

CHARACTERIZATION OF THE OLIGOMERIZATION STATE OF XRCC4

CHARACTERIZATION OF THE OLIGOMERIZATION OF THE HUMAN XRCC4
DNA REPAIR PROTEIN: IMPLICATIONS TO NON-HOMOLOGOUS END JOINING

By WILSON K.Y. LEE, B.Sc.

A Thesis submitted to the School of Graduate Studies in Partial Fulfilment of the
Requirements for the Degree of Master of Sciences

McMaster University © Copyright by Wilson K.Y. Lee, April 2013

McMaster University, Hamilton, Ontario

MASTER OF SCIENCES (2013)

TITLE: Characterization of the Oligomerization State of the Human XRCC4 DNA Repair Protein: Implications to Non-Homologous End Joining

AUTHOR: Wilson Kwang Yun Lee, B.Sc. (McMaster University)

SUPERVISOR: Murray S. Junop

PAGES: x; 92

ABSTRACT

If not efficiently repaired, DNA double-stranded breaks can result in cell death. A major contributor to the repair of this DNA damage is the non-homologous end joining pathway (NHEJ) which depends on the proteins: X-ray cross complementing protein 4 (XRCC4) and XLF. These proteins form a complex that can bridge DNA substrates *in vitro*. Analysis of these proteins has demonstrated that the C-terminal region of XRCC4 is necessary for this bridging function. However, this region is also critical for both tetramerization and DNA binding abilities of XRCC4, making the interpretation of XRCC4's role in the DNA-bridging unclear. Here, we intend to further characterize the tetramerization of XRCC4 and find a functionally independent mutant. Our studies suggest that regions in the N-terminus of XRCC4 may be important for the tetramerization of the protein but not for its DNA binding ability. These mutants were also analyzed by circular dichroism and mobility shift assays to verify for the integrity of their secondary structure composition and show that they are able to interact with its known binding partner, DNA Ligase IV. Additionally, we have shown that the XRCC4:XLF complex as well as XLF alone are able to interact with DNA substrates as short as 36 base pairs. Taking the data together, we expect to be able to construct a structural model for the XRCC4:XLF complex with DNA and obtain a better understanding on the role of XRCC4's tetramerization in the NHEJ pathway. As deficiency of XRCC4 has been implicated with tumorigenesis and immunodeficiency, understanding its role will be helpful for the development of treatments for such complications.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank Dr. Murray Junop whom I have the deepest gratitude for giving me the opportunity to do research for both my undergraduate and my graduate careers in the past four years since I have joined his laboratory. During this time, he was an excellent model of an intellectually curious researcher that I could set as a goal for me to strive for. In addition to that, I am extremely grateful for his patience and willingness to give me the freedom to pursue anything I found to be interesting, allowing me essentially to satiate my own scientific curiosities. Lastly, I would like to thank Dr. Junop for all the guidance he has given me in all these years.

Second, I would like to thank my committee members, Dr. Cecile Fradin and Dr. Xu-Dong Zhu, for their patience with me during every committee meeting. I am grateful for their understanding attitude while providing me guidance and reminding me of the areas I needed improvement on. Every meeting was very helpful to me as it forced me to step back and more carefully analyze my experiments.

Third, I would like to thank both Mac Mok and Simon Huang who greatly helped me by allowing me to bounce ideas and verify theories with them. I also greatly appreciate all their moral support in my times of frustrations. Specially, Mac Mok who, despite his own busy schedule, would make time to help me as many times as needed.

Fourth, I would like to thank Sara Andres and Seiji Sugiman-Marangos for all the guidance that they have provided. Although my time with Sara was short, she made sure to pass down as much of her knowledge of the project as she could. Seiji was one of the few in the laboratory who had the expertise and the knowledge and was willing to spare

some of his time to help me troubleshoot many of the problems that I have faced. I am deeply thankful for their guidance, as I am sure without it I would have faced many more hours of frustration.

I would also like to thank the undergraduate students who have worked with me in this project. Kritica Arora for beginning the DNA EMSA work and Joshua Chun for doing some of the protein crystal optimization. I was very impressed with their hardworking and resilient attitudes and could not have asked any more from any of them.

Finally, I would like to thank the other graduate students and my family for all their support.

TABLE OF CONTENTS

ABSTRACT	iii
ACKNOWLEDGEMENTS	iv
TABLE OF CONTENTS	vi
LIST OF FIGURES	viii
LIST OF TABLES	ix
LIST OF ABBREVIATIONS	ix
CHAPTER 1 – INTRODUCTION	1
1.1 DNA Double-Stranded Breaks	2
1.2 DNA damage response	5
1.3 Non-Homologous End Joining Pathway	6
1.3.1 Damage Recognition & End-Protection	8
1.3.2 End-Processing & Fill-in Synthesis	9
1.3.3 End-Ligation	10
1.4 XRCC4	11
1.4.1 The structure of XRCC4	12
1.4.2 The biochemistry of XRCC4	16
1.5 Thesis Objectives	20
CHAPTER 2 – TETRAMERIZATION STUDIES OF XRCC4	21
2.1 Rationale and Experimental Designs	22
2.1.1 XRCC4 Mutants	23

2.2 Materials & Methods	25
2.2.1 Mutagenesis	25
2.2.2 Expression & Purification of Proteins	26
2.2.3 Circular Dichroism	28
2.2.4 DNA Electrophoretic Mobility Shift Assay	29
2.2.5 Protein Electrophoretic Mobility Shift Assay	30
2.2.6 Analytical Ultracentrifugation	31
2.4 Results	32
2.3.1 Purification Profile of XRCC4 Proteins	32
2.3.2 XRCC4 Mutants retain their secondary structure	35
2.3.3 XRCC4 Mutants bind to DNA	35
2.3.4 XRCC4 Mutants interact with DNA Ligase IV	38
2.3.5 On-going Analytical Ultracentrifugation	40
2.5 Discussion	42
CHAPTER 3 – CHARACTERIZATION OF THE XRCC4:XLFI:DNA COMPLEX	50
3.1 Rationale and Experimental Design	51
3.2 Materials & Methods	56
3.2.1 Expression and Purification of Proteins	56
3.2.2 Preparation of DNA Substrates	58
3.2.3 DNA EMSAs	59
3.2.4 Crystallization	60

3.3 Results	62
3.3.1 Purification Profile of the XRCC4 ¹⁻²⁶⁵ and XLF Proteins	62
3.3.2 XRCC4 ¹⁻²⁶⁵ :XLF binds to Short DNA Substrates	64
3.3.3 XRCC4 ¹⁻²⁶⁵ :XLF binds to the Crystallization DNA Substrate	64
3.3.4 Crystallization attempts of XRCC4 ¹⁻²⁶⁵ :XLF and XRCC4 ¹⁻²⁶⁵ :XLF:DNA	67
3.4 Discussion	69
CHAPTER 4 – CONCLUDING REMARKS	77
4.1 Summary and Implications to NHEJ	78
4.2 Future Directions	84
REFERENCES	87

LIST OF FIGURES

1.1 Sources of DNA double-stranded breaks	3
1.2 Non-Homologous End Joining Pathway	7
1.3 Structural arrangement of XRCC4	13
1.4 Tetramerization modes of XRCC4	15
1.5 Crystal structure of XRCC4:DNA Ligase IV	17
1.6 Crystal structure of the XRCC4:XLF filament	18
2.1 XRCC4 mutants	24
2.2 Nickel Affinity Chromatography of XRCC4	33

2.3 Ion Exchange Chromatography of XRCC4	34
2.4 XRCC4 mutants retain their ability to bind to DNA	37
2.5 XRCC4 mutants retain their ability to interact with the BRCTs of DNA Ligase IV	39
3.1 Structure of XLF ¹⁻²²⁷	52
3.2 Purifications of XRCC4 ¹⁻²⁶⁵	63
3.3 Nickel Affinity Chromatography of XLF	65
3.4 XRCC4:XLF complex is able to bind to short DNA substrates	66
3.5 XRCC4 ¹⁻²⁶⁵ :XLF binds to the crystallization DNA substrate	68
3.6 Representative crystal observed for the XRCC4 ¹⁻²⁶⁵ :XLF crystallization	70
4.1 The possible roles of XRCC4 in the NHEJ pathway	83

LIST OF TABLES

2.1 XRCC4 mutants have a similar secondary structure content to the wildtype	36
2.2 Summarized AUC data on XRCC4 mutants	41
3.1 Summary of conditions for the best generated crystals	70
3.2 Summary of DNA binding for XRCC4 truncations	73

LIST OF ABBREVIATION

AUC	Analytical ultracentrifugation
DDR	DNA damage response
DNA	Deoxyribonucleic acid

DNA-PKcs	DNA protein kinase catalytic subunit
DSB(s)	Double-stranded break(s)
EMSA	Electrophoretic mobility shift assay
FPLC	Fast protein liquid chromatography
HR	Homologous recombination
IR	Ionizing radiation
LB	Luria Broth
NHEJ	Non-homologous end joining
PCR	Polymerase chain reaction
PEG	Polyethylene Glycol
PNK	Polynucleotide kinase
ROS	Reactive oxygen species
SFM	Scanning force microscopy
SSB(s)	Single-stranded break(s)
UV	Ultra-violet
XLF	XRCC4-like factor
XRCC4	X-ray cross complementing protein 4

CHAPTER 1 - INTRODUCTION

1.1 DNA Double-Stranded Breaks

DNA double-stranded breaks (DSBs) are an extreme form of DNA damages that if not efficiently repaired can quickly lead to cell death. In higher eukaryotes, apoptosis of cells with damaged DNA may at times be desired; as persistent unrepaired DNA can lead to DNA mutations, chromosomal translocations, immunodeficiency, or even tumorigenesis (Borek *et al.*, 1996; Cavazzana-Calvo *et al.*, 1993; Hei *et al.*, 1988). Many different factors can lead to the formation of double-stranded breaks, and they can be broadly categorized into two major groups: those caused by endogenous factors, and those caused by the exposure to exogenous agents (Figure 1.1).

Endogenous factors include the inadequate repair of DNA single-stranded breaks (SSBs) as well as failure in normal cellular processes involved in genetic recombination. In the first case, a cell with defects in the nucleotide excision repair or the base excision repair pathways is unable to efficiently repair a single-stranded lesion, eventually resulting in the formation of DSBs. This occurs by either replication fork collapse due to the replication machinery encountering the damaged site (Kuzminov *et al.*, 2001), or by a secondary SSB in close proximity to the initial damaged site. In the latter case, a secondary break on the complementary strand of a single-stranded damaged DNA molecule, within approximately 10 base pairs, has been shown to result in a lesion in which the overall structure of the chromatin is incapable of holding the DNA ends together, thus resulting in the formation a DSB (Michael *et al.*, 2000).

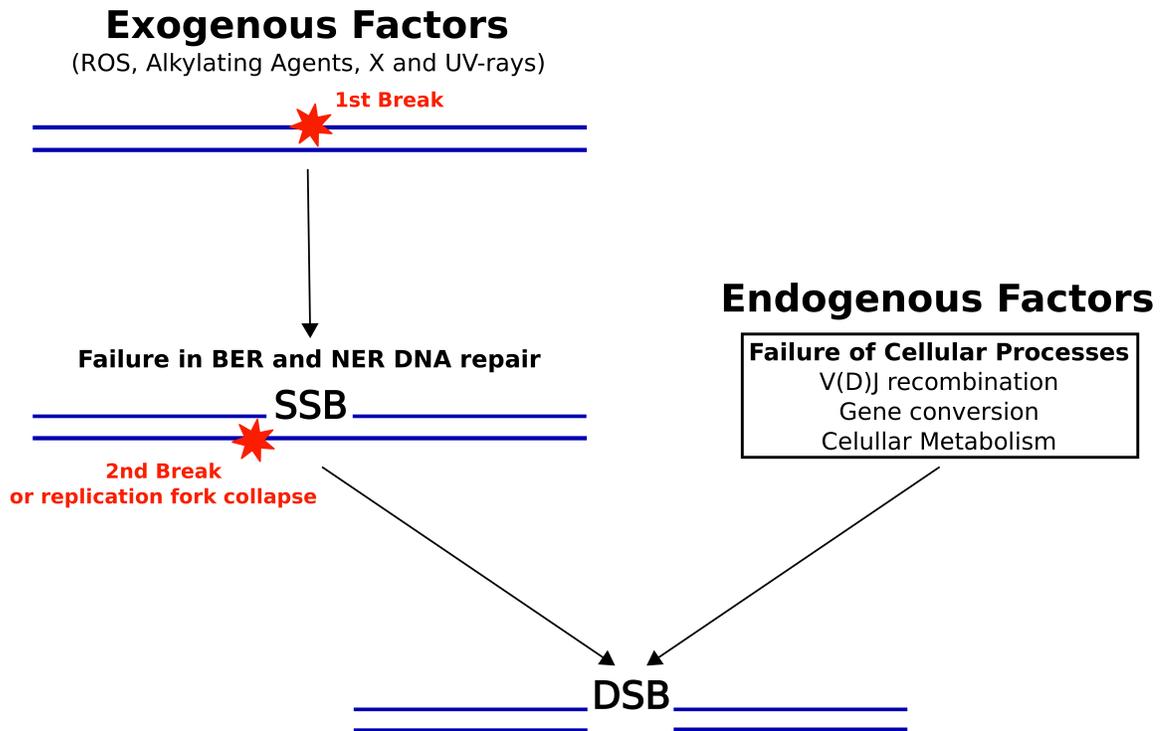


Figure 1.1 Sources of DNA Double-Stranded Breaks. The many factors responsible for DNA double-stranded breaks can be broadly categorized into two groups: Exogenous and endogenous factors. Exogenous factors include radiation causing single-stranded DNA lesions that if not repaired eventually leads to the formation of double-stranded breaks. Endogenous factors are failures in normal cellular processes that result in the direct or the indirect damage to DNA molecules forming DSBs.

Alternatively, DSBs can arise from failures or abortions of normal cellular processes that require the induced formation of such DNA damage in order to be successfully completed. These normal cellular processes include gene conversions, crossover events, and V(D)J recombination, and they occur to increase the diversity of the genetic material. For example, the V(D)J recombination of the heavy chain loci in lymphocytes is a method in which these cells use to create different combination of the V_H , D_H , and J_H gene segments and produce a diversity of immunoglobulin's, which in turn, is able to recognize the many different antigens that the organism might become exposed to (Early et al., 1980).

In addition to this, cellular metabolism is another potential source of DNA damage that if not regulated can result in the generation of detrimental amounts of reactive oxygen species (ROS). According to Chance *et al.* (1979), the cellular sources of ROS are the mitochondria, microsomes, peroxisomes, and certain cytosolic enzymes. In normal conditions, the cell has several different mechanisms to neutralize these ROS. As an example of this, one of the systems utilized by the mitochondria to prevent damage due to the accumulation of ROS is the superoxide dismutase enzyme that is able to catalyze the reaction of superoxides into oxygen and hydrogen peroxide. In doing so, this enzyme prevents any damage to the mitochondria itself as well as to the cell as a whole (Li et al., 1995). These systems regulate endogenously generated ROS and normally prevent them from causing significant damages to the nuclear DNA.

In contrast to this, ROS induced by exogenous sources are generated at a much higher rate and local concentrations than the endogenously generated ROS. These occur

due to the exposure to radiation, such as gamma, UV, X, and IR-rays, that can induce the formation of oxygen-derived free radicals, or more specifically hydroxyl radicals (Hutchinson, 1985). These free radicals, in turn, can act in a similar manner to that of the endogenously generated ROS to directly damage the DNA and form modified nitrogenous bases, SSBs, DSBs, or apurinic/apyrimidinic sites (Breimer & Lindahl, 1985a,b; Ward & Kuo, 1976). Exposure to these radiations can also damage the DNA by energy deposition and the direct ionization of the DNA bases or sugars.

1.2 DNA Damage Response

Damage to the DNA causes the activation a DNA damage response (DDR) that encompasses a signaling cascade mechanism to decide the fate of the cell. This signaling cascade begins with the localization of damage sensing proteins to the damaged site, followed by recruitment of mediator and transducer proteins that, in turn, transmit messages to effector proteins (Petrini and Stracker, 2003). These effector proteins are then responsible for eliciting appropriate responses to the damage such as apoptosis, cell cycle arrest, or DNA repair (Jackson, 2002). The decision for the outcome is determined based on factors such as the extent and the type of the DNA lesion as well as the stage of the cell cycle (Bartek, 2007). Although the exact mechanism of DDR is not known, current studies suggest that the initial sensing of the damage is done by the Mre11-RAD50-NBS1 complex (MRN complex) (Petrini, 2003). Following this, the recruitment of the phosphatidylinositol 3-kinase related kinases, ATM and ATR, occurs, and these act as the transducers of this pathway. From this point, many other mediators and effectors are recruited or activated, such as the tumour protein 53 binding protein1, mediator of

DNA damage checkpoint 1, breast cancer 1, and the MRN complex (Falck *et al.*, 2005; Paull *et al.*, 2000). Unfortunately, the responses associated with each of these proteins go beyond the scope of this thesis and will not be discussed further.

Despite the initial complexity of the DDR, mammals eventually repair DSBs through two major DNA repair pathways known as the homologous recombination (HR) and the non-homologous end joining (NHEJ) (Hoeijmakers, 2001). The exact mechanism that the cells employ to determine which pathway to use in the repair of the DNA is not fully characterized. However, the homologous recombination pathway utilizes a sister chromatid or a homologous chromosome as a template to accurately repair the damaged DNA. It then follows that this pathway occurs predominantly in the S and G2 phases of the cell cycle, where these DNA templates are available for the repair process (Helleday *et al.*, 2007; Mahaney *et al.*, 2009). Alternatively, the NHEJ pathway does not require any template to repair the damaged DNA and is able to be employed during any of the cell cycle phases, but it is a mutagenic process that results in a loss of integrity of the genetic material (Rothkamm *et al.*, 2003; Takata *et al.*, 1998). In spite of this, the NHEJ is the major pathway utilized by higher eukaryotes in the repair of DNA DSBs, and as the focus of my project is this pathway, it will be the only pathway that will be discussed in further detail.

1.3 Non-Homologous End Joining Pathway

The NHEJ pathway can be broken down into three distinct steps – End recognition and protection, end processing, and lastly end ligation (Figure 1.2). Each of these steps involves the participation of key DNA repair proteins to be efficiently completed. Despite

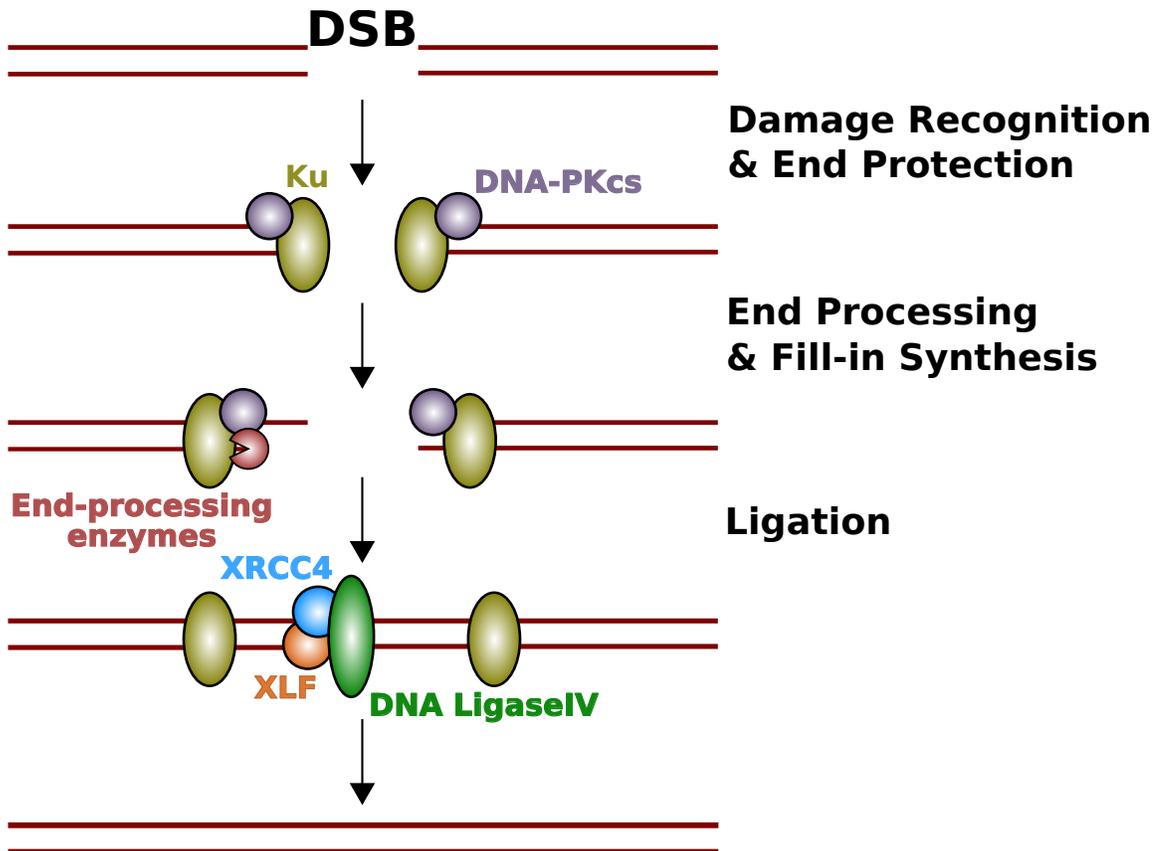


Figure 1.2 Non-Homologous End Joining Pathway. Once a double-stranded break forms, Ku70/80 quickly recognizes the damage and recruits DNA protein kinase catalytic subunit (DNA-PKcs) to the damaged site. This new protein complex, known as the DNA-PK, is in turn able to phosphorylate itself and many other DNA repair proteins in the pathway. Next, Ku slides inwards allowing end-processing enzymes to modify the ends into ligatable DNA ends, and DNA polymerases to fill in any missing nucleotides. Lastly, the XRCC4-DNA Ligase IV complex ligates the processed DNA ends together with the stimulation of XLF.

the many years of research done on these proteins, the exact role that each play in the pathway is not yet completely elucidated. However, their general role in specific steps of the repair has been determined and will be discussed below.

