ was capable of enhancing filament packing and interacting with XLF despite being an obligate dimer that also lacks DNA binding ability. This specifically implicates the head to head interaction of XRCC4/XLF in the remodelling of filaments. To further characterize the XRCC4/XLF/DNA complex, we sought to crystallize the filament and determine the structure by X-ray diffraction.

Although sizable crystals were obtained following considerable optimization of the growth parameters, the crystals diffracted to only 5Åbut a molecular replacement solution could not be obtained. Phasing problems may be overcome by experimental phasing methods such as Selenomethionine derivatization or other types of experimental phase determination. Tantalum Bromide soaking of crystals has been used successfully in the past to provide phasing information for crystals of an XRCC4/XLF complex. A possible explanation for the failure of molecular replacement is that filaments formed within the crystals were not composed exclusively of alternating dimers of XRCC4 and XLF. All of the structures of the filaments that have been solved thus far show an alternating arrangement of proteins, but these structures were solved at significantly higher protein concentrations than those one would find *in vivo*. The self association of XRCC4 has been shown to be concentration dependent and the impact of the presence of DNA on XLF self association shows a shift to larger structures when present. Furthermore, all existing structures are missing the

C-terminal tails of both proteins, both of which have been demonstrated as being necessary to form a multi-filament bundle capable of bridging DNA ends (Andres et al., 2011). Absent any structural information about how these tails interact with the rest of the structure, their positioning may alter the packing sufficiently to make existing models ineffective for molecular replacement. Another possible explanation for the lack of an MR solution is that the search models being used are based on a model wherein the protein complex forms a filamentous structure which wraps around the DNA akin to the insulation coating a wire. If the ternary complex were to adopt a structure similar to that of a nucleosome wherein the DNA wrapped around the protein, then this would create a sufficiently different structure that could not be solved by molecular replacement.

### 4.6 Conclusion

XRCC4 and XLF have now been shown to form DNA end bridging filaments, both *in vivo* and *in vitro* however the mechanism by which these filaments function remains unknown.

Results of work within this thesis suggest that XLF is capable of forming filaments in the absence of XRCC4, but only when DNA is present. These XLF filaments do not form with a DNA binding defective XLF mutant further supporting the idea that XLF forms DNA dependent filaments. Interestingly, XLF filaments result produce an end to end alignment of DNA molecules with the length of aligned DNA extending considerably longer than the expected 384 nm of a single 1000 bp DNA fragment.

Surprisingly, XLF filaments are susceptible to remodelling in their architecture by the addition of XRCC4 even though XRCC4 was not found to be required for either filament formation or DNA end bridging activity. XRCC4 truncations, point mutations and even the entire LigIV/XRCC4 complex all had a similar effect on the EM topology and EMSA migration pattern of XLF/DNA filaments.

Crystallization of various combinations of XRCC4, XLF and DNA complexes were attempted to further characterize this important NHEJ complex. Initial conditions were identified and verified to produce crystals containing both proteins in equivalent amounts. Following optimization of many parameters, in particular the DNA substrate used, a significant gain in the diffraction resolution was achieved. Following optimization, crystals diffracted to a resolution of 5Å, but further optimization is still required in order to determine the structure of the end-joining filament and answer once and for all how members of the XLF family bind their DNA substrates and what the molecular composition of filaments is at the DNA break site.

| Substrate Name | Sequence |
|---|---|
| Stack 7 | 5′ GCTTCGAGCTGCGTCTAGATTGTGTGCT<br>CGACGCAGATCTAACACACGACGAAGCT |
| Stack 8 | 5′ AGCTTCGAGCTGCGTCTAGATTGTGTGCT<br>CGACGCAGATCTAACACACGATCGAAGCT |
| Stack 9 | 5′ GAGCTTCGAGCTGCGTCTAGATTGTGTGCT<br>CGACGCAGATCTAACACACGACTCGAAGCT |
| Stack 10 | 5′ CGAGCTTCGAGCTGCGTCTAGATTGTGTGCT<br>CGACGCAGATCTAACACACGAGCTCGAAGCT |
| Stack 11 | 5′ TCGAGCTTCGAGCTGCGTCTAGATTGTGTGCT<br>CGACGCAGATCTAACACACGAAGCTCGAAGCT |
| 42 bp Blunt | 5′ CCTGATATGGCAGGTACCGTATGCAATAAGTTCGATGCAATT<br>GGACTATACCGTCCATGGCATACGTTATTCAAGCTACGTTAA |
| 36 bp Blunt | 5′ GATATGGCAGGTACCGTATGCAATAAGTTCGATGCA<br>CTATACCGTCCATGGCATACGTTATTCAAGCTACGT |

**Table 4.2 List of DNA substrates used in the crystallization of the XRCC4/XLF/DNA complex**

# Chapter 5 Crystallization and preliminary diffraction of the DNA ligase IV/ XRCC4/ DNA complex

## 5.1 Abstract

DNA ligaseIV and XRCC4 interact tightly to form a very stable complex. This complex is absolutely required to seal a DNA double strand break during repair by NHEJ. Despite a catalytic mechanism and domain architecture similar to other eukaryotic DNA ligases, the molecular mechanism by which two strands are repaired remains poorly understood in the context of NHEJ. In order to elucidate this mechanism, we sought to crystallize the ternary complex of DNA LigIV, XRCC4 and DNA. While initial crystals were obtained, low yields in protein complex and an intrinsically disordered linker region between domains of LigIV proved to be formidable barriers in crystallization of this complex. Furthermore, addition of other NHEJ factors (Ku, DNA, XLF) further de-stabilized the link between the catalytic core and the remainder of the complex making crystallization less amenable. Understanding the molecular mechanism of the coordination of the catalytic site during repair will greatly aid our understanding of this process, however, based on findings in this work reported here another structural approach is likely to be required.

## 5.2Introduction

DNA ligaseIV is the catalytic molecule which terminally ligates the sugar phosphate backbone of DNA during repair by NHEJ. It is the only human DNA ligase that is required for DSB repair via NHEJ (Conlin et al., 2017, Cotner-Gohara, E. et al., 2007). In humans, deficiencies in LigIV are associated with an extreme phenotype known as LigIV syndrome. The catalytic core of LigIV shares

a similar fold to all other known DNA ligases, but how the catalytic core is oriented relative to the C-terminus of LigIV and XRCC4 remains unknown (Ochi et al., 2011).

The interaction between LigIV and XRCC4 is one of the strongest protein-protein interactions ever observed (Mcfadden et al., 2014). Consequently, cells depleted of XRCC4 show no detectable levels of LigIV (Grawunder et al., 1998). While the structure of the C-terminal BRCT domains of LigIV bound to XRCC4 has been determined, there is currently no structural information for the complex as a whole. This information gap is surprising as the LigIV/XRCC4 complex represents a prime therapeutic target and key regulator of NHEJ. In this work, we report a 4-step purification protocol capable of yielding pure LigIV/XRCC4 in the appropriate 2:1 stoichiometry. Using this protocol, sufficient protein was obtained to identify crystallization conditions for the LigIV/XRCC4 complex. Despite extensive attempts at crystal growth optimization, including incorporation of other NHEJ core factors, no suitable diffraction quality crystals were obtained. We also sought to characterize how the presence of NHEJ factors affected the stability of the LigIV/XRCC4 complex and found the complex to be unstable and susceptible to spontaneous proteolysis.

## 5.3 Materials and Methods

### 5.3.1 Expression and purification of the LigIV/XRCC4 complex

*E. Coli* BL21-Rosetta cells were transformed with a plasmid encoding DNA LigIV and XRCC4. A single transformant was sub-cultured into LB media supplemented with 50 μg/ml chloramphenicol. Once cells reached an $OD_{600}$ of 0.4, cultures were cooled to 16°C in an ice bath. Protein expression was then induced by the addition of 1mM IPTG. Cells were incubated at 14°C overnight and cells were harvested by centrifugation at 3385 X g for 15 minutes immediately prior to purification. Cells were re-suspended in a nickel binding buffer containing 20 mM Tris (pH 8.0), 500 mM KCl, 3 mM BME and 3 mM imidazole. Following lysis, the lysate was clarified by centrifugation at 48000 X g for forty-five minutes. Clarified lysate was applied to a pre-equilibrated HIStrap HP column pre-equilibrated with lysis buffer. The column was then washed with 5 column volumes of binding buffer followed by a 12 mM imidazole wash of 2 column volumes. Owing to the low affinity of the LigIV/XRCC4 complex for the HIStrap column, the complex (eluted at 54 mM imidazole) resulting from IMAC was of low overall purity. Accordingly, the protein was further purified. Following IMAC, the sample was exchanged into an ion exchange buffer containing 20 mM Tris, pH 8.0, 1 mM EDTA, 100 mM KCl and 5 mM DTT, this sample was injected onto a pre-equilibrated 10/100 mono-Q column (GE healthcare). The protein was eluted over a linear gradient of 200-400 mM KCl over 120 minutes at a rate of 1ml/min. Fractions containing LigIV/XRCC4 were

diluted to a final salt concentration of 25 mM KCl and re-injected onto a pre-equilibrated 10/100 mono S column. The complex was eluted over a linear salt gradient of 25-400 mM KCl over 100 minutes at 1 ml/min. The fourth and final purification stage was SEC (GE S200 superdex column GE Healthcare). The SECcolumn was pre-equilibrated with a buffer composed of 50 mM bis-tris propane (pH 6.5), 150 mM KCl, 0.5 mM EDTA and 10 mM DTT. The LigIV/XRCC4 complex eluted at a volume consistent with an expected molecular mass of 180 kD. Purity was verified by SDS PAGE analysis (Figure 5.1). The complex was then concentrated to a final concentration of 5 mg/ml and was subsequently used for crystallization and biochemical characterization.

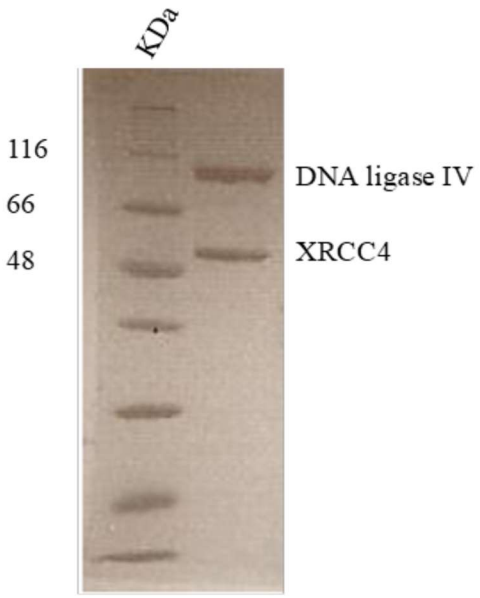



**Figure 5.1: SDS PAGE analysis of purified LigIV/XRCC4.** Recombinantly expressed LigIV and XRCC4 shown to be purified to homogeneity prior to use in crystallization trials.

### 5.3.2 Crystallization of the LigIV/XRCC4 complex with and without DNA

Purified LigIV/XRCC4 was mixed in a 1:1 stoichiometric ratio with precipitant and sealed over a reservoir containing 1.5M $NH_2SO_4$. Crystallization trays were equilibrated at either 4, 20 or 37 °C until precipitation or nucleation was observed. Crystals formed within 2 weeks. Initial hits were identified in 4 conditions and the crystals were verified to be protein by X-ray diffraction. Initial hits were optimized by a combination of additive screening (Hampton Industries) and controlled dehydration. Every 48 hours, $NH_2SO_4$ in the well was replaced with a more concentrated solution, at 250 mM increments.

For protein/DNA co-crystallization, protein and DNA were incubated in a 1:1.2 ratio for 30 minutes so as to ensure complete complex formation. Following this incubation, the complex was used for crystallization in an identical manner to the apo-protein.

### 5.3.3 Limited proteolysis of different NHEJ complexes.

LigIV/XRCC4 was incubated with either XLF, 1000 bp dsDNA or both. Reactions were then subdivided into 2 separate aliquots. One ng of trypsin was added to 1 of each reaction combination while the other received 1 µl of trypsin buffer. All tubes were incubated at 37°C for 30 minutes. Reactions were quenched by the addition of 2X SDS page load dye. Reaction products were then subjected to SDS page analysis and the resulting coomassie stained gels were visualized using a Bio-rad Geldoc imaging system.

### 5.3.4 X-ray diffraction of LigIV/XRCC4 crystals

Diffraction data was collected at the X-29 beamline of the National Synchrotron Light Source (NSLS) of Brookhaven National Lab. Data was collected in half degree oscillation wedges with an exposure time of 2 seconds per degree for a total of 360 seconds. Crystals rapidly succumbed to radiation damage.

