

XRCC4 AND XLF IN MAMMALIAN DNA DOUBLE-STRAND BREAK REPAIR

FUNCTIONAL STRUCTURES:
THE ROLE OF XRCC4 AND XLF IN MAMMALIAN
DNA DOUBLE-STRAND BREAK REPAIR

By
SARA N. ANDRES, B.Sc.

A Thesis
Submitted to the School of Graduate Studies
in Partial Fulfillment of the Requirements
for the Degree
Doctor of Philosophy

McMaster University

©Copyright by Sara N. Andres, September 2011

Descriptive Note

DOCTOR OF PHILOSOPHY (2011)

McMaster University

(Biochemistry and Biomedical Sciences)

Hamilton, Ontario

TITLE: Functional Structures: The Role of XRCC4 and XLF in Mammalian DNA Double-Strand Break Repair

AUTHOR: Sara Nicole Andres, B.Sc. (University of Guelph)

SUPERVISOR: Dr. Murray S. Junop

NUMBER OF PAGES: xvi, 161

Abstract

DNA double-strand breaks pose a serious threat to genomic integrity. Double-strand breaks can cause chromosomal rearrangement, leading to uncontrolled cell proliferation, or even cell death. However, mammalian systems have in place the non-homologous end-joining pathway for repair of DNA double-strand breaks, which requires a core group of proteins to function: Ku70/80, DNA-PKcs, and Artemis for recognition, protection, and processing of the DNA ends, and XLF, XRCC4, and DNA LigaseIV for ligation of the DNA break. The work presented here focuses on the specific roles of XLF and XRCC4 within non-homologous end-joining. Initially, the structure of the N-terminal 224 residues of XLF was determined and found to consist of a head and tail domain, structurally homologous to XRCC4. Furthermore, L115 of XLF and K63, K65 and K99 of XRCC4 were identified as essential for an interaction between both proteins. This interaction was then shown to be required for stimulating ligation of mismatched DNA ends. To further understand how XRCC4 and XLF enhance LigaseIV activity, an XRCC4-XLF complex was crystallized. Truncated XRCC4 (1-157) was co-crystallized with truncated XLF (1-224), grown under conditions of decreasing temperature and increasing dehydration. The resulting structure at 3.94Å confirmed the necessity of L115 (XLF) and K63, K65 and K99 (XRCC4) to the XRCC4-XLF interaction, but also illustrated that XRCC4-XLF exists as an extended helical filament. DNA binding regions in both XRCC4 and XLF were also identified and used to construct a structural XRCC4-XLF-DNA binding model. Interestingly, XRCC4-DNA binding occurs in the same region of XRCC4 required for homo-tetramerization and binding to LigaseIV. These results culminate in a proposed model of non-homologous end-joining where XRCC4-XLF is involved not only in ligation of the double-strand break, but also in initial protection of the DNA ends.

Acknowledgements

I wish to express my deepest thanks to my supervisor, Dr. Murray Junop, whose eternal optimism and enthusiasm for science continually encouraged me to persevere. I am so thankful he convinced me to remain on as a Ph.D. candidate – it's a choice I do not regret. I am truly grateful for his willingness to teach and discuss science, anytime, anywhere, as it has contributed greatly to my development as a scientist.

I would like to thank my committee members, Dr. Justin Nodwell and Dr. Paul Berti for their support and guidance of this project, which was much appreciated. I am grateful for their discussions and advice, which helped drive this project forward.

I am indebted to all the members of the Junop lab past and present, for their support and friendship – they have become like a family to me. I would particularly like to thank Margaret Sulek, Dr. Sean Jackson, Seiji Sugiman-Marangos, Tracy Tiefenbach and Mac Mok for their advice, willingness to listen, and ability to make me laugh especially on days when the research just wasn't working. To the students I had the privilege of mentoring, thank-you for all your patience while we learned together and for all your hard work. A special thank-you to Rishi Parihar and Margarita Blajeva for their work, which has contributed directly to the exciting results in this thesis. I would also like to thank Dr. Alba Guarné and her lab for their support and for being great lab neighbours.

I also wish to express my love and gratitude to my family and friends for all of their support, encouragement, and belief that I can accomplish anything. I could not have achieved this degree without all of you. To my parents, Fred and Patricia Andres, I love you both for always being there for me and supporting me in whatever path I choose. I

am so thankful that you taught me the value of hard work – it has helped me accomplish so much. To my brother and sister-in-law, Jason and Leanne Andres, thank-you for your love and reassurance (and for keeping my car so clean!). To my grandmothers Eva Andres and Helen Petryschuk, and my departed grandfathers, William Andres and William Petryschuk, thank-you for your love and encouragement (and cookies!) and for being remarkable examples of strength and grace when faced with life’s challenges. To my extended family, Andy, Claire, Marc and Eric Petryschuk, thank-you for letting me into your home when I first started out as a graduate student and for your constant support throughout the years. To the Guelph girls, I am so grateful for your friendship, and interest in my research. Also, a special thank-you to YuSun Chung and Andrew Gadsden, to whom I owe so much – thank you for keeping me sane and for being the best roommates ever.

Finally, I wish to thank the Lord, my God, through which all things are possible.

Table of Contents

Abstract	iv
Acknowledgements	v
Table of Contents	vii
List of Figures	xii
List of Tables	xiv
Abbreviations	xv

Chapter 1: Introduction	1
--------------------------------	----------

1.1 DNA Double-Strand Breaks	2
1.2 DNA Double-Strand Break Damage Response	4
1.3 A Brief Overview of Non-Homologous End-Joining	6
1.3.1 Initiating Non-Homologous End-Joining by Ku70/80 and the DNA-Dependent Protein Kinase Catalytic Subunit	8
1.3.2 Processing the DNA Double-Strand Break	16
1.3.3 Ligation of the DNA Double-Strand Break – XRCC4, LigaseIV and XLF	20
1.4 Thesis Objectives and Organization	30

Chapter 2: Crystal Structure of Human XLF: a Twist in Non-	
---	--

Homologous DNA End-Joining	31
-----------------------------------	-----------

2.1 Author's Preface	31
2.2 Summary	32
2.3 Introduction	32

2.4	Results and Discussion	34
2.4.1	Crystal Structure of XLF ¹⁻²²⁴ Shows Formation of a Stable Dimer	34
2.4.2	The XLF C-Terminal Domain is Required for DNA Binding and Mismatched End Ligation	39
2.4.3	XLF and XRCC4 Structures Exhibit Similarities and Differences	42
2.4.4	Model for the XLF-XRCC4-LigaseIV Complex	45
2.5	Experimental Procedures	49
2.5.1	Protein Expression and Purification	49
2.5.2	Electrophoretic Mobility Shift Assays	49
2.5.3	2D Gel Analysis of Protein-Protein Interactions	49
2.5.4	Crystallization and Data Collection of XLF ¹⁻²²⁴	49
2.5.5	Structure Determination and Model Refinement	50
2.5.6	In vitro End-Joining Reactions	50
2.6	Acknowledgements	50
2.7	Supplemental Data	51
2.7.1	Supplemental Experimental Procedures	54

Chapter 3: Crystallization and Preliminary X-ray Diffraction Analysis

<u>of the Human XRCC4-XLF Complex</u>		56
3.1	Author's Preface	56
3.2	Synopsis	57
3.3	Abstract	57
3.4	Introduction	57
3.5	Materials and Methods	59
3.5.1	Molecular Cloning and Protein Expression	59
3.5.2	Purification	60

3.5.3 General Crystallization	61
3.5.4 Crystallization and Diffraction Collection	61
3.6 Results and Discussion	62
3.7 Acknowledgements	67

Chapter 4: Crystal Structure of Human XRCC4-XLF Reveals an

Extended Helical Filament Important for Bridging DNA Ends **68**

4.1 Author's Preface	68
4.2 Summary	69
4.3 Introduction	69
4.4 Results	72
4.4.1 Structure of the XRCC4 ¹⁻¹⁵⁷ – XLF ¹⁻²²⁴ Complex	72
4.4.2 XRCC4 ¹⁻¹⁵⁷ – XLF ¹⁻²²⁴ Forms a Protein Filament	75
4.4.3 Identification of DNA Binding Regions in XLF and XRCC4	78
4.4.4 DNA Binding and the XRCC4-XLF Filament	79
4.4.5 LigaseIV and DNA Binding Regions of XRCC4 Modulate XRCC4-XLF Filament-DNA Interactions	81
4.4.6 Modelling DNA Binding in the XRCC4-XLF Filament	82
4.4.7 The XRCC4-XLF Filament Bridges DNA Molecules	83
4.4.8 Formation of a Higher-Order XRCC4-XLF Filament	85
4.5 Discussion	87
4.5.1 XRCC4-XLF Filament Bundles	88
4.5.2 Function of the XRCC4-XLF Filament in NHEJ	91
4.5.3 Conclusion	93
4.6 Experimental Procedures	96
4.6.1 XLF and XRCC4 Expression Vector Construction	96

4.6.2 Protein Expression and Purification	96
4.6.3 Crystallization and Data Collection of XLF ²²⁴ -XRCC4 ¹⁵⁷	96
4.6.4 Structure Determination and Model Refinement	96
4.6.5 Electrophoretic Mobility Shift Assays for DNA Binding	97
4.6.6 DNA Bridging	97
4.6.7 Scanning Force Microscopy	97
4.7 Acknowledgements	98
4.8 Supplemental Information	99
4.8.1 Supplemental Experimental Procedures	104
4.8.2 Supplemental Discussion	106
<u>Chapter 5: The Effects of XRCC4-XLF Filaments on LigaseIV Activity</u>	108
5.1 Author's Preface	108
5.2 Abstract	109
5.3 Experimental Procedures	109
5.3.1 Expression of Human XRCC4-LigaseIV	109
5.3.2 Purification of XRCC4-LigaseIV	110
5.3.3 Activity of XRCC4-LigaseIV	112
5.3.4 Crystallization of XRCC4-LigaseIV	112
5.4 Results	113
5.4.1 Purification of XRCC4-LigaseIV	113
5.4.2 Activity of XRCC4-LigaseIV	115
5.4.3 Crystallization of XRCC4-LigaseIV	118
5.5 Discussion	119
5.5.1 Crystallization of XRCC4-LigaseIV	119
5.5.2 Stimulation of Ligation Activity	120

Chapter 6: Summary and Conclusions	124
6.1 Summary	125
6.1.1 Challenges and Solutions in Crystallizing Multi-Protein Complexes	126
6.1.2 Protein Filaments are Not Unique to XRCC4-XLF	129
6.1.3 The Importance of the C-terminal Tails of XRCC4 and XLF	131
6.2 Future Directions	133
6.2.1 Structural Studies	133
6.2.2 Effect of XRCC4-XLF on Ligation Activity of LigaseIV	135
6.3 Significance and Conclusions	136
References	142
Appendix I	161
List of Publications by S.N. Andres	161

List of Figures

<u>Chapter 1: Introduction</u>	1
Figure 1.1 DNA double-strand break damage response.	5
Figure 1.2 DNA double-strand break repair by non-homologous end-joining.	7
Figure 1.3 Crystal structure of the Ku70/80 heterodimer.	10
Figure 1.4 Crystal structure of DNA-PKcs.	13
Figure 1.5 Crystal structure of a β -CASP domain.	18
Figure 1.6 BRCT domain of polymerase mu.	19
Figure 1.7 Crystal structure of XRCC4.	22
Figure 1.8 Crystal structure of XRCC4 bound to LigaseIV.	25
Figure 1.9 Crystal structure of the catalytic domains of human DNA ligases.	27
<u>Chapter 2: Crystal Structure of Human XLF: a Twist in Non-Homologous DNA End-Joining</u>	31
Figure 2.1 Structure and sequence conservation of the N-terminal region of human XLF.	36
Figure 2.2 Functional analysis of XLF ¹⁻²²⁴ .	41
Figure 2.3 Structure-based comparison of human XLF ¹⁻²²⁴ .	43
Figure 2.4 Mutational analysis of XRCC4 and XLF binding surfaces.	46
Supplemental Figure 2.1 Schematic of XLF residues.	51
Supplemental Figure 2.2 Concentration-dependent oligomeric states of XLF.	52
Supplemental Figure 2.3 Structure and sequence conservation of human XRCC4.	53
<u>Chapter 3: Crystallization and Preliminary X-ray Diffraction Analysis of the Human XRCC4-XLF Complex</u>	56
Figure 3.1 Diffraction image of XRCC4 ^{Δ157} /XLF ^{Δ224} crystals illustrating anisotropy.	65

<u>Chapter 4: Crystal Structure of Human XRCC4-XLF Reveals an</u>	
<u>Extended Helical Filament Important for Bridging DNA Ends.</u>	68
Figure 4.1 Crystal structure of the XLF ¹⁻²²⁴ - XRCC4 ¹⁻¹⁵⁷ Complex.	76
Figure 4.2 XRCC4-XLF DNA binding.	80
Figure 4.3 Bridging of DNA molecules by XRCC4-XLF.	84
Figure 4.4 SFM analysis of protein-DNA networks.	86
Figure 4.5 Model of XRCC4-XLF filaments bound to DNA.	89
Figure 4.6 Summary of structural states of XRCC4.	95
Figure S4.1 XRCC4 sequence alignment.	99
Figure S4.2 XLF sequence alignment.	100
Figure S4.3 DNA Binding Analysis of XLF and XRCC4 filament mutants.	101
<u>Chapter 5: The Effects of XRCC4-XLF Filaments on LigaseIV Activity</u>	108
Figure 5.1 Purified Human XRCC4 (~53kDa) and Human LigaseIV (~104kDa)	114
Figure 5.2 Effects of XLF and XRCC4 on Ligation	116
Figure 5.3 Effects of XLF L115A on Ligation	117
<u>Chapter 6: Summary and Conclusions</u>	124
Figure 6.1 Comparison of XRCC4 and SAS-6 homodimers.	130
Figure 6.2 Potential conformation of DNA ends at a double-strand break.	138
Figure 6.3 Proposed model for non-homologous end-joining.	140

List of Tables

<u>Chapter 2: Crystal Structure of Human XLF: a Twist in Non-Homologous DNA End-Joining</u>	31
Table 2.1. Crystallographic data and refinement statistics.	37
<u>Chapter 3: Crystallization and Preliminary X-ray Diffraction Analysis of the Human XRCC4-XLF Complex</u>	56
Table 3.1 Details of constructs used in crystallization trials.	60
Table 3.2 Crystals of XRCC4-XLF complexes and associated diffraction.	63
Table 3.3 Crystallographic data statistics.	66
<u>Chapter 4: Crystal Structure of Human XRCC4-XLF Reveals an Extended Helical Filament Important for Bridging DNA Ends.</u>	68
Table 4.1 Crystallographic Data and Refinement Statistics.	73
Table S4.1 XLF and XRCC4 interface residues with corresponding buried surface areas (BSA).	102
Table S4.2 DNA binding activity of XRCC4 mutants with mutations between amino acids 157-200.	103
<u>Chapter 5: The Effects of XRCC4-XLF Filaments on LigaseIV Activity</u>	108
Table 5.1 Putative protein crystals of XRCC4-LigaseIV	118

Abbreviations

Ala	alanine
Alt-NHEJ	alternative non-homologous end-joining
AMP	adenosine monophosphate
ATM	ataxia-telangiectasia mutated
ATP	adenosine triphosphate
bp	basepair
BRCT	BRCA1 C-terminal homology
CASP	CPSF-Artemis-Snm1-Pso2
DBD	DNA binding domain
DNA	deoxyribonucleic acid
DNA-PK	DNA protein kinase
DNA-PKcs	DNA protein kinase catalytic subunit
HEAT	Huntingtin-Elongation Factor 3-Protein Phosphatase 2A-PI3-kinase
HR	homologous recombination
MRN	Mre11/Rad50/Nbs1
MR-SAD	molecular replacement with single-wavelength anomalous diffraction
NHEJ	non-homologous end-joining
NTase	nucleotidyltransferase
OB	oligonucleotide/oligosaccharide
Phe	phenylalanine
PI3	phosphoinositide-3
PNK	polynucleotide kinase
Pol	polymerase
PPi	inorganic pyrophosphate

RAG1/2	recombination activating gene 1 and 2
RSS	recombination signal sequence
SAD	single-wavelength anomalous diffraction
SAS-6	spindle-assembly 6
SAXS	small-angle x-ray scattering
Ser	serine
SFM	scanning force microscopy
TdT	terminal deoxynucleotidyl transferase
Thr	threonine
V(D)J	variable, diverse and joining
XLF	XRCC4-like factor

Chapter 1: Introduction

1.1 DNA Double-strand breaks

DNA double-strand breaks represent a small, yet serious form of DNA damage occurring when single-strand breaks are generated on complementary strands of a DNA helix within close enough distance (~10 bp) that base-pairing and chromatin structure are incapable of keeping DNA ends physically together (Michael et al., 2000). While double-strand breaks are less common than other forms of DNA damage, the result can be severe, destabilizing genomic integrity. Only 1 to 2 double-strand breaks can be lethal for diploid cells of *S. cerevisiae* (Resnick et al., 1976). Strong evidence also exists showing that double-strand break formation leads to mutations and chromosomal translocations, which if left unrepaired, can lead to tumorigenesis or immunodeficiency (Borek et al., 1966; Cavazzana-Calvo, et al., 1993; Hei et al., 1988).