1.3.1 Damage Recognition and End-Protection

In the first step, the heterodimer Ku70/Ku80, or in short simply Ku, is the first to be localized to the damaged site. Ku is able to interact with the DNA by forming a ring shaped structure that allows the DNA to tread through the inner pore of the complex (Walker *et al.*, 2001). In doing so, it is able to bind to the DNA ends with a dissociation constant of about 50nM (Mimori and Hardin, 1986). Once bound in such a manner, Ku was also shown to be capable of translocating along the DNA molecule without an ATP requirement (de Vries *et al.*, 1989; Yoo and Dynan, 1999). This characteristic could later be necessary for Ku to slide inwards and away from the damage site in order to allow for other DNA repair proteins to access and process the DNA ends. Following Ku, the serine/threonine kinase, DNA-dependent protein kinase subunit (DNA-PKcs), is the next protein to be recruited to the site. This recruitment is promoted through Ku's ability to interact with the DNA-PKcs and increase its binding affinity to the DNA ends by approximately 100-fold (Hammarsten and Chu, 1998). Curiously, studies suggest that these proteins do not interact in the absence of DNA and that the complex formed by these two proteins, known as the DNA-PK complex, is only active with DNA as a cofactor (Downs and Jackson, 2004; Yaneva *et al.*, 1997). It then follows that Ku would first bind to the DNA-ends, creating a binding platform for DNA-PKcs. In doing so, an active DNA-PK complex would be assembled at the damage site, and from there, the

complex would be able to phosphorylate many downstream DNA repair proteins, as well as, itself (Mahaney *et al.*, 2009; Gu and Weinfeld, 1998). The phosphorylation done by DNA-PK is hypothesized to regulate many of the proteins involved in the repair of the DNA. In fact, the auto-phosphorylation of DNA-PKcs has been shown to result in the dissociation of this protein from the DNA-PK complex, potentially serving as a mechanism to regulate its own activity in the repair pathway (Chan and Lees-Miller, 1996).

1.3.2 End-Processing and Fill-in Synthesis

The next step of the NHEJ pathway, involves the processing of the DNA ends into suitable substrates for the final ligation of the strands. In most cases of DNA damage, the 5' and 3' ends are not a phosphate group and a hydroxyl group, respectively. As such, the recruitment of many different nucleases and other DNA end processing enzymes to the damaged ends is necessary to process the ends into ligatable forms. These nucleases must be able to recognize DNA substrates with the unusual conformations derived from the damage, and thus are recruited to the site depending on the type of lesion that is present. One important nuclease reported to be involved in the NHEJ pathway is the 5' exonuclease Artemis (Ma *et al.*, 2002). The role of this nuclease in the NHEJ pathway was identified through patients with defects in Artemis displaying severe immunodeficiency that was later associated with an incapability of performing proper V(D)J recombination (Moshous *et al.*, 2001). By forming a complex with DNA-PK, Artemis is able to act as an endonuclease with a 5' preference to blunt ended products and a 3' preference to products with 4-5 nucleotides overhangs (Ma *et al.*, 2002). Although

not necessary for the NHEJ process, this could allow for small and random microhomologies to be exposed between the ends, which could aid in the synapsis and ligation of the DNA ends. In addition to nucleases, other enzymes such as kinases and polymerases, also participate in the processing of the ends prior to ligation. Polynucleotide kinase (PNK) is an example of a kinase with both 3'-phosphatase and 5'-kinase activities that some evidence suggest have roles in the removal of non-ligatable groups from both the DNA ends, as well as, the addition of a phosphate group to the 5' end of the DNA (Chappell *et al.*, 2002). Finally, polymerase μ and polymerase λ have both been observed to be recruited to the damaged site by Ku (McElhinny *et al.*, 2005). These DNA polymerases would be recruited in situations where DNA gaps were present and, depending on the type of lesion, one polymerase would be recruited over the other to fill in the gap.

1.3.3 End-Ligation

Once the DNA ends have been processed into ligatable forms, the last step of the NHEJ pathway is the ligation of these ends. This is done primarily by DNA ligase IV with the aid of two other proteins known as the X-ray cross complementing protein 4 (XRCC4) and the XRCC4-like factor (XLF). Initial knockout studies of these three proteins in mammalian cells resulted in similar phenotypes where the cells became radiosensitive and demonstrated defects in the ability to repair DSBs (Giaccia *et al.*, 1990; Grawunder *et al.*, 1998a; Zha *et al.*, 2007). Later it was determined that XRCC4 is able to directly interact with DNA ligase IV to stimulate its activity (Grawunder *et al.*, 1998b; Wu *et al.*, 2009); while XLF interacts with the XRCC4:DNA ligase IV complex

through a interface on the XRCC4 protein to stimulate the activity of the complex. Grawunder *et al.* (1997) showed that the presence of XRCC4 in cell free-based assays resulted in a 5- to 8-fold increase in the DNA ligase IV activity towards the ligation of DSBs. Furthermore, the XRCC4:DNA ligase IV complex is able to ligate a variety of DSBs, ranging from those with blunt or fully incompatible ends to those with small microhomologies (Gu *et al.*, 2007). Although much is known about the activity of this protein complex, the actual mechanism as to how XRCC4 stimulates the activity of DNA ligase IV is largely undetermined.

In contrast, there are some studies done on the mechanism of stimulation done by XLF. As its name implies, XLF is a structurally related protein to XRCC4, or in other words, that adopts a very similar structure to that of XRCC4 (Andres *et al.*, 2007). Nevertheless, this protein is incapable of directly interacting with DNA ligase IV but it is able to stimulate its activity. Specifically, XLF stimulates the activity of the XRCC4:DNA ligase IV complex towards non-cohesive or mismatched DNA ends (Chu *et al.*, 2007). Although the exact mechanism of how this is accomplished is not known, recent studies suggest that XLF promotes the re-adenylation of the XRCC4:DNA ligase IV complex after the ligation, thus preparing it for the following ligation reaction (Riballo *et al.*, 2008).

The main focus of my work was to better characterize the XRCC4 protein in order to better understand how it interacts with the other DNA repair proteins in the NHEJ pathway. As such, a more detailed description of this protein will be given below.

1.4 XRCC4

1.4.1 The structure of XRCC4

The human XRCC4 is a protein composed of 336 residues, but is absent of any known enzymatic activity. Curiously, only the first 203 N-terminal residues have been shown to fold into structured domains while the remaining 133 residues were determined to be a largely unstructured and flexible region. Junop *et al.* (2000) showed that the structured region of XRCC4 adopts a fold with a distinct N-terminal head domain composed of a β -sandwich surrounding two α -helices, followed by a single long α -helical C-terminal tail region (Figure 1.3). This structure was determined by X-ray crystallography and in the observed crystal packing, XRCC4 formed homodimers (pdb 1FU1). This dimerization occurred predominantly through hydrophobic interactions mediated by residues at the N-terminal region of the α -helical tail and was further stabilized by hydrogen bonding and a salt bridge. By forming dimers through this interface, XRCC4 also forms a coiled-coil α -helical tail that has been shown to be important for its' ability to interact with DNA ligase IV, to bind to DNA molecules, and also to form homotetramers (Andres *et al.*, 2012; Junop *et al.*, 2000; Modesti *et al.*, 2003; Wu *et al.*, 2009). From this point onwards, this specific region outlined in orange on Figure 1.3 will be referred to as the critical tail region (CTR) of XRCC4.

The tetramerization of XRCC4 is done by the interaction of two dimers of XRCC4. Although different groups employing different techniques, such as X-ray crystallography, protein cross-linking, analytical ultracentrifugation (AUC) and small angle X-ray scattering (SAXS) experiments have confirmed the formation of these tetramers *in vitro*, the nature and the interface of this tetramerization has been a matter of much debate

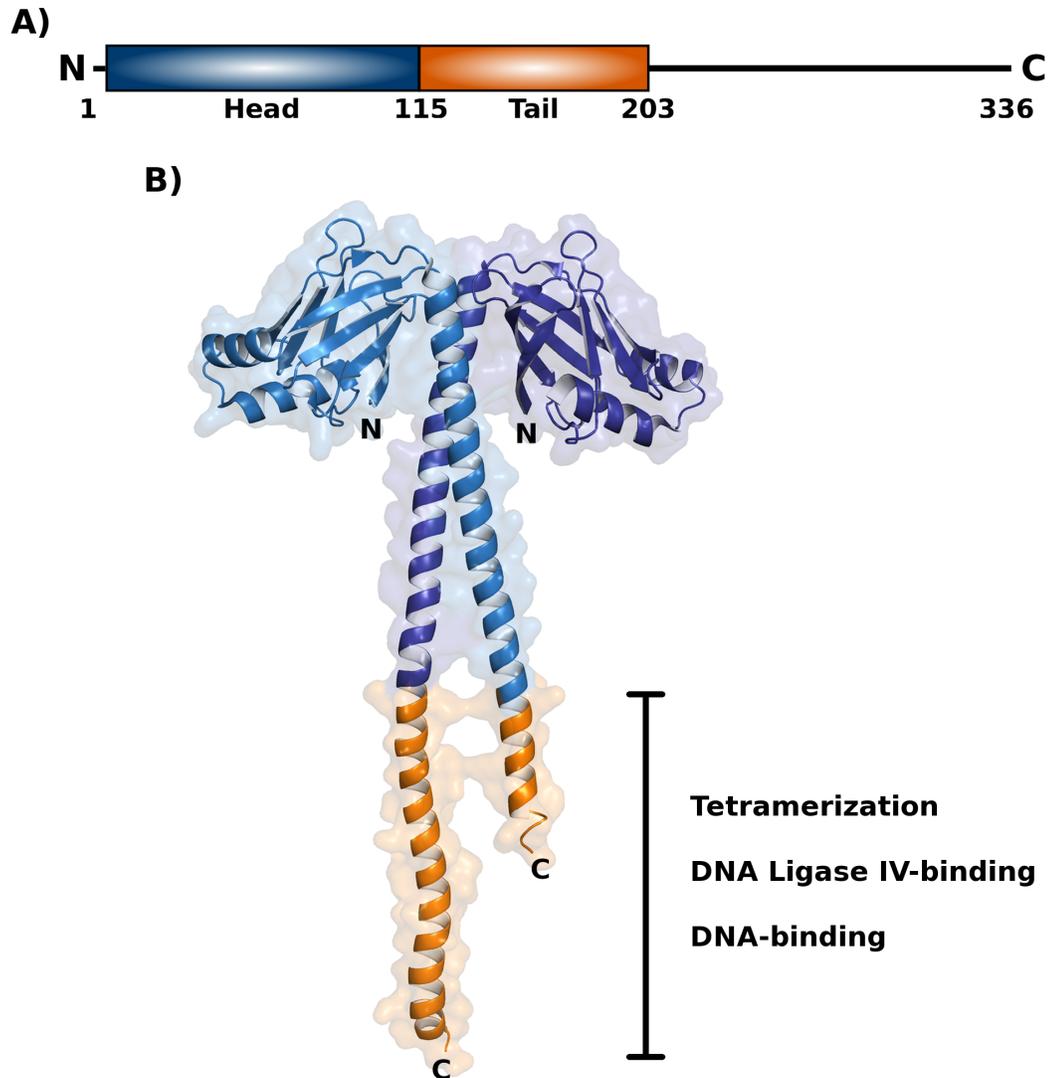


Figure 1.3 Structural arrangement of XRCC4. Panel A, Schematic of XRCC4's domains arrangement. Head and tail regions are coloured blue and orange, respectively; while residues from 203 to 336 forms an unstructured and flexible region represented by a single line. Panel B, the homodimer observed in the protein crystal structure of XRCC4 (pdb1FU1). This crystal structure spans up to residue K178 for one of the monomers, coloured blue, and up to residue Q203 for the other monomer, coloured purple-blue. The residues from 162-203, coloured orange, composes the region of the XRCC4 tail that has been shown to be important for the tetramerization, DNA ligase VI-binding and DNA-binding abilities of XRCC4.

(Hammel *et al.*, 2010; Junop *et al.*, 2000; Modesti *et al.* 2003). Currently, there is evidence for two different tetramerization modes for this protein. In the first mode, two dimers of XRCC4 interact through a large interface ($\sim 3300\text{\AA}^2$) in the C-terminal coiled-coil α -helical tail of the protein (Figure 1.4A). This mode is supported by X-ray crystallography as well as mutational studies based on the respective protein crystal structure. Residues that were mutated to disrupt the tetramerization of the XRCC4 through this mode demonstrated an inability to form tetramers as determined by analytical ultracentrifugation experiments (AUC) (Modesti *et al.*, 2003). However, recent SAXS data have suggested an alternative mode of tetramerization for the homodimers of XRCC4. This data was best fitted with a head-to-head interaction of the homodimers, where one of the dimers is horizontally rotated by 180° and vertically rotated by 90° to form a dimer-dimer interface with a buried surface of $\sim 700\text{\AA}^2$ (as calculated by pisa) in the head region of the dimers (Figure 1.4B). While this interface buries a much smaller surface area suggesting a weak argument for this mode of tetramerization, the flexible and unfolded C-terminal region of XRCC4 was suggested to fold back onto itself and redirect this region of the protein towards its N-terminal head (Hammel *et al.*, 2010). By doing so, interactions between the head region and the folded-back tails could occur to further stabilize this mode of tetramerization. At present, more work is required to conclusively determine which of the modes is preferred and a significant portion of my thesis was dedicated to better characterize the tetramerization of XRCC4.

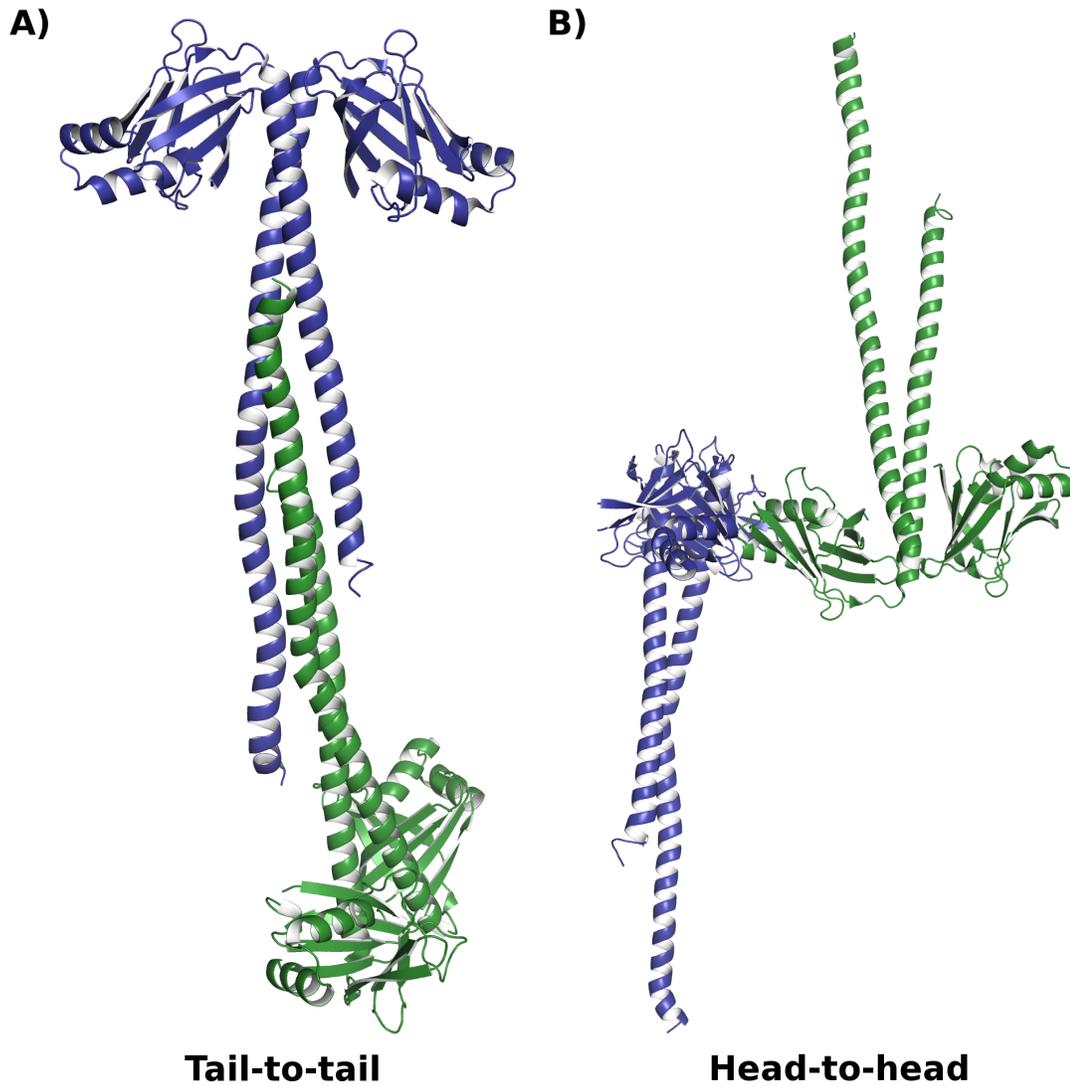


Figure1.4 Tetramerization modes of XRCC4. Tetramerization modes observed in the crystallographic packing from the pdb1FU1 crystal structure. Panel A, Tetramerization of XRCC4 via the tail-to-tail interaction. Panel B, Tetramerization of XRCC4 via the head-to-head interaction.

1.4.2 The Biochemistry of XRCC4

Despite its lack of enzymatic activity, XRCC4 has many biochemical functions attributing to its importance in NHEJ. Yeast two-hybrid studies as well as X-ray crystallography of XRCC4 have shown that, through direct contacts in its coiled coil α -helical tail region, the protein is able to interact with DNA ligase IV (Figure 1.5) (Grawunder *et al.*, 1997; Wu *et al.*, 2009). In doing so, XRCC4 promotes the stability and the activity of the ligase by, potentially, promoting its pre-adenylation (Grawunder *et al.*, 1997; Modesti *et al.*, 1999). In addition to this, XRCC4 is able to interact with XLF to further stimulate the activity of the ligase towards incompatible ends (Ahnesorg and Jackson, 2006; Lu *et al.*, 2007; Tsai *et al.*, 2007). Taken together, the functions of XRCC4 suggest that this protein plays a role in the late steps of the NHEJ pathway, or more specifically, during the DNA end-ligation step.