## 5.4 Results

### 5.4.1LigIV/XRCC4 crystal formation was very limited by nucleation

As the final complex for repair during NHEJ, LigIV/XRCC4 represents an optimal drug target. In order to further understand the molecular mechanism of the final step of repair, we sought to crystallize this complex. When mixed with various crystallization conditions, the LigIV/XRCC4 complex precipitated out of solution rapidly. Even at lower protein concentrations, which should favor crystal formation and growth, heavy precipitation was observed in the majority (>75%) of hanging drops. Nucleation was overcome in two ways: 1) controlled dehydration by gradually increasing ammonium sulfate concentrations and 2) performing crystallization experiments at 4°C. Reduction in nucleation came at the cost of crystals taking an extremely long time to form (~4 weeks from initial equilibration until crystal formation). Despite this optimization, crystals either did not diffract or diffracted to extremely low resolution (<20Å resolution). Although many strategies could be used to further optimize crystals (i.e. protein truncations, variations in precipitant components, removal of unstructured loops etc.),

crystallization of the LigIV/XRCC4 complex was not further pursued in order to focus on other crystallization studies (i.e. full length XRCC4, the XRCC4/XLF/DNA complex). During the completion of this thesis, a new purification strategy for the purification of LigIV/XRCC4 complex was developed. A newly optimized co-expression construct was generated that provided significantly higher yields of complex while requiring only two purification steps. Using this new construct, it may be worth going back to optimize the existing crystal forms identified here and also continuing to screen for new crystallization conditions that yield better diffracting LigIV/XRCC4 crystals.

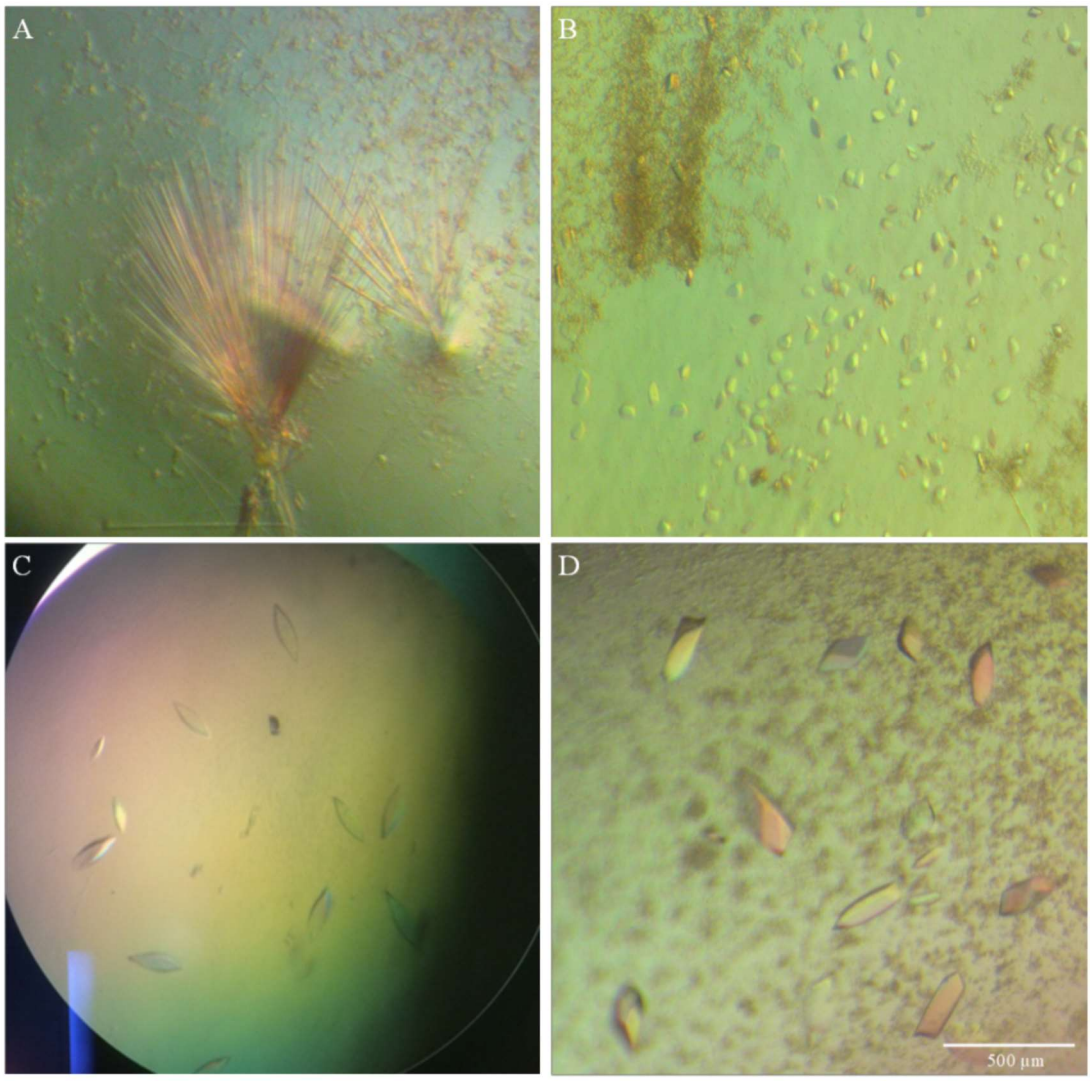



**Figure 5.2 Optimization of LigIV/XRCC4/DNA co-crystals** A) Initial hit of APO complex. B) Optimization of the crystals through controlled de-hydration. C) Crystals grown in complex with a 30 basepair duplex strand of DNA. D) Best crystals obtained through controlled dehydration coupled with complex formation with an 18 basepair non-ligatable strand of DNA with a nick in the middle of the 5' strand.

### 5.4.2 DNA substrates used during LigIV/XRCC4 crystallization do not facilitate crystal formation

Extensive trials to crystallize the apo-complex of LigIV/XRCC4 failed to yield diffraction quality crystals. Given the intrinsic flexibility of the linker region separating the OB-fold and the first BRCT domain of LigIV, it was thought that the APO complex may be too heterogeneous to crystallize into a well-ordered lattice. In order to overcome the intrinsic disorder between the catalytic core and the BRCT domains of LigIV, we added a DNA substrate with the hope of trapping the flexible domains in a more stable conformation. As DNA LigIV is pre-adenylated during purification, a substrate was designed to trap the ligation complex in the penultimate stage of repair. Based on homology modelling of other mammalian DNA ligases, an 18 bp substrate was designed. Composed of three separate synthetic oligonucleotides, the resulting structure contained a nick on one of the two strands with a 5'-phosphate to permit adenylation by LigIV. The corresponding nicked strand, however, possessed a 3' di-deoxy adenine at the site of the nick. This resulted in a complex that mimics the final stage of ligation without the capability of completing ligation. Trapped complex was prepared with this substrate and screened for crystal formation as described for apo LigIV/XRCC4 complex. Unfortunately, crystals identified in this screen also diffracted to very low resolution. Other DNA substrates with longer extensions designed to lock down the flexible catalytic core resulted in no improvement in

the diffraction of the crystals. Stacking substrates (as described in section 4.4.3) had no positive impact on crystallization of the complex.

### 5.4.3 Addition of ancillary NHEJ factors further destabilize the link between the catalytic core and the tandem BRCT domains of LigIV

As the addition of DNA alone was insufficient to stabilize the LigIV/XRCC4 complex, a limited proteolysis analysis performed in the presence different NHEJ factors. LigIV/XRCC4 was incubated with XLF, DNA and/or both factors to determine if a more stable complex could be generated for structural studies. Following tryptic digest and arrest of the reactions, the proteolytic products were visualized via SDS PAGE. The digest pattern revealed several distinct bands that were resistant to proteolysis. Interestingly, in the presence of additional binding partners, the LigIV/XRCC4 tryptic digest yielded a band migrating between LigIV and XRCC4 (Figure 5.3). The intermediate band was characterized by mass spectrometry to be the catalytic core of LigIV. Even in the absence of trypsin, in the undigested control lanes, there was an enrichment of the band corresponding to the catalytic core of LigIV. This finding suggests that the complex formed between XRCC4/XLF increases the exposure of the linker region between the BRCT domains and the catalytic core of LigIV. Although the biological significance of this observation is unclear, the overall effect of linker destabilization of XLF to LigIV/XRCC4 was not a good approach for further structural studies.

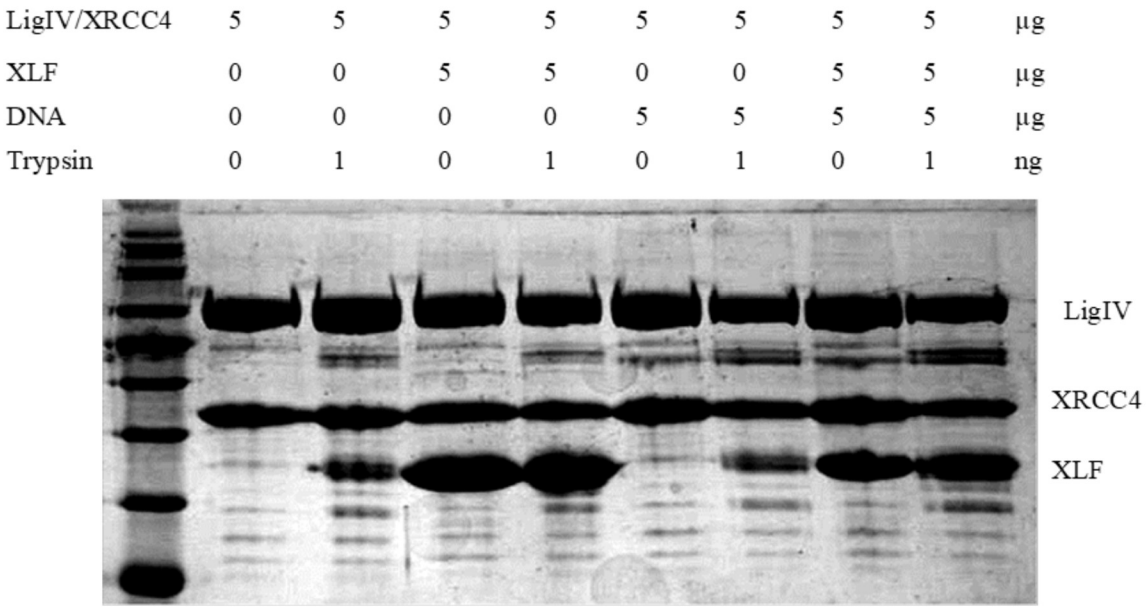



**Figure 5.3 SDS PAGE analysis of limited proteolytic digest of NHEJ complexes.** Different combinations of NHEJ components were incubated with a limited amount of trypsin in order to determine which factors stabilize the ternary complex.

## 5.5 Discussion

### 5.5.1 The link between the catalytic core and the tandem BRCT domains of LigIV is flexible and unstable.

Initially, extensive crystallization trials were carried out with the apo LigIV/XRCC4 complex. Despite screening 1000s of conditions, no diffraction quality crystals were obtained. While each of the 5 domains of LigIV have a defined structure, the linker region between the nucleotidal transferase domain and first BRCT domain (610-653) is predicted to be intrinsically disordered. Since a lack of crystallization is often associated with heterogeneity in the complex due

to instability or flexible regions, new approaches were attempted to overcome these limitations.

A frequent method used to stabilize protein assemblies is to add compounds (i.e. substrates), in this case DNA, to trap the complex in a rigid state. Specifically, we sought to stabilize the linker region between the catalytic core of LigIV and the tandem BRCT domains bound to XRCC4. A small, unligatable nicked substrate along with other smaller duplex DNA substrates were used. Extensive crystallization trials of the LigIV/XRCC4/DNA complex did not yield diffraction quality crystals, despite initial crystallization conditions being identified. This inability to obtain diffraction quality crystals may be due to suboptimal DNA. For instance, DNA that was to small would fail to fully engage both the catalytic core and the C-terminal XRCC4 interacting region. When diffraction of crystals did not improve, we sought to determine which combination of factors might yield a more stable complex amenable to X-ray diffraction.

### 5.5.2 Addition of NHEJ core factors occlude access of the LigIV catalytic core to a break site

As the unligatable substrate described above was not optimal to stably engage the interdomain region of LigIV, other factors were assayed by limited proteolysis to determine the most stable combination for crystallization trials.

Longer DNA substrates as well as XLF were added to the LigIV/XRCC4 complex. Interestingly, the undigested control lanes for each reaction containing either XLF, DNA or both showed an increase in accumulation of the catalytic core of LigIV. This finding suggests that the linker region between the catalytic core and the tandem BRCT domains is intrinsically unstable and highly susceptible to proteolytic attack

Of particular interest was the observation that addition of XLF also destabilized the linker region between the catalytic core and BRCT domains of LigIV, and that the effect is amplified in the presence of DNA. Filament bundles formed by LigIV/XRCC4/XLF and DNA have been observed in the cell by ultra-resolution microscopy studies (Reid et al., 2015). Despite being large, the structures appear homogeneous at medium to low resolution. Surprisingly, combining LigIV/XRCC4 with XLF and DNA had the largest apparent degradation of the linker region, resulting in the highest observed amount of free LigIV catalytic core. There is no evidence that XLF interacts with LigIV, *in vivo* or *in vitro,* but XLF has been shown to promote ligation of mismatched DNA ends as well as the re-adenylation of LigIV following initial ligation (Andres et al., 2007, Riballo et al., 2009). Based on the prior reports, it was thought that the effect of XLF on LigIV may be due to interaction of XLF with the catalytic core of LigIV. If so this would capture LigIV/XRCC4 to the DNA end site and trap it there, promoting a homogenous filament starting at one end of the DNA substrate. Instead, the XRCC4/XLF population of the filament was stabilized (less

susceptible to tryptic digest when in complex) and, the abundance of free catalytic core was increased.