Multiple causes, both exogenous and endogenous to the cell, are responsible for creating DNA double-strand breaks. Exogenous sources include ionizing radiation (x-rays, gamma-rays), which deposit energy, creating complex damage in a localized area of DNA (Ward, 1985). Results of this damage include the direct ionization of DNA bases and sugars, or indirectly, production of reactive oxygen species in the aqueous environment, such as the hydroxyl radical (Ward, 1973; Ward et al., 1976). Either single or multiple hydroxyl radicals can then damage the DNA, with 1 Gy of ionizing radiation (=1 J/kg of absorbed energy) creating 16 – 40 double-strand breaks (Siddiqi et al., 1987; Ward, 1988). Numerous anti-tumour agents also create double-strand breaks by free-radical damage. One such example is bleomycin, where 10% of the damage created results in DNA double-strand breaks (Bennet et al., 1993). Bleomycin damage is further complicated as it generates DNA ends that are not directly ligatable and contain a mixture of 5'-phosphates with 3'-phosphoglycolate ends (Giloni et al., 1981).

Interestingly, a human cell generates the majority of DNA double-strand breaks as part of endogenous cellular activities. Double-strand breaks can be formed unintentionally by exposure to reactive oxygen species generated through cellular metabolism. These free radicals damage DNA much like the free radicals generated by ionizing radiation (Chance et al., 1979; Richter et al., 1988). Furthermore, DNA replication machinery may encounter alternate secondary DNA structures, or a nick that polymerases cannot process through, causing the replication fork to collapse and form a double-strand break (Kuzminov 1995; Kuzminov, 2001).

In addition to double-strand breaks generated randomly through exposure to reactive oxygen species or blocked DNA replication, double-strand breaks are also created intentionally within the cell, usually for the purpose of introducing genetic variation. This is seen in meiosis, which requires the formation of double-strand breaks for gene conversion and crossovers, and in V(D)J recombination (Sun et al., 1989; Keeney et al., 1997). Genetic diversity is introduced into immunoglobulins during V(D)J recombination through the formation and resolution of double-strand breaks in immunoglobulin genes (Roth et al., 1993). V(D)J recombination is initiated by the Recombination Activating Gene (RAG)1 and RAG2 proteins, binding to and catalyzing cleavage at recombination signal sequences (RSS) located between variable (V) and joining (J) DNA segments (Oettinger et al., 1990; Schatz et al., 1989; McBlane et al., 1995). The sequence between the RSS that is released is termed the signal joint, and is ligated into circular DNA by an unknown mechanism. The chromosomal DNA ends, called coding joints, are left with DNA hairpin capped ends following removal of the signal sequences (Ramsden et al., 1995; Roth et al., 1992). After formation of hairpin structures on coding ends, the protein Artemis introduces nicks into these DNA hairpin structures, creating a DNA double-strand break (Ma et al., 2002). To generate a unique

DNA coding sequence in immunoglobulins, the DNA double-strand breaks are resolved by a mutagenic repair pathway called non-homologous end-joining (NHEJ), which will be discussed later in more detail (Section 1.3) (Biedermann et al., 1991; Early et al., 1980; Fukumura et al., 1998; Gu et al., 1997; Hendrickson et al., 1991; Schilling et al., 1980; Taccioli et al., 1994).

1.2 DNA Double-Strand Break Damage Response

The response to DNA double-strand break damage in cells can be thought of as a unique signal transduction cascade, initiated with the 'signal' being a DNA double-strand break. Double-strand break 'sensors' then recognize the damage and in turn activate protein 'transducers', which are responsible for signal amplification and diversification of the response, typically through a kinase cascade, leading to a range of outcomes (Figure 1.1; Jackson, 2002). Current evidence suggests that a complex of Mre11, Rad50 and Nbs1 (MRN) acts as a 'sensor' at the site of damage, with the resulting response causing chromatin remodelling over 2 megabase lengths of DNA away from the site of damage (Maser, et al., 1997; Rogakou, et al., 1998). Formation of MRN at the site of damage activates the protein ataxia-telangiectasia mutated (ATM), as a transducer (Lee et al., 2005). ATM is a member of the phosphoinositide-3-kinase-like kinase family, and controls the G1/S, intra-S, and G2/M checkpoints of the cell-cycle, which in the presence of DNA double-strand breaks phosphorylates a number of other 'transducer' proteins, including p53, Chk2, Nbs1, and CtIP, leading to a complex signalling cascade, whose mechanisms are not entirely understood (Banin et al., 1998; Canman et al., 1998; Matsuoka et al., 2000; Lim et al., 2000; Li et al., 2000). The end results of this cascade may include apoptosis, cell cycle arrest, and/or DNA double-strand break repair.

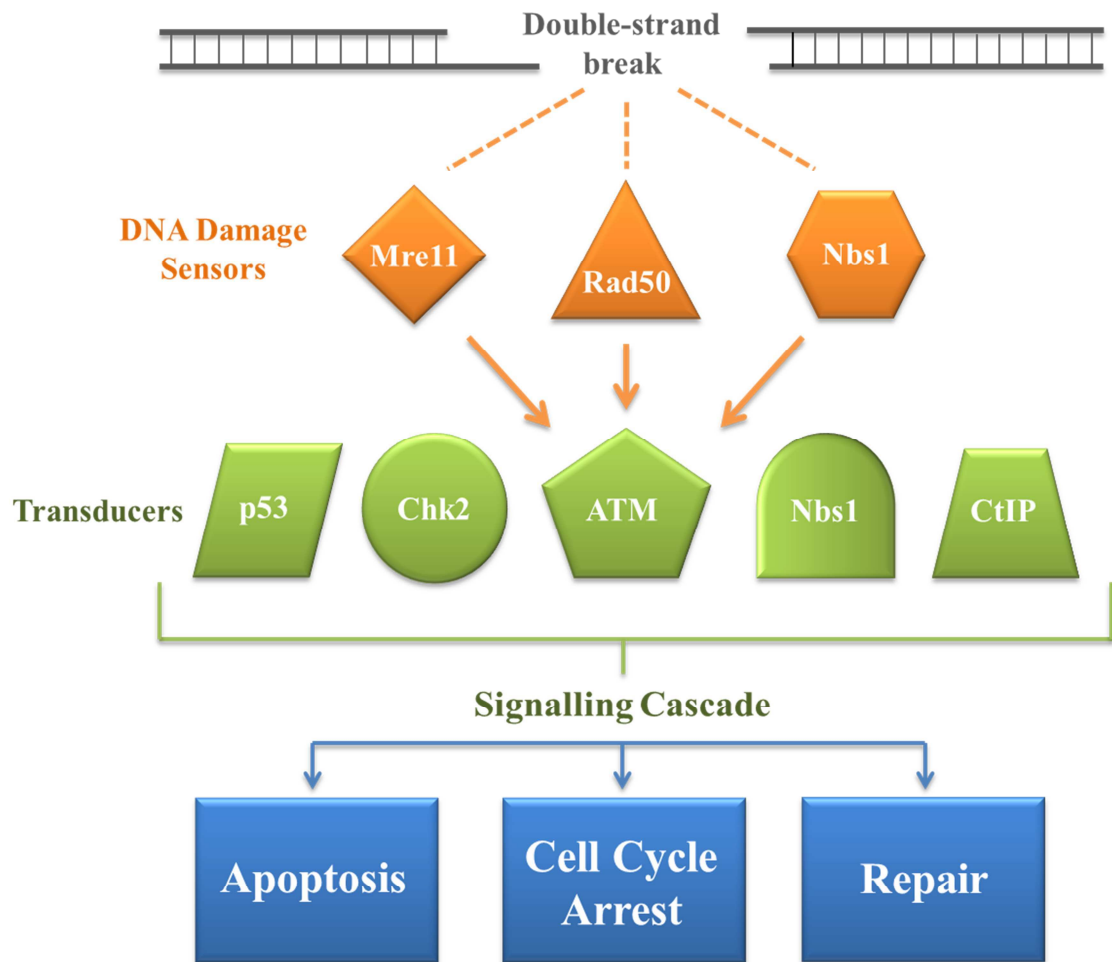


Figure 1.1 DNA double-strand break damage response. The cellular response to a DNA double-strand break involves a kinase signalling cascade, with Mre11, Rad50 and Nbs1 initiating the damage response by activation of proteins such as ATM (Lee et al., 2005; Maser et al., 1997; Rogakou et al., 1998). Through phosphorylation, ATM continues the signalling cascade, leading to outcomes such as apoptosis, cell cycle arrest, or DNA double-strand break repair. Figure adapted from Jackson, (2002).

The most favourable outcome when a cell encounters a DNA double-strand break is repair, and mammalian cells have evolved three pathways by which repair occurs: homologous recombination (HR), non-homologous end-joining (NHEJ), or alternative non-homologous end-joining (alt-NHEJ). How a cell determines which repair path to take is not well understood, but is partly based on the current stage of the cell cycle when damage occurs. HR, as the name implies, requires the use of a homologous DNA template in order to repair the damage. Therefore, HR is most efficient during S and G2 phases of the cell cycle, when a sister chromatid or homologous chromosome is present (Kadyk, 1992; Resnick, 1976). NHEJ, however, requires no template for repair, and is thus available throughout the entire cell cycle, with current research suggesting that approximately 80% of ionizing-radiation induced DNA double-strand breaks are repaired by NHEJ in both G1 and G2 (Beucher et al., 2009; Wilson et al., 1982). Differentiating between NHEJ and alt-NHEJ is even more uncertain as alt-NHEJ has only recently been identified and like NHEJ, does not require a template for repair. However, alt-NHEJ functions in the absence of key proteins required in NHEJ such as Ku 70/80, XRCC4 and, LigaseIV, and does require some microhomology between the broken DNA ends for repair (Bennardo et al. 2008; Yan et al., 2007). The work presented in this thesis deals with the proteins in NHEJ, and therefore only NHEJ will be discussed in detail.

1.3 A Brief Overview of Non-homologous End-Joining

NHEJ in mammals consists of 3 major parts: (1) initial recognition and protection of the DNA ends by NHEJ proteins; (2) processing the DNA ends to a ligatable form; and (3) final repair of the DNA double-strand break. Initial protection is carried out by the Ku70/80 heterodimer, which binds broken DNA ends to prevent degradation and

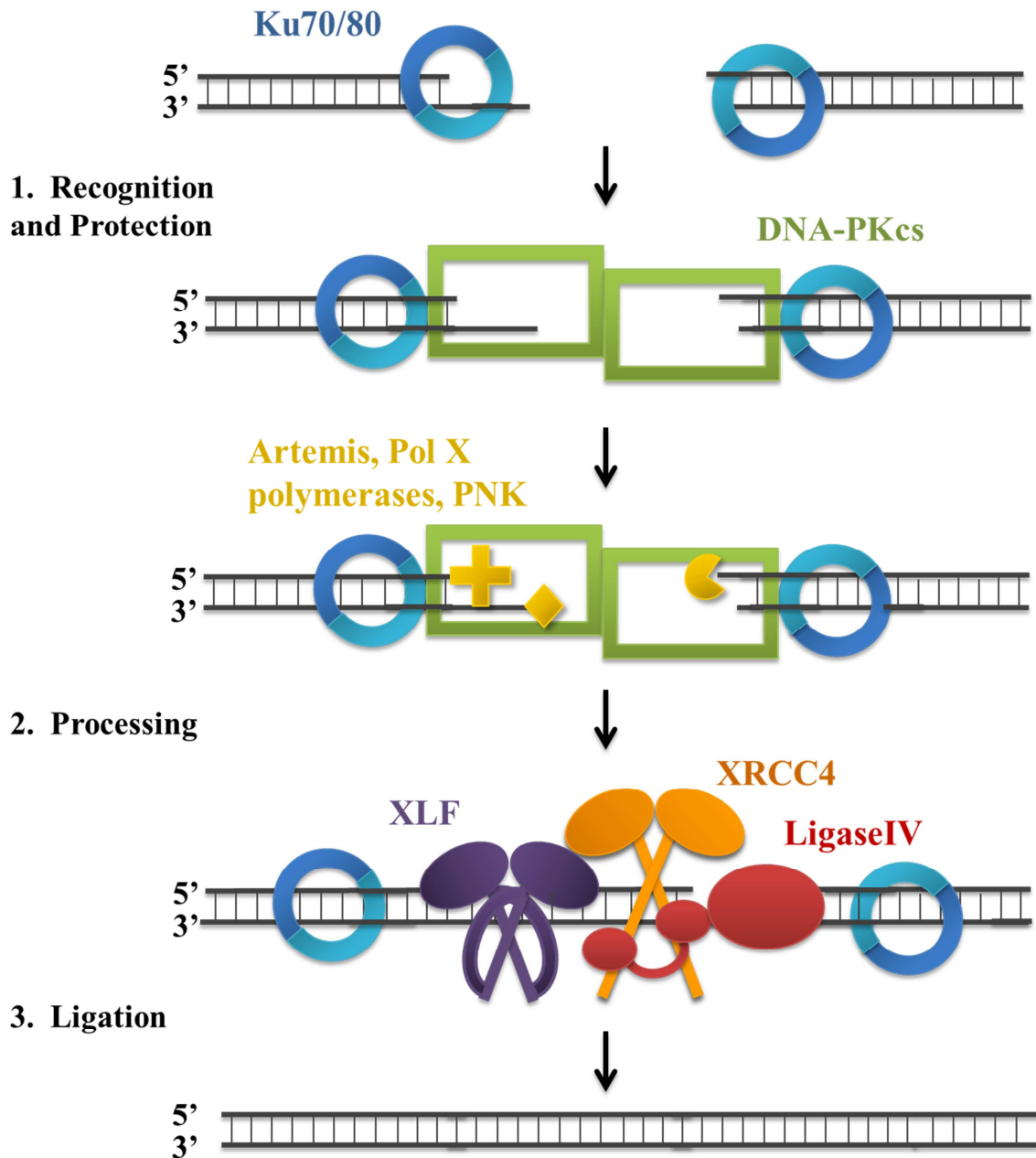


Figure 1.2 DNA double-strand break repair by non-homologous end-joining. DNA double-strand breaks are repaired by a core set of proteins in NHEJ, beginning with Ku70/80 and DNA-PKcs protecting and holding the DNA ends in place. Processing enzymes prepare the DNA ends for final ligation by XLF, XRCC4 and LigaseIV.

recruits other NHEJ repair proteins to the site of damage (Figure 1.2) (Mimori et al., 1986). These include XRCC4 and XRCC4-like factor (XLF), whose purpose at this step is not fully understood (Yano et al., 2008b). Ku70/80 also recruits the catalytic subunit of DNA protein kinase (DNA-PKcs), causing Ku70/80 to translocate approximately one helical turn inward along the DNA (Paillard et al., 1991; Gottlieb et al., 1993; Yoo et al., 1999). Ku70/80 and DNA-PKcs form the holoenzyme DNA-PK that synapses broken DNA ends, holding the DNA in place prior to ligation (DeFazio et al., 2002; Gottlieb et al., 1993). Based on the type of DNA ends encountered, processing enzymes including Artemis, polymerase mu and lambda, terminal deoxynucleotidyl transferase (TdT), and polynucleotide kinase (PNK) may also be recruited in order to generate DNA ends containing 5'-phosphate and 3'-hydroxyl groups (Chappell et al., 2002; Gilfillan et al., 1993; Komori et al., 1993; Ma et al., 2002; McElhinny et al., 2005). The resulting DNA ends are then ligated by an XRCC4-LigaseIV complex. XRCC4-LigaseIV may also interact with XLF, which has been shown to stimulate the joining of incompatible DNA ends by LigaseIV *in vitro* (Ahnesorg et al., 2006; Grawunder et al., 1997; Grawunder et al., 1998a; Robins et al., 1996; Tsai et al., 2007; Wilson et al., 1997).

1.3.1 Initiating Non-Homologous End-Joining by Ku70/80 and the DNA-Dependent Protein Kinase Catalytic Subunit

Ku70/80 and DNA-PKcs are two of the core proteins in NHEJ, suggested to be the initiators of NHEJ, with Ku70/80 binding the DNA double-strand break first (Griffith et al., 1992; Blier et al., 1993). Ku70 and Ku80 were first identified as a target of autoantibodies in patients suffering from the auto-immune disease scleroderma-polymyositis overlap syndrome (Mimori et al., 1981). Defects in either Ku protein prevent proper resolution of DNA double-strand breaks and in turn lead to severe

combined immunodeficiency in mice, along with a sensitivity to ionizing radiation in mammalian cells (Jackson et al., 1995; Lieber et al., 1988a).