However, a recent study done by Andres *et al.* (2012) has suggested additional roles for the interaction between XRCC4 and XLF. In this study, these proteins were shown to interact together to bridge DNA molecules *in vitro*. This was demonstrated through DNA bridging assays, DNA electrophoretic mobility shift assays (EMSAs) and also visualized by scanning force microscopy (SFM). Interestingly, structural studies were also done on this protein complex that have suggested the formation of helical protein filaments. These protein filaments were composed of alternating XRCC4 and XLF homodimers interacting through their N-terminal head regions (Figure 1.6) (Andres *et al.* 2012; Hammel *et al.*, 2011). Mutations and truncations of XRCC4 with compromised DNA binding and, due to the nature of the mutantions, possibly DNA ligase IV binding and tetramerization

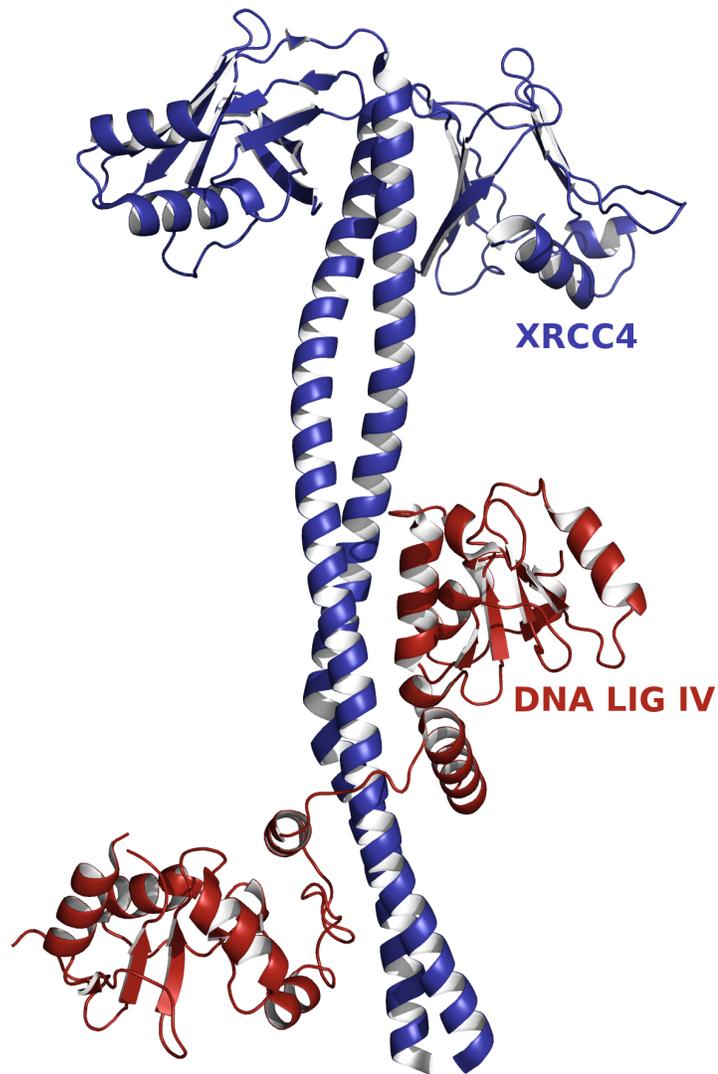


Figure 1.5 Crystal Structure of XRCC4:DNA Ligase IV. The structure of XRCC4¹⁻²⁰¹ in complex with the BRCT domains of DNA ligase IV was determined by X-ray crystallography (pdb3II6). This complex is formed in a 2:1 ratio of XRCC4:DNA ligase IV, where a homodimer of XRCC4 interacts with the BRCT domains of a DNA ligase IV monomer.

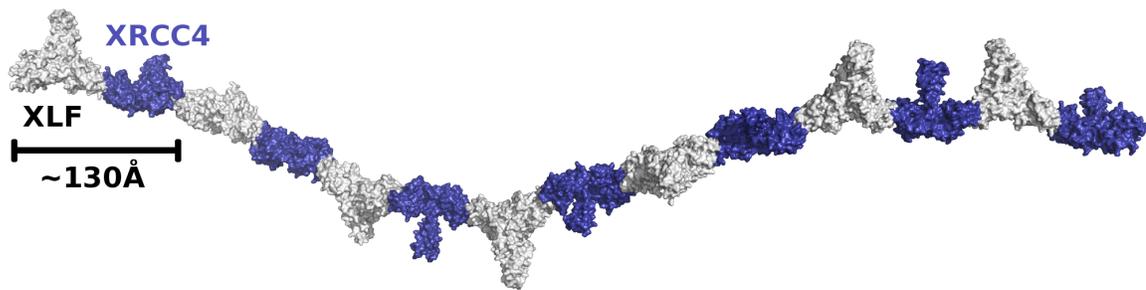


Figure 1.6 Crystal structure of the XRCC4:XLF filament. XRCC4:XLF filament observed in the solution of the crystal structure of the complex between XRCC4¹⁻¹⁵⁷, coloured in deep blue, and XLF¹⁻²²⁷, coloured in light gray. The filament is formed by alternating XRCC4 and XLF homodimers offset by approximately 30°, resulting in the formation of a coiled filament. The filament completes a full revolution after approximately 11 homodimers and a single XRCC4:XLF complex spans approximately 130Å. (pdb 3RWR)

abilities resulted in a loss of ability to bridge the DNA molecules. Additionally, XLF mutants unable to bind to DNA have also resulted in a disruption of the bridging function (Andres *et al.*, 2012). Although these results suggest that XLF's DNA binding ability plays a critical role to this bridging function, XRCC4's contribution to the bridging still remains to be elucidated due to the myriad of function that could have been affected with the mutants used in these assays.

Nevertheless, the bridging ability of this complex offers a new potential role for XRCC4 in the NHEJ DNA repair process. Recent studies using live-cell imaging have shown evidence for an early recruitment of XLF to the damaged site. This recruitment was dependent on Ku and while XRCC4 was not required, it stimulated and stabilized XLF's ability to bind to the DNA molecule (Yano *et al.*, 2008). Collectively, these results hint at a new potential function for XRCC4 and XLF in the early stages of the NHEJ, where the protein filaments could form to aid in the bridging of the damaged DNA ends.

Lastly, Modesti *et al.* (1999) performed an analysis of mutations and truncations of the XRCC4 characterizing the important regions for its DNA binding ability. It was found that the first 20 residues as well as the region between residues 160 and 203 were necessary for DNA-binding. In this study, XRCC4 was able to efficiently bind to DNA molecules longer than 1000 base pairs, demonstrating increasing binding affinities towards longer DNA substrates. Furthermore, the binding to DNA was shown to be stronger towards linear or nicked DNA molecules, suggesting a preferential binding to damaged DNA ends. Curiously, while the mutations causing defective DNA-binding affected the cells' ability to perform V(D)J recombination, it did not affect its ability to

stimulate the activity of DNA ligase IV. According to Modesti *et al.*, this supports the idea that XRCC4 may have other roles in NHEJ besides stimulation of DNA ligase IV activity.

1.5 Thesis Objective

Taking all the information on XRCC4 together, I have chosen to tackle two main aspects of this protein. The first one was to determine the most likely tetramerization mode of XRCC4 (head-to-head vs. tail-to-tail). As outlined in Figure 1.3B, many functions of XRCC4 has been associated with a particular region of the tails of this protein (CTR), amongst which is the tetramerization ability of XRCC4. A main problem that this generates is a difficulty in the analysis and interpretation of assays that probe for the importance of each of these functions to the actual repair of the DNA. This occurs because mutants of this protein used in assays would often have multiple compromised functions resulting in an ambiguous interpretation of the results. By determining the actual mode of tetramerization, one could potentially generate mutants affecting only tetramerization without disrupting the other functions. This work will be discussed in greater detail on Chapter 2 of this thesis. In addition to this, previous work done by Dr. Andres, a former member of our lab, has suggested the bridging ability of DNA molecules by XRCC4 and XLF. However, the mechanism of how this was accomplished has not yet been determined. Thus, the second goal of my work was to better determine the mechanism of DNA bridging by further characterizing the XRCC4:XLF:DNA complex and this work will be discussed in greater details on Chapter 3 of this thesis.

CHAPTER 2 – TETRAMERIZATION STUDIES OF XRCC4

2.1 Rationale and Experimental Designs

The main goal of this project was to determine if the tetramerization of XRCC4 occurs through the head or the tail regions of the XRCC4 homodimers. Since mutational studies targeting the tail-to-tail interactions observed in the crystal structure of the protein (pdb 1FU1) have already been done by Modesti *et al.* (2003), we have chosen to target the residues participating in potential head-to-head interactions of the homodimers. It is important to note here that the results observed by Modesti *et al.* (2003) do not necessarily exclude the possibility of a head-to-head tetramerization of this protein. If the flexible C-terminal end of XRCC4 folds back onto itself to further stabilize the head-to-head tetramerization of this protein, as suggested by Hammel *et al.* (2010), the mutations of the tails might prevent the folding back of these flexible ends and, in consequence, destabilize the head-to-head tetramerization mode of the protein. Interestingly, the head-to-head interaction of the protein has, in fact, been also observed in the same crystal structure of this protein (pdb 1FU1). As previously mentioned, this interaction however, buried a much smaller surface area than the tail-to-tail interaction.

We have chosen to approach this problem by performing mutational analysis coupled with analytical ultracentrifugation experiments. First, single or double point mutations of residues that could potentially play a role in the head-to-head interaction were generated based on the crystal structure of the protein. Following this, the integrity of the protein folding and its functions were verified for each mutant. This was done by circular dichroism (CD), DNA EMSA, and XRCC4:BRCT EMSA. Once this was accomplished the tetramerization ability of each mutant was assessed by sending the purified protein

samples to Dr. Ghirlando for the sedimentation equilibrium and velocity experiments. As a quick overview of this technique, the macromolecules in a solution are subjected to high gravitational fields and the concentration distribution of the different macromolecules are monitored in real-time by various optical measurements. In sedimentation velocity experiments, the rate in which the macromolecule of interest travels in solution is measured in Svedberg (S) and information based on the macromolecule's hydrodynamics can be obtained. While in sedimentation equilibrium experiments, the macromolecules are allowed to reach an equilibrium state where their flux of sedimentation is balanced by their flux of diffusion. Once this is reached, information on the macromolecule's molar mass and, in cases of interacting systems, on the stoichiometry and the affinity of the interaction can be obtained (Ghirlando, R., 2011; Shuck, P., 2000). For a more detailed explanation on both sedimentation experiments refer to the articles written by Ghirlando, R. (2011) and Shuck, P. (2000).

2.1.1 XRCC4 Mutants

The mutants generated for this project can be divided into two major groups: those participating directly in the head-to-head interaction of the XRCC4 homodimers, and those located at the N-terminal region of the XRCC4 tails (Figure 2.1). The first group of mutants was composed of the following mutations: V33E, V104E, F106A, and E121K, E125K. These mutants were designed to disrupt a specific hydrophobic pocket (between residues V33, F104 and F106) and an electrostatic interaction (between residues K65, E121, and E125). In this case, the hydrophobic residues were mutated to either glutamic acid, a large negative residue, or to alanine, a small hydrophobic residue; while charged

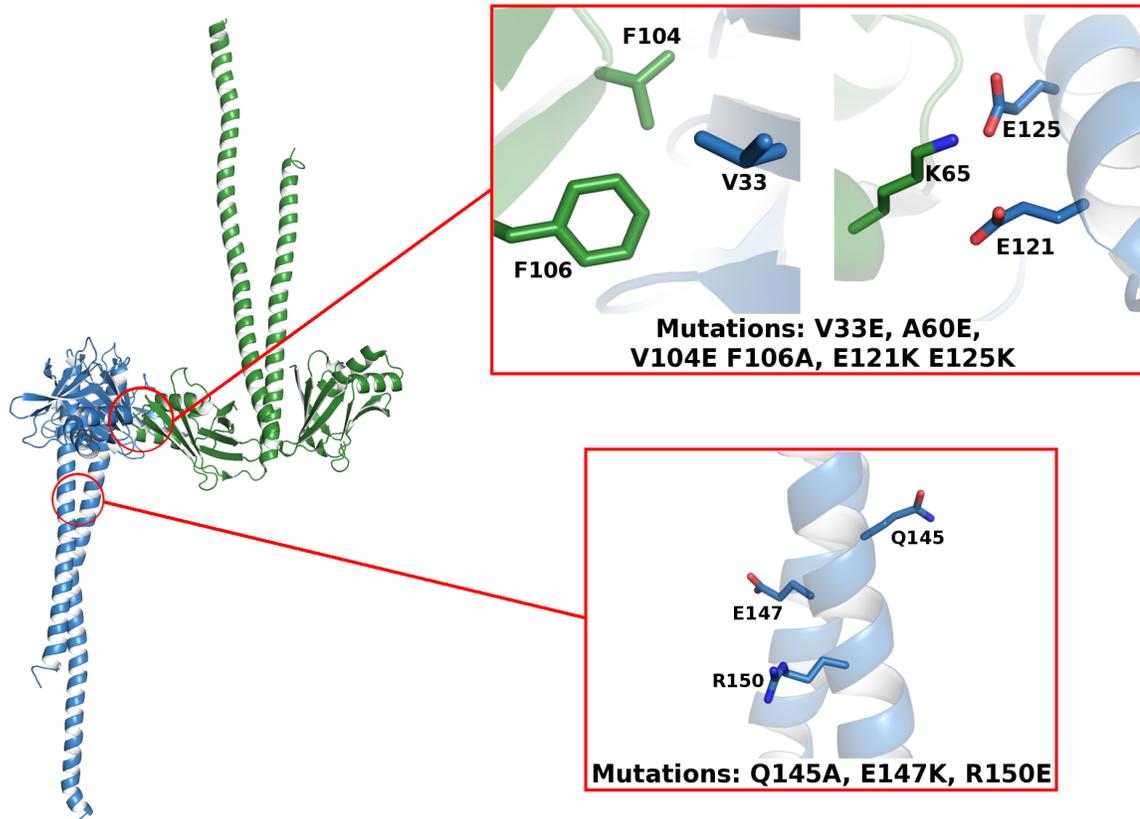


Figure2.1 XRCC4 Mutants. Visual special representation of each residue targeted for mutation in the head-to-head tetramerization of XRCC4. One of the homodimers composing the tetramer is coloured blue, while the other homodimer is coloured green. Mutants can be divided into two major groups: those located in the head region and those located in the coiled-coil tail region.

residues were mutated to residues with the opposite charge. In doing so, we aimed to disrupt the head-to-head interaction observed in the protein crystal structure and an effect on tetramerization would only be observed if this interaction is indeed important for the tetramer formation. The second set consisted of the mutants with the Q145A, E147K, and R150E mutations, and these were generated to account for the possibility of the C-terminal flexible region of XRCC4 folding back to stabilize the head-to-head interaction. Since these are highly conserved residues with no known specific function and are also exposed to solution, they could act as anchors to guide and lock the folded back tails towards the head region of the protein, potentially playing an important role in the head-to-head tetramerization of XRCC4. As a final note, all the generated mutants do not contain any mutations in the CTR of XRCC4 that were previously shown to be important for its tetramerization, DNA-binding, and DNA ligase IV-binding. Therefore, we would not expect to observe a disruption of any of these functions, excluding possibly the tetramerization of XRCC4.

2.2 Material & Methods

2.2.1 Mutagenesis for the Generation of the XRCC4 Mutants

Mutagenesis for all 7 mutants was done by designing the appropriate primers to introduce the desired point mutations. Primers were designed as instructed in the QuikChange™ site-directed mutagenesis kit instruction manual. The actual process was carried out by utilizing the Takara ExTaq™ DNA polymerase with the designed primers and the original plasmid containing the *xrcc4* gene (pWY1087) as the template. The reaction samples were prepared as instructed by the manufacturer's manual and the PCR

was performed with the following program: 94°C (5minutes), 30 cycles of 94°C (40seconds), 55°C (1minute), and 72°C (10minutes), and a final elongation step at 72°C (30minutes).

With the completion of the mutagenesis, DNA samples were treated with DpnI from ThermoScientific for 1 hour at 37°C before transformation into chemically competent *E.coli* TOP10 cells (Invitrogen). Cells were first plated onto Luria Broth (LB) agar plates containing 25µg/mL chloramphenicol and a single colony was picked for the inoculation of 10mL of LB media with the same concentration of antibiotic. Plasmids were then extracted from the liquid cultures as outlined in the High-Speed Plasmid Mini Kit Manual (GeneAid). Resulting plasmids were sent for sequencing at Mobix, McMaster University, to verify for the introduction of the mutations and the integrity of the rest of the gene. The mutants obtained through this process were the following: V33E (labeled as MJ4946), A60E (MJ4947), V0104E F106A (MJ4948), E121K E125K (MJ4949), Q145A (MJ4790), E147K (MJ4791), and R150E (MJ4792).

2.2.2 Expression & Purification of Proteins

The expression for all the *xrcc4*-related genes was done in the following manner. The plasmid containing the gene for the protein was introduced into chemically competent *E.coli* BL21 (DE3) cells (Invitrogen) by the heat-shock method. Following this, an overnight LB culture containing 25µg/mL chloramphenicol was prepared and used for the inoculation of larger volumes of LB media. This was done by inoculating 1L of LB containing 25µg/mL chloramphenicol with 10mL of the overnight culture. This was repeated to produce a total of 4L of LB culture. The inoculated 1L cultures were

incubated at 37°C with 225 rev min⁻¹ shaking (Certomat BS1-B. Braun Biotech Inc. USA) until the turbidity, as measured by the OD₆₀₀ (DU 530 Life Science UV/VIS Spectrophotometer, Beckman Coulter), reached approximately 0.450 units. At this point, expression of the genes were induced by the introduction of isopropyl β-D-1-thiogalactopyranoside, or more commonly referred as IPTG, to a concentration of 0.1M. Cells were allowed to grow at 37°C and 225 rev min⁻¹ for an additional 3 hours prior to their centrifugation at 3315 x G for 15min (Avanti J-30I Centrifuge with the JLA 9.1000 J-Lite Series Rotor, Beckman Coulter). The resultant pellet was then isolated, flash frozen in liquid nitrogen, and stored at -80°C until further use.

The same purification procedure was used for all the XRCC4-variants. It consisted of a two-step purification procedure with a nickel affinity chromatography followed by an anion exchange chromatography. First, frozen cells, corresponding to 2 liters of grown cell culture, were allowed to thaw on ice prior to their resuspension in 25mL of Nickel A buffer (20mM Tris pH8.0, 500mM KCl, 1mM β-mercaptoethanol, 10mM imidazole). Once these cells were resuspended, 3M KCl was introduced to bring the sample to a concentration of approximately 720mM KCl. Following this, cells were sonicated 3 times for 1 min with an amplitude of 18% (Sonic Dismembrator Mode 5000, Fisher Scientific) and the resultant sample was centrifuged at 48 384 x G for 40 minutes (Avanti J-30I Centrifuge with J.A-30.50 Rotor, Beckman Coulter). The supernatant was filtrated by vacuum filtration using the 25mm, 0.45µm filter from Metrical Membrane Filter, Pall Life Sciences. The resulting sample was then applied to a 5mL HiTrap nickel affinity column (GE Healthcare) that was set on a FPLC system (AKTA FPLC, Amersham

Pharmacia Biotech). Sequential step washes of the column with buffers containing 40mM and 70mM imidazole were done prior to the elution of the XRCC4 protein at 210mM imidazole. The buffers with different imidazole concentration were prepared by mixing of Nickel A buffer and Nickel B buffer (20mM Tris pH8.0, 500mM KCl, 1mM β -mercaptoethanol, 300mM imidazole) in the appropriate ratios. The obtained protein sample was then diluted in QA buffer (20mM Tris pH8.0, 10mM Dithiothreitol, 1mM Ethylenediaminetetraacetic acid, 10% Glycerol) to lower the KCl concentration to 150mM and then applied to a 5mL Q-sepharose HP anion exchange column (GE Healthcare) set on the FPLC system. A gradient from 150mM KCl to 400mM KCl over 100min and at a flowrate of 1mL/min was set by mixing the QA buffer with the QB buffer (20mM Tris pH8.0, 10mM DTT, 1mM EDTA, 10% Glycerol, 500mM KCl) at the appropriate ratios and rate. The protein was eluted and collected when the gradient reached a KCl concentration of approximately 290mM. Purified protein was flash frozen in liquid nitrogen and stored at -80°C until further use. In total, 9 XRCC4 constructs were purified in this manner and were used in the assays of this project; these constructs were: XRCC4 – V33E, A60E, V0104E F106A, E121K E125K, Q145A, E147K, R150E, WT, L184Q K187D I191S. The triple mutant is a previously reported XRCC4 mutant that has compromised DNA-binding, tetramerization, and DNA-ligase IV binding abilities and was used as a negative control for the assays.

2.2.3 Circular Dichroism

Purified XRCC4 samples were allowed to thaw on ice prior to extensive dialysis of samples in CD buffer (50mM Tris pH8.0, 250mM NaH_2PO_4). The dialysis was

performed as instructed in the manufacturer's manual at 4°C using the Slide-A-Lyzer® MINI Dialysis Unit 10K MWCO (Thermo Scientific). Following dialysis, samples were first diluted with CD buffer to a protein concentration of 0.1mg/mL to 0.5mg/mL and then treated with concentrated Tris(2-carboxyethyl)phosphine (TCEP) pH8.0 to achieve a final concentration of 1mM TCEP in each sample. Once the samples were prepared, the CD analysis for each XRCC4 construct was done by taking measurements of the difference in absorbed right- and left-handed circularly polarized light covering a wavelength range from 190nm to 260nm with the Model 410 Circular Dichroism Spectrophotometer (Aviv Biomedical, Inc., Lawewood, NJ). CD experiments were done in triplicates. The obtained data was processed with the CDPro software using the CONTINLL algorithm to determine the secondary structural content of each of the XRCC4 constructs.