## 5.6 Conclusion

The results presented here to crystallize human DNA ligaseIV in complex with XRCC4, DNA and other NHEJ factors were obtained using an extensive number of crystallization variables. While some initial crystal hits were obtained, further optimization including the use of many different DNA substrates failed to yield high resolution diffraction quality crystals. The addition of other NHEJ factors (DNA, Ku) were also shown to de-stabilize the linker region between the LigIV catalytic core and the tandem BRCT domains, resulting in a more heterogeneous population resistant to crystallization.

Modelling the structure of the catalytic core of LigIV bound to DNA suggested that the minimum DNA length required to bind LigIV is considerably longer than the length of DNA substrates employed for crystallization trials. Longer lengths of DNA were not used as the direction of the project had shifted considerably following the solving of the structure of the catalytic core of LigIV (Ochi et al., 2012).

Ultimately, another strategy may be required in order to determine the quaternary structure of the LigIV/XRCC4 complex bound to other NHEJ factors. Ultra-resolution and fluorescence microscopy have been previously employed for investigation of large complexes, however both lack the high-resolution

information that is obtained from a crystal structure, particularly the composition and relative arrangements of all the repair factors present at a DSB. NHEJ factors (LigIV in particular) remain prime therapeutic targets. Accordingly, understanding the spatial arrangement of the end repair complex would be expected to dramatically improve efforts in developing a small molecule capable of specifically inhibiting repair of DSBs by NHEJ.

# Chapter 6: Conclusions and future directions

## 6.1 Summary

Work conducted here has focused on structural and biochemical characterization of the mammalian DNA repair proteins XRCC4 and XLF as well as their yeast homologues, Lif1 and Nej1. The structure of full length XRCC4 was solved to 3.43Å. From this structure, we were able to ascertain several new secondary structure elements of the protein. Based on these findings, mutations were generated in an effort to determine the true tetramerization interface of XRCC4. Mutations that disrupted a multi-helix bundle were able to eliminate tetramerization. Interestingly this mutant retained all other XRCC4 functions including DNA end bridging when in complex with XLF. This result strongly suggests that the existing model whereby the tails of XRCC4 stabilize multi-filament bundles is incorrect. Truncated XRCC4(157-202) that is still containing the helical bundle region is still able to bridge DNA ends when in complex with XLF. As tetramerization or XRCC4 has been shown to be dispensable for end bridging, and that the contribution of this part of the XRCC4 tail is still absolutely required for end bridging, it would appear that this region of XRCC4 stabilizes end bridging filaments in a mechanism that is independent of XRCC4 tetramerization.

While XRCC4 induces a change in XLF/DNA filaments, this remodelling is not due to tail-based stabilization of filaments as the same changes are observed when filaments are incubated with C-terminally truncated (1-157, 202 and 265) or

tetramer defective XRCC4. This finding was corroborated by the ability of XRCC4 C-terminal truncations to bridge DNA fragments when XRCC4 posses the ability to form a multi-helix bundle. This finding suggests that the formation of this helical bundle is important for bridging and stabilizing a complex through a mechanism un-related to tetramerization.

While our studies of protein/protein interactions formed during NHEJ primarily focused on higher eukaryotic proteins, Nej1 (from yeast) was characterized biochemically and found to bind DNA very strongly through an extended, C-terminal interface. No individual mutations were identified that could disrupt DNA binding. Surprisingly even multiple mutations localized to this region of Nej1 failed to diminish DNA binding.

This type of tight, non-specific DNA binding is reminiscent of several other nucleo-protein structures that have been observed. Several non-specific DNA binding proteins, such as Rad51 or HMGA1 engage DNA by threading through or wrapping around DNA (Chen et al., 2007). AT hook motifs use an extended combination of structural (proline in particular) and basic residues to intercalate within the minor groove of DNA (Figure 6.1 A, B). Histones also use an extensive array of protein tails to bind and trap DNA as heterochromatin (Figure 6.1 C, D). These are but two examples of such non-specific, extensive mechanisms by which DNA can be bound. The tails of XLF (and Nej1) and XRCC4 have elements in common with both such modes of binding. As both are

meant to be non-specific DNA binders, the structural residues would allow for a structure-based DNA binding. The extensive binding, possibly even on both sides of a break, would provide additional stabilization of DNA ends to prevent their diffusion prior to repair.

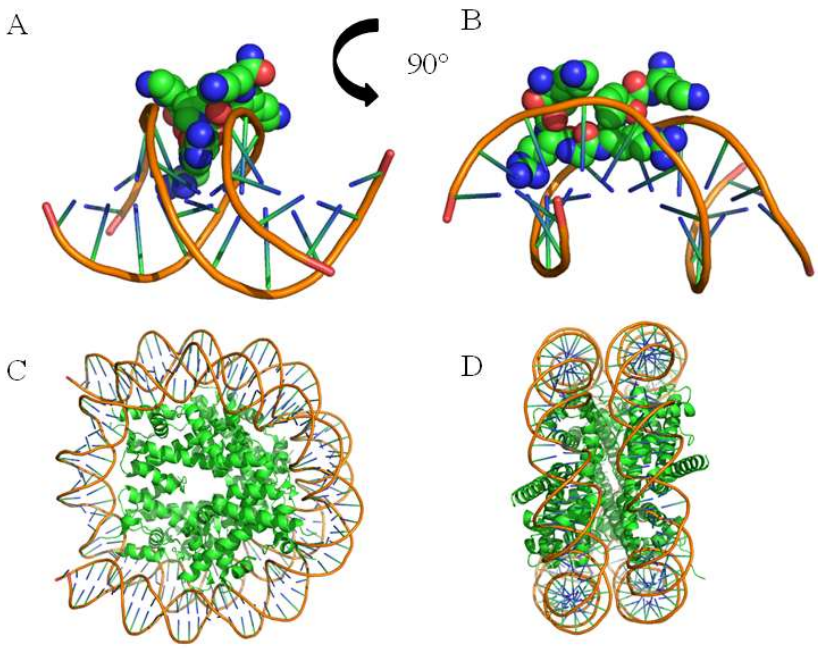



**Figure 6.1 Crystal structures of non-specific DNA binding proteins bound to their substrates.** A, B) the arginine rich AT-hook motif of HMGA1 bound to the minor groove of duplex DNA, pdb 3UXW. C, D) H3K56Q nucleosomes bound to genomic DNA showing the extensive interface formed between protein and DNA, pdb 3KXB.

Additional mutational analysis showed that phosphomimetic substitutions of Dun1 substrate residues in the Nej1 tail resulted in a reduction of DNA binding. Coupled with the fact that Dun1 phosphorylation of these residues is required for SRS helicase activity, this paints a picture wherein Nej1 plays a

regulatory role in repair pathway choice in addition to stimulation of Lig4 activity in *Saccharomyces cerevisiae*. As the C-terminal tails of NHEJ proteins appear to play a regulatory role in complex formation and dissociation, this decrease in affinity and helicase recruitment may be the mechanism by which a switch between repair pathways is mediated.

In addition to analysis of XRCC4 and XLF on their own, the end bridging complex was studied through several different structural means. Electron microscopy revealed that in the presence of XRCC4, individual XLF/DNA filaments were aligned in parallel multifilament arrangements. This type of behavior has never observed previously even at very high concentrations, when XLF alone or in the presence DNA. While all *in vivo* and ultra-resolution microscopy techniques confirmed that complex filaments are formed, their composition remains poorly understood. Here, we report for the first time that XLF is capable of forming simple filaments when bound to DNA. These filaments appear linear and show an end to end alignment of DNA strands, but do not show the complex filaments and filament junctions that were previously observed by atomic force microscopy (Andres et al., 2011). The results presented here suggest that it is the presence of XRCC4, XLF/DNA form more complex filaments and filament junctions are built up through a head to head interaction with XLF.

Finally, the end joining complex was studied in order to further characterize how different factors are coordinated at the site of the break and how

repair is ultimately achieved. Stability studies carried out through partial proteolysis analysis as well as crystallographic studies so as to determine the mechanism by which breaks are repaired.

### 6.1.1 Structural characterization of protein/DNA complexes within NHEJ repair

Structural biology (NMR, X-ray crystallography, cryo-em) remains an incredibly powerful tool for studying the mechanism of biological systems. Despite this power, there are numerous limitations. With respect to large, multi component complexes such as the NHEJ repairosome, identification of appropriate protein components, substrates and most importantly, crystallization conditions are challenging. The absolute requirement for a homogenous sample is compounded by the addition of each new factor to be analyzed. (McPhereson, A., 1976 and 2003)

While lower resolution techniques (Cryo em) are closing the gap, the only method of determining atomic level structural information for large complexes remains X-ray crystallography. Crystallizing apo-protein complexes can be challenging, and the addition of DNA creates another parameter that must be optimized in order to yield diffraction quality crystals. Here, we took a systematic approach to design a series of substrates that might allow the stable formation of NHEJ filament complexes. Rather than stochastically crystallizing complexes

with DNA substrates of varying lengths, a self stacking substrate was employed. While the length of the complementary core was held constant, the lengths of the palindromic overhangs were varied between 7-11 residues. Not only did this design allow for the protein to select an ideal length of DNA substrate, it also permitted protein assisted annealing of elongated DNA structures to further enhance the specificity of the protein complex for the perfect DNA substrate length. When stacking substrates were used in place of fixed length DNA substrates, crystal quality and diffraction resolution improved dramatically. This strategy has since been used to crystallize other DNA binding proteins in the lab with some success across different families of non-specific DNA binding proteins. Using these substrates made optimization of the DNA substrate less dependant on length of DNA substrates, effectively eliminating one large variable from the crystallization equation.

### 6.2 Current understanding of the dynamic NHEJ repair complex

When work on this thesis began, several novel factors (PAXX, APLF) had yet to be discovered and the role of XRCC4/XLF filaments was not widely accepted. Since then, several repair factors have been identified and our understanding of double strand break repair by Non-homologous end joining has grown considerably. The notion that repair is a simple, linear, flow of progressive steps has been replaced by a paradigm in which the entire repairosome is a dynamic complex with several factors playing multiple roles throughout the

process. Understanding where PAXX and APLF fit into the repairosome remains unclear (Balmus, et al., 2016).

Indeed, there has been a burst of new information in the literature about the roles of these new factors in NHEJ repair. Specifically, with respect to PAXX, there is mounting evidence that it is epistatic with XLF and could potentially serve a redundant function during repair. Like XLF, PAXX is able to supershift a Ku-DNA complex and has been pulled down with other single strand break repair factors (Tadi et al., 2016).

Interestingly, reports of rapid mobility of XRCC4/XLF end bridging sleeves is in stark contrast with some of our mutational and functional data. Optical tweezer experiments have shown an exponential increase in energy to increase the distance between two beads until a threshold is reached, at which point the process begins anew. This observation would be more consistent with XRCC4 and XLF forming protein ratchets that aligned DNA strands in parallel, similar to what was observed in the microscopy work presented here. Other high-resolution microscopy techniques have shown alignment and movement of filaments relative to one another suggesting that XRCC4/XLF filaments slide along one-another and not the DNA until the break site is reached by a LigIV.

### 6.3 The role of unstructured C-terminal tails of NHEJ proteins

A common theme among several non-catalytic NHEJ proteins is that they contain intrinsically disordered C-terminal regions (XLF, XRCC4, PAXX), all of which are required for repair *in vivo.* Recent work has focused on potential regulatory roles of the C-terminal tails in mediating complex dissociation. These tails are generally highly conserved despite having almost no regularly ordered secondary structure elements.

Given that the tails of XRCC4 are dispensable for end bridging as discussed in section 2.5.2, it stands to reason they may play some other roles in DSB repair. Phosphorylation of the XRCC4 tails by DNAPK has been shown to reduce end bridging, *in vitro* (Andres et al., 2012, Roy et. Al., 2012). This creates a scenario wherein the tails are not required for filament formation, but rather are required to dissociate the filaments from one another at the conclusion of the repair process. The DNA binding activity XRCC4, while observed *in vitro*, has never been shown to be biologically relevant despite the tails being highly conserved. A key question is therefore what roles do these C-terminal tails play in NHEJ?