Ku70 and Ku80 form a heterodimeric, ring-like structure (deVries et al., 1989; Walker et al., 2001). While only being 14% identical in a sequence-based alignment, the overall structure of each monomer is similar (Figure 1.3A). Both Ku70 and Ku80 are composed of 3 domains: an N-terminal α/β domain, a β -barrel, and a C-terminal α -helical arm (PDB 1JEQ) (Walker et al., 2001). While all three domains support the dimerization interface, the greatest contribution comes from the C-termini of both monomers. Each α -helical tail wraps around β -barrels of the opposing Ku monomer, forming the Ku70/80 ring (Walker et al., 2001; Wu et al., 1996). The pore of the ring is large enough to accommodate approximately 20bp of DNA, or two turns of a DNA helix, with DNA binding contacts occurring primarily through the central β -barrel domain (Figure 1.3B)(PDB 1JEY). Ku70/80-DNA interactions also include additional steric interactions between the protein and both major and minor grooves of the DNA's sugar-phosphate backbone, as illustrated by the positively charged electrostatic surface potential of Ku70/80 contacting the DNA substrate in Figure 1.3B (Walker et al., 2001). This DNA binding platform, combined with the ring shape of Ku70/80, allows for high-affinity ($1.5 \times 10^{-10} \text{M}^{-1}$), sequence-independent binding of blunt, 5'-overhang or 3'-overhang DNA ends (Blier et al., 1993; Falzon et al., 1993; Mimori et al., 1986; Paillard et al., 1991). These features are biologically significant, as DNA damage may occur anywhere within a chromosome, making Ku70/80 an ideal DNA binding partner and initiator of NHEJ.

More recently, it was discovered that Ku70/80 possessed enzymatic activity as a 5'-deoxyribose-5-phosphate lyase (Roberts et al., 2010). DNA double-strand breaks are rarely 'clean' (i.e. 5'-phosphate and 3'-hydroxyl groups), therefore having Ku70/80 as a

lyase would enable it to remove 5'-terminal abasic or apurinic/aprimidinic sites. The lyase activity occurs by creating a nick in the DNA, 3' of the abasic site, removing the

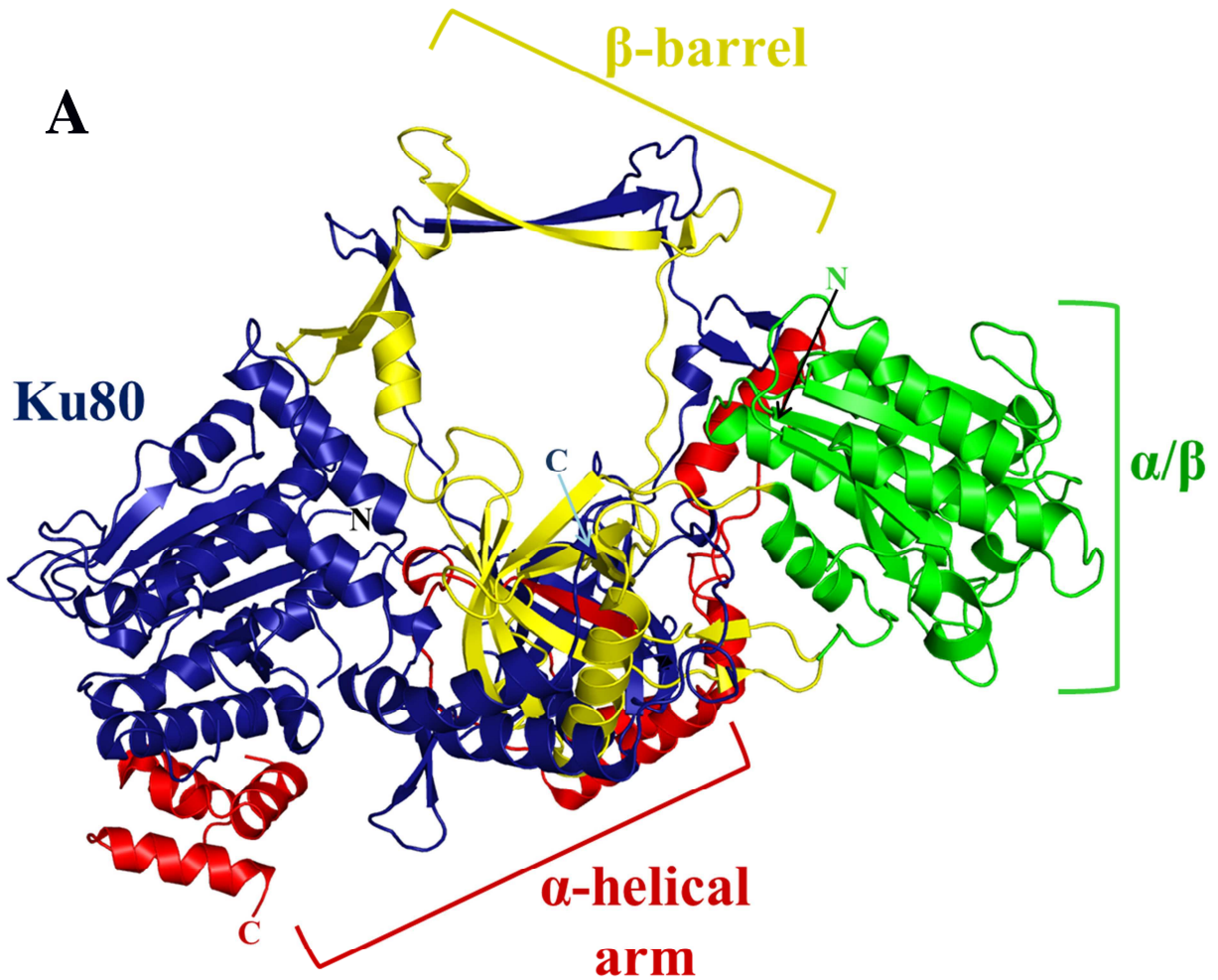


Figure 1.3 Crystal structure of the Ku70/80 heterodimer. (A) The crystal structure of Ku70/80, solved by Walker et al., (2001), consists of the Ku80 (blue) and the Ku70 monomer. The three domains of the Ku monomers are shown in Ku70 as the α/β domain (green), the β -barrel (yellow), and the α -helical arm (red). N and C-termini of Ku70 are in green and red, and in black for Ku80 (PDB 1JEQ). *Continued on page 11.*

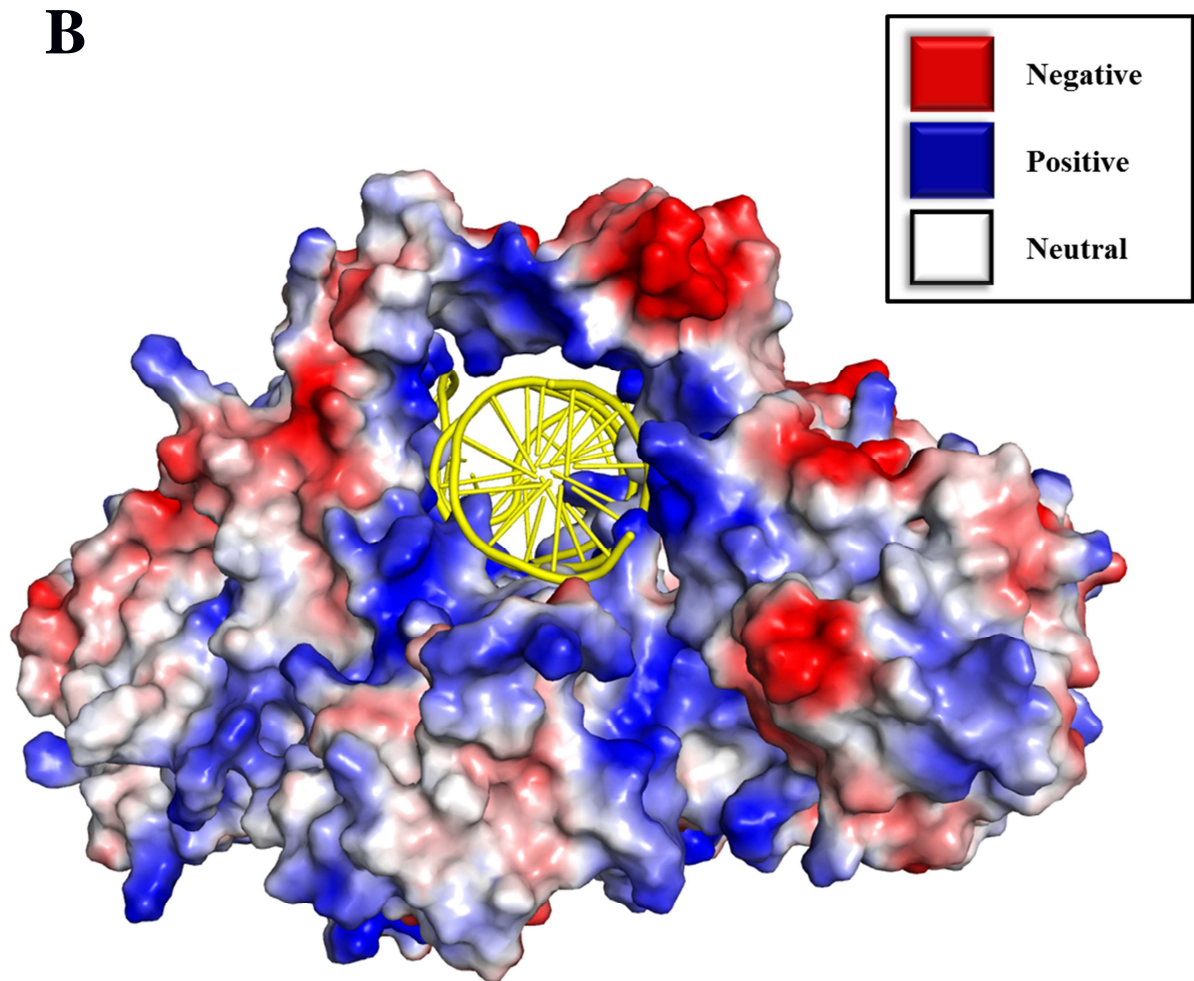


Figure 1.3 Crystal structure of the Ku70/80 heterodimer. (B) Crystal structure of Ku70/80 bound to DNA, solved by Walker et al., (2001), illustrating electrostatic surface potential. Regions of negative charge (red), positive charge (blue) and neutral (white) are shown, with a DNA substrate bound (yellow) (PDB 1JEY).

base and leaving behind a 5'-phosphate on the DNA end. The Ku70/80 lyase activity has been demonstrated as a necessary function in NHEJ *in vitro* and *in vivo*, giving Ku70/80 the added benefit of not just binding DNA ends, but also processing them (Roberts et al., 2010).

Another major role of Ku70/80 in DNA double-strand break repair is recruitment of other core NHEJ repair proteins. Ku70/80 has been reported to interact with XLF, XRCC4, LigaseIV, and DNA-PKcs (Costantini et al., 2007; Gottlieb et al., 1993; Suwa et al., 1994; Yano et al., 2008b). The purpose of Ku70/80 interaction with XLF, XRCC4 and LigaseIV is not well understood; however Ku70/80's interaction with DNA-PKcs has been studied extensively and shown to be particularly important during the initial stages of NHEJ. The C-terminal tail of Ku80 interacts with the C-terminus of DNA-PKcs to form the DNA-PK holoenzyme, greatly increasing the affinity of DNA-PKcs for DNA ends (Gell et al., 1999; Singleton et al., 1999; West et al., 1998). Within the context of DNA-PK, Ku70/80 is thought to function as the regulatory unit, while DNA-PKcs provides the catalytic kinase activity (Dvir, et al., 1992; Gottlieb et al., 1993).

DNA-PKcs is a serine/threonine protein kinase that is part of the phosphoinositide-3 kinase-related protein kinases family (Hartley et al., 1995). DNA-PKcs is a large, 470 kDa protein, structurally arranged with a head domain sitting atop a ring with two arms composed of anti-parallel HEAT repeats. These HEAT repeats fold back onto themselves, creating a gap at the bottom of the ring (Figure 1.4). The ring is 120Å in diameter, large enough to accommodate a double-stranded DNA helix (Sibanda et al., 2010). Small-angle x-ray scattering (SAXS) and cryo-electron microscopy data revealed similar, albeit lower-resolution, models (Hammel et al., 2010b; Rivera-Calzada et al., 2005). Similar to Ku70/80, it has been suggested that DNA-PKcs binds to each side of the double-strand break, forming a synaptic-like complex between the broken

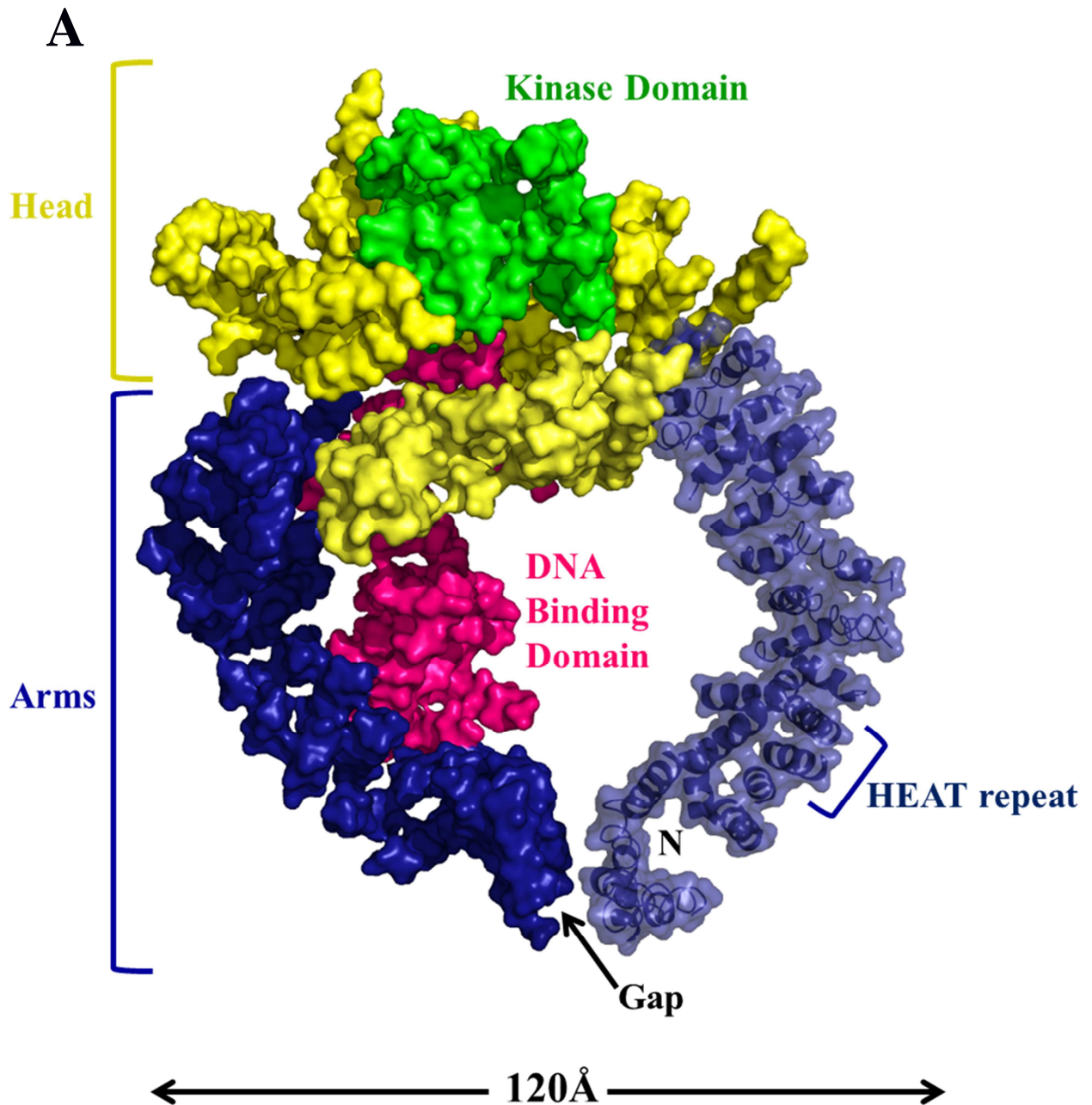


Figure 1.4 Crystal structure of DNA-PKcs. (A) Space-filling model of the crystal structure of DNA-PKcs, solved by Sibanda et al., (2010), consists of a head domain (yellow), encompassing the catalytic kinase domain (green). The arms of DNA-PKcs (blue), are made up of HEAT repeats, and also includes a putative DNA binding domain (pink). *Continued on page 14.*