2.2.4 DNA Electrophoretic Mobility Shift Assay

Protein samples were allowed to thaw on ice prior to their dialysis in DNA EMSA P. Buffer (20mM Tris pH 8.0, 120mM KCl, 1mM DTT, 0.5mM EDTA). Dialysis was done in the same manner as was described in the preparation of the protein samples for the CD analysis (2.2.3, pg. 28). Following dialysis, protein samples were diluted with DNA EMSA P. Buffer to a concentration of 48µM (or 1.837mg/mL). In parallel to this, the DNA sample was prepared from a polymerase chain reaction (PCR) using the pUC19 plasmid (Invitrogen) as the template, and primers designed to flank a 1000bp region of the plasmid (MJ3280 and MJ3281). The DNA product was gel extracted, purified (QIAEX II® Gel Extraction Kit QIAGEN) and subjected to a second round of PCR to

obtain the purified 1000bp DNA substrate used in this assay. The PCR was carried out with the EmeraldAmp[®] MAX PCR as outlined in the manufacturer's manual (Takara) and the DNA was extracted from a 1% agarose gel into TE buffer (20mM Tris pH8.0, 0.1mM EDTA). Purified DNA was diluted with TE buffer to a concentration of 30ng/μL prior to a further dilution with DNA EMSA D. Buffer (20mM Tris pH8.0, 240mM KCl, 2mM DTT, 30% glycerol) in a 1:1 ratio. The diluted protein and DNA samples were then mixed in a 2:1 ratio (protein:DNA) and allowed to incubate at room temperature for 45 minutes. Following this, the samples were loaded onto a 0.8% agarose TBE gel to 100ng of DNA in each lane and subjected to 170V for 5 minutes followed by 80V for 45 minutes in 1X TBE buffer (89mM Tris pH8.0, 89mM Boric Acid, 2.5mM EDTA). Once the electrophoresis was completed, the gel was stained in fresh 0.5μg/mL ethidium bromide (BioShop) followed by an overnight destain in ddH₂O. In addition to this, equal volumes from the reaction samples were used in a SDS-PAGE analysis using a 12% polyacrylamide gel. The electrophoresis of the samples was done at 140V for 70 minutes and the resulting gel was stained in Coomassie Brilliant Blue G-250 stain (Bio-Rad).

2.2.5 Protein Electrophoretic Mobility Shift Assay

Protein samples (XRCC4 constructs and the BRCT domains of DNA ligase IV) were allowed to thaw on ice prior to their dialysis in Protein EMSA Buffer (20mM Tris pH 8.0, 150mM KCl, 1mM DTT, 10mM EDTA). Dialysis was done in the same manner as was described in the preparation of the protein samples for the CD analysis (2.2.3, pg. 26). Following dialysis, protein samples were diluted with Protein EMSA buffer to a concentration of 20μM. Reaction samples were then prepared by mixing of the XRCC4,

BRCT, and Protein EMSA buffer in a 1:1:2 ratio (XRCC4:BRCT:Buffer) and these reaction samples were incubated at room temperature for 30 minutes. Next, concentrated glycerol solution was added to the reaction samples to a concentration of 10%, effectively diluting the sample by a factor of 0.8, prior to their loading onto a 6% Native polyacrylamide gel. Electrophoresis was done in 1X TBE buffer at 150V for 5 minutes followed by 120V for 50 minutes. Gels were stained in Coomassie Brilliant Blue G-250 stain (Bio-Rad).

2.2.6 Analytical Ultracentrifugation

XRCC4 samples used in AUC experiments were not frozen after the anion exchange chromatography, but instead, the proteins were transferred into AUC buffer (100mM KCl, 20mM Tris-HCl pH8.0, 1mM EDTA pH8.0, 1mM TCEP). This was done by applying the purified protein samples to the HiPrep™ 26/10 Desalting column (GE Healthcare Lifesciences) previously equilibrated with AUC buffer and collecting the elution fractions containing the protein in the AUC buffer, determined by the monitoring of the A_{280} . Following this, the concentration of the proteins was increased using the Corning® Spin-X® UF 20mL Concentrator with a 30kDa molecular weight cutoff to an absorbance at 280nm of 4.0. These samples were flash frozen in liquid nitrogen and sent to Dr. Roldofo Ghirlando from LMB NIDDK, where the AUC experiments were carried on.

The AUC experiments were carried in two steps. In the first step, samples were subjected to sedimentation velocity experiments at a single point concentration to obtain initial data on the behaviour of the protein during these experiments. Following this, the

samples were subjected to sedimentation equilibrium experiments at multiple speeds and concentrations to obtain information on the oligomer state of the different XRCC4 constructs. In general the samples with concentrations ranging from 20-40 μ M were subjected to sedimentation velocity analysis at 20°C and rotor speed of 50krpm, and absorbance data was collected at 280nm while interference data was collected at 655nm. Samples with concentrations ranging from 9 to 140 μ M were subjected to sedimentation equilibrium analysis at 4°C and rotor speeds of 9, 14, and 19krpm, and the absorbance data was collected at 280nm.

2.3 Results

2.3.1 Purification Profile of XRCC4 Proteins

The purification profiles for the XRCC4 mutants demonstrated similar behaviours to that of the wildtype protein. Following lysis of the induced cells, the proteins were first purified by nickel affinity chromatography and fractions were collected throughout the purification procedure to be analyzed by SDS-PAGE. As indicated in Figure 2.2, the protein was eluted off the column at an imidazole concentration of 210mM and, at that point, was at a purity of about 80%. This purified sample was subjected to a second purification step by anion exchange chromatography (Figure 2.3). The protein was first applied to the anion exchange column and a salt gradient from 150mM to 400mM KCl was then set. The elution of the XRCC4 proteins began at approximately 290mM KCl and the highest concentration of the proteins was found at 310mM KCl. In general, 3 to 5 mg of purified protein was obtained from 4L of cell preparation (variation was mostly dependent on the construct being purified) at the end of the purification steps. These

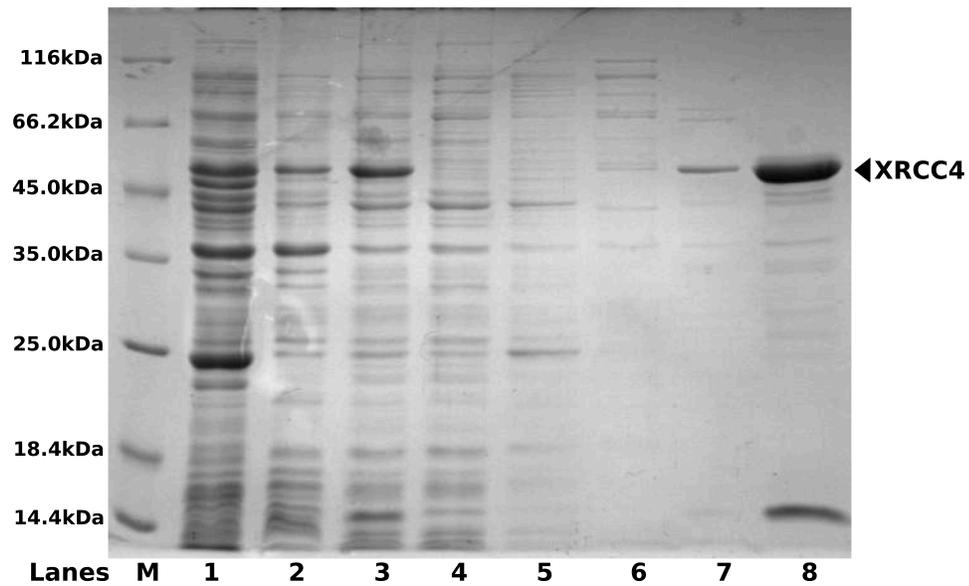


Figure 2.2 Nickel Affinity Chromatography of XRCC4. 12% SDS-PAGE of the samples collected during nickel affinity purification. The labeled lanes are 1) marker, 2) insoluble fraction after lysis, 3) soluble fraction after lysis, 4) flowthrough, 5) 40mM imidazole wash, 6) 70mM imidazole wash, and 7) elution (210mM imidazole). The protein of interest XRCC4 is indicated by the black arrow and is represented by a band of approximately 50kDa.

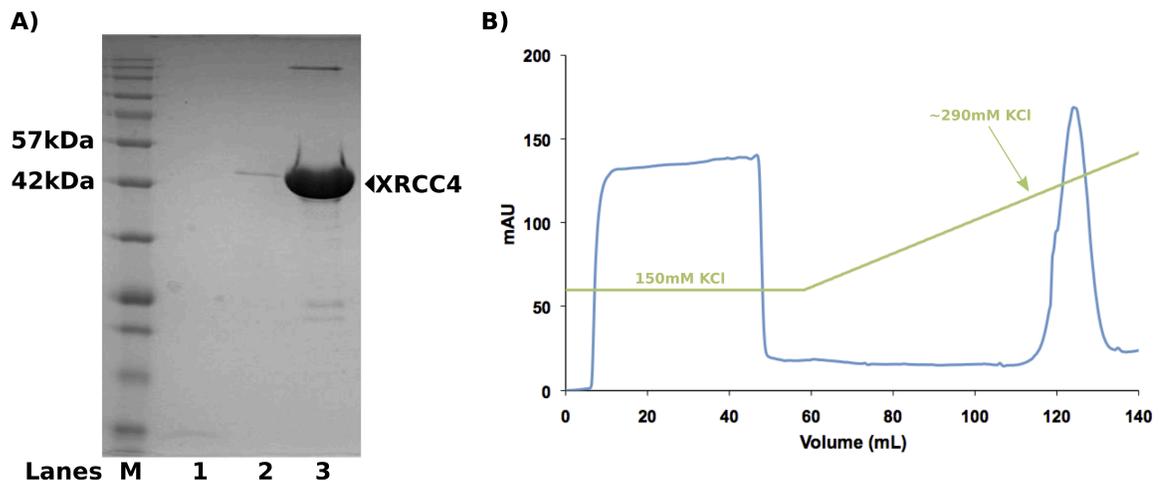


Figure 2.3 Ion Exchange Chromatography of XRCC4. Panel A, 12% SDS-PAGE of samples collected during the anion exchange purification step. The lanes are labeled as M) marker, 1) flowthrough, 2) fraction collected at 290mM KCl, 3) fraction collected at 310mM KCl. Black arrow indicates the band representing the purified XRCC4. Panel B, chromatogram generated from the anion exchange chromatography. Detection of A_{280} is represented in blue, while salt concentration of the buffers used during the purification is indicated in green. The green arrow, at approximately 290mM KCl, marks the point where XRCC4 begins to be eluted.

protein samples were at approximately 95% purity containing a few impurities at a lower molecular weight and single band at a higher molecular weight than that of the XRCC4 band.

2.3.2 XRCC4 Mutants retain their Secondary Structure

The secondary structural content of the XRCC4 proteins was verified by circular dichroism. The measurements obtained by the CD analysis were further processed with the CDPro software using the CONTINLL algorithm to obtain calculated information on the structural content of each XRCC4 protein (Table 2.1). Most mutants contained a similar structural content to that of the wildtype XRCC4 with a 34% α -helical content, 13% β -sheet content, 20% turns, and 33% unordered.

2.3.3 XRCC4 Mutants bind to DNA

The ability of the mutants to bind to DNA was probed by DNA electrophoretic mobility shift assay. In this assay, the proteins were incubated with a 1000 base pair blunt DNA substrate for 30 minutes at room temperature before being loaded onto an agarose gel. Once electrophoresis was performed on this gel, the gel was incubated with ethidium bromide solution to stain the DNA substrate. Figure 2.4 shows that the incubation of the DNA substrate with any of the generated mutants resulted in an upward shift of the band in relation to the band found in the DNA substrate alone lane. While XRCC4 – V33E, V104E F106A, Q145A, and R150E resulted in shifts to similar positions to that of the XRCC4 – WT, XRCC4 – A60E resulted in a shift to a lower position and XRCC4 – E121K E125K, and E147K resulted in shifts to higher positions than that of the XRCC4 – WT. The triple mutant (XRCC4 – L184Q K187D I191S) demonstrated an incomplete

XRCC4	α-helical (%)	β-sheets (%)	Turns (%)	Disordered (%)
Wildtype	34 (2)	13 (3)	20 (2)	33 (3)
L184Q K187D I181S	34 (4)	16 (3)	19 (1)	30 (0.5)
V33E	32 (8)	16 (4)	20 (2)	32 (3)
A60E	33 (5)	17 (5)	20 (2)	30 (1)
V104E F106A	36 (4)	13 (2)	21 (1)	30 (4)
E121K E125K	36 (6)	12 (4)	21 (0.4)	32 (3)
Q145A	25 (6)	22 (6)	22 (0.2)	32 (1)
E147K	35 (1)	16 (2)	21 (1)	29 (2)
R150E	35 (2)	12 (3)	19 (1)	34 (2)

Table2.1 XRCC4 mutants have a similar secondary structure content to the wildtype. XRCC4 mutants were purified and exchanged into the appropriate CD buffer. CD spectra was collected from 190-260nm and analyzed by CONTINLL to generate the secondary structure content for each mutant. Experiment was done in triplicates and the calculated secondary composition percentage and their associated standard deviations, in brackets, are listed on the table.

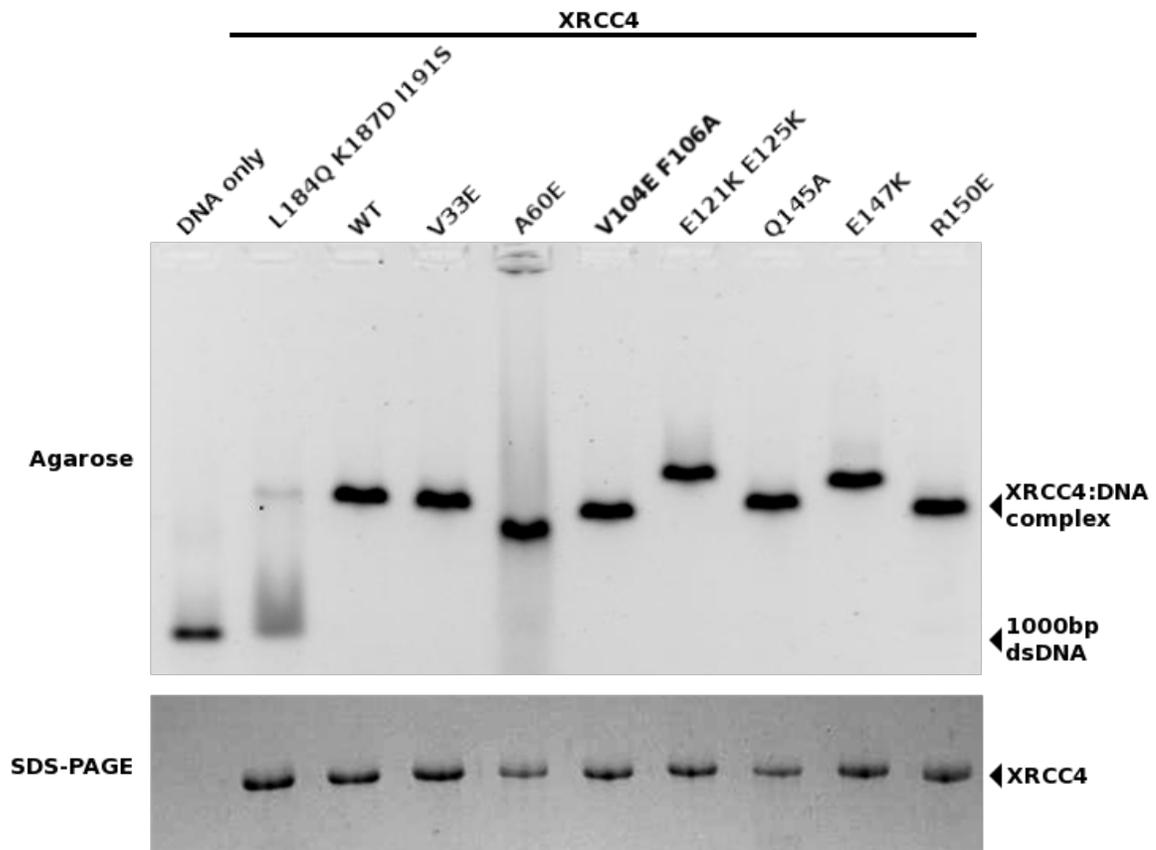


Figure 2.4 XRCC4 mutants retain their ability to bind to DNA. Top gel, 0.8% agarose gel used for the DNA EMSA. 100ng of a 1000bp dsDNA substrate was incubated with the XRCC4 protein samples for 30 minutes at room temperature prior to their electrophoresis at 80V for 70min. Bottom gel, SDS-PAGE of the samples used in the DNA EMSA to verify for the levels and presence of the XRCC4 proteins. The SDS-PAGE was performed at 120V for 70min.

shift of the DNA alone band. In addition to this, the samples used in the DNA assay were also used in a SDS-PAGE analysis to verify for the presence and the relative levels of the XRCC4 mutants in each of the reaction sample. As shown in the bottom panel of Figure 2.4, the intensity of the bands were similar to all the reaction samples containing the XRCC4 proteins, with the exception of the XRCC4 – A60E and Q145A, which demonstrated a weaker intensity of the bands.

2.3.4 XRCC4 Mutants interact with DNA Ligase IV

Protein EMSA was performed in order to probe for the ability of the mutants to interact with DNA ligase IV. This assay was set up by incubating the XRCC4 proteins either in the absence or presence of the BRCTs domain of DNA Ligase IV in equimolar concentrations (5 μ M). The incubation was done at room temperature and allowed to occur for 30 minutes prior to loading onto a 6% TBE-based native polyacrylamide gel. Once electrophoresis was completed, gel was incubated in Coomassie Brilliant Blue stain in order to detect the protein bands. Figure 2.5 shows that all XRCC4 bands are shifted upwards relative to their respective XRCC4 alone lanes in the presence of the BRCTs domains, with the exception of the triple mutant (L184Q K187D I191S) that does not demonstrate a shift with the addition of the BRCTs domain. Similarly to the DNA EMSA, the samples were also subjected to SDS-PAGE analysis in order to verify for the presence and levels of the proteins in each reaction sample. As shown in the bottom panel of figure 2.5, the intensity of the bands for all proteins was similar for all the reaction samples. Additionally, the bands representing the BRCTs domain were only present in the reaction samples that contained the BRCTs.

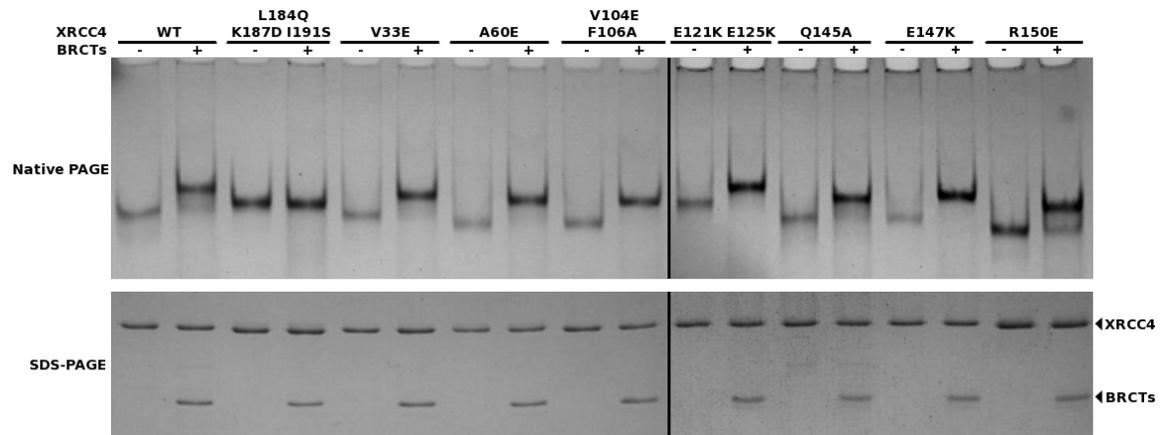


Figure 2.5 XRCC4 mutants retain their ability to interact with the BRCTs of DNA Ligase IV. Top gel, protein EMSA of the XRCC4 mutants in the presence or absence of the BRCTs region of DNA ligase IV. XRCC4 mutants were incubated with the BRCTs at an equimolar concentration for 30min at room temperature prior to their loading onto a 6% native polyacrylamide gel. The electrophoresis was performed at 120V for 30min and stained for proteins by Coomassie Brilliant Blue stain. Bottom gel, SDS-PAGE of the samples to verify for the level and the presence of the proteins in each binding reaction. SDS-PAGE was done at 140V for 70min.