Here, we have shown that DNA binding in the C-terminal tails of XLF (Nej1) is required for XLF filament formation and that these filaments align multiple pieces of duplex DNA in a continuous strand, far longer than the length of the DNA alone. Elimination of the DNA binding ability of XLF resulted in

abolishment of filament formation. As a DNA binding mutant of Nej1 could not be identified, this finding remains limited to higher eukaryotes for the time being. That being said, however, phosphorylation of Nej1, the yeast XLF homologue, resulted in a decrease in DNA binding and a shift towards HRR repair for DSBs. Both XRCC4 and XLF are known to be substrates of DNAPK and its activity has been associated with a decrease in end bridging activity *in vitro.* Coupled with existing knowledge, we can propose a model wherein the contributions of the N-terminal head domains and C-terminal tails are invoked differentially throughout repair. The requirement of head to head interactions to stabilize initial filament formation has been well documented. Work presented in this thesis (section 4.4.1) has shown that the C-terminal DNA binding domain of XLF binds a DNA substrate and begins to form a simple filament. Unclear, however, is the mechanism by which these filaments are disassembled following repair of the DSB. The newly discovered decrease of DNA binding affinity by Nej1 when phosphorylated in the C-terminus is consistent with the effect of DNAPK on end bridging. This would suggest that the C-terminal tails of these repair factors (XRCC4, XLF, Nej1) serve as regulators of complex formation and disassembly through phosphorylation events. This kind of global regulation would be necessary as a large, end bridging filaments could be bound to DNA through dozens, if not hundreds of C-terminal tails, resulting in a very stable complex. Mutational and biochemical analysis of Nej1 bound to DNA suggests that this interaction occurs through a sequence independent wrapping of the DNA which

may require a considerable structural change to abolish. Once C-terminal tails are phosphorylated following repair, the resulting decrease in affinity for DNA would begin the process of filament disassembly and release of repaired DNA. A crystal structure of these proteins bound to DNA, such as the one currently being pursued, would dramatically improve the quality of this model for filament formation and dissociation (figure 6.2)

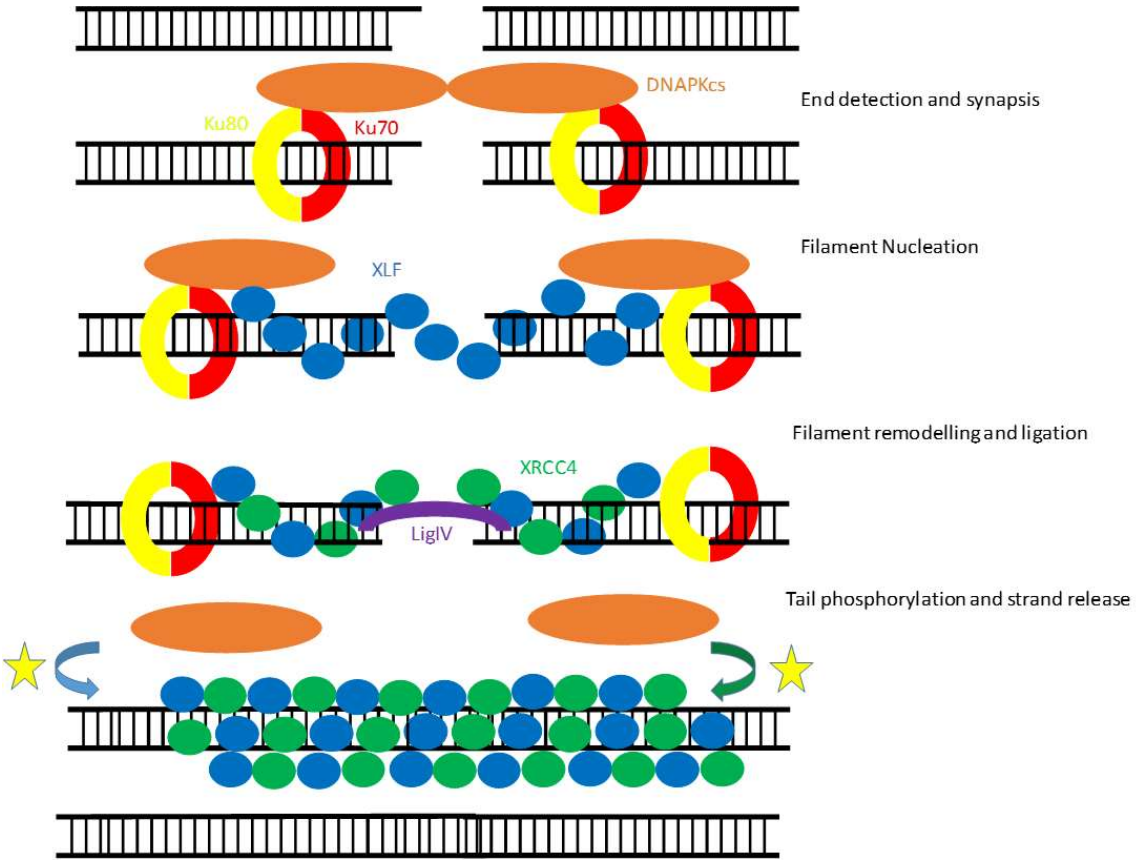



**Figure 6.2 Model of filament dissociation by phosphorylation of filament protein tail.** Following initial end detection and synapsis by Ku 70/80 and DNAPK (Red/Yellow and Orange), initial filament formation is nucleated by XLF (Blue) binding to DNA. This initial filament is then remodelled by XRCC4(green), resulting in a filament aligning ends for ligation by LigIV (purple). Following repair, the complex filament remains bound to DNA until it is

phosphorylated by DNAPK resulting in a decrease in the filament affinity for DNA, thereby mediating strand release.

### 6.4 Future directions

End alignment, bridging and stimulation of ligation remain the three most important phases of the NHEJ repair pathway. Work within this thesis has been focused on elucidation of the molecular mechanism of DNA end joining during NHEJ, structure determination of full length XRCC4 and characterization of the mechanism Nej1 DNA binding. The finding that DNA binding of XLF, but not XRCC4 is required for filament formation, and identification of crystallization conditions for the terminal ligation complex represent the chief findings of this work. Despite these new advances into protein/protein and protein/DNA interactions, many questions remain unanswered.

With the revelation that tetramerization of XRCC4 is not required for bridging of DNA ends, an outstanding question remains as to whether or not this defect is accompanied by a repair deficiency in cells. To that end, V(D)J recombination assays and *in vivo* complementation studies will need to be carried out.

Determination of the specific DNA binding residues by which Nej1 and XLF bind their DNA substrates is also important. While work reported here has not determined the specific residues by which such interactions are mediated, it has eliminated a number of potential interfaces and interaction mechanisms.

While HDX and initial crystallization trials have not identified the direct interacting residues, other techniques such as NMR or SAXS studies could provide useful knowledge of the Nej1 or XLF tails bound to DNA. Based on the overall topology of such an envelope, it would be possible to determine if the protein is adopting a distinct structure or is just wrapping the DNA non-specifically.

One of the most important aspects of this work relates to XLF DNA binding and its role in filament formation. Much of the previous work in the field has focused on how XLF interacts primarily with other proteins during repair. Here, we examined how the DNA binding of XLF affected its ability to form different complexes. The DNA binding activity of XLF was shown to be necessary for simple filament formation, a key first step in successful end bridging. Like the role of XRCC4 self association, this finding needs to be further probed to assay the impact of this early event during NHEJ. VDJ recombination and *in vivo* testing of DNA binding defective mutants are therefore a high priority.

Perhaps the most important future step to be taken remains structural characterization of the XLF/XRCC4/DNA co-crystal structure. While cryo-electron microscopy is ongoing with collaborators, it still remains a lower resolution technique. While it would inform on the overall topology of the filament, it would be lacking in the high resolution and atomic level detail that a high-resolution crystal structure would provide. Such a structure would provide a comprehensive view of how the end bridging complex is composed and oriented

at the site of the break, including the manner by which the DNA is engaged and the residues which mediate this interaction. Although crystallization conditions were identified, the resulting crystals diffracted to low resolution. Systematic optimization of the DNA substrate will likely make the largest difference to the optimization of diffracting crystals. As the current crystals have been verified to contain both proteins in equal amounts, and crystals don't form in the absence of DNA, improving diffraction even slightly beyond the existing 5Å threshold, will likely result in the ability to determine the XLF/XRCC4/DNA ternary structure. Since extensive mutational analysis of Nej1 has yet to shed much insight into the mechanism by which DNA is bound, capturing this complex in-crystallo still provides the highest likelihood of determining the mechanism of protein/DNA interaction. This complex still remains a highly important target for both our general understanding of the mechanism of NHEJ and as a drug target for chemo-therapy intervention.

Mutations in the head domains of either XRCC4 or XLF which abolish head to head interaction are associated with a significant sensitivity to double strand break inducing agents. Development of a small molecule that could interrupt this interaction would therefore be able to re-sensitize chemo-therapeutic resistant tumor cells whereby resistance is mediated by up-regulation of NHEJ to counteract the damage being caused. As actively dividing somatic cells would eventually be able to invoke HRR, the rapidly dividing tumor cells would therefore succumb to the damage. Multiple potential targets exist within NHEJ

and as our mechanistic understanding of the pathway grows, so to do the opportunities to modulate it through small molecule intervention. This thesis has focused on the underlying protein/protein and protein/DNA interactions which not only govern repair by NHEJ, but also represent prime therapeutic targets.

## References

Abramoff, M.D., Magelhaes, P.J., Ram, S.J. (2004). Image Processing with ImageJ. *Biophotonics International.***11:** 36-42.

Adams, P.D., Grosse-Kunstleve, R.W., Hung, L.W., Ioerger, T.R., McCoy, A.J., Moriarty, N.W., Read, R.J., Sacchettini, J.C., Sauter, N.K., and Terwilliger, T.C. (2002). PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr.* **58:** 1948-1954.

Ahnesorg, P., Smith, P., and Jackson, S.P. (2006). XLF interacts with the XRCC4-DNA ligase IV complex to promote DNA nonhomologous end-joining. *Cell* **2:** 301-313.

Ahnesorg, P. and Jackson, S.P. (2007) The non-homologous end-joining protein Nej1P is a target of the DNA damage checkpoint. *DNA repair* **6:** 190-201

Alt, F.W., and Baltimore, D. (1982). Joining of immunoglobulin heavy chain gene segments: implications from a chromosome with evidence of three D-JH fusions. *Proc Natl Acad Sci USA* **79:** 4118-4122.

Andres, S.N., Modesti, M., Tsai, C.J., Chu, G., and Junop, M.S. (2007). Crystal structure of XLF: a twist in non-homologous end-joining. *Mol Cell* **28:** 1093-1101.

Andres, S.N., Vergnes, A., Ristic, D., Wyman, C., Modesti, M., and Junop, M (2011). A human XRCC4-XLF complex bridges DNA. *Nucl Acids Res* **40:** 1868-1678

Aoufouchi, S., Flatter, E., Dahan, A., Faili, A., Bertocci, B., Storck, S., Delbos, F., Cocea, L., Gupta, N., Weill, J.C., and Reynaud, C.A. (2000). Two novel human and mouse DNA polymerases of the PolX family. *Nucl Acids Res* **28:** 3684-3693.

Balmus, G., Barros, A.C., Wijnhoven, P.W., 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. and Jackson, S.P. (2016) Synthetic lethality between PAXX and XLF in mammalian development. *Genes Dev* **19:** 2152-2157.

Ban, N., Nissen, P., Hansen, J., Capel, M., Moore, P.B. and Steitz, T.A. (1999). Placement of protein and RNA structures into a 5 A-resolution map of the 50S ribosomal subunit. *Nature* **400:** 841-847

Banumathi, S., Dauter, M., & Dauter, Z. (2002). SAD/MAD phasing at high resolution using tantalum bromide cluster *Acta Cryst* **A58:** c81.

Battye, T.G.G., Kontogiannis, L., Johnson, O., Powell, H.R. and Leslie, A. G.W. (2011) IMosflm: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Cryst* **D76:** 271-281

Bennardo, N., Cheng, A., Huang, N., and Stark, J.M. (2008). Alternative-NHEJ is a mechanistically distinct pathway of mammalian chromosome break repair. *PLoS Genet* **4:** e1000110.

Bennet, R.A., Swerdlow, P.S., and Povirk, L.F. (1993). Spontaneous cleavage of bleomycin-induced abasic sites in chromatin and their mutagenicity in mammalian shuttle vectors. *Biochemistry* **32:** 3188-3195.

Beucher, A., Birraux, J., Tchouandong, L., Barton, O., Shibata, A., Conrad, S., Goodarzi, A.A., Krempler, A., Jeggo, P.A., and Löbrich, M. (2009). ATM and Artemis promote homologous recombination of radiation-induced DNA double-strand breaks in G2. *EMBO J* **28:** 3413-3427.

Biedermann, K.A., Sun, J.R., Giaccia, A.J., Tosto, L.M., and Brown, J.M. (1991). Scid mutation in mice confers hypersensitivity to ionizing radiation and a deficiency in DNA double-strand break repair. **88:** 1394-1397.

Bjornsti, M.A., and Megonigal, M.D. (1999). Resolution of DNA molecules by onedimensional agarose-gel electrophoresis. *Methods Mol Biol* **94:** 9-17.

Blier, P.R., Griffith, A.J., Craft, J., and Hardin, J.A. (1993). Binding of Ku protein to DNA. Measurement of affinity for ends and demonstration of binding to nicks. *J Biol Chem* **268:** 7594-7601.

Borek, C., and Sachs, L. (1966). *In Vitro* cell transformation by X-irradiation. *Nature* **210:** 276-278.

Bork, P., Hofmann, K., Bucher, P., Neuwald, A.F., Altschul, S.F., and Koonin, E.V. (1997). A superfamily of conserved domains in DNA damage-responsive cell cycle checkpoint proteins. *FASEB J* **11:** 68-76.