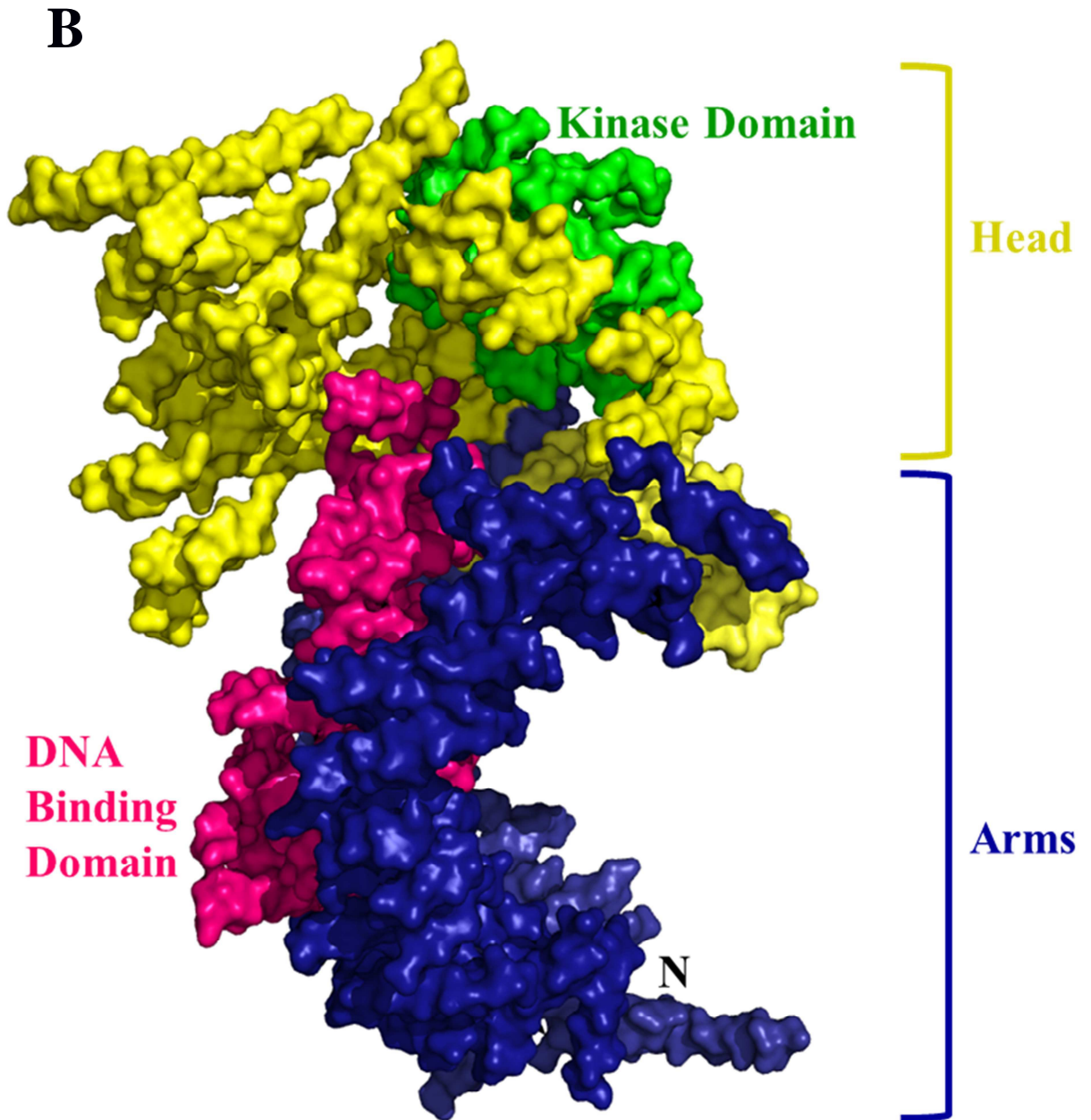


Figure 1.4 Crystal structure of DNA-PKcs. (B) Side-view of DNA-PKcs, rotated by 90°. Predicted N-terminus is in black. C-terminus is predicted to be within the head domain (PDB 3KGV).

ends that would align and bridge them for repair (Spagnolo et al., 2006).

Given the low resolution structure of DNA-PKcs (6.6Å), it is only predicted that the C-terminus lies within the head domain, and this is where Ku80 is predicted to bind (Sibanda et al., 2010). The DNA-PKcs head region also contains the functional kinase domain, which has been shown to phosphorylate itself at Serine (Ser) 2612 and Ser 2624, along with Threonine (Thr) 2609, 2620, 2638, and 2647 in response to DNA double-strand break damage *in vivo*. While the purpose of autophosphorylation is unknown, it has been shown to be necessary for NHEJ and to induce structural changes within the arm regions of DNA-PKcs, corresponding to the N-terminal 2880 residues (Ding et al., 2003). SAXS analysis of phosphorylated DNA-PKcs also indicates that large conformational changes occur, corresponding to the arm regions on either side of the head domain, while *in vitro* experiments show that autophosphorylation of DNA-PKcs results in reduced kinase activity and release from the Ku70/80 subunit (Gottlieb et al., 1993; Hammel et al., 2010b). In addition, electron microscopy studies suggest that DNA-PKcs ‘arms’ are flexible and mobile. Therefore, it has been postulated that autophosphorylation acts as a regulator for disassembly of the repair complex (Chan et al., 1996; Chan et al., 1999; Douglas et al., 2001; Rivera-Calzada, et al., 2005). Mechanistically, the flexibility in the arm regions of DNA-PKcs permits them to move apart, widening the gap at the bottom of the ring, so that DNA-PKcs can be removed from DNA (Rivera-Calzada, et al., 2005).

Within NHEJ, DNA-PKcs is also known to phosphorylate other DNA repair proteins. *In vitro*, Ku70/80, XRCC4, and LigaseIV have been identified as phosphorylation targets of DNA-PKcs, but these phosphorylation events have not been shown to correlate with biological function (Douglas et al., 2005; Wang et al., 2004; Yu et al., 2003). It has been proposed that phosphorylation sites on these NHEJ proteins act like those of DNA-PKcs to regulate protein-protein or protein-DNA associations. *In vivo*,

DNA-PKcs phosphorylates XLF through SQ/TQ motifs, though this modification does not appear to be required for repair (Yu et al., 2008). Therefore, the only known biologically relevant phosphorylation activities of DNA-PKcs are those of self-phosphorylation, and phosphorylation of the nuclease Artemis. Artemis requires phosphorylation by and interaction with DNA-PKcs to activate its latent endonuclease function, an important enzymatic activity required for processing of DNA ends (Goodarzi et al., 2006; Ma et al., 2002; Ma et al., 2005; Meek et al., 2007).

1.3.2 Processing the DNA Double-Strand Break

DNA double-strand breaks generated *in vivo* are rarely chemically compatible for ligation. As discussed previously, DNA damage does not always produce 3'-hydroxyl and 5'-phosphate ends, therefore the DNA must be processed in order for ligation to occur. This can involve removal or addition of bases, or phosphorylation of DNA ends. For this, polymerases, kinases and nucleases are required, one of which is the Artemis nuclease.

Nucleolytic activity in NHEJ has been primarily attributed to Artemis. Artemis was first identified in patients with radiosensitive-severe combined immunodeficiency, and like Ku70/80^{-/-} and DNA-PKcs^{-/-} mammalian cells, Artemis^{-/-} mammalian cells are also highly sensitive to ionizing radiation, indicating the requirement for Artemis in NHEJ (Moshous, D. et al., 2000; Moshous, D. et al., 2001; Rooney et al., 2003). Artemis belongs to the CPSF-Artemis-Snm1-Pso2 (CASP) family of proteins and functions as a 5' - 3' exonuclease. However, when Artemis' C-terminus associates with DNA-PKcs and is phosphorylated by DNA-PKcs, Artemis also takes on a structure-specific endonuclease activity (Goodarzi et al., 2006; Ma et al., 2002; Ma et al., 2005; Niewolik et al., 2006). This makes Artemis useful in not only removing damaged nucleotides, such as 3'-

phosphoglycolates, but also for opening up DNA hairpin structures generated during V(D)J recombination (Ma et al., 2002). No structural solution exists yet for Artemis, but it is predicted to be structurally homologous to the conserved catalytic domain from another CASP protein, CPSF-73. The catalytic region of CPSF-73 consists of a β -CASP domain, subdivided into a metallo- β -lactamase domain and CASP motif. The metallo- β -lactamase domain consists of a β -sandwich, flanked by α -helices, with the CASP motif inserted in the middle. The CASP motif contains a parallel β -sheet surrounded by α -helices. Two zinc ions are located at the interface of the CASP motif and metallo- β -lactamase domain. (Figure 1.5) These zinc ions are proposed to be necessary for Artemis' endo- and exonuclease activity, a function that makes Artemis a versatile player in NHEJ and processing of a DNA double-strand break (Callebaut et al., 2002; Carfi et al., 1995; Mandel et al., 2006; Moshous et al., 2001).

During NHEJ processing, polymerases, kinases, and phosphatases are also required to create compatible DNA ends for ligation. The Polymerase (Pol) X family is associated with NHEJ, and includes Pol μ , Pol λ , and TdT, with Pol λ activity stimulated by interaction with LigaseIV (Fan et al., 2004). These polymerases are small, single unit enzymes, with 5'-3' polymerase activity only. They contain BRCA1 C-terminal homology (BRCT) domains, a characteristic domain found in DNA repair proteins including LigaseIV, and associated with protein-protein interactions. The BRCT domain of Pol μ is composed of four parallel β -strands, surrounded by three α -helices, and is structurally homologous to all other known BRCT domains (Figure 1.6)(Alt et al., 1982;

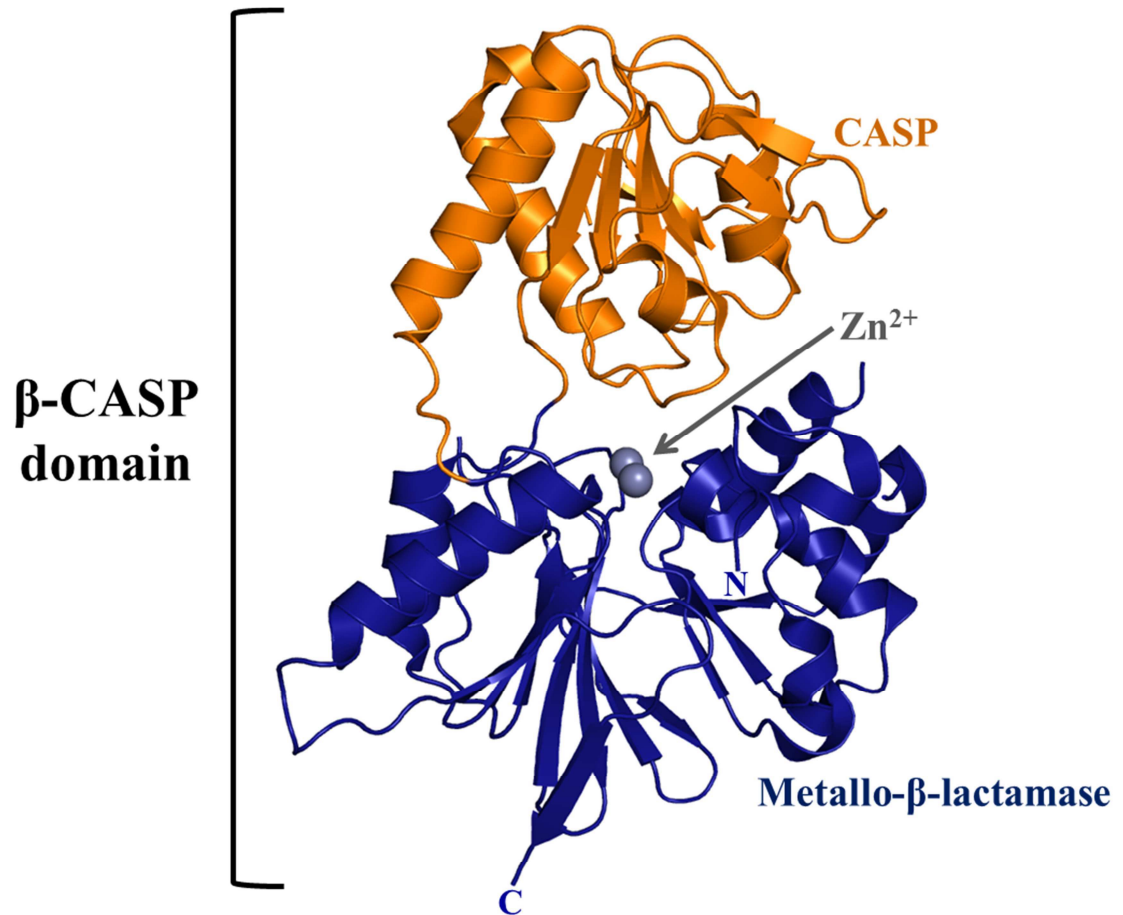


Figure 1.5 Crystal structure of a β -CASP domain. The crystal structure of the β -CASP domain from CASP family member, CPSF-73 (Mandel et al., 2006). The metallo- β -lactamase domain is coloured blue, and is interrupted by the CASP motif (orange). Two zinc ions are located at the interface of both domains. N- and C-termini are in blue (PDB 2I7T).

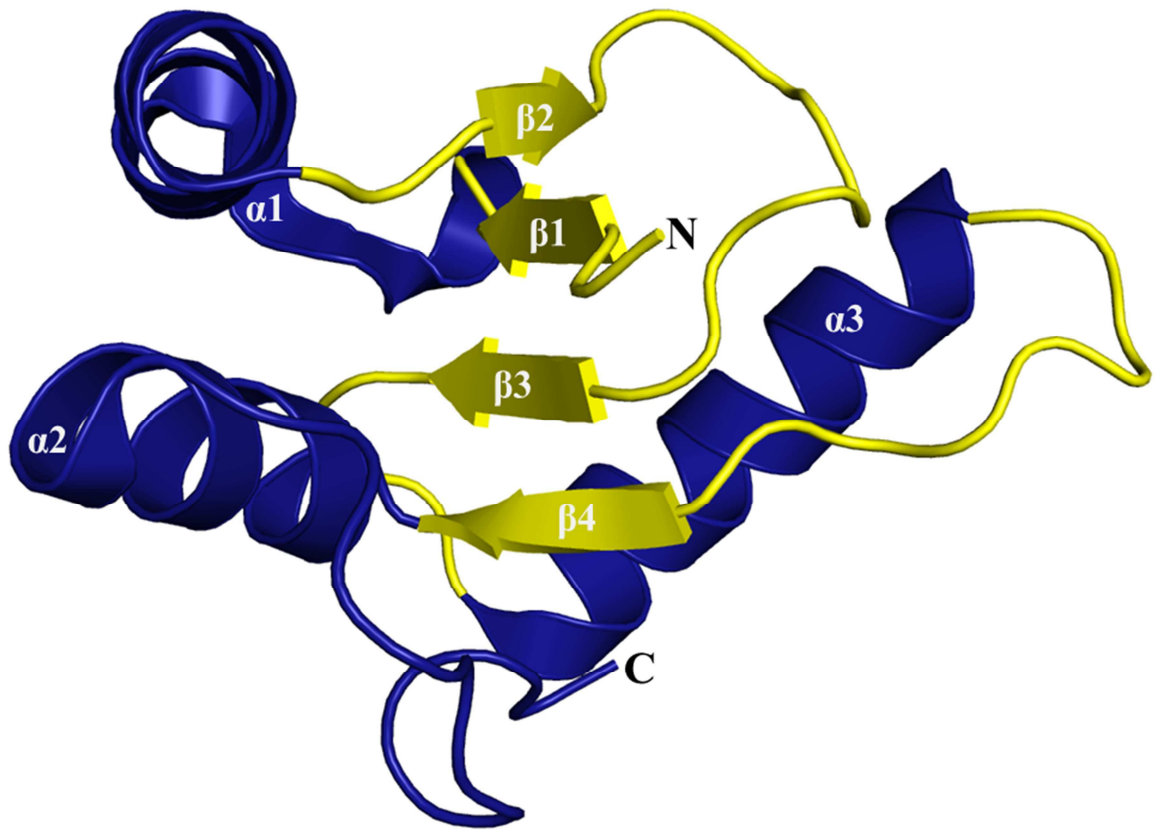


Figure 1.6 BRCT domain of polymerase mu. The solution structure of the BRCT domain from polymerase mu, solved by nuclear magnetic resonance, consists of four anti-parallel β -strands (yellow), surrounded by three α -helices (blue). N- and C-termini are in black (DeRose et al., 2007; PDB 2HTF).

Aoufouchi et al., 2000; Bork et al., 1997; DeRose et al., 2007; Garcia-Diaz et al., 2002; Ito et al., 1991; Lee et al., 2003; Mahajan et al., 2002; Mueller et al., 2008; Zhang et al., 1998). In addition to polymerization, the final processing method involves PNK, which modifies the phosphate groups on DNA ends. PNK has both 5'-kinase and 3'-phosphatase activities, to ensure that DNA ends consist of a 5'-phosphate and 3'-hydroxyl, which is chemically compatible for ligation (Chappell et al., 2002; Habraken et al., 1983; Koch et al., 2004; Pfeiffer et al., 1982). Therefore, Artemis, the PolX family of polymerases and PNK are all available in NHEJ to process all forms of a DNA double-strand break that may occur.

1.3.3 Ligation of the DNA Double-Strand Break – XRCC4, LigaseIV and XLF

The final step in NHEJ is ligation of the DNA backbone through formation of a phosphodiester bond. The three proteins responsible for mediating this activity are XRCC4, LigaseIV and XLF (Ahnesorg et al., 2006; Buck et al., 2006; Giaccia et al., 1990; Li et al., 1995). The necessity for these proteins in DNA repair has been well established through deletion and mutational studies in mammals. Cells lacking any of these three proteins show increased levels of radiosensitivity and defects in the ability to repair double-strand breaks, while knockout mouse models of XRCC4 and Ligase IV are embryonic lethal (Frank et al., 1998; Gao et al., 1998; Gao et al., 2000; Giaccia et al., 1990; Zha et al., 2007). In humans, mutations in Ligase IV or XLF genes result in radiosensitivity, immunodeficiency, and developmental delays, all characteristics of double-strand break repair defects (Buck et al., 2006; O'Driscoll et al., 2001; Riballo et al., 1999). Thus, these three proteins are essential to productive NHEJ.