2.3.5 On-going Analytical Ultracentrifugation

Currently, the analytical ultracentrifugation experiments probing for the oligomerization state of the XRCC4 proteins are still in progress. In general the proteins are being subjected to two analyses by this method: the sedimentation velocity experiments and the sedimentation equilibrium experiments. Unfortunately, the results obtained so far are in conflict with the expected results based on other available and published data. These results are summarized on Table 2.2. The issue being observed for all the XRCC4 proteins is that the data collected is not fit to be modeled as reversible dimer-tetramer species. Instead, the best models that fit the data are the ones that include irreversible dimer-tetramer species containing either additional smaller species or large aggregates. Noteworthy results are those for the wildtype, the triple mutant, and the V104E F016A mutant. Although displaying evidence for both dimer and tetramer species, the wildtype XRCC4 is displaying a much smaller amount of tetramers than is expected (only ~10% of the species is found in a tetramer state). The triple mutant is displaying a predominant evidence for dimers but the estimated sedimentation coefficient (3.62S) and molecular weight (54.4kDa) are much smaller than expected when comparing to the wildtype XRCC4 with a sedimentation coefficient of 4.00S and a dimer molecular weight of 76.5kDa. Finally, the data collected for mutant V104E F106A is, currently, the only data able to be fit with a reversible dimer-tetramer equilibrium model with an approximate K_d of 33 μ M, which is higher than what has been reported for the wildtype XRCC4 with a K_d of 22 μ M.

XRCC4	Sedimentation Velocity	Sedimentation Equilibrium
Wildtype	Evidence for mostly dimer (4.00S) but some evidence for higher order species (5.60S)	Modeling of data only works with the inclusion of incompetent monomers or smaller species
L184Q K187D I181S	Evidence for mostly dimer	Best modeled as a 54.4kDa species
	Smaller than expected sedimentation coefficient (3.62S)	
V33E	Evidence for mostly dimer	No data collected
	Best modeled as a dimer in the presence of large aggregates	
A60E	Evidence for mostly dimer (3.90S)	Same issue as in wildtype case
V104E F106A	Evidence for mostly dimer (3.97S)	Best modeled as a 73.1kDa species with a reversible dimer-tetramer Kd of 35 μ M
E121K E125K	Evidence for mostly dimer	No data collected
	Best modeled as a dimer in the presence of large aggregates	
Q145A	Evidence for mostly dimer	No data collected
	Best modeled as a dimer in the presence of large aggregates	
E147K	Evidence for species at 4.00S and 4.75S	Same issue as in wildtype case
R150E	Evidence for mostly dimer (3.90S)	Best modeled as a 73.9kDa species

Table 2.2 Summarized AUC Data on XRCC4 Mutants. The AUC experiments were done in two different methods for each of the mutants: Sedimentation velocity and Sedimentation equilibrium. The table above outlines the analysis of the results for each of the experiments done for the individual XRCC4 proteins.

2.4 Discussion

XRCC4 is a structural protein that does not have any known enzymatic activity, but curiously, has been shown to be critical for the repair of DSBs by the NHEJ pathway. Although this protein has been extensively studied and characterized for almost 20 years, its exact role in the repair of the DNA still remains to be elucidated (Li *et al.*, 1995). Several studies have demonstrated that XRCC4 is able to interact with many DNA repair proteins, such as XLF and DNA Ligase IV, to directly interact with DNA substrates, and to form homodimers as well as homotetramers (Andres *et al.*, 2012; Junop *et al.*, 2000; Modesti *et al.*, 2003; Wu *et al.*, 2009). Additionally, by interacting with such partners, XRCC4 is capable of bridging DNA molecules and stimulating the ligase activity of DNA ligase IV (Andres *et al.*, 2012; Grawunder *et al.*, 1998b; Wu *et al.*, 2009). Despite all that has been learned about XRCC4, the direct relationship between its functions and the repair of DNA continues to be a mystery, in part due to the structure: function relationship of this protein. As mentioned previously, the CTR of XRCC4, more specifically ranging from residues 162 to 203, have been shown to be important for XRCC4's ability to form tetramers and to interact with DNA ligase IV and DNA molecules. Due to the nature of the CTR of XRCC4, mutants generated to study these individual functions have so far affected multiple functions of the protein, complicating the analyses of XRCC4's functions in respect to the actual repair of the DNA. In light of this and the current debate of XRCC4's tetramerization mode, we have aimed to further characterize the tetramerization of this protein. In doing so, mutants only affecting this

function could then be generated to individually study the importance of XRCC4's tetramerization to the actual repair of the DNA.

To accomplish this, mutants targeting the head-to-head tetramerization were generated and verified by DNA sequencing. Following the expression of these genes and the purification of the proteins, the proteins were subjected to circular dichroism, and DNA and protein EMSAs analyses. This was done with a dual purpose: i) to confirm that the introduced mutations did not disrupt the overall folding of the XRCC4 protein and ii) to verify that the other functions of XRCC4 were not affected by the mutations. As shown in Table 2.1, CD analysis of the mutants suggests that the mutations did not grossly modify the secondary structural content of the XRCC4 protein. In general the XRCC4 proteins adopt a secondary structural content of about 34% α -helices, 13% β -sheets, 20% turns, and 33% unordered. Although the secondary structures of the mutants do not significantly change from the wildtype, the secondary structural content of the wildtype XRCC4 itself as determined by CD slightly deviates from what is observed in the crystal structure of this protein. The crystal structure of XRCC4 (pdb1FU1) only spans residues 1 to 203, while the entire protein is composed of 336 residues. If the remaining 133 C-terminal residues were assumed to be unordered, the protein would be composed of about 23% α -helices and 16% β -sheets, while the remaining portion would be either turns or unordered regions. This discrepancy could be attributed to either the rigid nature of the protein structure in the crystals as opposed to its more labile nature in solution, or to the possible formation of additional α -helical structures beyond residue 203 that is not observed in the crystal. Despite this discrepancy, XRCC4 mutants behaved in a much

similar manner to that of XRCC4 wildtype suggesting that the mutations did not introduce any gross change in the structure of XRCC4. Therefore, the results observed for these mutants are not likely to be due to gross misfoldings of the proteins but rather to be associated with the specific mutations that we have generated for this protein.

Following this, DNA EMSAs were completed to test the DNA-binding ability of the XRCC4 mutants. As expected, the XRCC4 mutations generated for this project did not significantly affect the ability of XRCC4 to bind to DNA (Figure 2.3). Interestingly, the triple mutant was able to interact with the DNA substrate, albeit, to a much weaker affinity than the other mutants. Another noteworthy point to be considered is the relative shifts of the protein:DNA complexes in the gel. Particularly, the A60E mutation resulted in a lower shift while the E121K E125K and the E147K mutations resulted in a higher shift of the protein:DNA complexes. The mobility of the entities in an EMSA generally reflects the conformation or the charge to mass ratio of the species in the gel. Considering that the mutations introduced to XRCC4 does not significantly change the molecular weight or the theoretical pI of the molecule, one would expect that the charge to mass ratio of the mutants:DNA species would not differ much from that of the wildtype:DNA specie. However when considering the fact that XRCC4 does not recognize a specific sequence on the DNA but rather binds DNA promiscuously, there is a possibility that multiple XRCC4 proteins would bind to and fully coat each single DNA substrate. With this in mind, the additive impact of the introduced charged amino acids, via the mutations, could be significant enough to explain the observed altered shifts of the mutants:DNA complexes. In fact, A60E introduces a negative charge and thus results in

the observed faster migration of the complex, while E121K E125K and E147K introduces positive charges and thus result in the observed slower migration of the complexes. Although this is a feasible explanation of our observed results, it becomes less likely as this same logic is applied to the other mutants. Specifically, the V33E and R150E mutants also introduced negative charges to the protein and thus should have resulted in faster migration of the protein:DNA complexes. However as it is shown in figure 2.3, these mutants migrated to positions similar to that of the wildtype:DNA complex. Therefore, this result suggest that although the mutations did not grossly affect the folding of the protein, they may have affected the overall conformation of the XRCC4:DNA complex. In such case, our result would suggest a possible bending of the DNA substrate by the XRCC4 protein. As the bending of DNA molecules can significantly affect their electrophoretic mobility in gels, the observed anomalous shifts of the DNA bands for each mutant could be due to the different degrees of DNA bending caused by the binding of these mutants to the DNA. Further work would need to be done with these mutants to verify this hypothesis and, such work, potentially provide a better understanding of how XRCC4 interacts with DNA.

Next, a XRCC4:BRCT EMSA was performed to verify the ability of the mutants to interact with DNA ligase IV. Since the mutations generated for this project were not located at the XRCC4:DNA ligase IV interaction region, it is expected that our mutations would not interfere with this interaction. Indeed, all the mutants were able to interact with the BRCTs region of DNA ligase IV as it is shown in figure 2.5. Again, the triple mutant is the reported mutant unable to interact with DNA ligase IV and also behaved as

expected in this EMSA. Similarly to the DNA EMSA, the XRCC4 proteins also migrated to slightly different positions. In this case, the introduced charges do not appear to dictate the migration of these proteins, as there is no evident pattern of a change in charge and a corresponding slower or faster migration of the XRCC4 protein. For this reason, this result suggests the mutants of XRCC4 are adopting different conformations than that of the wildtype XRCC4. In addition to this, the XRCC4 R150E mutant appears to have a weaker affinity towards the BRCTs. However, previous work on this mutant (not shown in this thesis report) suggests that this weaker affinity is due to the use of an older preparation of this protein and not due to the mutation.

Lastly, the XRCC4 proteins were subjected to both sedimentation velocity and sedimentation equilibrium experiments to determine their oligomerization state. Sedimentation velocity experiments are generally done at high speeds and provide information on the behaviour of the species as it transverses the solution, while sedimentation equilibrium experiments are done at lower speeds and allowed to reach an equilibrium where the flux of sedimentation is balanced by the flux of diffusion of the species. In this manner, we are able to obtain information on the hydrodynamic nature, the molecular weight of the macromolecules, and the affinity and stoichiometry of their interactions (Ghirlando, R., 2011). Currently we are facing difficulties in regards to obtaining good quality data. As summarized in table 2.2, the major issue that we are encountering during the AUC experiments is the possible contamination of the samples with, in some cases, smaller contaminants or in other cases larger aggregates. These contaminants make it difficult to properly fit models to the data collected during the

sedimentation equilibrium experiments, ultimately, interfering with our ability to determine the proper K_d of the interactions. A reasonable explanation for the formation of aggregates could be the freeze/thaw process that the protein samples undergo as they are being prepared to be sent to our collaborator. In order to verify for this possibility, the protein could be handled in the same manner and then analyzed by size exclusion chromatography (SEC). Previous SEC experiments on this protein revealed an inability to completely resolve the tetramer from the dimer during the chromatography. However in this case, we would expect to see a peak at the void volume representing the aggregates while the dimer would elute at a later point. If this is confirmed, we will need to more carefully handle the samples in order to avoid the freeze/thaw of the samples and it will also allow us to take these large aggregates into account during our analysis of the AUC data. In addition to this issue, the triple mutant displayed a much smaller sedimentation coefficient of 3.62S than that of the wildtype at 4.00S. These results suggest that the triple mutant could have undergone degradation during the course of the experiment. Analyzing the protein samples by SDS-PAGE after the AUC experiments might help us determine if that is the case.

Despite the difficulties that we have encountered in our AUC experiments, the results obtained from sedimentation velocity experiments suggest the presence of mostly dimers for the majority of the XRCC4 proteins, with the exception of the wildtype, V104E F106A, and E147K. The species at approximately 4.00S represent the dimer form of the protein, while those at 5.60S represent the tetramer form of the protein, as observed for the wildtype protein. In the case of the wildtype, the relative signal at 5.60S was much

smaller than expected (only 10% of the signal) considering the reported K_d at $22\mu\text{M}$. The V104E F106A mutation also showed evidence for some tetramers and through sedimentation equilibrium experiments a calculated K_d of $35\mu\text{M}$ was obtained. However as the negative (triple mutant) and positive (wildtype) controls are not yet behaving as previously reported in literature, this estimate for the tetramerization K_d of the V104E F106A mutant is somewhat premature. At this moment, the issue with the controls must first be resolved before we are able to confidently analyze the results we obtain for the mutants. Lastly, the observed sedimentation coefficient of 4.75S for the XRCC4 E147K mutant is much smaller than that observed for the tetramer form of the wildtype XRCC4. Although unlikely, this could represent a different conformation for the XRCC4 protein as sedimentation coefficients are also dependent on the hydrodynamic nature of the specie. If we are able to improve the data collected by sedimentation equilibrium, we will be able to determine the molecular weight of the observed species independently from its shape and therefore better characterize the specie observed at 4.75S . Overall, the AUC experiments suggest that the mutated residues in both the head region and the upper tail region of XRCC4 play a role in the tetramerization of the protein. It is interesting to note that while mutations of the residues in the electrostatic interaction observed in the head-to-head tetramerization resulted in disruption of tetramerization, the mutations of the residues participating in the hydrophobic pocket did not necessarily result in such disruption. (Figure 2.1, Table 2.2). Upon closer inspection of the hydrophobic interactions between the homodimers of XRCC4, we were able to find the residue L101 (a hydrophobic residue) located near the V104E F106A mutated residues (on the green

homodimer shown in figure 2.1) that could potentially compensate for the mutation that we have introduced in the XRCC4 protein. Conversely when looking at the region of the interacting homodimer (coloured blue in figure 2.1), we were unable to find any nearby located hydrophobic residue that could compensate for the V33E mutation. Thus, the mutation of V33E resulted in a greater effect of tetramerization disruption than the V104E F106A mutation. Further analysis, potentially using a triple mutant of L101 V104 F106, would be required to test this hypothesis.

Nevertheless, if our results are confirmed, it becomes interesting to consider what the implications of this tetramerization mode would be to the NHEJ pathway and this will be discussed in greater detail in Chapter 4 of this thesis report.

**CHAPTER 3 – CHARACTERIZATION OF THE
XRCC4:XLF:DNA COMPLEX**

3.1 Rationale and Experimental Design

Another aspect of my work was on the characterization of the XRCC4:XLF:DNA complex. As a recent development in the field, XRCC4 was shown to interact with XLF and to be able to bridge long DNA substrates *in vitro* (Andres *et al*, 2012). At the same time, the XRCC4:XLF complex was also crystallized and its structure was solved by X-ray crystallography. Through this technique, the complex was shown to be composed of alternating homodimers of XRCC4 and XLF (XRCC4-like factor) forming a coiled protein filament (Andres *et al*, 2012). As its name suggests, XLF is a structurally related protein to XRCC4 that forms a homodimer with a head and a tail region. However, unlike XRCC4, XLF's C-terminal tail has been shown to fold back onto itself and coil around its tails (Andres *et al*, 2006; Li *et al.*, 2008; Figure 3.1). In the observed crystal structure of the XRCC4:XLF filament, the proteins interacted through residues on the head regions of both proteins. Additionally, studies such as scanning force microscopy and SAXS have shown evidence for the localization of multiple filaments to a single site, potentially forming a filament bundle (Andres *et al.*, 2012; Hammel *et al.*, 2011). Taken these together, models for the DNA interaction with the XRCC4:XLF complex were proposed where the formation of this protein filament:DNA structure occurs with the treading of the DNA through the inner pore of the coiled protein filament or by its coiling around the filament bundle (Andres *et al*, 2012; Hammel *et al*, 2011).

In order to determine the method in which XRCC4:XLF complex is bridging the DNA molecule, we have set out to first characterize the binding of this complex to DNA.



Figure 3.1 Structure of XLF¹⁻²²⁷. Crystal structure of XLF¹⁻²²⁷ solved by X-ray crystallography (pdb 2R9A). The structure is composed of a homodimer with a N-terminal head region followed by a coiled-coil tail region. The C-terminus region of the protein is flexible and fold back onto itself coiling around the tails of the XLF's homodimer. N- and C- terminus are labeled to demonstrate their position in the overall structure of the protein.

The method in which we have chosen to tackle this problem was by crystallizing this complex in the presence of a DNA substrate to then solve for its structure by X-ray crystallography. However, a major limitation to crystallography is determining the protein boundaries and DNA lengths to be used during the crystallization. Proteins with flexible domains or DNA substrates with inappropriate lengths can result in the physical interference of the packing of the proteins, which in turn prevents the formation of good quality crystals. While we have chosen to use a truncated form of XRCC4 containing only residues 1 to 265, we have decided to use the full-length version for the XLF protein. The boundaries for XRCC4 were decided based on previous work demonstrating that this truncation is soluble, stable, and still able to interact with XLF. The full length of XLF was used in the crystallization attempts because the bridging of the XRCC4:XLF complex have been shown to be disrupted by the removal of the C-terminus region of XLF (Andres *et al.*, 2012). Moreover, this region has been shown to be important for XLF's binding to DNA – a characteristic that becomes more significant as we consider the possibility of the protein complex:DNA interaction occurring by the treading of the DNA through the coiled protein filament. In this model of interaction, the tails regions of XRCC4 and XLF point outwards and away from the inner pore. However since XLF's tail has been shown to fold back onto itself, the C-terminus of this protein is redirected towards the inner pore of the coiled protein filament. This allows for a possible interaction of the C-terminus of XLF with the DNA located within the inner section of the protein complex.

With this in mind, we have also designed a short dsDNA substrate of 36 base pairs with complementary 2 nucleotides 5' overhangs. This was done in an attempt to use the minimal DNA length that would be found in the asymmetric unit in the unit cell of the protein:DNA crystal. More specifically, the length of the DNA was determined based on the crystal structure of the XRCC4:XLF complex. By taking into consideration that a single XRCC4:XLF complex spans approximately 130Å and a single nucleotide in a DNA molecule spans about 3.3Å, we have approximated the minimum DNA substrate for this complex to be approximately 38 nucleotides long. The overhangs at the ends of the DNA substrate were introduced to account for the repetitive nature observed in protein crystals to, if necessary, allow for or aid in the formation of longer repetitive units of this short DNA substrate in the crystal.

Once the protein and DNA substrates were chosen, the next challenge was to find the appropriate protein and solvent conditions to promote nucleation followed by crystal growth. This was done by testing several commercially available broad screens for their ability to form protein crystals and then optimizing these conditions to generate crystals that could give good X-ray diffraction patterns. These crystallization conditions often include buffer reagents (to set the pH of the condition) and a precipitant reagent (to induce the crystallization of the protein). The precipitant reagent, in turn, can be subdivided into many classes including salts, polyethylene glycol (PEG) reagents, and organic solvents, which can induce crystallization or precipitation of the protein via different methods. For example, salts induce crystallization by the “salting-out” effect, where the salt ions compete against the protein molecules for the water molecules, where

PEGs induce the crystallization by competition as well as by volume-excluded effects (Drenth, J., 2007). With this knowledge, one could apply rational thinking to guide which conditions to attempt first depending on the nature of the protein used for crystallization. However, by doing so one could also accidentally apply a bias towards their crystallization attempts by using a set of specific precipitants that are unfavourable for the crystallization of the protein of interest. As such, we have used fairly broad screens to avoid falling into such problem and reserved this consideration only for promising conditions that were obtained.

The next step was the optimization of the conditions, which was done by making small variations on temperature, pH, and concentrations of the reagents found in the crystallization condition. The “fine-tuning” of these parameters influences the phase diagram of the solubility of the protein in the crystallization solution. In this graph, we observe two major regions, the undersaturated (soluble) region and the supersaturated region. The supersaturated region is further divided into three smaller regions with the metastable zone, the labile zone, and the precipitation zone. The labile zone of this graph represents the condition in which the nucleation of the protein occurs and the metastable zone is where the growth of the crystals occurs (Asherie, 2004). Ideally, we manipulate the conditions in such manner that the proteins are initially present at the labile/nucleation zone but close to the metastable zone. As the protein nucleates its concentration decreases shifting the protein into the metastable zone, where it begins to grow. As the concentration of the protein continues to drop it reaches the boundary between the undersaturated and saturated regions. At this point the precipitant in the wells lowers the

solvent content in the drops, effectively increasing the precipitant and the protein concentrations in the drop, and as a result promotes further growth of the crystals. Through iterative cycles of growth and concentration, the crystal grow until the concentration of precipitant in the drop becomes equivalent to that in the wells or until a poisoning of the crystal occurs, where protein molecules begin interact with the crystals in the wrong conformation.

Prior to the actual crystallization of the protein:DNA complex, we have also performed EMSAs to verify for the ability of this protein complex to bind to the short DNA substrate that we have designed. Following that, multiple crystallization attempts were done to crystallize the XRCC4¹⁻²⁶⁵:XLF^{FL} complex with and without the DNA substrate. Obtaining information on the structure of this protein complex without DNA can also be useful as the reported structure for the XRCC4:XLF complex was accomplished with truncations of these proteins (XRCC4¹⁻¹⁵⁷: XLF¹⁻²²⁴).