Brouwer, I., Sitters, G., Candelli A., Heerema, S.J., Heller, I., Melo De, A.J., Zhang, H., Normanno, D., Modesti, M., Peterman, E.J., and Wuite, G.J. (2016) Sliding sleeves of XRCC4-XLF bridge DNA and connect fragments of broken DNA. *Nature* **535:** 566 -569

Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve,
R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S. *et al*. (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta*
*Crystallogr D Biol Crystallogr* **Pt 5:** 905-921.

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., and Revy, P. (2006). Cernunnos, a novel nonhomologous end-joining factor, is mutated in human immunodeficiency with microcephaly. *Cell* **124:** 287-299.

Budman, J., and Chu, G. (2005). Processing of DNA for nonhomologous end-joining by cell-free extract. *EMBO J*. 4, 849-860.

Callebaut, I., Malivert, L., Fischer, A., Mornon, J.P., Revy, P. and de Villartay, J.P. (2006) Cernunnos interacts with the XRCC4x DNA-ligase IV complex and is homologous to yeast nonhomologous end-joining factor Nej1. *J Biol Chem* **281:** 13857-13860

Callebaut, I., Moshous, D., Mornon, J.P., and de Villartay, J.P. (2002). Metallo-betalactamase fold within nucleic acids processing enzymes: the beta-CASP family. *Nucleic Acids Res* **30:** 3592-3601.

Callen, E., Jankovic, M., Wong, N., Zha, S., Hua-Tang, C., Difilippantonio, S., Di Virgilio, M., Heidkamp. G., Alt, F.W., Nussenzweig, A. and Nussenzweig, M. (2009) Essential role for DNA-PKcs in DNA double strand break repair and apoptosis in ATM deficient lymphocytes. *Mol Cell* **34:** 285-297

Calsou, P., Delteil, C., Frit, P., Drouet, J., and Salles, B. (2003). Coordinated assembly of Ku and p460 subunits of the DNA-dependent protein kinase on DNA ends is necessary for XRCC4-ligase IV recruitment. *J Mol Biol* **326:** 93-103.

Carfi, A., Pares, S., Duée, E., Galleni, M., Duez, C., Frère, J.M., and Dideberg, O. (1995). The 3-D structure of a zinc metallo-beta-lactamase from Bacillus cereus reveals a new type of protein fold. *EMBO J* **14:** 4914-4921.

Carter, S.D., Vigasova, D., Chen, J., Chovanec, M. and Astrom, S.U. (2009) Nej1 recruits the Srs2 helicase to DNA double-strand breaks and supports repair by a single-strand annealing-like mechanism. *Proc Natl Acad Sci* **106:** 12037-12042

Cavero, S., Chahwan, C., and Russell, P. (2007). Xlf1 is required for DNA repair by nonhomologous end joining in Schizosaccharomyces pombe. *Genetics* **2:** 963-967.

Cavazzana-Calvo, M., Le Deist, F., De Saint Basile, G., Papadopoulo, D., De Villartay, J.P., and Fischer, A. (1993). Increased radiosensitivity of granulocyte macrophage colony-forming units and skin fibroblasts in human autosomal recessive severe combined immunodeficiency. *J Clin Invest* **91:** 1214-1218.

Chan, D.W., and Lees-Miller, S.P. (1996). The DNA-dependent protein kinase is inactivated by autophosphorylation of the catalytic subunit. **271:** 8936-8941.

Chan, D.W., Ye, R., Veillette, C.J., and Lees-Miller, S.P. (1999). DNA – dependent protein kinase phosphorylation sites in Ku70/80 heterodimer. *Biochemistry* **38:** 18191828.

Chance, B., Sies, H., and Boveris, A. (1979). Hydroperoxide metabolism in mammalian organs. *Physiol Rev* **59:** 527-605.

Chappell, C., Hanakahi, L.A., Karimi-Busheri, F., Weinfeld, M., and West, S.C. (2002). Involvement of human polynucleotide kinase in double-strand break repair by nonhomologous end joining. *EMBO J* **21:** 2827-2832.

Chen, X., and Tomkinson, A.E. (2011). Yeast Nej1 is a key participant in the initial end binding and final ligation steps of nonhomologous end joining. *J Biol Chem* **286:** 49314940.

Chirgadze, D.Y., Ascher, D.B., Blundell, T.L. and Sibanda, B.L. (2017) DNA-PKcs, allostery and DNA double-strand break repair: Defining the structure and setting the stage. *Methods Enzymol* **592.** 145-157

Conlin, M.P., Reid, D.A., Small, G.W., Chang, H.H., Watanabe. G., Lieber. M.R., Ramsden, D.A. and Rothenberg. E. (2017). DNA ligase IV guides end-processing choice during nonhomologous end joining. *Cell Rep* **20.** 2810-2819

Conway, A.B., Lynch, T.W., Zhang, Y., Fortin, G.S., Fung, C.W., Symington, L.S. and Rice, P.A. (2004). Crystal structure of a Rad51 filament *Nat Struct Mol Biol* **11:** 791-796.

Costantini, S., Woodbine, L., Andreoli, L., Jeggo, P.A., and Vindigni, A. (2007). Interaction of the Ku heterodimer with the DNA ligaseIV/XRCC4 complex and its regulation by DNA-PK. *DNA Repair (Amst)* **6:** 712-722.

Cotner-Gohara, E., Kim, I.K., Hammel, M., Tainer, J.A., Tomkinson, A.E., and Ellenberger, T. (2010). Human DNA ligase III recognizes DNA ends by dynamic switching between two DNA-bound states. *Biochemistry* **49:** 6165-6176.

Critchlow, S.E., Bowater, R.P., and Jackson, S.P. (1997). Mammalian DNA doublestrand break repair protein XRCC4 interacts with DNA ligase IV. *Curr Biol* **7:** 588-598.

Daley, J.M., Palmbos, P.L., Wu, D., and Wilson, T.E. (2005) Nonhomologous end joining in yeast. *Annu Rev. Gener* **39:** 431-451

Datta, S., Prabu, M.M., Vaze, M.B., Ganesh, N., Chandra, N.R., Muniyappa, K., and Vijayan, M. (2000). Crystal structures of Mycobacterium tuberculosis RecA and its complex with ADP-AlF (4): implications for decreased ATPase activity and molecular aggregation. *Nucleic Acids Res* **28:** 4964-4973.

DeFazio, L.G., Stansel, R.M., Griffith, J.D., and Chu, G. (2002). Synapsis of DNA ends by DNA-dependent protein kinase. *EMBO J* **21:**3192-3200.

DeLano, W.L. (2002). The PyMOL User's Manual (Palo Alto, CA, USA.: DeLano
Scientific).

DeRose, E.F., Clarkson, M.W., Gilmore, S.A., Galban, C.J., Tripathy, A., Havener, J.M., Mueller, G.A., Ramsden, D.A., London, R.E., and Lee, A.L. (2007). Solution structure of polymerase mu's BRCT domain reveals an element essential for its role in nonhomologous end joining. *Biochemistry* **46:** 12100-12110.

Deshpande, R.A. and Wilson, T.E. (2007). Modes of interaction among yeast Nej1, Lif1 and Dnl4 proteins and comparison to human XLF, XRCC4 and Lig4. *DNA Repair (Amst)* **6:** 1507-1516.

Desiderio, S.V., Yancopoulos, G.D., Paskind, M., Thomas, E., Boss, M.A., Landau, N., Alt, F.W., and Baltimore, D. (1984). Insertion of N regions into heavy-chain genes is correlated with expression of terminal deoxytransferase in B cells. *Nature* **311:** 752-755.

deVillartay, J.P., Shimazaki, N., Charbonnier, J.B., Fischer, A., Mornon, J.P., Lieber, M.R., and Callebaut, I. (2009). A histidine in the beta-CASP domain of Artemis is critical for its full *In Vitro* and *In Vivo* functions. *DNA Repair (Amst)* **8:** 202-208.

deVriest, E., van Driel, W., Bergsma, W.G., Arnberg, A.C., and van der Vliet, P.C. (1989). HeLa nuclear protein recognizing DNA termini and translocating on DNA forming a regular DNA-multimeric protein complex. *J Mol Biol* **208:** 65-78.

Ding, Q., Reddy, Y.V., Wang, W., Woods, T., Douglas, P., Ramsden, D.A., Lees-Miller, S.P., and Meek, K. (2003). Autophosphorylation of the catalytic subunit of the DNAdependent protein kinase is required for efficient end processing during DNA doublestrand break repair. *Mol Cell Biol* **23:** 5836-5848.

Dobbs, T.A., Tainer, J.A., and Lees-Miller, S.P. (2010). A structural model for regulation of NHEJ by DNA-PKcs autophosphorylation. *DNA Repair (Amst)* **9:** 1307-1314.

Dolman, M.E., van der Ploeg. I., Koster, J., Bate-Eya, L.T., Versteeg, R., Caron, H.N. and Molenar, J.J. (2015) DNA-Dependent Protein Kinase As Molecular Target for Radiosensitization of Neuroblastoma Cells. *PLos One* **10:** e145744

Dore, A.S., Furnham, N., Davies, O.R., Sibanda, B.L., Chirgadze, D.Y., Jackson, S.P., Pellegrini, L., Blundell, T.L. (2006). Structure of an Xrcc4-DNA LigaseIV yeast ortholog complex reveals a novel BRCT interaction mode. *DNA Repair*, **5:** 362-368.

Douglas, P., Moorhead, G.B., Ye, R., and Lees-Miller, S.P. (2001). Protein phosphatases regulated DNA-dependent protein kinase activity. *J Biol Chem* **276:** 18992-18998.

Douglas, P., Gupta, S., Morrice, N., Meek, K. and Lees-Miller, S. P. (2005) DNA-PKdependent phosphorylation of Ku70/80 is not required for non-homologous end joining. *DNA Repair* **4:** 1006–1018.

Dragan, A.I., Liggins, J.R., Crane-Robinson, C. and Privalov, P.L. (2003) The energetics of specific binding of AT-hooks from HMGA1 to target DNA. *J Mol Bio* **327:** 393-411.

Dudasova, Z., Dudas, A., and Chovanec, M. (2004) Non-homologous end joining factors of saccharomyces cerivisiae, *FEMS Microbiol rev* **28:** 581-601

Dvir, A., Peterson, S.R., Knuth, M.W., Lu, H., and Dynan, W.S. (1992). Ku autoantigen is the regulatory component of a template-associated protein kinase that phosphorylates RNA polymerase II. *Proc Natl Acad Sci USA* **89:** 11920-11924.

Early, P., Huang, H., Davis, M., Calame, K., and Hood, L. (1980). An immunoglobulin heavy chain variable region gene is generated from three segments of DNA: VH, D and JH. *Cell* **19:** 981-992.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr **60:** 2126-2132.

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. and Ehl, S. (2006) A severe form of human combined immunodeficiency due to mutations in DNA ligase IV. *J Immunol* **176:** 5060-5068

Evans, P.R. and Murshudov, G.N. (2013) How good are my data and what is the resolution *Acta Cryst* **D76:** 1204-1214

Falzon, M., Fewell, J.W., Kuff, E.L. (1993). EBP-80, a transcription factor closely resembling the human autoantigen Ku, recognizes single- to double-strand transitions in DNA. *J Biol Chem* **268:** 10546-10552.

Fan, W., and Wu, X. (2004). DNA polymerase lambda can elongate on DNA substrates mimicking non-homologous end joining and interact with XRCC4-ligase IV complex. *Biochem Biophys Res Commun* **323:** 1328-1333.

Frey, M. (1994). Water structure associated with proteins and its role in crystallization. *Acta Cryst* **D50:** 663-666.

Frank, K.M., Sekiguchi, J.M., Seidl, K.J., Swat, W., Rathbun, G.A., Cheng, H.L., Davidson, L., Kangaloo, L., and Alt, F.W. (1998). Late embryonic lethality and impaired V(D)J recombination in mice lacking DNA ligase IV. *Nature* **396:** 173-177.

Fukumura, R., Araki, R., Fujimori, A., Mori, M., Saito, T., Watanabe, F., Sarashi, M., Itsukaichi, H., Eguchi-Kasai, K., Sato, K., Tatsumi, K., and Abe, M. (1998). Murine cell line SX9 bearing a mutation in the dna-pkcs gene exhibits aberrant V(D)J recombination not only in the coding joint but also in the signal joint. *J Biol Chem* **273:** 13058-13064. Galkin, V.E., Wu, Y., Zhang, X.P., Qian, X., He, Y., Heyer, W.D., Luo, Y., and Egelman, E.H. (2006). The Rad51/RadA N-terminal domain activates nucleoprotein filament ATPase activity. *Structure* **14:** 983-992.

Gao, Y., Sun, Y., Frank, K.M., Dikkes, P., Fujiwara, Y., Seidl, K.J., Sekiguchi, J.M., Rathbun, G.A., Swat, W., Wang, J., et al. (1998). A critical role for DNA end-joining proteins in both lymphogenesis and neurogenesis. *Cell* **95:** 891-902.

Gao, Y., Ferguson, D.O., Xie, W., Manis, J.P., Sekiguchi, J., Frank, K.M., Chaudhuri, J., Horner, J., DePinho, R.A., and Alt, F.W. (2000). Interplay of p53 and DNA-repair protein XRCC4 in tumorigenesis, genomic stability and development. *Nature* **6780:** 897-900.