XRCC4 takes on a structural role in ligation, binding to and stimulating DNA LigaseIV (Critchlow et al., 1997; Grawunder et al., 1997; Grawunder et al., 1998a). It is

a 336 amino acid protein, whose structure consists of an N-terminal head domain and a long, extended C-terminal tail. The head is made up of a 7-stranded anti-parallel beta sandwich, interrupted by a helix-turn-helix motif between β -strands four and five, while the C-terminal tail extends away from the head in a single long alpha-helix (Figure 1.7A). The biological unit of XRCC4 is a dimer, as determined by analytical ultracentrifugation, with the primary interface occurring between the head domains and upper portion of the tail (residues 119-155) (Figure 1.7B) (Junop et al., 2000; Leber et al., 1998). Only the structure of XRCC4 1-203 has been determined via x-ray crystallography, leaving the C-terminal 133 residue structure unknown. Electron microscopy studies suggest that the C-terminus may form a small globular domain at the end of the helical tail region. However, SAXS analysis of full-length XRCC4 suggests that the C-terminus is unstructured and may fold back towards the head domain (Hammel et al., 2010a; Recuero-Checa et al., 2009).

XRCC4 does not harbour any known enzymatic activity and functions as a scaffold protein, interacting with numerous repair proteins and DNA during NHEJ. XRCC4 only binds efficiently to long DNA substrates of 100 bp or more, that are either nicked or contains double-stranded breaks *in vitro*. XRCC4-DNA binding has also been shown to be highly cooperative, suggesting XRCC4 may form extended protein-DNA complexes (Modesti et al., 1999). In addition to DNA, XRCC4 also binds proteins XLF and LigaseIV (Ahnesorg et al., 2006; Critchlow et al., 1997). Furthermore, recent evidence from fluorescence recovery after photobleaching analysis suggests that XRCC4 is recruited to DNA double-strand breaks earlier in repair, perhaps by Ku70/80, and is stabilized at the break site through interaction with DNA-PKcs, which is known to phosphorylate the C-terminus of XRCC4 *in vitro* (Calsou et al., 2003; Yano et al., 2008a; Yu et al., 2003). However, the nature of these protein-protein and protein-DNA

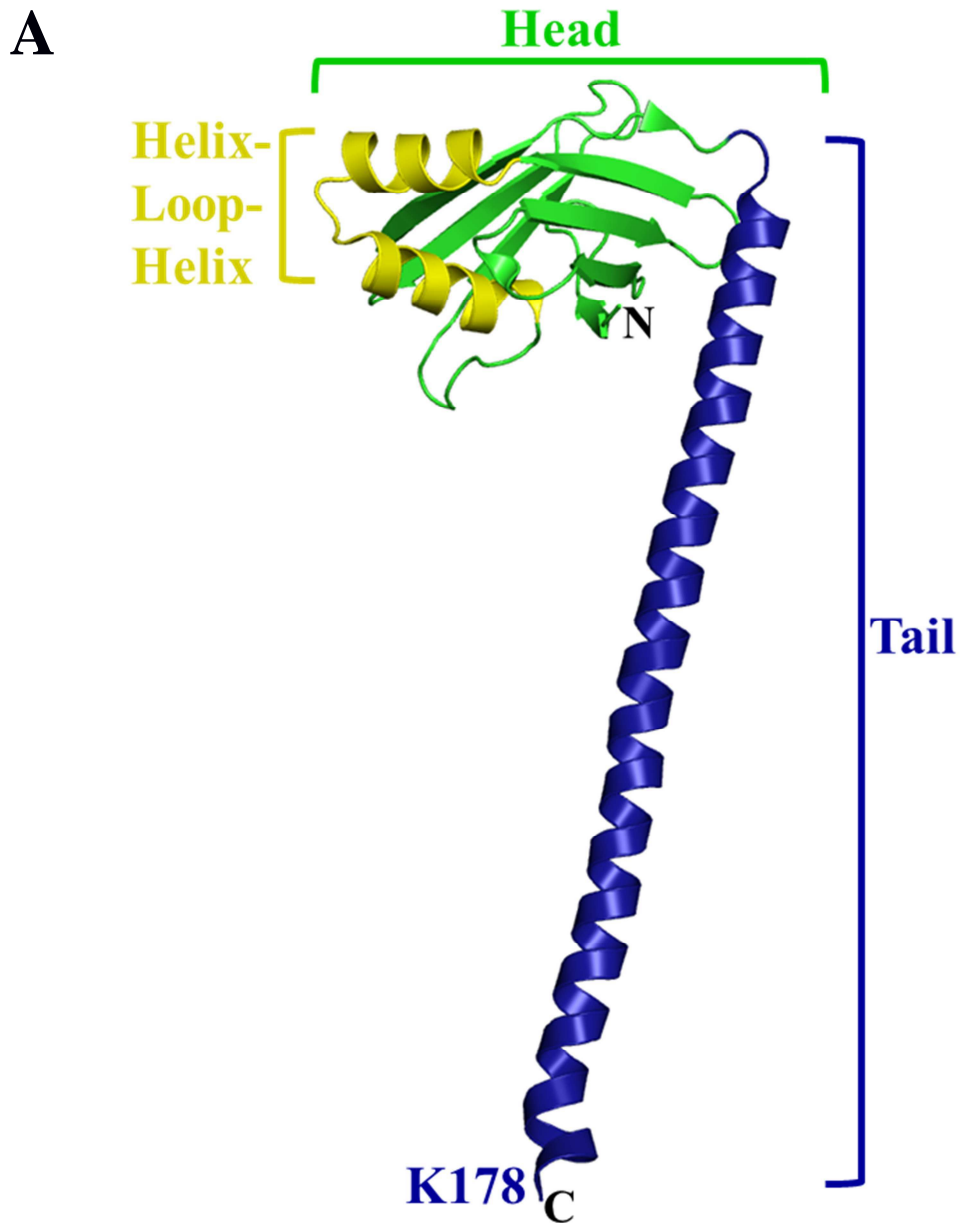


Figure 1.7 Crystal structure of XRCC4. (A) The crystal structure of an XRCC4 monomer to amino acid K178 (blue). XRCC4 consists of a head domain (green) made up of a 7-stranded anti-parallel β -sandwich, interrupted by a helix-loop-helix motif (yellow). The C-terminus consists of a long α -helical tail (blue) (Junop et al., 2000). *Continued on page 23.*

B

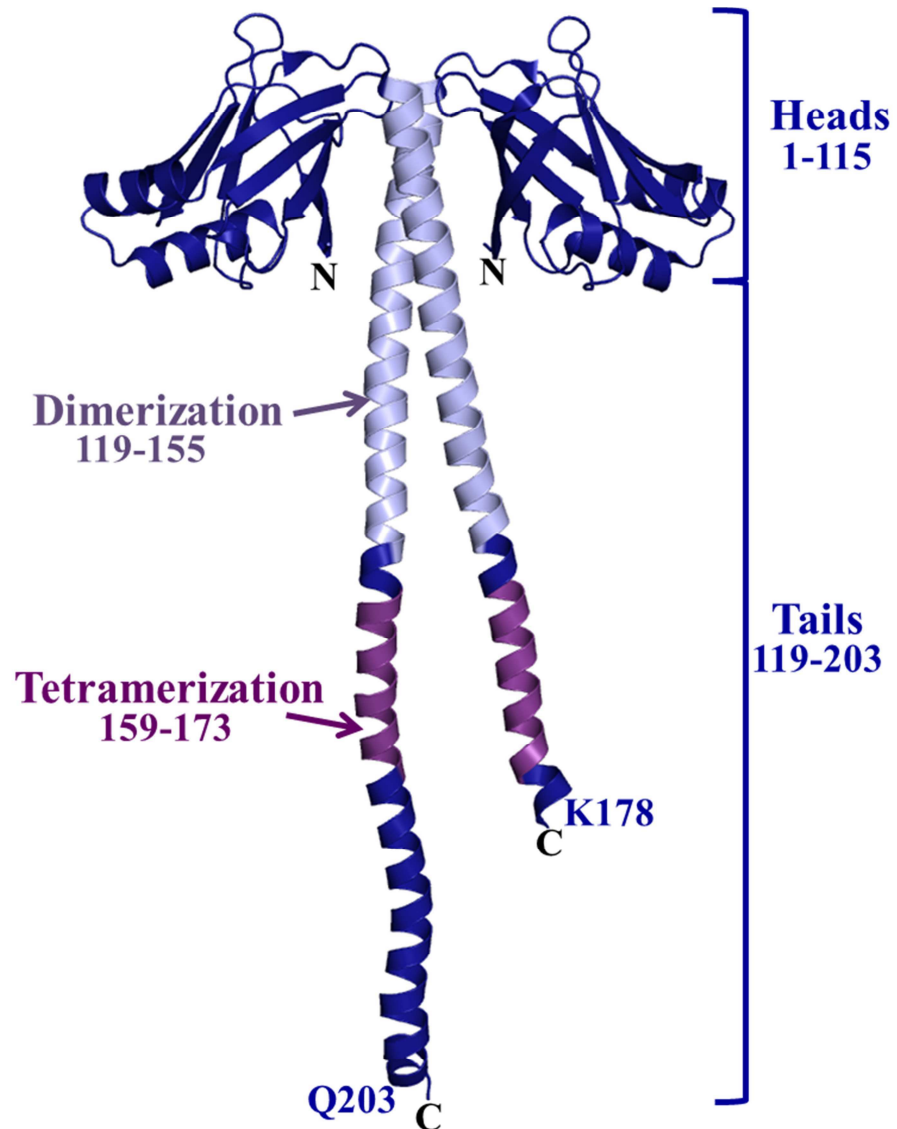


Figure 1.7 Crystal structure of XRCC4. (B) XRCC4 dimer, ending at amino acids Q203, and K178. The dimerization interface is coloured light purple, while the homotetramerization interface is coloured dark purple. The head and tail amino acid boundaries are indicated below the label for each respective domain. N- and C-termini are in black (PDB 1FU1).

interactions and their mechanistic purpose are not well understood. Of all these interactions, the most studied is the interaction between XRCC4 and Ligase IV.

One of the primary roles of XRCC4 is to stabilize and stimulate LigaseIV. Grawunder et al. (1997) first demonstrated that LigaseIV co-immunoprecipitates with XRCC4, while a yeast two-hybrid assay also pulled out XRCC4 and LigaseIV as interacting partners. More importantly, though, they noted *in vitro* that when co-expressed with XRCC4, LigaseIV activity was stimulated 7-8 fold higher when compared to LigaseIV activity on its own (Grawunder et al., 1997). Attempts to understand the mechanism of stimulation took on a structural approach. The interaction interface between XRCC4 and the C-terminus of LigaseIV, consisting of two tandem BRCT domains with a small unstructured region between them was initially defined biochemically (Critchlow et al., 1997; Grawunder et al., 1998a). These regions of XRCC4 (1-203) and LigaseIV (654-911) were used for structural studies. The resulting crystallized complex then found that structurally, there are two key interfaces between XRCC4 and LigaseIV. One is the helix-loop-helix clamp that wraps around the tails of XRCC4, while the other is a three helix bundle, formed by the helical tails of the XRCC4 dimer, and a third helix from the second BRCT domain of LigaseIV (Figure 1.8) (Wu et al., 2009). Unfortunately, this structure lacks the C-terminus of XRCC4 (203-336), the N-terminal 653 residues of LigaseIV, and does not provide a clear mechanism for why XRCC4 stimulates LigaseIV activity.

The ligation mechanism of LigaseIV itself is common among both DNA and RNA ligases and is a condensation reaction requiring a nucleotide cofactor. In the case of LigaseIV, this is adenosine triphosphate (ATP), which forms a phosphoamide bond

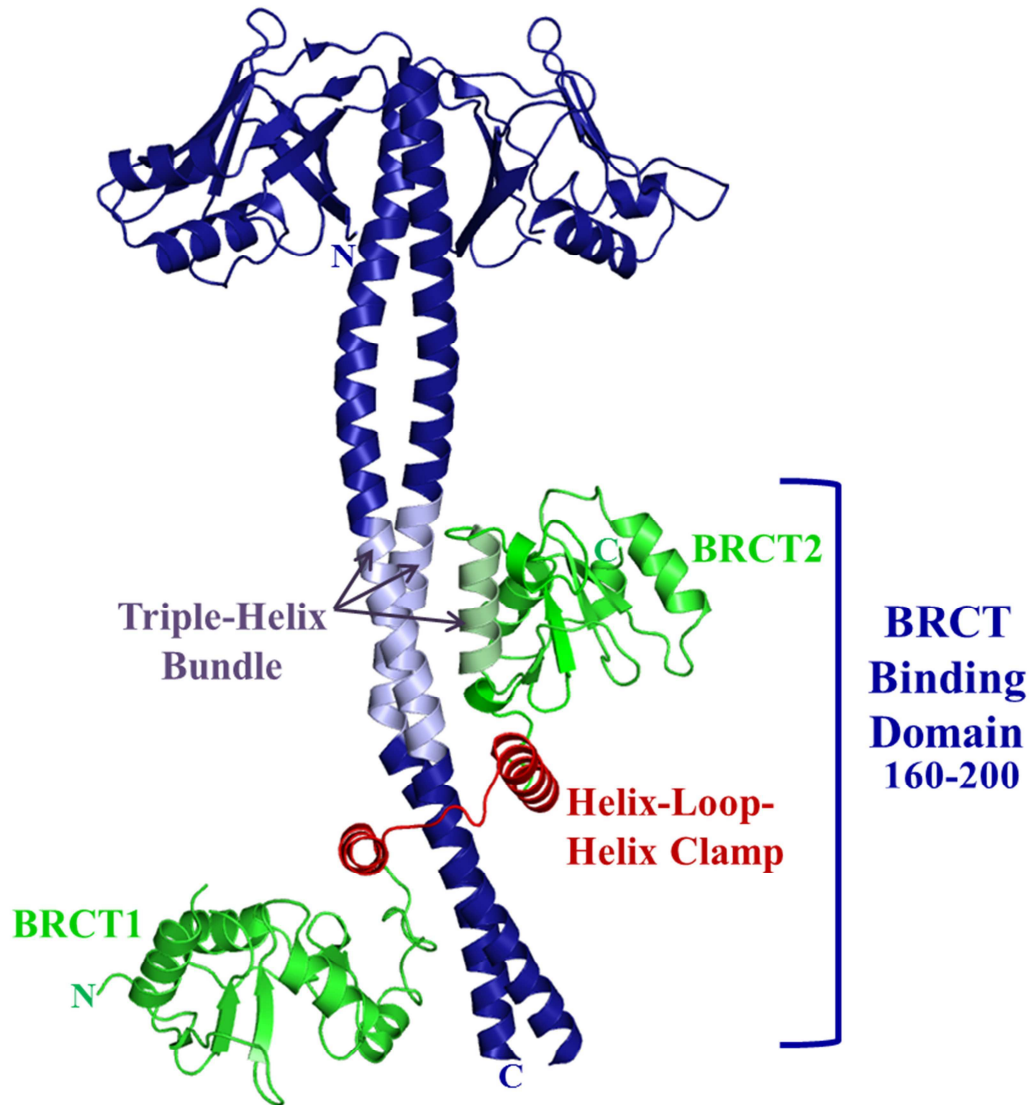


Figure 1.8 Crystal structure of XRCC4 bound to LigaseIV. The crystal structure of XRCC4 (blue) bound to the tandem BRCT domains of LigaseIV (green). XRCC4-LigaseIV interactions include a helix-loop-helix clamp (red), and a triple helix bundle (light purple, light green). The BRCT binding domain of XRCC4 spans amino acids 160-200, as indicated. N- and C-termini are in blue (Wu et al., 2009; PDB 3II6).

between the alpha-amino group of K273 with the α -phosphate of ATP, releasing inorganic pyrophosphate (PPi). The adenosine monophosphate (AMP) is then transferred from K273 to a free 5'-phosphate at a DNA end creating a pyrophosphate bond. Finally, the 3'-hydroxyl group performs a nucleophilic attack on the 5'-phosphorylated DNA end, removing the AMP and covalently joining the DNA strands in a phosphodiester bond (Lehman, 1974). The N-terminal residues of LigaseIV are key to this reaction, as they contain the catalytic core, with the active site residue K273 responsible for ligation (Tomkinson et al., 1991). The catalytic core of LigaseIV is composed of a 3-domain subunit conserved amongst eukaryotic ligases, including LigaseI and LigaseIII. The catalytic core consists of a DNA-binding domain (DBD), a nucleotidyltransferase (NTase) domain and an oligonucleotide/oligosaccharide-binding (OB) domain (Cotner-Gohara et al., 2010; Martin et al., 2002; Murzin 1993; Pascal et al., 2004). While there is currently no structure of LigaseIV's catalytic domain, the catalytic domain structure of LigaseI bound to DNA has been solved. The catalytic domain of LigaseI forms a ring around the DNA, with both the DBD and NTase binding DNA and the active site lysine residing in the NTase domain. The OB domain aids in the transfer of AMP to the active site lysine (Figure 1.9A) (Pascal et al., 2004). Given the conservation of the catalytic domain between human DNA Ligases (I, III, IV), the LigaseI structure (residues 1-607), was used in SWISS-MODEL to predict the structure of the catalytic domain of LigaseIV. The resulting model of LigaseIV suggests that the LigaseIV N-terminus would be able to form a ring around DNA, and function in a similar manner to LigaseI (Figure 1.9B) (Arnold et al., 2006; Guex et al., 1997; Schwede et al., 2003; Shuman et al., 1995). More recently, electron microscopy studies of the full-length XRCC4 bound to full-length LigaseIV suggested that the catalytic subunit of LigaseIV resides near the N-terminal head domains of the XRCC4 dimer. This interaction occurs in a distinct region