3.2 Materials & Methods

3.2.1 Expression and Purification of Proteins

Expression of the gene coding for XRCC4¹⁻²⁶⁵ was completed in a similar manner as described on Chapter 2, section 2.2.2 for the expression of the *xrcc4*-genes. The modifications were as follow: *E. coli* BL21 (DE3) cells were transformed with the pWY1107 plasmid (containing the *xrcc4*¹⁻²⁶⁵ gene) and the antibiotic used in every step was kanamycin at a concentration of 50µg/mL. Purification of the XRCC4¹⁻²⁶⁵ protein was done as described on Chapter2, section 2.2.3 for the purification of the XRCC4 proteins.

The expression for the xlf gene was performed in a similar way to the expression of the xrcc4-genes described on Chapter 2, section 2.2.2. Modifications involved the transformation of *E. coli* BL21 (DE3) cells with the MJ4489 plasmid (containing the gene for XLF) and the use of ampicillin at a concentration of 100 µg/mL as the selective pressure factor. Once this was completed, the complete purification of the XLF protein was done in two purification steps: nickel affinity chromatography and heparin affinity chromatography. First, the frozen induced cell pellets were allowed to thaw on ice prior to its resuspension in Nickel A buffer (20mM Hepes pH8.0, 2M NaCl, 10mM imidazole, 1mM β-mercaptoethanol). Following this, cells were sonicated 3 times for 1 min with an amplitude of 18% (Sonic Dismembrator Mode 5000, Fisher Scientific) and then centrifuged at 48 384 x G for 40 minutes (Avanti J-30I Centrifuge with J.A-30.50 Rotor, Beckman Coulter). The supernatant was filtrated by vacuum filtration using the 25mm, 0.45 µm filter from Metricel Membrane Filter, Pall Life Sciences. The filtered sample was applied to a 5mL HiTrap nickel affinity column (GE Healthcare) that was set on a FPLC system (AKTA FPLC, Amersham Pharmacia Biotech). Sequential step washes of the column with buffers containing 40mM and 90mM imidazole were done prior to the elution of the XLF protein at 240mM imidazole. The buffers with different imidazole concentration were prepared by mixing of Nickel A buffer and Nickel B buffer (20mM Hepes pH8.0, 2M NaCl, 1mM β-mercaptoethanol, 240mM imidazole) in the appropriate ratios. The elution sample containing XLF was applied to a HiPrep™ 26/10 Desalting column (GE Healthcare Lifesciences) that was previously equilibrated with 3 column volumes of HepA buffer (20mM Hepes pH8.0, 10mM BME, 1mM EDTA, 150mM NaCl,

10% glycerol) and elution of the protein was monitored by A_{280} . This step was done to transfer the protein into HepA buffer prior to its further purification by heparin affinity chromatography. Following this, the sample was applied to a 5mL HITrap Heparin HP column (GE Healthcare) previously equilibrated with HepA buffer. After extensive wash of the column with HepA buffer, the protein was eluted at 400mM NaCl using a buffer that was prepared by mixing of HepA buffer with HepB buffer (20mM Hepes pH8.0, 10mM BME, 1mM EDTA, 1M NaCl, 10% glycerol) in the appropriate ratio.

Lastly, both proteins were transferred into the crystallization buffer (20mM Tris pH8.0, 1mM EDTA, 10% glycerol, 10mM DTT, 200mM KCl) by applying it to a previously equilibrated HiPrep™ 26/10 Desalting column (GE Healthcare Lifesciences). Protein samples were then flash frozen in liquid nitrogen and stored at -80°C until further use.

3.2.2 Preparation of DNA Substrates

The DNA substrate used for the crystallization of the XRCC4¹⁻²⁶⁵:XLF:DNA complex was prepared by annealing two oligos of 38 nucleotides: O1 – 5' TTAA(32XG)AA 3' & O2 – 5' AATT(32XC)TT 3' (IDT). Prior to the annealing procedure, oligos were purified by gel purification. This was done by loading the oligos onto a 20% polyacrylamide gel containing 7M urea and performing the electrophoresis at 700V for 4hours using the vertical electrophoresis system from CBS Scientific. Following this, the appropriate bands were cut out and the oligos were eluted out of the gel by an overnight incubation of the gel pieces in 200mM NaCl, 5mM Tris pH7.5, 1mM EDTA at 37°C. Next, the oligos were ethanol precipitated and re-suspended in TE buffer (20mM

Tris pH8.0, 0.1mM EDTA). Once oligos were purified, they were annealed by mixing them in equimolar concentration and incubating them in boiling water. This system was removed from the heat source and allowed to equilibrate to room temperature, completing the annealing process of the oligos.

The DNA substrates used to probe for the ability of the proteins to interact with short DNA molecules were prepared in a similar manner to that of the crystallization DNA substrate. In total two DNA substrates were prepared for the EMSA: the Blunt36DNA and the Stagger36DNA substrates. These were prepared by the annealing of oligo 3601 (5'– TTGGTG(9XGGT)GGG –3') and oligo 3602 (5'– CCC(9XACC)CACCAA –3') for the generation of the Blunt36DNA substrate and by the annealing of the oligos 3501 and 3603 (5'–AACCC(9XACC)CACC –3') for the generation of the Stagger36DNA substrate. The annealing of the oligos was done as described for the crystallization DNA, but the oligos were not gel purified.

3.2.3 DNA EMSAs

Prior to the crystallization attempts, EMSAs were performed to verify if the XRCC4¹⁻²⁶⁵:XLF complex was capable of interacting with the crystallization DNA. The first EMSA was prepared by incubating 0.1µM of the DNA substrate with decreasing concentration of the proteins in the EMSA buffer (20mM Tris pH8.0, 50mM KCl, 1mM DTT, 5% glycerol). The gradient started with a high concentration of 17.1µM for each of the proteins and decreased by a factor of 2 for each step-down. The prepared reaction samples were then incubated at room temperature for 30 minutes prior to their loading onto a 5% native polyacrylamide gel and the electrophoresis was done at 100V for 40min.

The visualization of the DNA was done by staining with the GelStar™ Nucleic Acid Gel Stain (Lonza) and exposure to UV light. A second EMSA was done in a similar manner where the concentrations of the proteins were fixed at 4.27μM and the DNA concentration was at 0.1μM. In this EMSA, the DNA was incubated with XRCC4¹⁻²⁶⁵ only, XLF only, or XRCC4¹⁻²⁶⁵ and XLF. Electrophoresis, staining, and visualization was done as previously described.

In addition to this, we have also performed two other EMSAs to further probe for the ability of the XRCC4^{FL} or XRCC4¹⁻²⁶⁵:XLF to interact with a 36 bases long double stranded DNA molecule with either blunt ends or 2 nucleotide overhang ends. Reaction samples were prepared by incubating 0.0625μM of either of the DNA substrates (Blunt36DNA or Stagger36DNA) with XRCC4^{FL} alone, XRCC4¹⁻²⁶⁵ alone, XLF alone, XRCC4^{FL} and XLF, or XRCC4¹⁻²⁶⁵ and XLF. The incubation was done at room temperature for 30 minutes prior to loading of the samples onto a 5% native polyacrylamide gel and electrophoresis at 100V for 40 minutes. Visualization was done by staining with GelStar™ Nucleic Acid Gel Stain (Lonza) and exposure to UV light.

3.2.4 Crystallization

Crystallization attempts for the XRCC4¹⁻²⁶⁵:XLF complex with the crystallization DNA were done by the hanging drop method. First the proteins and the DNA substrate were mixed together to prepare a mastermix solution with concentrations of 100μM XRCC4, 50μM XLF, 37.5μM DNA. The crystallization drops were then set by mixing 1μL of the mastermix with 1μL of a precipitant condition over 800μL of 1.5M (NH₄)₂SO₄. The 96-well crystallization system (Nextal) was then incubated at 18°C until

the growth of crystals were observed. In total 288 conditions were tested using the solutions provided in the The PEGS (Nextal), Classic (Nextal), and CRYOS (Nextal) broadscreen kits. From these, the best three precipitant conditions were selected for further optimization and these were condition 46 (0.1M Tris-HCl pH8.5, 25%(w/v) PEG 8000) from PEGS, condition 47 (0.1M sodium acetate pH4.6, 2.0M sodium formate) from Nextal Classics, and condition 3 (0.17M magnesium chloride, 0.085M Tris pH8.5, 2.89M 1,6-Hexanediol, 15%(v/v) glycerol) from CRYOS. Each of these conditions was repeated for 96 times in the same method as described above and an additional 0.2 μ L of different additive conditions (provided by the additive screen kit from Hampton Research) were added to each drop.

Additionally, crystallization of the XRCC4¹⁻²⁶⁵:XLF without the DNA substrate was attempted by the hanging drop method. The crystallization procedure was done in the same manner as described for the crystallization of this complex with the DNA substrate with a change of the XRCC4¹⁻²⁶⁵ and the XLF concentrations to 150 μ M in the mastermix. In total 192 conditions were tested using the MCSGI and MCSGII Suite broadscreen kits (Microlytic). From these, three conditions were chosen for further optimization and these were the MSGII condition 41 (0.1M Tris-HCl pH8.5, 3.0M Sodium chloride), MSCGII condition 50 (0.1M Bis-Tris Propane-HCl pH7.0, 3.2M Sodium chloride), MSCGII condition 63 (0.1M HEPES-NaOH pH7.5, 3.0M Sodium Chloride). All three conditions were subjected to an additive screen (Hampton Research) in the same manner described previously. Furthermore, protein:precipitant volume ratios were also tested by setting drops with different combination of 1-3 μ L of protein mastermix sample and 1-3 μ L of

each of the three chosen precipitant condition. Lastly, further dehydration of selected crystals was attempted. This was done by, first, allowing the crystals to grow at 18°C for one day, second, allowing it to grow at 4°C for another day, and third, replacing the dehydration solution in the wells with increasing concentration of $(\text{NH}_4)_2\text{SO}_4$, in 0.5M incremental steps per day, until a concentration of 4.0M was achieved. Promising crystals were also screened using the Rigaku X-ray system with a copper rotating anode and the R-AXIS IV area detector plate with a wavelength of 1.5Å and exposure times of 1 minute under a stream of liquid nitrogen.

3.3 Results

3.3.1 Purification Profile of the XRCC4¹⁻²⁶⁵ and XLF Proteins

The XRCC4¹⁻²⁶⁵ protein was purified in two steps: the nickel affinity chromatography and the anion exchange chromatography. As shown in Figure 3.2A, the majority of the XRCC4¹⁻²⁶⁵ protein is eluted off the nickel column at 300mM imidazole at a purity of approximately 80%. Other bands at the molecular weight of XRCC4¹⁻²⁶⁵ are also observed in the flowthrough, 41mM and 62.2mM imidazole washes. The 300mM imidazole fraction collected from the nickel affinity chromatography was applied to a Q-Sepharose column to further purify the protein. By applying a gradient with increasing concentration of salt, the protein is eluted off the column and collected in several 1.5mL fractions. The fractions containing the XRCC4¹⁻²⁶⁵ protein at purity above 95% were pooled together, flash frozen in liquid nitrogen and stored at -80°C until further use. The purification of XLF was done by a nickel affinity chromatography followed by a heparin affinity chromatography. Cell lysate of induced cells were applied to a nickel column and

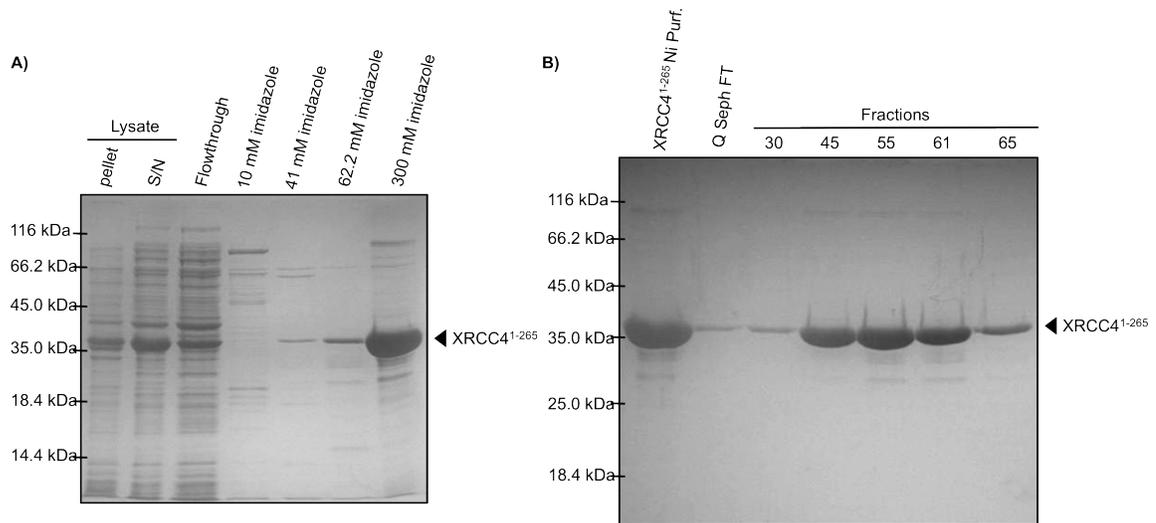


Figure 3.2 Purifications of XRCC4¹⁻²⁶⁵. Panel A, 12% SDS-PAGE of the samples collected during nickel affinity chromatography. Panel B, 12% SDS-PAGE of the samples collected during the Q-sepharose ion exchange purification. The arrowhead marks the band representing the XRCC4¹⁻²⁶⁵ protein.

the column was washed at 48.5mM imidazole prior to the elution of the protein at 96.2mM imidazole (Figure 3.3). Further purification of the protein by heparin affinity chromatography was done to ensure that the protein was at 95% purity (data not shown).

3.3.2 XRCC4¹⁻²⁶⁵:XLF binds to Short DNA Substrates

To better characterize the DNA binding ability of XRCC4 and XLF, an EMSA with 36 nucleotide long double-stranded DNA substrates with either blunt or staggered ends in the presence of XRCC4^{FL/1-265} alone, XLF alone, or XRCC4^{FL/1-265} and XLF was done. At this point, this assay has only been performed once and repetition of the assay is required to properly verify the validity of the results. The results for both DNA substrates were similar. Figure 3.4 shows that the addition of XLF alone and XLF:XRCC4^{FL} resulted in an incomplete shift of the DNA alone bands, as observed by the decreased intensity of the bands and an appearance of a shifted species in the top region of the gel. The shifted species observed in the XLF alone lane was a much darker and sharper band than that observed for the XLF:XRCC4^{FL} lane, which shifted the DNA but resulted in some smearing of the lane. In contrast, the XRCC4^{FL/1-265} alone lane was incapable of shifting the DNA alone band. Finally, the XLF:XRCC4¹⁻²⁶⁵ lane did not show a distinct shifted species but resulted in a decreased intensity of the DNA substrate band.

3.3.3 XRCC4¹⁻²⁶⁵:XLF binds to the Crystallization DNA Substrate

An EMSA was done to verify for the ability of the XRCC4¹⁻²⁶⁵:XLF to bind to the crystallization DNA. This was done to ensure that the crystallization of the protein complex with the DNA was feasible. In this EMSA, the crystallization DNA substrate was incubated with XRCC4¹⁻²⁶⁵ alone, XLF alone, or XRCC4¹⁻²⁶⁵:XLF. Additionally, this

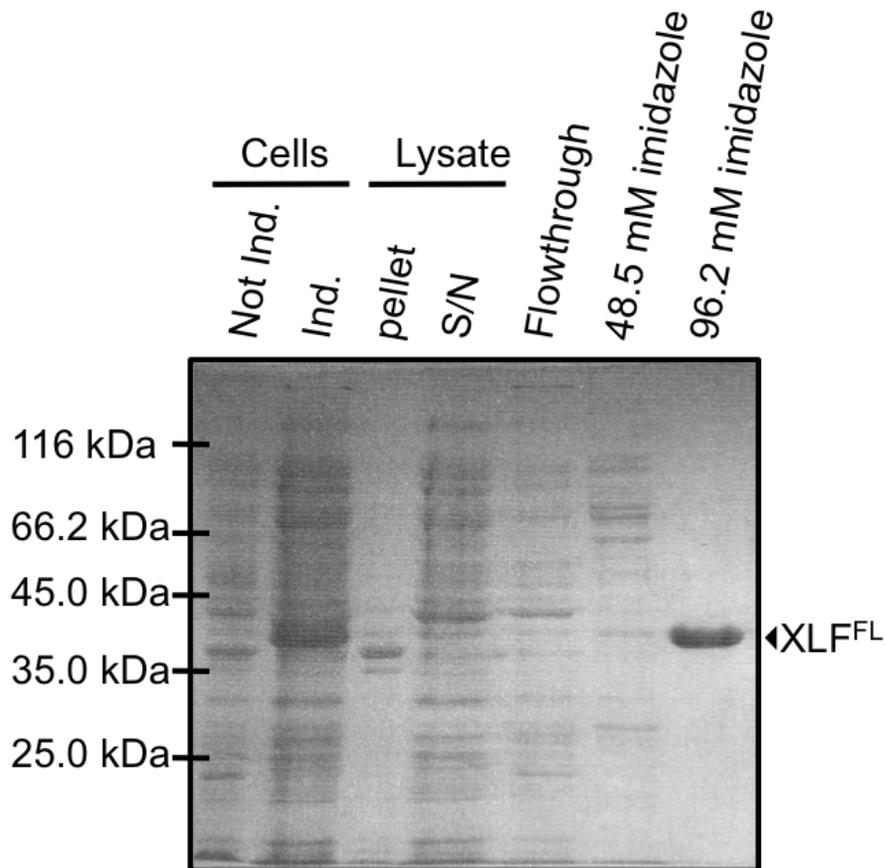


Figure 3.3 Nickel Affinity Chromatography of XLF. 12% SDS-PAGE of the samples collected during nickel affinity chromatography. The arrowhead marks the band representing the XLF protein.

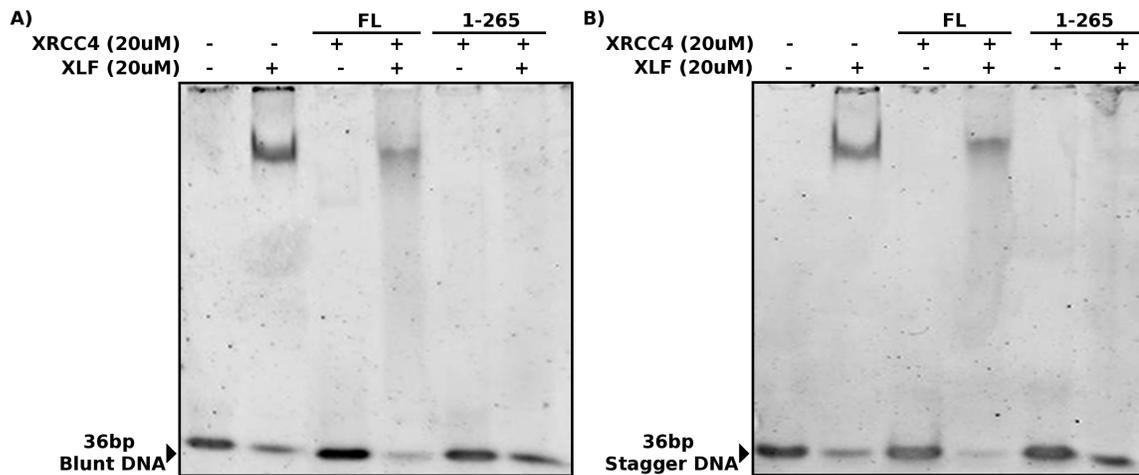


Figure 3.4 XRCC4:XLF complex is able to bind to short DNA substrates. Panel A, 5% non-denaturing polyacrylamide gel used in the EMSA for the 36bp blunt DNA substrate. The DNA substrate (0.0625 μ M) was incubated in the absence or presence of XLF, XRCC4^{FL}, XRCC4¹⁻²⁶⁵ or the presence of XLF:XRCC4^{FL/1-265}. Panel B, 5% non-denaturing polyacrylamide gel used in the EMSA for the 36bp staggered DNA substrate. The DNA substrate was incubated in the same way as it was done for the blunt DNA substrate. Both gels were incubated in Lonza Gel Star stain to detect the position of the DNA or DNA:protein complexes.

DNA substrate was also incubated with increasing concentrations of the XRCC4¹⁻²⁶⁵:XLF complex at equimolar concentrations of XRCC4¹⁻²⁶⁵ and XLF. Figure 3.5B shows that while XLF alone is able to interact with the crystallization DNA substrate, XRCC4¹⁻²⁶⁵ is unable to do so. The XRCC4¹⁻²⁶⁵:XLF resulted in a shifted band with higher intensity than that of the XLF alone. However, the shifts of the DNA only band was incomplete in both cases, where both resulted in DNA only bands of about the same intensity. Figure 3.5C shows the binding of DNA with increasing concentrations of XRCC4¹⁻²⁶⁵:XLF. A shifted species begin to be observed at a total protein concentration of 4.2μM (2.1μM XRCC4¹⁻²⁶⁵ + 2.1μM XLF) and becomes more significant at 8.53μM. At 17.1μM the shifted specie is positioned higher on the gel than observed in the other lanes. Lastly, the proteins concentrations used were not able to fully shift the DNA only band.