Geiger, S.R., Kuhn, C.-D., Leidig, C., Renkawitz, J., and Cramer, P. (2008). Crystallization of RNA polymerase I subcomplex A14/A43 by iterative prediction, probing and removal of flexible regions. *Acta Cryst* **F64:** 413-418.

Gell, D., and Jackson, S.P. (1999). Mapping of protein-protein interactions within the DNA-dependent protein kinase complex. *Nucleic Acids Res* **27:** 3494-4502.

Giaccia, A.J., Denko, N., MacLaren, R., Mirman, D., Waldren, C., Hart, I., and Stamato, T.D. (1990). Human chromosome 5 complements the DNA double-strand break-repair deficiency and gamma-ray sensitivity of the XR-1 hamster variant. *Am J Hum Genet* **47:** 459-469.

Gilfillan, S., Dierich, A., Lemeur, M., Benoist, C., and Mathis, D. (1993). Mice lacking TdT: mature animals with an immature lymphocyte repertoire. *Science* **261:** 1175-1178.

Giloni, L.M., Takeshita, F., Johnson, Iden, C., and Grollman, A.P. (1981). Bleomycininduced strand scission of DNA: mechanism of deoxyribose cleavage. *J Biol Chem* **256:** 8608-8615.

Glusker, J.P., and Trueblood, K.N. (1985). Crystal Structure Analysis: A Primer (Second edition), pp.269, Oxford University Press, New York.

Goodarzi, A. A., Yu, Y., Riballo, E., Douglas, P., Walker, S. A., Ye, R., Harer, C., Marchetti, C., Morrice, N., Jeggo, P. A. and Lees-Miller, S. P. (2006). DNA-PK autophosphorylation facilitates Artemis endonuclease activity. *EMBO J* **25:** 3880–3889.

Gottlieb, T.M., and Jackson, S.P. (1993). The DNA-dependent protein kinase: requirement for DNA ends and association with Ku antigen. *Cell* **72:** 131-142.

Grawunder, U., Wilm, M., Wu, X., Kulesza, P., Wilson, T.E., Mann, M., and Lieber, M.R. (1997). Activity of DNA ligase IV stimulated by complex formation with XRCC4 protein in mammalian cells. *Nature* **388:** 492-495.

Grawunder, U., Zimmer, D., Kulesza, P., and Lieber, M.R. (1998a). Requirement for an interaction of XRCC4 with DNA ligase IV for wild-type V(D)J recombination and DNA double-strand break repair *In Vivo*. *J Biol Chem* **273:** 24708-24714.

Grawunder, U., Zimmer, D., Fugmann, S., Schwarz, K., and Lieber, M.R. (1998b). DNA ligase IV is essential for V(D)J recombination and DNA double-strand break repair in human precursor lymphocytes. *Mol Cell* **2:** 477-484.

Griffith, A.J., Blier, P.R., Mimori, T., and Hardin, J.A. (1992). Ku polypeptides synthesized *In Vitro* assemble into complexes which recognized ends of double-stranded DNA. *J Biol Chem* **267:** 331-338.

Grosse-Kunstleve, R.W., and Adams, P.D. (2003). Substructure search procedures for macromolecular structures. *Acta Crystallogr D Biol* Crystallogr **59:** 1966-1973.

Gu, Y., Jin, S., Gao, Y., Weaver, D.T., and 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. *Proc Natl Acad Sci USA* **94:** 8076-8081.

Habraken, Y., and Verly, W.G. (1983). The DNA 3'-phosphatase and 5'-hydroxyl kinase of rat liver chromatin. *FEBS Lett* **160:** 46-50.

Hammel, M., Rey, M., Yu, Y., Mani, R.S., Classen, S., Liu, M., Pique, M.E., Fang, S., Mahaney, B., Weinfeld, M., Schreimer, D.C., Lees-Miller, S.P., and Tainer, J.A. (2011). XRCC4 interactions with XRCC4-like factor (XLF) create an extended grooved scaffold for DNA ligation and double-strand break repair. *J Biol Chem* **37:** 32638-32650.

Hammel, M., Yu, Y., Fang, S., Lees-Miller, S.P., Tainer, J.A. (2010a). XLF regulates filament architecture of the XRCC4-ligase IV complex. *Structure* **18:** 1431-1442.

Hammel, M., Yu,Y., Mahaney, B.L., Cai, B., Ye, R., Phipps, B.M., Rambo, R.P., Hura, G.L., Pelikan, M., So, S., Abolfath, R.M., Chen, D.J., Lees-Miller, S.P., and Tainer, J.A. (2010b). Ku and DNA-dependent protein kinase dynamic conformations and assembly regulate DNA binding and the initial non-homologous end joining complex. *J Biol Chem* **285:** 1414-1423.

Hammel, M., Yu, Y., Radhakrishnan, S.K., Chokshi, C., Tsai M.S., Matsumoto, Y., Kuzdovich, M., Remesh, S.G., Fang, S., Tomkinson, A.E., Lees-Miller, S.P. and Tainer, J.A. (2016) An intrinsically Disordered APLF Links Ku, DNA-PKcs, and XRCC4-DNA Ligase IV in an Extended flexible Non-homologous End Joining Complex. *J Biol Chem* **291:** 26987-27006

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., and Jackson, S.P. (1995). DNA-dependent protein kinase catalytic subunit: A relative of phosphatidylinositol 3-kinase and the ataxia telangiectasia gene product. *Cell* **82:** 849-856.

Hei, T.K., Komatsu, K., Hall, E.J., and Zaider, M. (1988). Oncogenic transformation by charged particles of defined LET. *Carcinogenesis* **9:** 747-750.

Hendrickson, E.A., Qin, X.Q., Bump, E.A., Schatz, D.G., Oettinger, M., and Weaver, D.T. (1991). A link between double-strand break-related repair and V(D)J recombination: the scid mutation. *Proc Natl Acad Sci USA* **88:** 4061-4065.

Hendrickson, W.A., Horton, J.R., and LeMaster, D.M. (1990). Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *EMBO J* **5:** 1665-1672.

Hentges, P., Ahnesorg, P., Pitcher, R.S., Bruce, C.K., Kysela, B., Green, A.J., Bianchi, J., Wilson, T.E., Jackson, S.P., and Doherty, A.J. (2006). Evolutionary and functional conservation of the DNA non-homologous end-joining protein, XLF/Cernunnos. *J Biol Chem* **49:** 37517-37526.

Heras, B., Edeling, M. A., Byriel, K. A., Jones, A., Raina, S. & Martin, J. L. (2003). Dehydration converts double-strand break crystal diffraction from low to high resolution Structure **11: 139–145.**

Heras, B., and Martin, J.L. (2005). Post-crystallization treatments for improving diffraction quality of protein crystals. *Acta Cryst* **D61:** 1173-1180.

Hung, P.J., Chen, B.R., George, R., Liberman, C., Morales, A.J., Colon-Ortiz, P., Tyler, J.K., Sleckman, B.P. and Bredermeyer, A.L. (2017) Deficiency of XLF and PAXX prevents DNA double-strand break repair by non-homologous end joining in lymphocytes.
*Cell Cycle* **16:** 286-295.

Huxford, T., Malek, S., and Ghosh, G. (2000). Preparation and crystallization of dynamic
NF-kappa B.Ikappa B complexes. *J Biol Chem* **275:** 32800-32806

Iles, N., Rulten, S., El-Khamisy, S.F. and Caldecott, KW. (2007) APLF (C2orf13) is a novel human protein involved in the cellular response to chromosomal DNA strand breaks. *Mol Cell* **27:** 3793-3803

Ito, J., and Braithwaite, D.K. (1991). Compilation and alignment of DNA polymerase sequences. *Nucleic Acids Res* **19:** 4045-4057.

Jackson, S.P. (2002). Sensing and repairing DNA double-strand breaks. *Carcinogenesis* **23:** 687-696.

Jackson, S.P., and Jeggo, P.A. (1995). DNA double-strand break repair and V(D)J recombination: involvement of DNA-PK. *Trends Biochem Sci* **20:** 412-415.

Janin, J., and Chothia, C. (1990). The structure of protein-protein recognition sites. *J Biol Chem* **265:** 16027-16030.

Jeggo, P.A. (1998). Identification of genes involved in repair of DNA double-strand breaks in mammalian cells. *Radiat Res* **5:** Suppl, S80-91.

Jiang, J., Tang, W., An, Y., Tang, M., Wu, J., Qin, T. and Zhao, X. (2016) Molecular and immunological characterization of DNA ligase IV deficiency *Clin Immunol* **163:** 75-83

Jones, T.A., Zou, J.Y., Cowan, S.W., and Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr A* **47:** 110-119.

Junop, M.S., Modesti, M., Guarne, A., Ghirlando, R., Gellert, M., and Yang, W. (2000).
Crystal structure of the Xrcc4 DNA repair protein and implications for end joining.
*EMBO J* **22:** 5962-5970.

Kadyk, L.C., and Hartwell, H. (1992). Sister chromatids are preferred over homologs as substrates for recombinational repair in Saccharomyces cerevisiae. *Genetics* **132:** 387402.

Keeney, S., Giroux, C.N., and Kleckner, N. (1997). Meiosis-specific DNA double-strand breaks are catalyzed by Spo11, a member of a widely conserved protein family. *Cell* **88:** 375-384.

Kitagawa, D., Vakonakis, I., Olieric, N., Hilbert, M., Keller, D., Olieric, V., Bortfeld, M., Erat, M.C., Flückiger, I., Gönczy, P., and Steinmetz, M.O. (2011). *Cell* **144:** 364-375.

Knäblein, J., Neuefeind, T., Schneider, F., Bergner, A., Messerschmidt, A., Löwe, J., Steipe, B., & Huber, R. (1997). Ta6Br(2+)12, a tool for phase determination of large biological assemblies by X-ray crystallography *J Mol Biol* **270:** 1-7.

Koch, C.A., Agyei, R., Galicia, S., Metalnikov, P., O'Donnell, P., Starostine, A., Weinfeld, M., and Durocher, D. (2004). Xrcc4 physically links DNA end processing by polynucleotide kinase to DNA ligation by DNA ligase IV. *EMBO J* **23:** 3874-3885.

Komori, T., Okada, A., Stewart, V., and Alt, F.W. (1993). Lack of N regions in antigen receptor variable region genes of TdT-deficient lymphocytes. *Science* **261:** 1171-1175.

Krissinel, E., and Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. *J Mol Biol* **372:** 774-797.

Kuzminov, A. (1995). Collapse and repair of replication forks in Escherichia coli. *Mol Microbiol* **16:** 373-384.

Kuzminov, A. (2001). Single-strand interruptions in replicating chromosomes cause double-strand breaks. *Proc Natl Acad Sci USA* **98:** 8241-8246.

Leber, R., Wise, T.W., Mizuta, R., and Meek, K. (1998). The XRCC4 gene product is a target for and interacts with the DNA-dependent protein kinase. *J Biol Chem* **273:** 17941801.

Lee, J.W., Blanco, L., Zhou, T., Garcia-Diaz, M., Bebenek, K., Kunkel, T.A., Wang, Z., and Povirk, L.F. (2003). Implication of DNA polymerase lambda in alignment-based gap filling for nonhomologous DNA end joining in human nuclear extracts. *J Biol Chem* **279:** 805-811.

Lee, J.H., and Paull, T.T. (2005). ATM activation by DNA double-strand breaks through the Mre11-Rad50-Nbs1 complex. *Science* **308:** 551-554.

Lee, J.H. and Paul, T.T. (2007) Activation and regulation of ATM kinase activity in response to DNA double-strand breaks *Oncogene* **26:** 7741-7748

Lee, K.J., Huang, J., Takeda, Y., and Dynan, W.S. (2000). DNA Ligase IV and XRCC4 form a stable mixed tetramer that functions synergistically with other repair factors in a cell-free end-joining system. *J Biol Chem* **275:** 34787-34796.

Lee, K.W.Y (2013) Characterization of the oligomerization of the human XRCC4 DNA repair protein: Implications to Non-Homologous End Joining. Masters Dissertation

Lehman, I.R. (1974). DNA ligase: structure, mechanism and function. *Science* **186:** 790-797.

Leidel, S., Delattre, M., Cerutti, L., Baumer, K,. and Gönczy, P. (2005). SAS-6 defines a protein family required for centrosome duplication in *C. elegans* and in human cells. *Nat Cell Biol* **7:** 115-125.

Li, S., Ting, N.S., Zheng, L., Chen, P.L., Ziv, Y., Shiloh, Y., Lee, E.Y., and Lee, W.H. (2000). Functional link of BRCA1 and ataxia telangiectasia gene product in DNA damage response. *Nature* **406:** 210-215.

Li, Y.H., Wang, X., Pan, Y., Lee, D.H., Chowdhurry, D. and Kimmelman, A.C. (2012). Inhibition of non-homologous end joining impairs pancreatic cancer growth and enhances radiation response. *Plos One* **7:** e39588.

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. *EMBO J*. **27:** 290-300.

Li, Z., Otevrel, T., Gao, Y., Cheng, H.L., Seed, B., Stamato, T.D., Taccioli, G.E., and Alt, F.W. (1995). The XRCC4 gene encodes a novel protein involved in DNA double-strand break repair and V(D)J recombination. *Cell* **83:** 1079-1089.