A

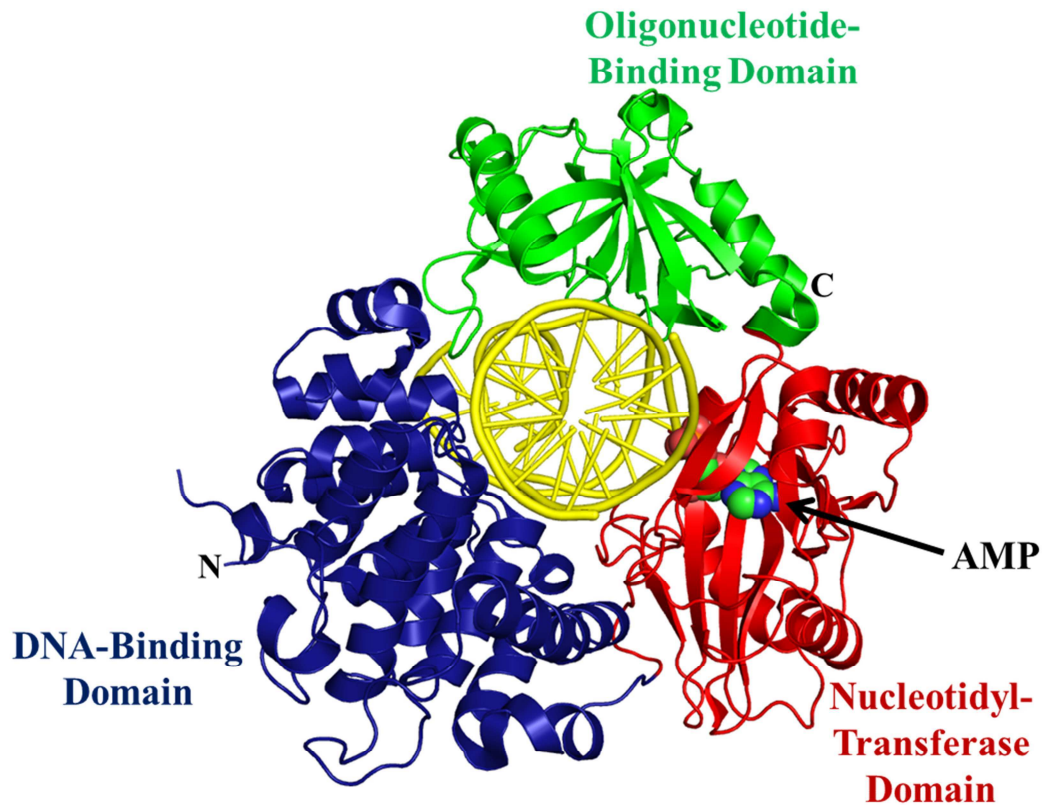


Figure 1.9 Crystal structure of the catalytic domains of human DNA ligases. (A) The crystal structure of human LigaseI bound to a DNA substrate (yellow), includes the DNA-binding domain (blue), a nucleotidyl-transferase domain (red), and the oligonucleotide-binding domain (green). AMP (blue/green spheres) is bound in the nucleotidyl-transferase domain. N- and C-termini are in black (Pascal et al., 2004; 1X9N). *Continued on page 28.*

B

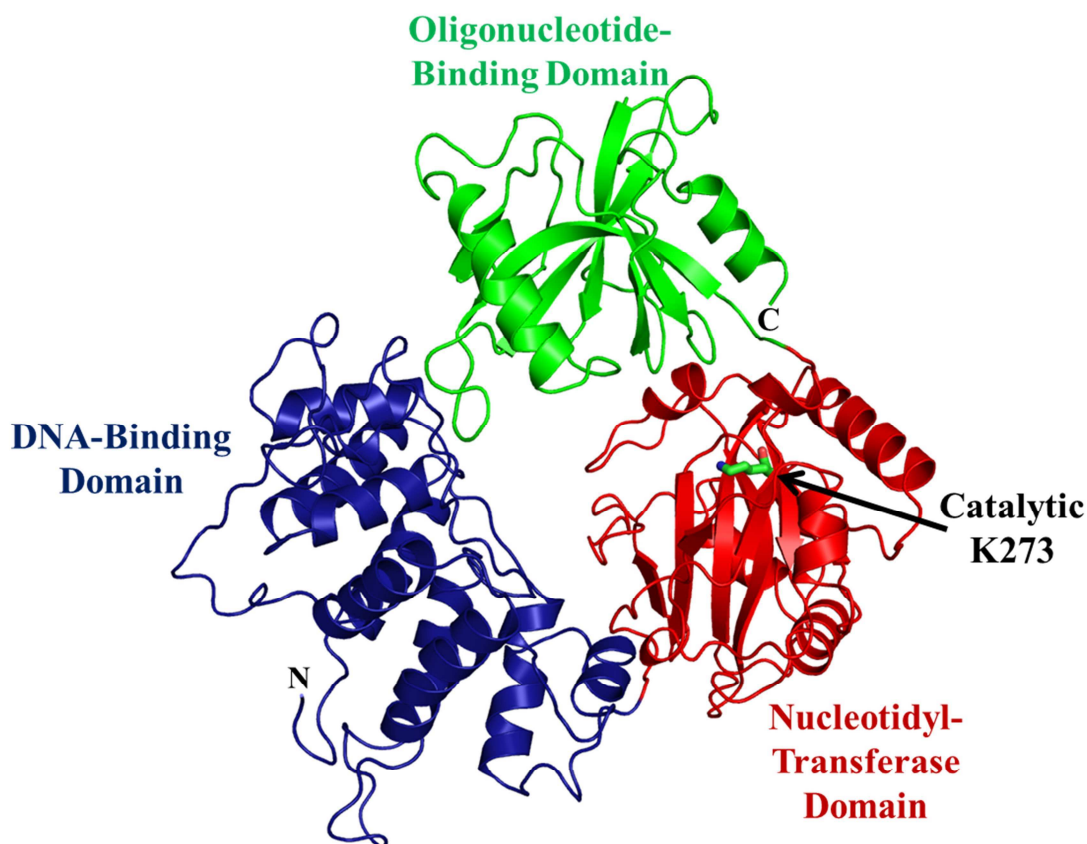


Figure 1.9 Crystal structure of the catalytic domains of human DNA ligases. (B) Model of human LigaseIV catalytic domain, based on LigaseI, and generated using SWISS-MODEL. The predicted DNA-binding (blue), nucleotidyl-transferase (red) and oligonucleotide-binding (green) domains suggest a similar ring structure to LigaseI. The catalytic K273 (green) is located in the nucleotidyl-transferase domain of the model. N- and C-termini are in black (Arnold et al., 2006; Guex et al., 1997; Schwede et al., 2003).

from where the tandem BRCT domains of LigaseIV interact with XRCC4 (Recuero-Checa et al., 2009). From this data, it has been postulated that the LigaseIV catalytic subunit may require movement for function, and that the BRCT domains serve primarily to anchor LigaseIV to XRCC4, while still allowing for movement of the LigaseIV catalytic subunit. Although this model is compelling and able to account for how XRCC4 stimulates LigaseIV activity, further structural analysis is required to truly understand how full-length XRCC4 associates with LigaseIV and stimulates its ligation activity (Recuero-Checa et al., 2009).

The third protein involved in ligation, and the most recently identified member of the NHEJ pathway, is the 299 amino acid protein Cernunnos or XLF, so named for its prediction to be similar in structure to XRCC4, even though sequence identity between the two proteins is low. XLF was identified independently by two groups, through yeast-two hybrid screening for interactions with XRCC4, and identified in human patients exhibiting defective DNA repair phenotypes in conjunction with cDNA functional complementation studies (Ahnesorg et al., 2006; Buck et al., 2006). The primary role identified thus far for XLF is stimulation of LigaseIV ligation activity, however the mechanism is unknown. *In vitro*, XLF binds DNA, and like XRCC4, requires long DNA substrates (>80bp) for stable association (Lu et al., 2007). In the yeast homolog of XLF, DNA binding has been localized to the C-terminus, suggesting a similar location for DNA binding in mammalian XLF (Sulek et al., 2007). XLF has also been shown to promote re-adenylation of LigaseIV after ligation, as well as stimulate ligation of incompatible DNA ends *in vitro* (Riballo et al., 2009; Tsai et al., 2007). Furthermore, it was established that in whole cell extracts lacking XLF, broken DNA ends that required gap-filling by polymerases μ and λ could not be filled, and were only able to do so once XLF was present (Akopiants et al., 2009). Therefore, from data currently available it would

appear that XLF functions in NHEJ primarily through interactions in the final steps of ligation.

1.4 Thesis Objectives and Organization

In the context of NHEJ, XLF's role is unclear, except that it is required and appears to be involved in repairing DNA double-strand breaks with chemically incompatible DNA ends. This thesis takes a structural approach to understanding XLF, its interaction with XRCC4, and how XLF fits into the ligation mechanism. As a 'sandwich' thesis, the following chapters consist of prepared manuscripts describing the research undertaken to answer these questions. Chapter 2 begins with a structural solution of XLF and identification of functional residues involved in DNA binding, XRCC4 interaction, and stimulating LigaseIV. Chapter 3 describes the technical challenges encountered when determining the structure of XRCC4 bound to XLF, while Chapter 4 presents the structure of XRCC4 bound to XLF, the necessity of the XRCC4-XLF interaction in DNA binding, bridging of DNA ends and the significance of the XRCC4-XLF complex to NHEJ. Chapter 5 describes initial work in determining the structure of XRCC4-LigaseIV, and how this complex interacts with XLF. The final chapter (6) presents a summary of all the work presented and discusses its significance to the field of DNA double-strand break repair.

Chapter 2: Crystal structure of human XLF: a twist in non-homologous DNA end joining

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

Reproduced with permission from Elsevier Ltd.

2.1 Author's Preface

The research presented in chapter 2 has been published in the peer-reviewed journal *Molecular Cell*, and appears in its published format. This article presented the crystallographic structure of human XLF for the first time, identified amino acids through which XRCC4 and XLF interact, and identified amino acids required to stimulate ligation in non-homologous end-joining. S.N. Andres conducted crystallization and structural determination of XLF¹⁻²²⁴, along with creation of all XLF mutants. S.N. Andres also carried out all DNA binding and protein interaction analysis on the XRCC4 and XLF mutants, and was involved in writing of the manuscript. Dr. M. Modesti supplied the original XLF clone, along with all expression plasmids of XRCC4 mutants. Dr. M. Tsai and Dr. G. Chu were responsible for the ligation assay and manuscript revision. Dr. M.S. Junop was involved in structural determination of XLF¹⁻²²⁴ and writing of the manuscript.

2.2 Summary

DNA double-strand breaks represent one of the most severe forms of DNA damage in mammalian cells. One pathway for repairing these breaks occurs via non-homologous end-joining (NHEJ) and depends on XRCC4, LigaseIV and Cernunnos, also called XLF. While XLF stimulates XRCC4/LigaseIV to ligate mismatched and non-cohesive DNA ends, the mechanistic basis for this function remains unclear. Here we report the structure of a partially functional 224 residue N-terminal fragment of human XLF. Despite only weak sequence similarity, XLF¹⁻¹⁷⁰ shares structural homology with XRCC4¹⁻¹⁵⁹. However, unlike the highly extended 130 Å helical domain observed in XRCC4, XLF adopts a more compact, folded helical C-terminal region involving two turns and a twist, wrapping back to the structurally conserved N-terminus. Mutational analysis of XLF and XRCC4 reveals a potential interaction interface, suggesting a mechanism for how XLF stimulates the ligation of mismatched ends.

2.3 Introduction

Generation of DNA double-strand breaks pose a serious threat to chromosomal integrity. If left unrepaired, such breaks can generate destabilizing chromosomal rearrangements that may lead to tumorigenesis (Gao, et al., 2000; Lieber, 1998). Double-strand breaks occur in response to exogenous genotoxic agents, and as intermediates in the genomic rearrangements associated with V(D)J recombination. Eukaryotic cells maintain two systems for repairing double-strand breaks: homologous recombination, which facilitates accurate restoration of DNA (Jeggo, 1998), and non-homologous end

joining (NHEJ), which facilitates error-prone repair without a homologous DNA template.

NHEJ requires a core set of five proteins: Ku, DNA-PKcs, XRCC4, LigaseIV and Cernunnos, an XRCC4-like factor hereafter called XLF. Ku initially binds broken DNA ends, preventing nucleolytic degradation (Liang and Jasin, 1996), and then recruits DNA-PKcs, which stabilizes synapsis of the DNA ends (DeFazio et al., 2002). DNA end processing enzymes, including Artemis, enhance NHEJ by making DNA ends compatible for ligation by XRCC4/LigaseIV (Ma et al., 2002; McElhinny et al., 2005). However, in the presence of Ku, XLF stimulates XRCC4/LigaseIV to ligate mismatched and non-cohesive DNA ends (Tsai et al., 2007). XLF can bind DNA and interact with the XRCC4-LigaseIV complex; however, the nature of these interactions and the mechanism by which XLF functions remain poorly understood (Hentges et al., 2006; Callebaut et al., 2006; Ahnesorg et al., 2006; Lu et al., 2007).

XLF was identified as the gene mutated in patients exhibiting immune deficiencies and microcephaly, symptoms consistent with impaired NHEJ (Ahnesorg et al., 2006; Buck et al., 2006). Secondary structure predictions suggested that XLF and XRCC4 are structurally related, despite limited sequence similarity (Ahnesorg et al., 2006). In addition, XLF and XRCC4 exhibit other similarities. Both proteins exist as dimers and have DNA binding activity dependent on the presence of long (> 80 bp) DNA fragments (Hentges et al., 2006; Lu et al., 2007). Although XLF and XRCC4 are able to interact, the nature and functional importance of this quaternary structure remains unknown. Current data suggest that XLF interacts with XRCC4/LigaseIV primarily through contact

with XRCC4, but the role of DNA and LigaseIV in this interaction has not been rigorously evaluated (Deshpande and Wilson, 2007; Lu et al., 2007).

Here we report the crystal structure of a partially functional N-terminal fragment of human XLF¹⁻²²⁴ at 2.5 Å resolution. This fragment fails to efficiently bind DNA or stimulate ligation of non-cohesive DNA ends, but retains the ability to directly interact with XRCC4. The N-terminal region of XLF (amino acids 1-170) adopts a structure almost identical to XRCC4, as predicted from the amino acid sequences (Ahnesorg, et al., 2006; Callebaut et al., 2006). However, amino acids 170-224 diverge greatly from the expected elongated helix seen in XRCC4. Two helices within this region of XLF fold back onto and twist around the main helical stalk of XLF occluding the analogous LigaseIV binding interface observed in XRCC4. The structure of XLF therefore explains why LigaseIV is unable to bind XLF in the same manner observed for XRCC4. Based on the structures of XLF and XRCC4, we then conducted mutational analysis aimed at further defining surfaces involved in the protein-protein interactions. These studies suggest that the N and C-terminal domains of XLF interact with XRCC4 and DNA, respectively. Taken together, our results provide new insight into the assembly of a functional DNA repair complex.

2.4 Results and Discussion

2.4.1 Crystal structure of XLF¹⁻²²⁴ shows formation of a stable dimer

Human XLF contains 299 residues (Ahnesorg et al., 2006; Buck et al., 2006). Comparison with other eukaryotic homologs reveals a high degree of sequence similarity within the N-terminal ~220 residues. Interestingly, this does not extend to the C-terminal

portion of XLF. As expected, similarity within the N-terminal region is most pronounced amongst hydrophobic amino acids (Figure 2.1A). Several clusters of more highly conserved residues, including both hydrophobic and charged/polar amino acids occur in the N-terminal region (57-65, 108-123, 160-186), as well as at a single basic region in the extreme C-terminus (residues 288-295) (Figure 2.1A and C). Based on these observations, we created a truncation to remove the less conserved C-terminal 75 residues of human XLF with the hope of reducing structural flexibility and thereby facilitating crystallization.

Although native XLF¹⁻²²⁴ formed crystals, the quality and resolution of diffraction data were poor; thus, all data used for both structure solution and refinement came from selenomethionine-substituted protein. Experimental phases were determined using single-wavelength anomalous diffraction (SAD). Crystals formed in space group P2₁2₁2₁, with 2 monomers of XLF in each asymmetric unit. Regions of disorder within the crystal prevented modeling of residues 85-92 in chain A and residues 80-92 in chain B. These residues are illustrated as dashed loops in the final structure (Figure 2.1C). The overall structure was refined to R and R_{free} values of 25.3 and 28.7 %, respectively (PDB 2R9A). Table 2.1 contains a complete list of data collection and model refinement statistics.