3.3.4 Crystallization attempts of XRCC4¹⁻²⁶⁵:XLF and XRCC4¹⁻²⁶⁵:XLF:DNA

Broad screen attempts of the XRCC4¹⁻²⁶⁵:XLF with the crystallization DNA substrate resulted in small crystals with varying morphology (thin rods, clusters of needles, and hexagonal bipyramids) that were not suitable for data collection. Additive screening of the selected conditions (Nextal PEGS #46, Nextal Classics #47, Nextal CRYOS #3) were not able to improve the growth of the crystals and in most cases resulted in no nucleation occurring. Broad screen attempts of the XRCC4¹⁻²⁶⁵:XLF without the DNA proved to be more fruitful than with DNA. Most of the conditions used for screening resulted in the formation of varying sizes of crystals with the hexagonal bipyramidal shape.

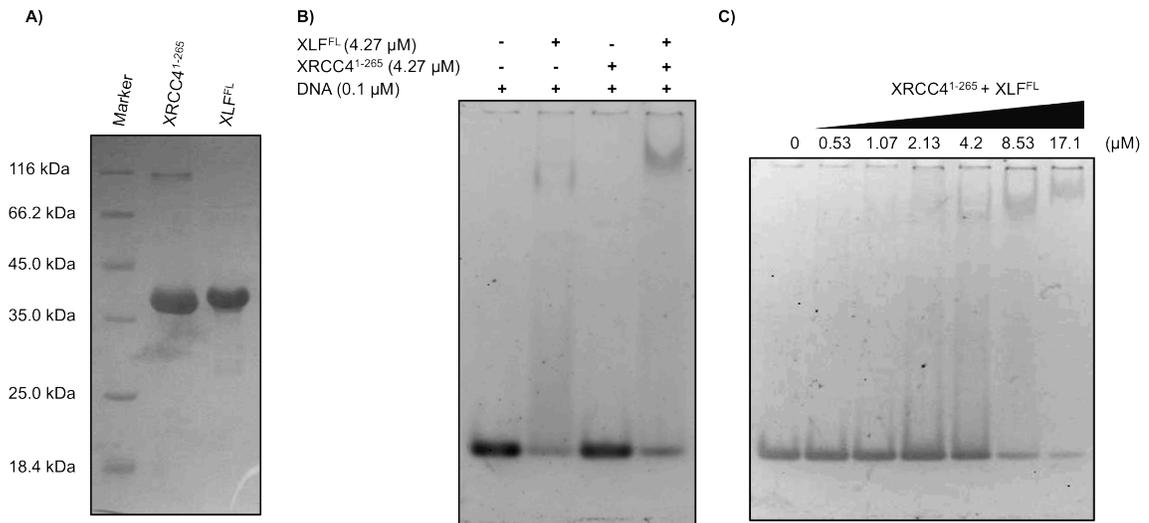


Figure 3.5 XRCC4¹⁻²⁶⁵:XLF binds to the crystallization DNA substrate. Panel A, 12% SDS-PAGE of protein samples used for the EMSAs. Panel B, 5% PAGE performed in the EMSA of the individual proteins and DNA. Presence and absence of the proteins in each lane is indicated by the ± signs. Panel C, 5% PAGE performed in the EMSA for the increasing concentrations of XRCC4¹⁻²⁶⁵:XLF. Proteins were mixed at a 1:1 molar ratio and the concentration shown is the total protein concentration. The concentration of the DNA in both assays was 0.1 μM.

The conditions that generated the best crystals were further manipulated by additive screening, further dehydration, variation of drop ratios, and growth temperature (Table 3.1). Although crystals were grown to a considerable size (Figure 3.6), data collection returned with poor data containing ice rings that were not sufficient for the determination of the crystal structure.

3.4 Discussion

The bridging of DNA molecules by the XRCC4:XLF complex conveys new implications to the function of these proteins in the NHEJ DNA repair pathway. As previously mentioned, both proteins have been shown to be involved in the later steps of the repair by interacting with DNA ligase IV to stimulate its ligation activity towards the DNA ends (Ahnesorg and Jackson, 2006; Grawunder *et al.*, 1997; Lu *et al.*, 2007; Tsai *et al.*, 2007). However, the bridging of the DNA and the early recruitment of these proteins to the damaged site by Ku suggest that they may also play a role in the early stages of the repair by aiding in the bridging of the damaged ends (Yano *et al.*, 2008). In order to study this bridging ability of these proteins, we aimed to better characterize the interaction of the protein complex with the DNA.

Here, we have shown that the XRCC4^{FL}:XLF complex is able to bind to DNA substrates as short as 36 base pairs long (Figure 3.4). Additionally, XLF alone is able to bind to the short DNA substrate with both blunt and staggered ends while XRCC4^{FL} or XRCC4¹⁻²⁶⁵ alone are unable to do so. The blunt ended DNA substrate ensured that DNA molecules would not form concatemers due to the short complementary overhangs that are present in the staggered DNA substrate. Thus, the observed binding to the short DNA

Drop #	BroadScreen Condition	AdditiveScreen Condition	XRCC4	XLF	DNA	Well-Solution	Temp.
1	0.1 M Tris-HCl pH 8.5, 25%(w/v) PEG 8000	---	100 μ M	50 μ M	37.5 μ M	1.5 μ M $(\text{NH}_4)_2\text{SO}_4$	18°C
2	0.1M Sodium Acetate pH4.6, 2.0M Sodium Formate						
3	0.17M Magnesium Chloride, 0.085M Tris pH8.5, 2.89 1,6-Hexanediol, 15%(v/v) glycerol						
4	0.1M Tris-HCl pH8.5, 3.0M Sodium Chloride	1.0M Potassium Chloride	150 μ M	150 μ M	---	1.5 μ M \rightarrow 3.0 μ M $(\text{NH}_4)_2\text{SO}_4$	18°C \rightarrow 4°C
5	0.1M Bis-Tris Propane-HCl pH7.0, 3.2M Sodium Chloride	1.0M Sodium Iodide					
6	0.1M HEPES-NaOH pH7.5, 3.0M Sodium Chloride	0.1M Strontium Chloride hexahydrate					

Table3.1 Summary of conditions for the best-generated crystals. Drops 1 to 3 are the conditions that generated the best three crystals for the crystallization of XRCC4¹⁻²⁶⁵:XLF with DNA; while drops 4 to 6 are the conditions that generated the best three crystals for the crystallization of XRCC4¹⁻²⁶⁵:XLF without DNA. The concentrations are from the solutions used to set the drops and do not account for the dilution factor when setting the trays.

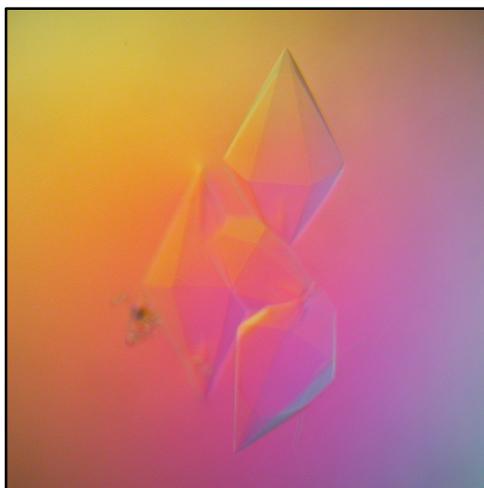


Figure3.6 Representative crystal observed for XRCC4¹⁻²⁶⁵:XLF. The hexagonal bipyramidal crystals observed in most conditions for the crystallization of XRCC4¹⁻²⁶⁵:XLF.

is not due to a potentially long stretch of DNA molecules generated from interaction between the short DNA substrates. These results suggest that XRCC4 alone is unable to bind to DNA substrates as short as 36bps; however with the aid of XLF, the XRCC4:XLF complex is able to bind to short DNA substrates. XLF seems to be the main contributor to this protein complex:DNA interaction. Curiously, EMSA studies from other groups suggest that XLF is only able to bind to DNA substrates longer than 80bps (Lu *et al.*, 2007). In their study, XLF was unable to bind to a DNA substrate as short as 60bps but was able to bind to DNA substrates longer than 83bps. The contradiction between the observed DNA binding behaviours may, in part, be due to the levels of the XLF protein used in the assays. Assuming that the K_d of this interaction is significantly greater than the concentrations used in the EMSAs, the protein:DNA molar ratios would have to be much higher for the interaction to be observed. While in their assay the protein:DNA mol/mol ratio was pushed to a maximum of 10:1, our assay pushed the ratio to 320:1. Considering that to be the case, the K_d of the interaction would be well above the micromolar range indicating that this interaction would likely not be biologically relevant. Additionally, it is interesting to note that the XRCC4¹⁻²⁶⁵:XLF complex was not able to interact with the DNA as effectively as the XRCC4^{FL}:XLF complex. This suggests that the binding of the protein complex to short DNA substrates is not solely dependent on XLF, but that the flexible C-terminal region of XRCC4 may play an important role in this interaction by either directly interacting with the DNA molecule or by stimulating XLF's binding to DNA. In doing so, the C-terminus region of XRCC4 may play a role in the regulation of the XRCC4:XLF DNA binding. In support of this argument, this flexible C-

terminal region of XRCC4 has been shown to contain several potential phosphorylation sites, some of which shown to be phosphorylated by DNA-PK (Yu *et al.*, 2003). These sites when phosphorylated resulted in the disruption of the DNA-bridging ability of the XRCC4:XLF complex as reported by Sunetra *et al.* (2012), indicating that the C-terminus end of XRCC4 is important for the interaction between the XRCC4:XLF complex and DNA.

Alternatively, this peculiar inability of the XRCC4¹⁻²⁶⁵:XLF protein complex to bind to the short DNA substrate while its full length version is able to bind to such DNA substrate could suggest a completely different mode of interaction between the XRCC4:XLF complex with DNA than that observed for the XRCC4 alone or XLF alone proteins with DNA (Fig. 3.4). In this case, XRCC4 is not regulating the ability of XLF to interact with this short DNA substrate but is directly participating in this new mode of XRCC4:XLF complex interaction with the DNA. It is worthwhile to then begin considering these results in context with what has been reported by Andres *et al.* (2011) in regards to the ability of the XRCC4:XLF complex to bind to and bridge long DNA substrates of 1000 bps. In their study, the XRCC4¹⁻²⁶⁵:XLF complex was not only able to interact with a long DNA substrate, but was also able to bridge the DNA molecules. This different behaviour of DNA binding by the protein complex with short and long DNA substrates could be reflective of its mode of interaction with DNA molecules (Table 3.2). Taking into consideration the DNA binding model where the DNA is wrapped around the XRCC4:XLF filament as suggested by Hammel *et al.* (2011), the inability of the XRCC4¹⁻²⁶⁵:XLF complex to bind to the short DNA substrate might hint at the

XRCC4	DNA-Binding				Filament		DNA Bridging
	Short DNA		Long DNA		Intermediate	SuperShift	
	XLF		XLF				
	-	+	-	+			
FL	✗	✓	✓	✓	✗	✓	✓
1-265	✗	✗	✓	✓	?	?	✓
1-157	?	?	✗	✓	✓	✗	✗

Table3.2 Summary of DNA binding for XRCC4 truncations. DNA binding behaviours of different XRCC4 truncations in the presence or absence of XLF. DNA-binding to long DNA substrates, filament formation, and DNA bridging data were obtained from Andres *et al.* (2011). “?” represents situations that have not yet been tested or reported in literature.

importance of the tails in stabilizing the interaction of the complex with the DNA. In this scenario, multiple XRCC4:XLF protein filaments would form a protein bundle via parallel interactions of the filament and the DNA would interact with the bottom of the head regions of both proteins (Hammel *et al.*, 2011). While short DNA substrate would require the further stabilization of the DNA interaction by XRCC4's flexible C-terminal region, long DNA substrates would wrap around this protein filament bundle increasing its affinity towards the XRCC4:XLF filament and, thus, not requiring the flexible C-terminal region of XRCC4 for further stabilization. Curiously, Hammel *et al.* (2011) was not only able to observe binding of XRCC4 alone to a short DNA substrate of 40bps but also able to obtain results by hydrogen/deuterium exchange mass spectrometry as well as SAXS suggesting binding of the 40bps DNA substrate with truncated forms of XRCC4 and XLF lacking their flexible C-terminal region. This discrepancy could be due to the minimum DNA length in which the complex is able to bind to DNA without the aid of the C-terminal flexible region of XRCC4. While 40bps long DNA substrates are able to interact with 4 parallel XRCC4:XLF filaments, our 36bps long DNA substrate would only be able to interact with 3 parallel XRCC4:XLF filament and, thus, unable to strongly interact with this DNA substrate.

Similarly to the Stagger36DNA substrate used in the previous EMSA, the crystallization DNA was a 38-nucleotide long double-stranded DNA molecule with 2 nucleotide overhangs. Despite this, the binding of the XRCC4¹⁻²⁶⁵:XLF complex to this DNA was verified prior to crystallization attempts to ensure its feasibility. As shown in Figure 3.4B, the XRCC4¹⁻²⁶⁵:XLF complex and XLF alone are able to bind to this DNA

substrate while XRCC4¹⁻²⁶⁵ alone is not able to do so. While this result is in agreement with the previous EMSA (Figure 3.3), it differs at the observed ability of the truncated XRCC4:XLF complex to interact with the crystallization DNA substrate but not the 36 nucleotides long DNA substrates. These results suggest that the XRCC4:XLF filament lacking the flexible C-terminal tail of XRCC4 is able to interact with DNA substrate as short as 38 bps long but unable to interact with DNA substrates shorter than 36 bps. Having compared all our results with those reported in literature, it should be noted that these experiments were carried out in different conditions and methods; thus, the observed results could be dependent on the experimental setup rather than the actual nature of DNA binding of the XRCC4:XLF filament. In order to better investigate the DNA binding of the XRCC4:XLF complex, we would need to standardize the DNA binding assay and use the same DNA substrates, protein truncations, and conditions. By doing so, we would be able to directly compare the data to conclusively infer on the binding affinities of the proteins to the DNA. Nevertheless, the data here suggests that the XRCC4¹⁻²⁶⁵:XLF complex is able to bind to the crystallization DNA and the crystallization experiments were then carried out.

Unfortunately, the crystallization attempts for the complex with the DNA were not successful. By performing a broad screen with 288 conditions, we were able to obtain three conditions where nucleation events were observed. These crystals were mostly hexagonal bipyramidal in shape and they did not grow to larger sizes with longer dehydration times of the drops. Further optimization of these conditions with additive screens proved to be unsuccessful with no apparent aid in growth or nucleation and, in

some cases, resulted in the reduction of the nucleation event. In contrast, the crystallization attempts for the XRCC4¹⁻²⁶⁵:XLF complex were found to be more promising. The crystallization of this complex without the DNA could still provide new insights on the C-terminus flexible region of XRCC4 and XLF, as these regions have never been crystallized before. By performing a broad screen of 192 conditions with this protein complex, we found many conditions that generated crystals. The top 3 conditions were selected based on the nucleation levels (lower nucleation events), and the size and sharpness of the crystals. Further optimization of these conditions included additive screening, changes in temperature for the growth of the crystals, further dehydration of the crystals, and small variations of the concentration of the compounds and pH found in the selected broad screen solutions. Table 3.1 summarizes the conditions found to produce the crystals with the greatest size and sharpness.

The data collection of these crystals yielded poor data with no visible spots and evidence for the presence of ice in the crystals. By manipulating the compounds in condition 4 of Table 3.1 to 2.8M NaCl, 0.1M Tris pH8.0, and 20% (v/v) glycerol with no additive screen, we were able to improve the collection of the data to the point where a spot was seen at 8.18Å and ice rings in the data were significantly decreased. At this point, the conditions must be further optimized before we are able to obtain data of sufficient quality to solve the crystal structure. The optimization must be done based on a direct collection of data because, in this case, the size and sharpness of the crystals does not correlate with an improvement of the collected data.

CHAPTER 4 – CONCLUDING REMARKS

4.1 Summary and Implications to NHEJ

The NHEJ field of research has grown very rapidly in the past two decades and many pathways and their respective players have been discovered in this period. Each of these proteins in the repair pathways have, in turn, been associated with many potential functions involving a complex interplay between each of the DNA repair proteins themselves and the DNA. Additionally, the DNA repair pathways also communicate with many other factors associated with the DNA damage response to ensure a holistic cellular response towards a potentially critical damage to the DNA (Petrini and Stracker, 2003). Amidst all the complex network of processes involved in this repair, we have chosen to study the NHEJ pathway, as it is the predominant mode of repair in mammals (Hoeijmakers, 2001). More specifically, my master's work was mostly invested in understanding the tetramerization mode of the DNA repair protein XRCC4 as well as its interaction with its structurally related partner XLF and also DNA molecules.

Here we have shown that mutations of residues at the head region and of conserved residues at the N-terminal region of the tails of XRCC4 do not impact its ability to interact with long DNA molecules, nor does it affect its ability to bind to DNA ligase IV BRCTs region. Furthermore, we have ensured that the observed behaviours for the mutants were not due to gross changes in the structure of the protein by performing circular dichroism. The CD data suggested that the XRCC4 proteins' secondary structure are in general arranged as 34% α -helices, 13% β -sheets, 20% turns, and 33% unordered. With the functional aspect of the mutants verified, the oligomerization states of these mutants were studied by analytical ultracentrifugation. Unfortunately, good quality data

for these mutants have not yet been obtained due to experimental difficulties, such as the presence of large aggregates in the sample after the analysis. However, the data collected thus far suggest for the disruption of the tetramerization with specific mutation at both the head and the upper tail region of XRCC4. If these data can be verified, this would indicate a head-to-head tetramerization mode for the XRCC4 proteins without excluding the tail-to-tail tetramerization mode of the protein. It then becomes interesting to imagine what the possible implications of a head-to-head tetramerization would bring. The residues E121 and E125 are in the same interface shared between the XRCC4:XLF interaction. Thus, a head-to-head tetramerization would not be physically feasible to occur during a XRCC4:XLF interaction. Assuming that the problems with the AUC data are resolved, we can test this by performing AUC, or alternatively by SAXS, in the presence or absence of XLF and detect for the formation of tetramers in the solution. Another interesting implication to this head-to-head interaction is the folding back of the tails of XRCC4 to further stabilize this tetramerization mode. As the R150E mutation shows evidence for the formation of mostly dimers, it suggests that this conserved residue at the upper region of the tails is important for the tetramerization. A potential reason as to why this residue could be important is through the anchoring and guiding of the flexible C-terminal region towards the head of the XRCC4. If that is the case, the observed disruption of the tetramerization by mutations at the CTR tail region of XRCC4 can be explained as a prevention of the folding back of the flexible C-terminal region of XRCC4, and thus a weaker head-to-head interaction that is not stabilized by the flexible C-terminal region.

Besides the potential implications to the NHEJ pathway, we have potentially generated mutants that only affect the tetramerization of XRCC4 without affecting its other functions. Due to the fact that many functions are affected with mutations in the CTR tail region of XRCC4, many assays involving these mutants are difficult to interpret in regards to the actual role of XRCC4 in the phenomena observed. Thus, the mutants we have generated will allow for future studies on the effects of tetramerization alone to the bridging of the DNA and the importance of the tetramerization to the overall repair of DNA.

Additionally, we have shown that the XRCC4:XLF complex and XLF alone are able to bind to DNA molecules as short as 36 base pairs, albeit with a weak affinity towards it. The interaction of XLF with the short DNA substrate contradicts what has been reported by another study in the field but such contradiction is likely to be due to the difference in concentrations of protein to DNA used in the EMSAs (Yu *et al.*, 2003). In our EMSA, we have pushed the protein:DNA concentration ratio much higher than what they have used, possibly explaining the difference in our observation of the XRCC4:XLF binding to 36 bps DNA molecules against their observation of no binding towards a 60bps DNA molecule. Nevertheless, the observed interaction is a good indication that the pursuit of the crystallization of the XRCC4:XLF complex with short DNA molecules is feasible. In fact by attempting to crystallize this proteins:DNA complex with short DNA substrates, we promote a tighter and better packing of the molecules, which in turn will result in crystals of higher quality for the solution of the structure. Unfortunately, such attempts have so far been unsuccessful and only the protein complex without the DNA

has resulted in the formation of large crystals. At the same time, this drastic change in crystallographic behaviour of the protein complex and the proteins:DNA complex suggest that the DNA is interacting with the proteins and thus preventing the normal formation of the protein crystal. Therefore, continuing with the crystallization with a similar or equal DNA molecule should eventually result in the formation of crystals with the DNA incorporated in them rather than the proteins alone.