Liang, F., and Jasin, M. (1996). Ku80-deficient cells exhibit excess degradation of extrachromosomal DNA. *J Biol Chem* **24:** 14405-14411.

Lieber, M.R. (1998). Warner-Lambert/Parke-Davis Award Lecture. Pathological and physiological double-strand breaks: roles in cancer, aging, and the immune system. *Am J Pathol* **5:** 1323-1332.

Lieber, M.R., Gu, J., Lu, H., Shimazaki, N., & Tsai, A.G. (2010). Nonhomologous DNA end joining (NHEJ) and chromosomal translocations in humans. *Subcell Biochem* **50:** 279-296.

Lieber, M.R., Hesse, J.E., Lewis, S., Bosma, G.C.,. Rosenberg, N., Mizuuchi, K., Bosma, M.J., and Gellert, M. (1988a). The defect in murine severe combined immune deficiency: joining of signal sequences but not coding sequences in V(D)J recombination. *Cell* **55:** 7-16.

Lieber, M.R., Hesse, J.E., Mizuuchi, K., and Gellert, M. (1988b). Lymphoid V(D)J recombination: nucleotide insertion at signal joints as well as coding joints. *Proc Natl Acad Sci USA* **85:** 8588-8592.

Lieber, R., Wise, T. W., Mizuta, R., and Meek, K. (1998) The XRCC4 gene product is a target for and interacts with the DNA-dependent protein kinase. *J. Biol. Chem.* **273:** 1794-1801.

Liu, X., Shao, Z., Jiang, W., Lee, B.J. and Zha, S. (2017) PAXX promotes KU accumulation at DNA breaks and is essential for end-joining in XLF-deficient mice. *Nat Commun.*

Lu, H., Pannicke, U., Schwarz, K., and Lieber, M.R. (2007). Length-dependent binding of human XLF to DNA and stimulation of XRCC4.DNA Ligase IV activity. *J Biol Chem* **282:** 11155-11162.

Ma, Y., Pannicke, U., Schwarz, K., and 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* **6:** 781-794.

Ma, Y., Pannicke, U., Lu, H., Niewolik, D., Schwarz, K., and Lieber, M.R. (2005). The DNA-dependent protein kinase catalytic subunit phosphorylation sites in human Artemis. *J Biol Chem* **280:** 33839-33846.

Mahajan, K.N., McElhinny, N., Mitchell., B.S., and Ramsden, D.A. (2002). Association of DNA polymerase mu (pol mu) with Ku and ligase IV: role for pol mu in end-joining double-strand break repair. *Mol Cell Biol* **22:** 5194-5202.

Mahaney, B. L., Meek, K., and Lees-Miller, S.P. (2009) Repair of ionizing radiation-induced DNA double strand breaks by Non-homologous end joining. *Biochem J* **417**: 639-650)

Malivert, L., Ropars, V., Nunez, M., Drevet, P., Miron, S., Faure, G., Guerois, R., Mornon, J.P., Revy, P., Charbonnier, J.B., Callebaut, I., & de Villartay, J.P. (2010). Delineation of the Xrcc4-interacting region in the globular head domain of cernunnos/XLF. *J Biol Chem* **285:** 26475-26483.

Mandel, C.R., Kaneko, S., Zhang, H., Gebauer, D., Vethantham, V., Manley, J.L., and Tong, L. (2006). Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease. *Nature* **444:** 953-956.

Martin, I.V., and MacNeill, S.A. (2002). ATP-dependent DNA ligases. *Genome Biol* **3:**
reviews3005-3005.7.

Maser, R.S., Monsen, K.J., Nelms, B.E., and Petrini, J.H. (1997). hMre11 and hRad50 nuclear foci are induced during the normal cellular response to DNA double-strand breaks. *Mol Cell Biol* **17:** 6087-6096.

Matsuoka, S., Rotman, G., Ogawa, A., Shiloh, Y., Tamai, K., and Elledge, S.J. (2000). Ataxia telangiectasia-mutated phosphorylates Chk2 *In Vivo* and *In Vitro*. *Proc Natl Acad Sci USA* **97:** 10389-10394.

McBlane, J.F., van Gent, D.C., Ramsden, D.A., Romeo, C., Cuomo, C.A., Gellert, M., and Oettinger, M.A. (1995). Cleavage at a V(D)J recombination signal requires only RAG1 and RAG2 proteins and occurs in two steps. *Cell* **83:** 387-395.

McElhinny, S.A., Havener, J.M., Garcia-Diaz, M., Juárez, R., Bebenek, K., Kee, B.L., Blanco, L., Kunkel, T.A., and Ramsden, D.A. (2005). A gradient of template dependence defines distinct biological roles for family X polymerases in nonhomologous end joining. *Mol Cell* **19:** 357-366.

McElhinny, N.S.A., Snowden, C.M., McCarville, J., and Ramsden, D.A. (2000) Ku recruits the XRCC4-Ligase IV complex to DNA ends. *Mol Cell Biol* **20:** 2996-2003.

McFadden, M.J., Lee, W.K., Brennan, J.D., and Junop, M.S. (2014) Delineation of keu XRCC4/Ligase IV interfaces for targeted disruption of non-homologous end joining DNA repair. *Proteins* **82:** 187-194

McPherson, A. Jr. (1976). The growth and preliminary investigation of protein and nucleic acid crystals for X-ray diffraction analysis. *Methods Biochem Anal* **23:** 249-345.

McPherson, A. (2003). Introduction to Macromolecular Crystallography, pp.237, John Wiley & Sons, Inc., New Jersey.

Meek, K., Douglas, P., Cui, X., Ding, Q. and 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. *Mol Cell Biol* **27:** 3881–3890.

Mehrotra, P.V., Ahel, D., Ryan, D.P., Weston, R., Wiechens, N., Kraehenbuehl, R., Owen-Hughes, T. and Ahel, I. (2011) DNA repair factor APLF is a histone chaperone *Mol Cell* **41:** 46-55

Michael, B.D. and O'Neill, P. (2000). A sting in the tail of electron tracks. *Science* **287:** 1603-1604.

Mimori, T., Akizuki, H., Yamagat, S., Inada, S., Yoshida, S., and Homma, M. (1981). Characterization of a high molecular weight acidic nuclear protein recognized by autoantibodies in sera from patients with polymyositis-scleroderma overlap. *J Clin Invest* **68:** 611-620.

Mimori, T., and Hardin, J.A. (1986). Mechanism of interaction between Ku protein and DNA. *J Biol Chem* **261:** 10375-10379.

Modesti, M., Hesse, J.E., and Gellert, M. (1999). DNA binding of Xrcc4 protein is associated with V(D)J recombination but not with stimulation of DNA ligase IV activity. *EMBO J* **18:** 2008-2018.

Modesti, M., Junop, M.S., Ghirlando, R., van de Rakt, M., Gellert, M., Yang, W., and Kanaar, R. (2003). Tetramerization and DNA LigaseIV interaction of the DNA doublestrand break repair protein XRCC4 are mutually exclusive. *J Mol Biol* **2:** 215-228.

Moshous, D., Li, L., Chasseval., R., Philippe, N., Jabado, N., Cowan, M.J., Fischer, A., and de Villartay, J.P. (2000). A new gene involved in DNA double-strand break repair and V(D)J recombination is located on human chromosome 10p. *Hum Mol Genet* **9:** 583-588.

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., and 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:** 177-186.

Mueller, G.A., Moon, A.F., DeRose, E.F., Havener, J.M., Ramsden ,D.A., Pedersen, L.C., and London, R.E. (2008). A comparison of BRCT domains involved in nonhomologous end-joining: Introducing the solution structure of the BRCT domain of polymerase lambda. *DNA Repair* **7:** 1340-1351.

Murshudov, G. (2005). Refmac5 (CCP4: Supported Program).April 15, 2.

Murzin, A.G. (1993). OB(oligonucleotide/oligosaccharide binding)-fold: common structural and functional solution for non-homologous sequences. *EMBO J* **12:** 861-867.

Niewolik, D., Pannicke, U., Lu, H., Ma, Y., Wang L.C., Kulesza, P., Zandi, E., Lieber, M.R., and Schwarz, K. (2006). DNA-PKcs dependence of Artemis endonucleolytic activity, differences between hairpins and 5' or 3' overhangs. *J Biol Chem* **281:** 33900-33909.

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., and Jackson, S.P. (2015) DNA repair: PAXX, a paralog of XRCC4 and XLF, interacts with Ku to promote DNA double-strand break repair. *Science* **347:** 185-188

Ochi, T., Wu Q., and Blundell, T.L. Structure of the catalytic region of DNA ligase IV in complex with an Artemis fragment sheds light on double-strand break repair. *Structure* **21:** 672-679

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., and Concannon, P. (2001). DNA ligase IV mutations identified in patients exhibiting developmental delay and immunodeficiency. *Mol Cell* **8:** 1175-1185.

Oettinger, M.A., Schatz, D.G., Gorka, C., and Baltimore, D. (1990). RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination. *Science* **248:** 15171523.

Ooi, S.L., Shoemaker, D.D., and Boeke, J.D. (2001). A DNA microarray-based genetic screen for nonhomologous end-joining mutants in Saccharomyces cerevisiae. *Science* **5551:** 2552-2556.

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray Diffraction Data Collected in Oscillation Mode. *Methods Enzymol.* **276:** 307-326.

Paillard, S., and Strauss, F. (1991). Analysis of the mechanism of interaction of simian Ku protein with DNA. *Nucleic Acids Res* **19:** 5619-5624.

Pascal, J.M., O'Brien, P.J., Tomkinson, A.E., and Ellenberger, T. (2004). Human DNA ligase I completely encircles and partially unwinds nicked DNA. *Nature* **432:** 473-478.

Pavlicek, A., and Jurka, J. (2006). Positive selection on the nonhomologous end-joining factor Cernunnos-XLF in the human lineage. *Biol. Direct* **15:**. doi:10.1186/1745-6150-115

Pflugrath, J.W. (1999). The finer things in X-ray diffraction data collection. *Acta Crystallogr D Biol Crystallogr* **55:** 1718-1725.

Pheiffer, B.H., and Zimmerman, S.B. (1982). 3'-phosphatase activity of the DNA kinase from rat liver. *Biochem Biophys Res Commun* **109:** 1297-1302.

Pomeranz Krummel, D.A., Oubridge, C., Leung, A.K., Li, J., and Nagai, K. (2009). Crystal structure of human spliceosomal U1 snRNP at 5.5 A resolution. *Nature* **458:** 475480.

Radhakrishnan, S.K. and Lees-Miller, S.P. (2017) DNA requirements for interaction of the C-terminal region of Ku80 with the DNA-dependent protein kinase catalytic subunit. *DNA Repair*B **57:** 17-28

Ramsden, D.A., and Gellert, M. (1995). Formation and resolution of double-strand break intermediates in V(D)J rearrangement. *Genes Dev* **9:** 2409-2420.

Recuero-Checa, M.A., Doré, A.S., Arias-Palomo, E., Rivera-Calzada, A., Scheres, S.H., Maman, J.D., Pearl, L.H., and Llorca, O. (2009). Electron microscopy of Xrcc4 and the DNA ligase IV-Xrcc4 DNA repair complex. *DNA Repair Amst* **8:** 1380-1389.

Reid, D.A., Keegan, S., Leo-Macias, A., Watanabe, G., Strande, N.T., Chang, H.H., Oksuz, B.A., Fenyo, D., Lieber, M.R., Ramsden, D.A. and Rothenberg, E. (2015) Organization and dynamics of the nonhomologous end-joining machinery during DNA double-strand break repair. *Proc Nat Acad Sci*. **112**: 2575-2584

Resnick, M.A., and Martin, P. (1976). The repair of double-strand breaks in the nuclear DNA of Saccharomyces cerevisiae and its genetic control. *Mol Gen Genet* **143:** 119-129.

Riballo, E., Critchlow, S.E., Teo, S.H., Doherty, A.J., Priestley, A., Broughton, B., Kysela, B., Beamish, H., Plowman, N., Arlett, C.F., Lehmann, A.R., Jackson, S.P., and Jeggo, P.A. (1999). Identification of a defect in DNA ligase IV in a radiosensitive leukaemia patient. *Curr Biol* **9:** 699-702.

Riballo, E., Woodbine, L., Stiff, T., Walker, S.A., Goodarzi, A.A., and Jeggo, P.A. (2009). XLF-Cernunnos promotes DNA ligase IV-XRCC4 re-adenylation following ligation. *Nucleic Acids Res* **37:** 482-492.

Rich, T., Allen, R.L., and Wyllie, A.H. (2000). Defying death after DNA damage. *Nature* **407:** 777-783.

Richter, C., Park, J.W., and Ames, B.N. (1988). Normal oxidative damage to mitochondrial and nuclear DNA is extensive. *Proc Natl Acad Sci USA* **85:** 6465-6467.

Ristic, D., Wyman, C., Paulusma, C., and Kanaar, R. (2001). The architecture of human Rad54-DNA complex provides evidence for protein translocation along DNA. *Proc Natl Acad Sc. USA* **98:** 8454-8460.