XLF¹⁻²²⁴ is comprised of two principal domains. The globular N-terminal head domain consists of a 7-stranded anti-parallel β -sandwich with a single helix at the N-terminus and a helix-turn-helix motif inserted between strands 4 and 5 (Figure 2.1B). The loop region connecting α 3 and β 5 is not well-ordered and therefore not present in the final model. Three α -helices combine to form an elongated C-terminal domain, the first

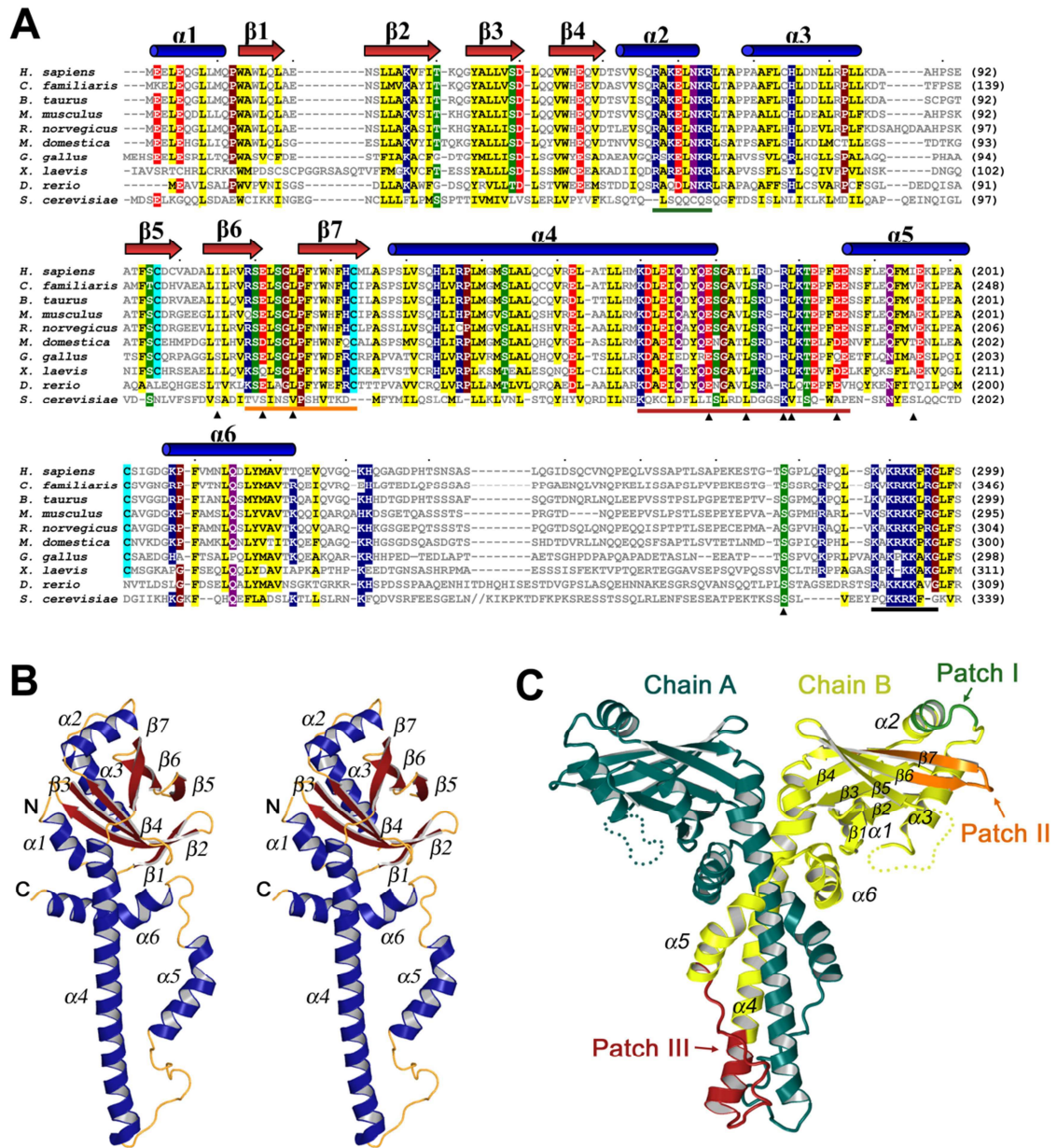


Figure 2.1. Structure and sequence conservation of the N-terminal region of human XLF. (A) Sequence alignment of XLF homologs. Conserved residues are colored as follows: hydrophobic, yellow; negative charge, red; positive charge, blue; proline, glycine, brown; threonine, serine, green; cysteine, light blue; glutamine, asparagine, purple. Regions of highly conserved residues are underlined in green, orange, red, and black. For clarity, XLF residues 249-282 from *S. cerevisiae* are not shown in the alignment. (B) Stereo image of a single XLF¹⁻²²⁴. β -strand and α -helix in red and blue, respectively. (C) XLF dimer observed in crystal asymmetric unit. Chain A and B are shown in teal and yellow. Conserved patches are labeled and colored to correlate with (A).

Table 2.1. Crystallographic data and refinement statistics.

Data Collection	
Wavelength (Å)	0.9800
Space group	P2 ₁ 2 ₁ 2 ₁
Cell parameters (Å)	a=63.41, b=86.93, c=91.87 $\alpha = \beta = \gamma = 90$
Molecules in A.U.	2
Resolution range (Å)^a	25.0 – 2.50 (2.59 – 2.50)
Unique reflections	33,944
Data Redundancy^a	7.3 (7.1)
Completeness (%)^a	99.9 (99.7)
I/σ(I)^a	9.1 (3.0)
R_{merge} (%)^a	15.0 (38.3)
Mosaicity	0.75
Wilson scaling B factor (Å²)	37.0
Model and refinement	
Resolution range (Å)	25.0 – 2.50
R_{work} (%)	25.3
R_{free} (%)	28.7
Refl. observed	14110
Refl. test set	1031
No. of protein atoms	3,464
No. of waters	288
rmsd bond lengths (Å)	0.016
rmsd bond angles (Å)	1.613
Average B factor (Å²)	40.3

^a Statistics for the highest data resolution shell are shown in parentheses.

of which ($\alpha 4$) extends $\sim 60 \text{ \AA}$ away from the N-terminal head. This rather long helix terminates with a conserved glycine at position 171. The following extended coiled region (residues 171-185) reverses the overall trajectory of the polypeptide chain, bringing the second ($\alpha 5$) and third ($\alpha 6$) helices of the C-terminal domain back into contact with $\alpha 4$. The last helix ($\alpha 6$) of the C-terminal domain wedges itself between $\alpha 4$ and $\beta 1$, forcing the N-terminal head domain to adopt an elevated position (Figure 2.1C). In addition, the orientation of $\alpha 6$ suggests that the terminal 75 residues of XLF reside near the N-terminal head domain.

XLF formed a dimer within the crystallographic asymmetric unit. This quaternary structure has also been observed for XLF homologs in *S. pombe* and *S. cerevisiae* (Cavero et al., 2007; Sulek et al., 2007) suggesting that human XLF forms a homodimer in the absence of DNA and other NHEJ proteins. XLF¹⁻²²⁴ monomers are related by a 2-fold non-crystallographic axis, and interact primarily through conserved hydrophobic regions within $\alpha 4$, listed in detail in Supplementary Figure 2.1. The dimer is also held in place by compensatory interactions between $\alpha 4$ of one monomer and $\alpha 5$ of the opposing monomer. Further stabilization occurs from the clamp formed by $\alpha 5$ and $\alpha 6$ as they twist upward and wrap around $\alpha 4$ from the opposite subunit. The dimer interface buries a surface area of $\sim 6500 \text{ \AA}^2$, the greatest contribution coming from interactions that result from the folding back and clamping of $\alpha 5$ and $\alpha 6$ onto $\alpha 4$. The combined effects of burying extensive hydrophobic surface and structural support gained by wrapping $\alpha 5$ and $\alpha 6$ onto $\alpha 4$, suggest that the XLF dimer is highly stable. However, the possibility remains that opposing tails ($\alpha 5$ and $\alpha 6$) from two XLF dimers could interchange to form a tetramer

similar to XRCC4 (Junop et al., 2000). Preliminary data from gel filtration analysis suggests that XLF can form concentration-dependent higher-order complexes (Supplementary Figure 2. 2).

Interestingly, three patches of highly conserved residues in XLF¹⁻²²⁴ do not contribute significantly to the dimer interface (Figure 2.1). Many of the conserved residues in these regions are exposed and likely involved in other interactions. Patch III (residues 160-186) is N-terminal to the region in XLF (residues 186-209) corresponding to the region in XRCC4 that binds to LigaseIV. However, Patch III forms a hinge that folds back to occlude the potential LigaseIV binding surface. Although Patch I and II are non-contiguous in primary structure, these regions form a continuous surface in the tertiary structure of XLF (Figure 2.1C). Many of the conserved residues in these two patches consist of exposed polar/charged residues that are not expected to contribute to protein folding. A stretch of 3 exposed hydrophobic residues (amino acid 115, 116 and 118; see supplemental Figure 2.1, section 2.7) is also found in this region, specifically within the short loop joining β 6 and β 7. These features raise the possibility that the Patch I-II surface mediates interactions with XRCC4 or other NHEJ proteins.

2.4.2 The XLF C-terminal domain is required for DNA binding and mismatched end ligation

To determine the role of the conserved N-terminal region of XLF in NHEJ, we tested XLF¹⁻²²⁴ for its ability to stimulate ligation of mismatched DNA ends, bind DNA, and interact with XRCC4/LigaseIV. Tsai et al., (2007) recently used an *in vitro* end-joining assay to demonstrate that XLF stimulates ligation of non-cohesive and

mismatched ends. Using this assay, we found that full length XLF and XLF¹⁻²²⁴ stimulated ligation of mismatched ends by ~ 300 and 3-fold, respectively (Figure 2.2A). The decreased activity of XLF¹⁻²²⁴ was surprising given the general absence of conservation in the C-terminal 75 amino acids deleted from XLF¹⁻²²⁴. However, the extreme C-terminus of XLF contains a small conserved basic cluster, which was proposed to represent an NLS (Ahnesorg et al., 2006).

We compared the DNA binding activities of XLF¹⁻²²⁴, full length XLF and XRCC4, and found that the ability of XLF¹⁻²²⁴ to bind DNA was dramatically decreased (Figure 2.2B), suggesting that the C-terminal 75 residues of XLF are necessary for DNA binding. In agreement with these findings, biochemical studies of the XLF homolog in *S. cerevisiae* indicate that the analogous C-terminal region is necessary and sufficient for DNA binding (Sulek et al., 2007).

XLF interacts with XRCC4 and the XRCC4-LigaseIV complex (Ahnesorg et al., 2006). Using a two-step approach, we examined the protein-protein interactions retained by XLF when its C-terminal 75 residues are absent. In the first step, individually purified proteins were mixed and then resolved by native PAGE (Figure 2.2C). In the second step, protein bands were excised from the gel and resolved by electrophoresis under denaturing conditions (Figure 2.2D). XRCC4 and Ligase IV⁶⁵⁴⁻⁹¹¹ were included as a positive control, since they form a stable complex (Modesti et al., 2003). When XRCC4 and Ligase IV⁶⁵⁴⁻⁹¹¹ were combined and resolved under native conditions a new band was observed with altered mobility compared to either individual protein on its own (Figure

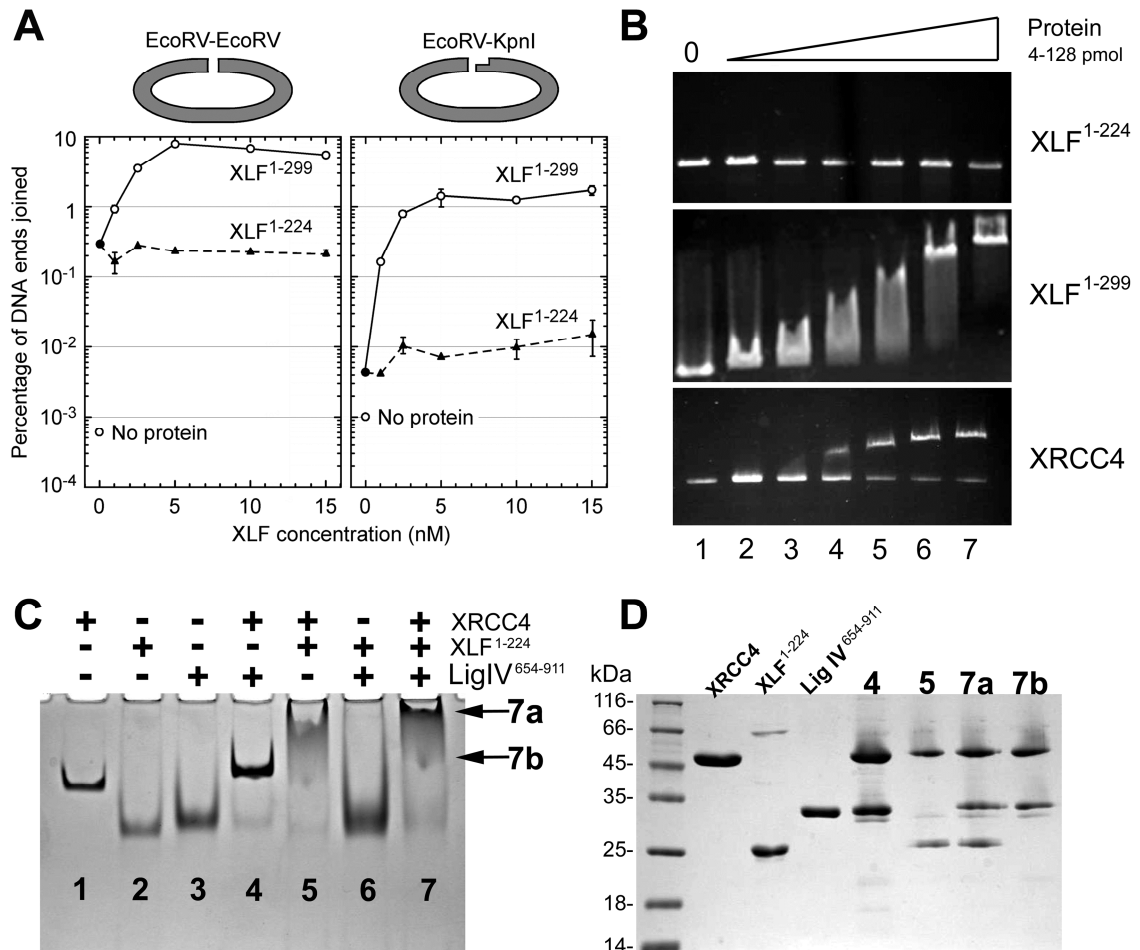


Figure 2.2 Functional analysis of XLF¹⁻²²⁴. (A) Stimulation of non-cohesive end-joining requires the C-terminus of XLF. (B) DNA binding of XLF requires the C-terminal 75 amino acids. XLF¹⁻²²⁴ and/or XRCC4 were tested by mobility gel shift using increasing amounts of protein (2-fold increase starting at 4 pmol in lane 2) and 100 ng of linearized dsDNA (2.6 kbp). (C&D) Protein-Protein Interactions with XLF¹⁻²²⁴, XRCC4, and Ligase IV⁶⁵⁴⁻⁹¹¹. Native PAGE (C). Bands from lane 4, 5 and 7 in (C) were cut out and resolved by SDS-PAGE (D).

2.2C, compare lane 4 to lanes 1 and 3). When this band was resolved under denaturing conditions, both XRCC4 and Ligase IV⁶⁵⁴⁻⁹¹¹ were present as expected (Figure 2.2D, lane 4). We then tested various combinations of XRCC4, Ligase IV⁶⁵⁴⁻⁹¹¹ and XLF¹⁻²²⁴ for their ability to form stable complexes (Figure 2.2C and 2.2D). The conserved N-terminal fragment of XLF¹⁻²²⁴ retains the ability to interact stably with XRCC4 and to the complex of XRCC4 bound to Ligase IV⁶⁵⁴⁻⁹¹¹. On the other hand, XLF¹⁻²²⁴ and Ligase IV⁶⁵⁴⁻⁹¹¹ failed to form a stable complex, consistent with studies examining the interaction between full length XLF and LigaseIV (Lu et al., 2007; Deshpande and Wilson, 2007).

In summary, the C-terminal 75 residues of human XLF are required for stimulation of mismatched DNA ligation and DNA binding, but not for interaction with XRCC4. Interestingly, this region of XLF exhibits little sequence conservation with other homologs, except for the extreme C-terminus.