By obtaining a better understanding in both the tetramerization of XRCC4 and the interaction between the XRCC4:XLF complex and the DNA, we can more appropriately investigate the DNA bridging ability of the XRCC4:XLF. Andres *et al.* (2012) showed that in the cases with XRCC4 or XLF mutants unable to bind to DNA, the XRCC4:XLF complex was unable to bridge DNA molecules. In their study, the importance of XLF's DNA binding to both bridging and the interaction of the protein complex to DNA was clearly demonstrated. While they have used DNA-binding mutants of XRCC4 to show that XRCC4's DNA binding was important for the bridging function but not the interaction of the protein complex with the DNA, these mutants contained mutations in or lacked the CTR tail region of XRCC4 that have previously been shown to be also important for BRCT binding and the homo-tetramerization of the protein. Thus, the interpretation of XRCC4's DNA-binding being important for the bridging could potentially be misleading as tetramerization of this protein might also have been affected in these mutants. It is possible that while XLF is important for the protein complex binding to DNA, XRCC4's tetramerization and not DNA binding is critical for the bridging to occur. Elucidating the structural-functional relationship of XRCC4 will allow

us to generate mutants affecting only one function of XRCC4 to clarify many results and build better mechanistic models of how XRCC4 aids in the repair of DSBs.

Taking all the information together, it is tempting to form speculations on the overall role of XRCC4 in the NHEJ pathway. (Figure 4.1) As previously mentioned, this pathway is divided into three major steps – the recognition and protection of the damaged ends, the processing of these ends, and finally the ligation of the strands. In the first step, Ku would recognize the DNA damage and recruit XRCC4, XLF and DNA-PKcs (Costantini *et al.*, 2007; Gottlieb and Jackson, 1993; Yano *et al.*, 2008). Following this, XRCC4 and XLF would form a protein filament to hold the damaged DNA ends together and protect it from further degradation. In doing so, Ku would have to translocate along the DNA molecule and away from the damaged site in order to physically allow for this protein filament to form. In fact, Ku has indeed been shown to be able to translocate along a DNA molecule without an ATP requirement by de Vries *et al.* (1989). At this point, multiple XRCC4/XLF protein filaments could be recruited to form a protein filament bundle and further strengthen the synapsis of the DNA ends. Next, DNA-PK would recruit DNA end-processing enzymes and at the same time phosphorylate the XRCC4:XLF complex to promote the complete or partial disruption of the protein filament bundle (Andres *et al.*, 2012; Sunetra *et al.*, 2012). In this manner, DNA-PK could promote the access to the damaged sites by DNA-end processing enzymes. Finally, DNA Ligase IV would act to ligate the processed DNA ends and complete the repair of the DSB.

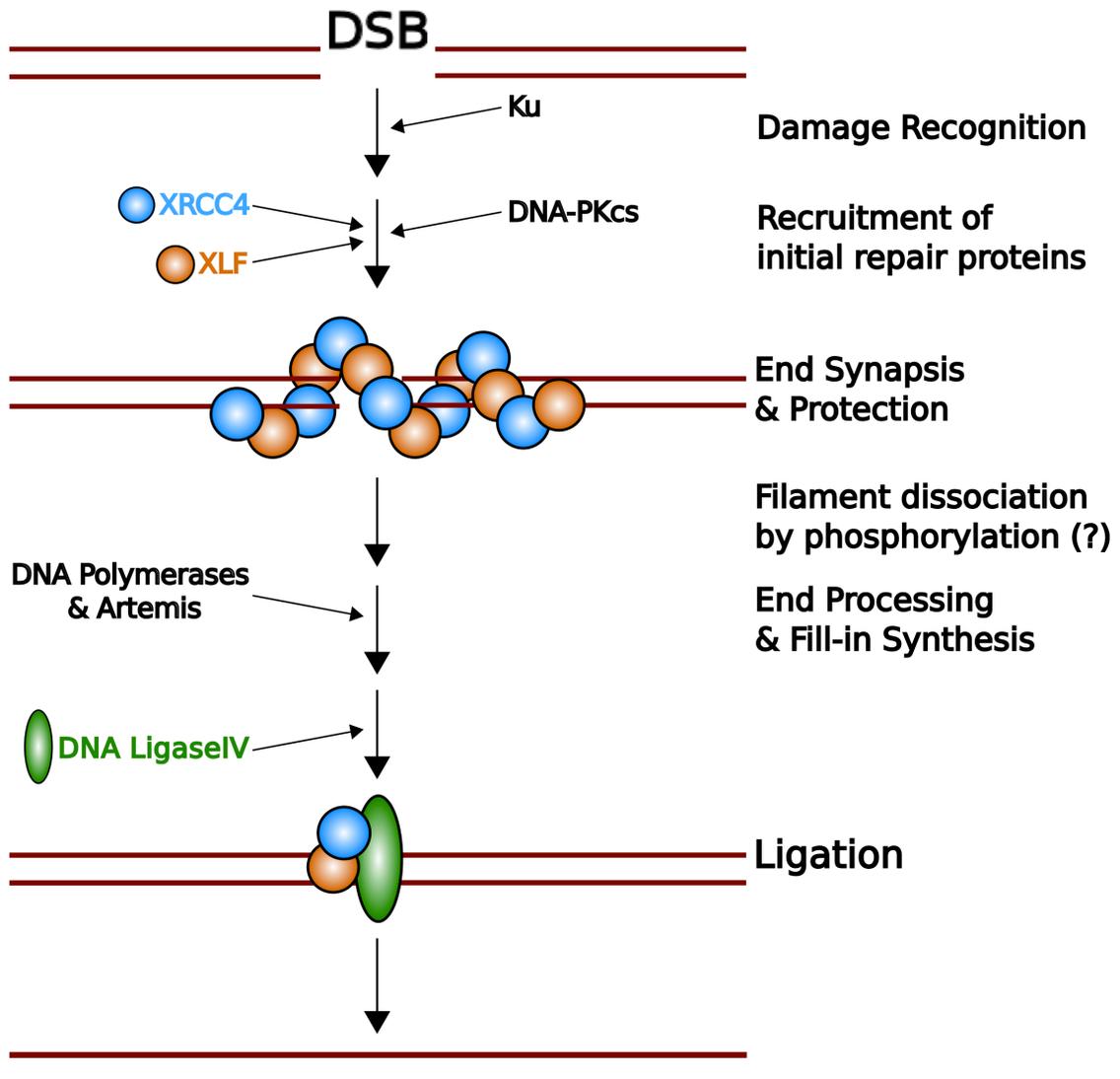


Figure4.1 **The possible roles of XRCC4 in the NHEJ pathway.** Recent findings have provided some evidence suggesting an early role, in addition to the already established late role, of XRCC4 in the NHEJ. In the early role, XRCC4 can interact with XLF to form a protein filament (or a bundle of this filament) to synapse and protect the damaged DNA ends. The figure above is one model of how this filament interacts with the DNA, where the DNA treads through the inner section of the filament. In the second model of how this interaction is occurring, the DNA wraps around the coiled filament bundle. In the late role, XRCC4 interacts with XLF and DNA Ligase IV to ligate the ends together.

4.2 Future Directions

The first task that must be completed is the clarification of the issues involving the large species found in the analytical ultracentrifugation experiments. As previously mentioned, this should be investigated through size exclusion chromatography of the XRCC4 samples prior to and after a freeze and thaw process of the samples. If the results returns as no formation of aggregation due freeze/thaw, the next attempt should be a more stringent purification procedure of the protein samples by possibly subjecting the purified samples to a size exclusion chromatography to remove the large aggregates from the sample. Although not ideal, alternatives to studying the oligomerization state of the mutants could be employed. Unfortunately, size exclusion chromatography have been demonstrated by other groups and through our own efforts to not be enough to resolve the dimer from the tetramer form of the protein, but another alternative would be to attempt cross-linking experiments. Cross-linking experiments could provide us with informative data if employed with the proper positive and negative controls and also with careful use of the cross-linking agent to not induce artificial protein-protein interactions. Lastly, SAXS analysis could be done in a similar way as was described by Hammel *et al.* (2010). As this technique have provided recent evidence for this mode of tetramerization, subjecting these mutants to the same experimental procedures would be of great interest. By doing this, one could directly compare the results to determine if the mutants have affected the proposed head-to-head tetramerization observed by SAXS and verify the validity of this mode of tetramerization.

The structural study of XRCC4:XLF:DNA complex should also be pursued as our results have shown promising potential for the crystallization of this complex. We have demonstrated that the XRCC4:XLF complex is able to bind to DNA substrate as short as 36bps and that this binding occurs at a higher affinity with the full length XRCC4 than its 1-265 truncated form. Furthermore, the presence of DNA have shown to dramatically decrease the nucleation of the XRCC4:XLF complex suggesting that this complex is attempting to incorporate the DNA substrate. With this in mind, the crystallization should be pursued by using full length XRCC4 and XLF in the presence of short DNA substrates. Additionally, the short DNA substrates could be generated in a variety of lengths with complementary overhangs to allow the DNA substrate to accommodate for different conformations as it finds to be necessary within the crystal. A huge limiting step in crystallization is finding the suitable protein boundaries and DNA lengths; thus, taking a broader approach to this crystallization by utilizing varying lengths of the DNA substrates will likely reduce the time required to find the appropriate DNA length for crystallization. Finally, using XRCC4 and XLF mutants with the known phosphorylation sites mutated to alanine could prove to be useful in the crystallization attempts of these protein with DNA as the phosphorylation of these proteins have been shown to negatively affect their binding to DNA. (Sunetra *et al.*, 2012) By using these mutants, we would prevent any phosphorylation of these proteins from occurring and, in this way, ensure that these phosphorylation are not present to disrupt the interaction between the XRCC4:XLF complex with DNA during the crystallization attempts.

Furthermore, these recent studies in the field also suggest important functions for the C-terminal flexible region of XRCC4. This region has been shown to contain nine potential phosphorylation sites that have been proposed to be targeted by DNA-PKcs (Yu *et al.*, 2003). This phosphorylation of XRCC4 was shown to affect its ability to bind to DNA as well as to disrupt the DNA-bridging ability of the XRCC4:XLF complex (Sunetra *et al.*, 2012). Given these recent findings, investigating the effects of the phosphorylation of the C-terminal flexible region in respect to the overall structure and the tetramerization of XRCC4 would be an interesting venue to pursue, as there is amounting evidence for its role in the regulation of XRCC4's functions. As the functions of XRCC4 becomes better characterized, the next step to the understanding of its overall role in the pathway will likely be the elucidation of how this protein's functions are regulated during the course of DNA repair.

A greater understanding of XRCC4 will aid in the mechanistic elucidation of the NHEJ pathway. As this is the major repair pathway in higher eukaryotes, it will give us a more comprehensible grasp on the repair process of double-stranded breaks in humans. The underlying goal being that this knowledge can be used in the treatment of human disorders associated with deficient repair of DNA. Such disorders could range from those directly involved with a deficient NHEJ, such as specific cases of immunodeficiency, to those where manipulation of DNA repair could be employed to aid in the treatment of the disorder, such as in certain cases of cancer treatment where apoptosis of cancerous cells is induced by massive DNA damage. Before this can occur, much more work remains to be done to better characterize, not only, XRCC4, but also the NHEJ pathway as a whole.

REFERENCES:

- 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* **124**: 301-313.
- Andres, S.N., Modesti, M., Tsai, C. J., Chu, G., and Junop, M.S. (2007) Crystal structure of human XLF: A twist in non-homologous DNA end-joining. *Mol. Cell*. **28**: 1093-1101.
- Andres, S.N., Vergnes, A., Ristic, D., Wyman, C., Modesti, M., and Junop, M.S. (2012) A human XRCC4-XLF complex bridges DNA. *Nucl. Acid Res.* **40**: 1868-1878.
- Asherie, N. (2004) Protein crystallization and phase diagrams. *Methods* **34**: 266-272.
- Bartek, J., and Lukas, J. (2007) DNA damage checkpoints: from initiation to recovery or adaption. *Curr. Opin. Cell Biol.* **19**: 238-245.
- Borek, C., and Sachs, L. (1966) *In vitro* cell transformation by X-irradiation. *Nature* **210**: 276-278.
- Breimer, L.H., and Lindahl, T. (1985) Thymine lesions produced by ionizing radiation in double-stranded DNA. *Biochemistry*, **24**: 4018-4022.
- Breimer, L.H., and Lindahl, T. (1985) Enzymatic excision of DNA bases damaged by exposure to ionizing radiation or oxidizing agents. *Mutat. Res.* **150**: 85-89.
- 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 unites and skin fibroblasts in human autosomal recessive severe combined immunodeficiency. *J Clin Invest* **91**: 1214-1218.
- Chan, D.W., and Lees-Miller, S. (1996) The DNA-dependent protein kinase is inactivated by autophosphorylation of the catalytic subunit. *J. Biol. Chem.* **271**: 8936-8941.
- Chance, B., Sies, H., and Boveris, A. (1979) Hydroperoxide metabolism in mammalian organs. *Phys. 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 non-homologous end joining. *EMBO J.* **21**: 2827-2832.
- Chun, J.T., Kim, S.A., and Chu, G. (2007) Cernunnos/XLF promotes the ligation of mismatched and noncohesive DNA ends. *PNAS* **104**: 7851-7856.
- Costantini, S., Woodbine, L., Andreoli, L., Jeggo, P.A., and Vindigni, A. (2007) Interaction of the Ku heterodimer with DNA ligase IV/XRCC4 complex and its regulation by DNA-PK. *DNA Repair (Amst)* **6**: 712-722.

Dianov, G.L., O'Neil, P., and Goodhead, D.T. (2001) Securing genome stability by orchestrating DNA repair: removal of radiation-induced lesions in DNA. *Bioessays* **23**: 745-749.

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

Downs, J.A., and Jackson, S.P. (2004) A means to a DNA end: The many roles of Ku. *Nature Review Mol. Cell Biol.* **5**: 367-378.

Drenth, J. (2007) *Principles of protein X-ray Crystallography*. (3rd ed.) New York, NY: Springer.

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.

Falck, J., Coates, J. & Jackson, S. P. (2005) Conserved modes of recruitment of ATM, ATR and DNA-PKcs to sites of DNA damage. *Nature* **434**: 605–611.

Ghirlando, R. (2011) The analysis of macromolecular interactions by sedimentation equilibrium. *Methods* **54**: 145-156.

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.

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

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

Grawunder, U., Zimmer, D., Fugmann, S., Schwarz, K., and Lieber, M.R. (1998a) 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.

Grawunder, U., Zimmer, D., and Lieber, M.R. (1998b) DNA ligase IV binds to XRCC4 via a motif located between rather than within its BRCT domains. *Curr. Biol.* **8**:873-876.

Gu, J., Lu, H., Tippin, B., Shimazaki, N., Goodman, M.F., and Lieber, M.R. (2007) XRCC4:DNA ligase IV can ligate incompatible DNA ends and can ligate across gaps. *EMBO J.* **26**: 1010-1023.

Gu, X.Y., Weinfeld, M.A., and Povirk, L.F. (1998) Implication of DNA-dependent protein kinase in an early, essential, local, phosphorylation event during end-joining of DNA double-stranded breaks in vitro. *Biochemistry* **37**: 9827-9835.

Hammarsten, O., and Chu, G. (1998) DNA-dependent protein kinase: DNA binding and activation in the absence of Ku. *Proc. Natl. Acad. Sci. USA* **95**: 525-530.

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

Hammel, M., Rey, M., Yu, Y., Mani, R.S., Liu, M., Pique, M.E., Fang, S., Mahaney, B.L., Weinfeld, M., Schriemer, D.C., Lees-Miller, S.P., Tainer, J.A. (2011) XRCC4 protein interactions with Xrcc4-like factor (XLF) create an extended grooved scaffold for DNA ligation and Double Strand Break Repair. *J. Biol. Chem.* **286**: 32638-32650.

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

Helleday, T., Lo, J., van Gent, D.C., and Engelward, B.P. (2007) DNA double-stranded break repair: from mechanistic understanding to cancer treatment. *DNA Repair Amst* **6**: 923-935.

Hoeijmakers, J. H. (2001) Genome Maintenance Mechanisms for Preventing Cancer. *Nature* **411**: 366-374.

Hutchinson, F. (1985) Chemical changes induced in DNA ionizing radiation. *Prog. Nucleic Acid Res. Mol. Biol.* **32**:115-154.

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

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.

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

Li, Y. *et al.* (1995) Dilated cardiomyopathy and neonatal lethality in mutant mice lacking manganese superoxide dismutase. *Nature Genet.* **11**: 376-381.

Li, Y., Chirgadze, D.Y., Bolanos-Garcia, V.M., Sibanda, B.L., Davies, O.R., Ahnesorg, P., Jackson, S.P., and Blundell, T.L. (2008) Crystal structure of human XLF/Cernunnos reveals unexpected differences from XRCC4 with implications for NHEJ. *EMBO J.* **27**: 290-300.

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., Lieber, M.R. (2002) Hairpin opening and overhang processing by an Artemis/ DNA-dependent protein kinase complex in nonhomologous end joining and V(D)J recombination. *Cell* **108**: 781–794.

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.

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

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

Mimori, T. and Hardin, J.A. (1986) Mechanism of interaction between Ku protein and DNA. *J. Biol. Chem.* **268**: 10546-10552.

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* **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 ligase IV interaction of the DNA double-strand break repair protein XRCC4 are mutually exclusive. *J. Mol. Biol.* **334**: 215-228.

Moshous, D., Callebaut, I., de Chasseval, R., Corneo, B., Cavazzana-Calvo, M., Le Deist, F., Paull, T. T., Rogankou, E.P., Yamazaki, V., Kirchgessner, C.U., Gellert, M., and Bonner, W.M. (2000) A critical role for histone H2AX in recruitment of repair factors to nuclear foci after DNA damage. *Curr. Biol.* **10**: 886–895.

Petrini, J.H., and Stracker, T.H. (2003) The cellular response to DNA double-strand breaks: defining the sensors and mediators. *Trends Cell Biol.* **13**: 458-462.

Riballo, E., Woodbine, L., Stiff, T., Walker, S.A., Goodarzi, A.A. and Jeggo, P.A. (2008) XLF-Cernunos promotes DNA ligase IV-XRCC4 re-adenylation following ligation. *Nucl. Acids. Res.* **37**: 482-492.

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.

Rothkamm, K., Kruger, I., Thompson, L.H. (2003) and Lobrich, M. Pathways of DNA double-strand break repair during the mammalian cell cycle. *Mol Cell Biol.* **23**: 5706-5715.

Schuck, P. (2000) Size-Distribution Analysis of Macromolecules by Sedimentation Velocity Ultracentrifugation and Lamm Equation Modeling. *Biophysical Journal* **78**: 1606-1619.

Sunetra, R., 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. *Nucl. Acid Res.* **40**: 1684-1694.

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 double-stranded DNA. *Proc. Natl. Acad. Sci. USA* **91**: 6904-6908.

Takata, M., Sasaki M.S., Sonoda, E., Morrison, C., Hashimoto, M., Utsumi, H., Yamaguchi-Iwai, Y., Shinohara, A., and Takeda, S. (1998) Homologous recombination and non-homologous end-joining pathways of DNA double-stranded break repair have overlapping roles in the maintenance of chromosomal integrity in vertebrate cells. *EMBO J.* **17**: 5497-5508.

Tezcan, I., Sanal, O., Bertrand, Y., Philippe, N., Fischer, A., de Villartay, J.P. (2001) Artemis, a novel DNA double-strand break repair/V(D)J recombination protein, is mutated in human severe combined immune deficiency. *Cell* **105**: 177-186.

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. U.S.A.* **104**: 7851-7856.

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

Ward, J.F. & Kuo, I. (1976). Strand breaks, base release, and postirradiation changes in DNA y-irradiated in dilute O₂-saturated aqueous solution. *Radiat. Res.* **66**: 485-489.

Wu, P.Y., Frit, P., Meesala, S., Dauvillier, S., Modesti, M., Andres, S.N., Huang, Y., Sekiguchi, J., Calsou, P. Salles, B., Junop, M.S. (2009) Structural and functional interaction between the human DNA repair proteins DNA ligase IV and XRCC4. *Mol. Cell. Biol.* **11**: 3163-3172.

Yaneva, M., Kowalewski, T., and Lieber, M.R. (1997) Interaction of DNA-dependent

protein kinase with DNA and with Ku: biochemical and atomic-force microscopy studies. *EMBO J.* **16**: 5098-5112.

Yano, K., Morotomi-Yano, K., Wang, S., Uematsu, N., Lee, K.J. (2008) Ku recruits XLF to DNA double-strand breaks. *EMBO J.* **9**: 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 inwards translocation of Ku protein. *Nucleic Acids Res.* **27**: 4679-4686.

Yu, Y., Wang, W., Ding, Q., Ye, R., Chen, R., 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., and 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.