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

Robins, P., and Lindahl, T. (1996). DNA ligase IV from HeLa cell nuclei. *J Biol Chem* **271:** 24257-24261.

Rogakou, E.P., Pilch, D.R., Orr, A.H., Ivanova, V.S., and Bonner, W.M. (1998). DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. *J Biol Chem* **273:** 5858-5868.

Rooney, S., Alt, F.W., Lombard, D., Whitlow, S., Eckersdorff, M., Fleming, J., Fugmann, S., Ferguson, D.O., Schatz, D.G., and Sekiguchi, J. (2003). Defective DNA repair and increased genomic instability in Artemis-deficient murine cells. *J Exp Med* **197:** 553565.

Ropars, V., Drevet, P., Legrand, P., Baconnais, S., Amram, J., Faure, G., Márquez, J.A., Piétrement, O., Guerois, R, Callebaut, I., LeCam, E., Revy, P., deVillartay, J.P., and Charbonnier, J.B. (2011). Structural characterization of filaments formed by human XRCC4-Cernunnos/XLF complex involved in nonhomologous DNA end-joining. *Proc Natl Acad Sci USA* **31:** 12663-12668.

Roth, D.B., Menetski, J.P., Nakajima, P.B., Bosma, M.J., and Gellert, M. (1992). V(D)J recombination: broken DNA molecules with covalently sealed (hairpin) coding ends in scid mouse thymocytes. *Cell* **70:** 983-991.

Roth, D.B., Zhu, C., Gellert, M. (1993). Characterization of broken DNA molecules associated with V(D)J recombination. *Proc Natl Acad Sci USA* **90:** 10788-10792.

Roy, S., Andres, S.N., Vergnes, A., Neal, J.A., Xu, Y., Yu, Y., Lees-Miller, S.P., Junop, M., Modesti, M., and Meek, K. (2012) XRCC4's interaction with XLF is required for coding (but not signal) end joining. *Nucleic Acids Res* **40:** 1684-1694

Salunke, D.M., Veerapandian, B., Kondandapani, R., and Vijayan, M. (1985). Watermediated transformations in protein crystals. *Acta Cryst* **B41:** 431-436.

Schatz, D.G., Oettinger, M.A., and Baltimore, D. (1989). The V(D)J recombination activating gene, RAG-1. *Cell* **59:** 1035-1048.

Schilling, J., Clevinger, B., Davie, J.M., and Hood, L. (1980). Amino acid sequence of homogeneous antibodies to dextran and DNA rearrangements in heavy chain V-region gene segments. *Nature* **283:** 35-40.

Schroeder, G.F., Levitt, M., and Brunger, A.T. (2010). Super-resolution biomolecular crystallography with low-resolution data. *Nature* **464:** 1218-1222.

Sheriff, S., & Hendrickson, W.A. (1987). Description of overall anisotropy in diffraction from macromolecular crystals. *Acta Cryst.* **A43:** 118-121.

Sibanda, B.L., Chirgadze, D.Y., and Blundell, T.L. (2010). Crystal structure of DNAPKcs reveals a large open-ring cradle comprised of HEAT repeats. *Nature* **463:** 118121.

Sibanda, B.L., Chirgadze, D.Y., Ascher, D.B., and Blundell, T.L. (2017). DNA-PKcs structure suggests an allosteric mechanism modulating DNA double-strand break repair. *Science* **6324:** 520-524

Sibanda, B.L., Critchlow, S.E., Begun, J., Pei, X.Y., Jackson, S.P., Blundell, T.L., and Pellegrini, L. (2001). Crystal structure of an Xrcc4-DNA LigaseIV complex. *Nat Struct Biol* **12:** 1015-1019.

Siddiqi, M.A., and Bothe, E. (1987). Single- and double-strand break formation in DNA irradiated in aqueous solution: dependence on dose and radical scavenger concentration. *Radiat Res* **112:** 449-463.

Singleton, B.K., Torres-Arzaus, M.I., Rottinghaus, S.T., Taccioli, G.E., and Jeggo, P.A. (1999). The C terminus of Ku80 activates the DNA-dependent protein kinase catalytic subunit. *Mol Cell Biol* **19:** 3267-3277.

Smith, J., Riballo, E,. Kysela, B., Baldeyron, C., Manolis, K., Masson, C., Lieber, M.R., Papadopoulo, D., and Jeggo, P. (2003). Impact of DNA ligase IV on the fidelity of end joining in human cells. *Nucleic Acids Res* **31:** 2157-2167.

Spagnolo, L., Rivera-Calzada, A., Pearl, L.H., and Llorca, O. (2006). Three-dimensional structure of the human DNA-PKcs/Ku70/Ku80 complex assembled on DNA and its implications for DNA double-strand break repair. *Mol Cell* **22:** 511-519.

Sulek, M., Yarrington, R., McGibbon, G., Boeke, J.D., and Junop, M. (2007). A critical role for the C-terminus of Nej1 protein in Lif1p association, DNA binding and nonhomologous end-joining. *DNA Repair (Amst)* **6:** 1805-1818.

Sun, H., Treco, D., Schultes, N.P., Szostak, J.W. (1989). Double-strand breaks at an initiation site for meiotic gene conversion. *Nature* **338:** 87-90.

Suwa, A., Hirakata, M., Takeda, Y., Jesch, S., Mimori, T., and Hardin, J.A. (1994). DNA-dependent protein kinase and (Ku protein-p350 complex) assembles on doublestranded DNA. *Proc Natl Acad Sci USA* **91:** 6904-6908.

Taccioli, G.E., Gottlieb, T.M., Blunt, T., Priestley, A., Demengeot, J., Mizuta, R., Lehmann, A.R., Alt, F.W., Jackson, S.P., and Jeggo, P.A. (1994). Ku80 is the product of the XRCC5 gene: a direct link with DNA repair and V(D)J recombination. *Science* **265:** 1442-1445.

Tadi, S.K., Tellier-Lebégue, C., Nemoz.C., Drevet. P., Audebert. S., Roy. S., Meek. K., Charbonier. J.B., and Modesti. M., (2016) PAXX is an accessory c-NHEJ factor that associates with Ku70 and has overlapping functions with XLF. *Cell Rep* **2:** 541-555

Teo, S.H., and Jackson, S.P. (2000). Lif1p targets the DNA ligase Lig4p to sites of DNA double-strand breaks. *Curr Biol* **10:** 165-168.

Tomkinson, A.E., Totty, N.F., Ginsburg, M., and Lindahl, T. (1991). Location of the active site for enzyme-adenylate formation in DNA ligases. *Proc Natl Acad Sci USA* **88:** 400-404.

Tsai, C.J., Kim, S.A., and Chu, G. (2007). Cernunnos/XLF promotes the ligation of mismatched and noncohesive DNA ends. *Proc Natl Acad Sci USA* **104:** 7851-7856.

Tsukuda, T., Fleming, A.B., Nickoloff, J.A., and Osley, M.A. (2005). Chromatin remodelling at a DNA double-strand break site in Saccharomyces cerevisiae. *Nature* **438:** 379-383.

Valencia, M., Bentele, M., Vaze, M.B., Hermann, G., Kraus, E., Lee, S.E., Schar, P. and Haber, J.E. (2001) NEJ1 controls non-homologous end joining in *Saccharomyces cerevisiae Nature* **414:** 666-669

van Breugel, M., Hirono, M., Andreeva, A., Yanagisawa, H.A., Yamaguchi, S., Nakazawa, Y., Morgner, N., Petrovich, M., Ebong, I.O., Robinson, C.V., Johnson, C.M., Veprintsev, D., and Zuber, B. (2011). Structures of SAS-6 suggest its organization in centrioles. *Science* **331:** 1196-1199.

Walker, J.R., Corpina, R.A., and Goldberg, J. (2001). Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair. *Nature* **412:** 607-614.

Wang, Y. G., Nnakwe, C., Lane, W. S., Modesti, M. and Frank, K. M. (2004) Phosphorylation and regulation of DNA ligase IV stability by DNA-dependent protein kinase. *J Biol Chem* **279:** 37282–37290.

Ward, J.F., and Kuo, I. (1973). Models for studying mechanisms of strand breakage in DNA. II. Thymidine 3'5'-diphosphate. *Int J Radiat Biol* **23:** 543-557.

Ward, J.F., and Kuo, I. (1976). Strand breaks, base release, and postirradiation changes in DNA gamma-irradiated in dilute $O_2$-saturated aqueous solution. *Radiat Res* **66:** 485498.

Ward, J.F., (1981). Symposium on radical processes in radiobiology and carcinogenesis. *Radiat Res* **86:** 185-195.

Ward, J.F. (1988). DNA damage produced by ionizing radiation in mammalian cells: identities, mechanisms of formation and reparability. *Prog Nucleic Acids Res Mol Biol* **35:** 95-125.

Wei, Y.F., Robins, P., Carter, K., Caldecott, K., Pappin, D.J., Yu, G.L., Wang, R.P., Shell, B.K., Nash, R.A., Schar, P. Et al. (1995). Molecular cloning and expression of human cDNAs encoding a novel DNA ligase IV and DNA ligase III, an enzyme active in DNA repair and recombination. *Mol Cell Biol* **15:** 3206-3216.

West, R.B., Yaneva, M., and Lieber, M.R. (1998). Productive and nonproductive complexes of Ku and DNA-dependent protein kinase at DNA termini. *Mol Cell Biol* **18:** 5908-5920.

Wilson, J.H., Berget, P.B., and Pipas, J.M. (1982). Somatic cells efficiently join unrelated DNA segments end-to-end. *Mol Cell Biol* **2:** 1258-1269.

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

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 XRC62C4. *Mol Cell Biol* **29:** 3163-3172.

Wu, X., and Lieber, M.R. (1996). Protein-protein and protein-DNA interaction regions within the DNA end-binding protein Ku70-Ku86. *Mol Cell Biol* **16:** 5186 -5193.

Xing, M., Yang, M., Huo, W., Feng, F., Wei, L., Jiang, W., Ning, N., Yan, Z., Li, W., Wang, Q., Hou, M., Dong, C., Guo, R., Gao, G., Ji, J., Zha, S.,

Lan, L. Liang, H., and Xu,D. (2015) Interactome analysis identifies a new paralogue of XRCC4 in Non-homologous end joining DNA repair pathway. *Nat Comm* **6233:**

Yan, C.T., Boboila, C., Souza, E.K., Franco, S., Hickernell, T.R., Murphy, M., Gumaste, S., Geyer, M., Zarrin, A.A., Manis, J.P., Rajewsky, K., and Alt, F.W. (2007). IgH class switching and translocations use a robust non-classical end-joining pathway. *Nature* **449:** 478-482.

Yano, K., and Chen, D.J. (2008a). Live cell imaging of XLF and XRCC4 reveals a novel view of protein assembly in the non-homologous end-joining pathway. *Cell Cycle* **7:** 1321-1325.

Yano, K, Morotomi-Yano, K., Lee, K.J., & Chen, D.J. (2011). Functional significance of the interaction with Ku in DNA double-strand break recognition of XLF *FEBS Lett* **6:**
841-846.

Yano, K., Morotomi-Yano, K., Wang, S.Y., Uematsu, N., Lee, K.J., Asaithamby, A., Weterings, E., and Chen, D.J. (2008b). Ku recruits XLF to DNA double-strand breaks. *EMBO Rep* **1:** 91-96.

Yoo, S., and 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 Res* **27:** 4679-4686.

Yu, Y., Mahaney, B.L., Yano, K., Ye, R., Fang, S., Douglas, P., Chen, D.J., and LeesMiller, S.P. (2008). DNA-PK and ATM phosphorylation sites in XLF/Cernunnos are not required for repair of DNA double strand breaks. *DNA Repair (Amst).* **7:** 1680-1692.

Yu, Y., Wang, W., Ding, Q., Ye, R., Chen, D., Merkle, D., Schriemer, D., Meek, K. and 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:** 1239–1252.

Zha, S., Alt, F.W., Cheng, H.L., Brush, J.W., Li, G. (2007). Defective DNA repair and increased genomic instability in Cernunnos-XLF-deficient murine ES cells. *Proc Natl Acad Sci USA* **104:** 4518-4523.

Zha S., Guo C., Boboila C., Oksenych V., Cheng H.L., Zhang Y., Wesemann D.R., Yuen, G., Patel H., Goff P.H., Dubois R.L., and Alt F.W. (2011). ATM damage response and XLF repair factor are functionally redundant in joining DNA breaks. *Nature* **13:** 250-254.

Zhang, Y., Wu, X., Yuan, F., Xie, Z., and Wang, Z. (2001). Highly frequent frameshift DNA synthesis by human DNA polymerase mu. *Mol Cell Biol* **21:** 7995-8006.

Zhang, X., Moréra, S., Bates, P.A., Whitehead, P.C., Coffer, A.I., Hainbucher, K., Nash, R.A., Sternberg, M.J., Lindahl, T., and Freemont, P.S. (1998). Structure of an XRCC1 BRCT domain: a new protein-protein interaction module. *EMBO J* **17:** 6404-6411.

Zhorov, B.S. (1981). Vector method for calculating derivatives of energy of atom-atom interactions of complex molecules according to generalized coordinates. *J Struct Chem* **22:** 4-8.