2.4.3 XLF and XRCC4 structures exhibit similarities and differences

Figure 2.3A presents a structure based sequence alignment of XLF¹⁻²²⁴ and XRCC4¹⁻²¹¹. Although the first 120 residues of XLF and XRCC4 exhibit only 13.4% sequence identity, the predicted secondary structures display almost perfect homology with an overall root mean square deviation of 1.2 Å for the C α atoms (SuperPose, Maiti et al., 2004). However, examination of the entire tertiary structure reveals unexpected and dramatic differences between XRCC4 and XLF. Two key differences become evident upon alignment of XLF and XRCC4 using either the tail domains (Figure 2.3B) or the head domains (Figure 2.3C). The first difference is a large 45° outward rotation of the XLF helical tail domain (Figure 2.3C). This structural difference is due to insertion of α 1

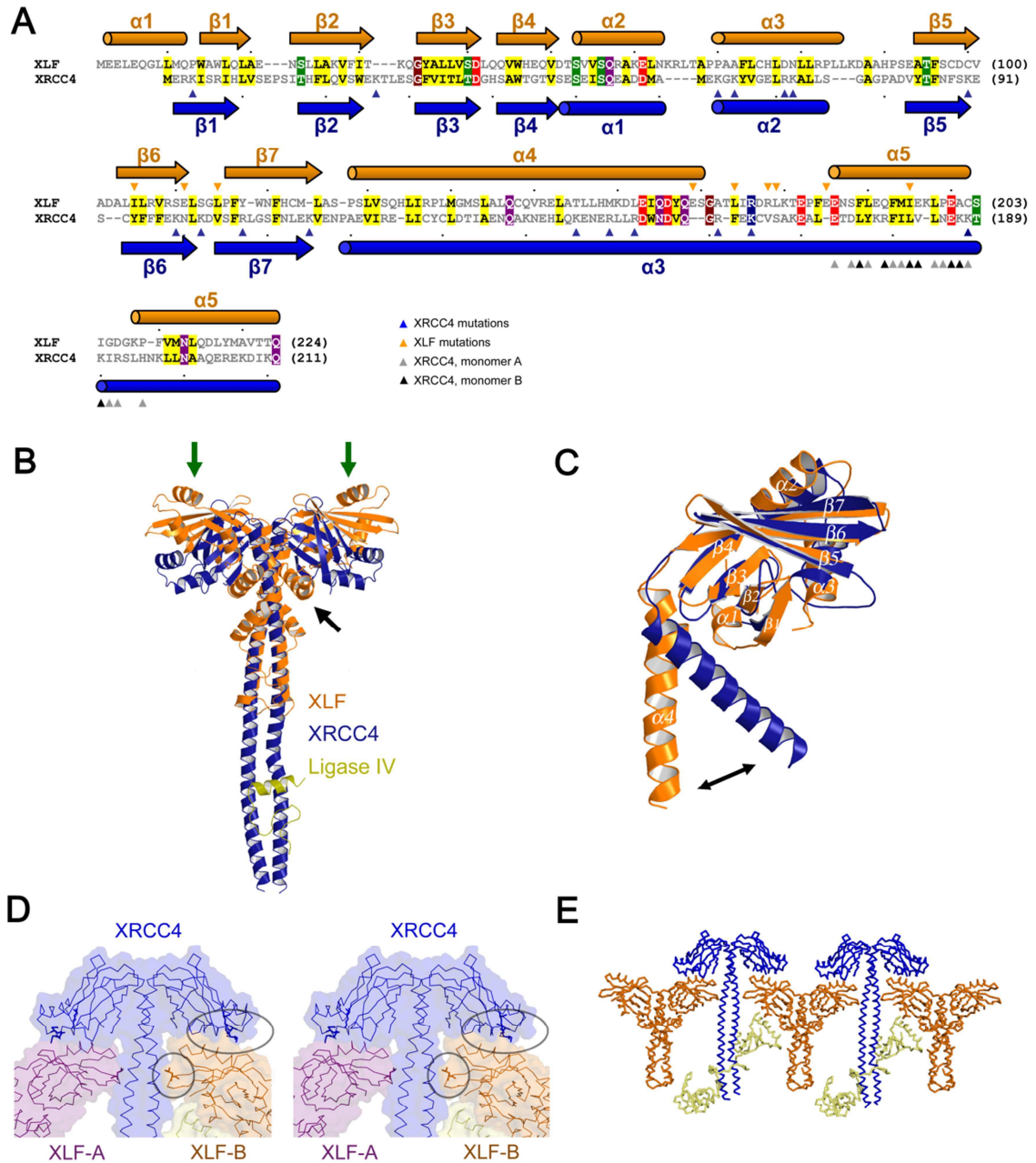


Figure 2.3. Structure-based comparison of human XLF¹⁻²²⁴. (A) Structure-based sequence alignment of XLF¹⁻²²⁴ and XRCC4¹⁻²¹¹. Conserved residues are colored according to the designation in Figure 1. Based on the *S. cerevisiae* structure of Lif1-Dnl4(tandem BRCT), triangles indicate putative residues of XRCC4 that mediate LigaseIV interaction. Reflecting asymmetric interactions in Lif1-Dnl4, triangles are colored either black or grey for interaction with subunits A or B of XRCC4. XRCC4 and XLF mutations are indicated by triangles in blue and orange, respectively.

Continued on page 44.

Continued from page 43.

Figure 2.3. Structure-based comparison of human XLF¹⁻²²⁴. (B) Overlay of XLF¹⁻²²⁴ and XRCC4¹⁻²⁰¹ bound to Ligase IV⁷⁵⁵⁻⁷⁸², in orange, blue, and yellow respectively. The helical tail region of XLF (residues 128-171) was used to structurally align the corresponding region of XRCC4. The structure of XRCC4 bound to LigaseIV (PDB 1Z56) is reported in Dore *et al.*, (2006). (C) Overlay of N- terminal head domains of XLF¹⁻¹⁵² and XRCC4¹⁻¹⁴², in orange and blue, respectively. Arrow indicates a 45° difference in the trajectory of XLF and XRCC4 C-terminal tail domains. (D) Stereo image of XRCC4-XLF interface suggested by mutational analysis. XLF-A and XLF-B are contributed from separate dimers. Mutations inhibiting XRCC4-XLF interaction are circled in grey. (E) Model for assembly of an XRCC4-XLF-LigaseIV filament.

and $\alpha 6$ between the head and stalk domains, elevating both heads within the XLF dimer (Figure 2.3B, black arrow). This creates a flattened, elongated surface ($\sim 25 \text{ \AA} \times 70 \text{ \AA}$) extending from either end of the dimer (Figure 2.3B green arrows). We speculate that the newly formed surface may serve as a binding interface.

The second major difference is the folding of the C-terminal helical domain of XLF. In XRCC4 the C-terminal domain exists as a single helix that extends away from the head domain with a small but important conformational deviation at the point that interacts with LigaseIV (Figure 2.3B). The analogous region of XLF is broken into three helices, with $\alpha 5$ and $\alpha 6$ folding back onto $\alpha 4$ of the opposing dimer subunit. XLF helix $\alpha 5$ aligns with the LigaseIV binding region of XRCC4 (Figure 2.3A). Interestingly, XLF helix $\alpha 5$ is less conserved than Patch III of the XLF hinge (Figure 2.1A), suggesting that $\alpha 5$ might not interact with LigaseIV. In addition, the helical stalk of XLF folds back onto itself to produce a more compact structure with a greatly increased dimer interface. Importantly, the folding and twisting of $\alpha 5$ and $\alpha 6$ back onto $\alpha 4$ prevents LigaseIV from binding to XLF in the same way it binds to XRCC4 (Modesti *et al.*, 2003; Sibanda *et al.*, 2001). This result supports our biochemical data showing that XLF and LigaseIV do not interact strongly (Figure 2.2C), and is consistent with data from other groups (Lu *et al.*,

2007; Deshpande and Wilson 2007). Nevertheless the folded helical region of XLF may still interact with other NHEJ proteins, or with itself to form tetramers, as discussed above.

2.4.4 Model for the XLF-XRCC4-LigaseIV Complex

Recent studies by Deshpande and Wilson (2007) suggest that the head region of human XRCC4 is necessary and sufficient for stable interaction with the N-terminal region of human XLF. We conducted a mutational analysis of both XRCC4 and XLF to define their interacting surfaces in greater detail and thereby provide data for a model of the XRCC4-XLF complex.

We targeted mutations in XRCC4 and XLF to exposed, conserved residues not predicted to be involved in dimerization or LigaseIV binding (Figure 2.3A and Supplemental Figure 2.3, section 2.7). In total, 17 XRCC4 mutants and 10 XLF mutants were analyzed for their ability to bind XLF, XRCC4, LigaseIV and DNA. XLF mutants were also tested for stimulation of mismatched end ligation. Figure 2.4 summarizes the results. While most mutations did not alter function, amino acid substitutions at K63, K65 and K99 of XRCC4 disrupted interaction with XLF. These mutations nevertheless preserved full LigaseIV and DNA binding activity. Within the tertiary structure of XRCC4, these residues cluster at the base of the head domain (Figure 2.3D). Of the 10 XLF mutants, three (L174A, R178A, L179A located in Patch III) were shown to be essential for XLF activity, but not for binding to DNA or XRCC4. This phenotype is consistent with these residues mediating interaction with Ku70/80 and/or LigaseIV. Only a single XLF mutant,

XRCC4	Binding			Stimulation of joining
	XLF	LIGIV	DNA	EcoRV-KpnI
Wildtype	+	+	+	nd
K90E	+	+	+	nd
R107E	+	+	+	nd
K115E	+	+	+	nd
K4E, K99E	-	+	+	nd
K26E, K99E	-	+	+	nd
K26E, K102E	+	+	+	nd
K65E, K99E	-	+	+	nd
K65E, K102E	-	+	+	nd
R71E, K99E	-	+	+	nd
R71E, K102E	+	+	+	nd
K72E, K99E	-	+	+	nd
K72E, K102E	+	+	+	nd
K99E, K102E	+/-	+	+	nd
R161Q, K164Q	+	+	+	nd
K146Q, R150Q R153Q	+	+	+	nd
K63Q, K65Q R71Q, K72Q	-	+	+	nd
K63Q, K65Q R71Q, K72Q K102E	+/-	+	+	nd
XLF	Binding			Stimulation of joining
	XRCC4	LIGIV	DNA	EcoRV-KpnI
Wildtype	+	-	+	115
Δ 224	+	-	-	3
I105S	+	nd	+	70
E111A	+	nd	+	48
L115A	-	nd	+	6
E169A	+	nd	+	37
L174A	+	nd	+	24
R178A	+	nd	+	7
L179A	+	nd	+	8
E185A	+	nd	+	89
I195A	+	nd	+	122
S278A	+	nd	+	165

Figure 2.4. Mutational analysis of XRCC4 and XLF binding surfaces. XRCC4 and XLF mutants either bound (+) or did not bind (-) DNA, DNA LigaseIV, XLF or XRCC4. Stimulation of end joining, illustrated as a bar graph, was measured from 0 to 150-fold for the joining of an *EcoRV-KpnI* digested DNA substrate.

L115A, disrupted XRCC4 binding. The L115A mutant retained DNA binding activity and showed no alteration in the circular dichroism spectra, suggesting that the L115 residue is important for protein-protein interaction, but not for the folding of XLF (data not shown). L115 is located within the small cluster of three exposed hydrophobic residues found in the β 6- β 7 loop of Patch II (Figure 2.1C and 2.3D). Mutations that disrupted the interaction between XRCC4 and XLF were all located in the head domains of XRCC4 and XLF. This result suggests that the head domains are essential for interaction, in agreement with the conclusions of Deshpande and Wilson (2007). As shown in Figure 2.3D, the β 6- β 7 loop region of XLF is predicted to interact with the α 3 helices of XRCC4 (residues 138-150). Conversely, residues K63, K65 and K99 of XRCC4 appear to make contacts with the conserved Patch I and II regions of XLF. Further investigation is required to verify these interactions.

Our structural and biochemical data further define the XRCC4-XLF interacting regions and suggest a working model for how these proteins assemble during NHEJ. The most probable way of bringing XLF into contact with XRCC4, without introducing steric clashes with the BRCT domains of LigaseIV, is to assemble the proteins in a stacked head to head fashion (Figure 2.3D). This model satisfies the current biochemical data indicating that head regions of XRCC4 and XLF interact with each other, and conforms to the structural restrictions imposed by shape complementarity. The model also predicts an interaction of XLF with the C-terminal BRCT domains of LigaseIV (Figure 2.3D and 2.3E). Such an interaction would be weaker than the highly stable binding mode between XRCC4 and LigaseIV, in agreement with a recent report (Deshpande and Wilson, 2007).

In this model, an XRCC4 dimer potentially binds two dimers of XLF (Figure 2.3D). Furthermore, a single dimer of XLF potentially binds two dimers of XRCC4, thus generating a continuous structure or filament of alternating XRCC4 and XLF dimers (Figure 2.3E). Formation of such a filament would be expected to cooperatively stabilize binding of XRCC4 and XLF to DNA. In support of this proposed arrangement of XLF and XRCC4, these proteins have been shown to exhibit highly cooperative DNA binding and are only able to interact stably with unusually large DNA substrates (~100 bp) (Modesti et al., 1999; Lu et al., 2007; Sulek et al., 2007).

To explain how XLF stimulates ligation of blunt and mismatched ends, we propose the following model. Ku stabilizes binding of the first XRCC4/LigaseIV complex to the DNA ends (McElhinny et al., 2000). For cohesive ends, alignment by base pairing would facilitate immediate ligation by LigaseIV even in the absence of XLF. For blunt or mismatched ends, an XLF/XRCC4 filament would assemble on the DNA. Both XLF and XRCC4 can bind to internal sites on DNA (Figure 2.2B), and assembly of the XLF/XRCC4 filament would stabilize DNA binding. The DNA ends could slide with respect to each other within the filament. Since each XLF and XRCC4 subunit of the filament includes a DNA binding domain, juxtaposition of the DNA ends would optimize the energy of DNA binding. Mismatched or blunt ends could slide into a position that would permit LigaseIV to join one or both strands, depending on the structure of the DNA ends (Tsai et al., 2007). We are currently conducting additional experiments to confirm this model.

2.5 Experimental Procedures

2.5.1 Protein expression and purification

See supplemental section 2.7 for details of vector construction. Purification was achieved by Ni²⁺ affinity and gel filtration chromatography (see supplemental section 2.7 for further details). A description of XRCC4 and XLF mutagenesis and mutant purification is given in Supplemental Data, section 2.7.

2.5.2 Electrophoretic Mobility Shift Assays

DNA binding reactions were assembled with EMSA buffer (20 mM Tris, pH 8.0, 50 mM KCl, 0.1 mM dithiothreitol, 10 µg/mL BSA, 5% glycerol), 100 ng of *HindIII* (NEB) linearized pUC-18, and varying amounts of protein (Figure 2.2B). Reactions were resolved by electrophoresis on a 0.8% TBE agarose gel, for 1 hour at 80 V.

2.5.3 2D Gel Analysis of Protein-Protein Interactions

Purified proteins were combined as indicated in Figure 2.2B in buffer (20 mM Tris, pH 8.0, 50 mM KCl, 0.1 mM DTT, 10 µg/mL BSA, 5% glycerol). Reactions were resolved by 6% native PAGE in TBE. Shifted gel bands were excised and boiled in SDS-PAGE gel loading dye for 5 minutes prior to 12% SDS-PAGE gel electrophoresis.

2.5.4 Crystallization and Data collection of XLF¹⁻²²⁴

Crystals were grown using the hanging drop vapor diffusion method. Equal volumes of protein (3.4 mg/mL XLF, 150 mM KCl, 10 mM Tris pH 8.0, 5 mM DTT, 1 mM EDTA) and crystallization solution (0.1 M Hepes, pH 7.5, 20% PEG 10000, 200 mM NDSB-201) were dehydrated over 800 µL of 1.8M (NH₄)₂SO₄. Crystals (700 x 50 x 20

μm) grew after 1-2 days at 4°C. Diffraction data was collected at NSLS, X8C (Brookhaven, NY).

2.5.5 Structure Determination and Model Refinement

SAD data was processed using *d*TREK* (Pflugrath, 1999) to 2.5 Å. All 18 SeMet sites were located using HYSS (Adams et al., 2002; Grosse-Kunstleve and Adams, 2003). Phasing and density modification were carried out with *CNS* (Brunger et al., 1998). Iterative rounds of manual model building and refinement were carried out using WinCoot (Emsley and Cowtan, 2004) and REFMAC (Murshudov, 2005), until R and Rfree values converged (Jones et al., 1991). Structural figures were generated using PyMOL (DeLano, 2002).

2.5.6 In vitro end-joining reactions

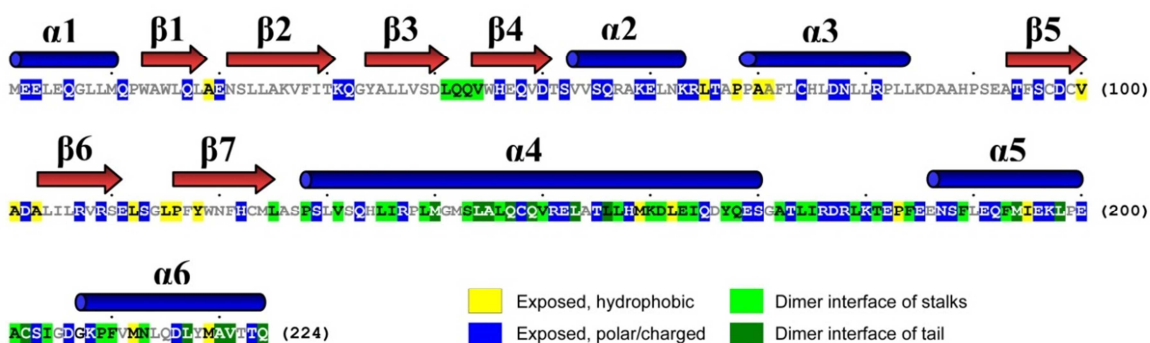
End-joining reactions were performed as previously described using *EcoRV-EcoRV* or *EcoRV-KpnI* substrate (Tsai et al., 2007). See supplemental section 2.7 for further details.

2.6 Acknowledgements

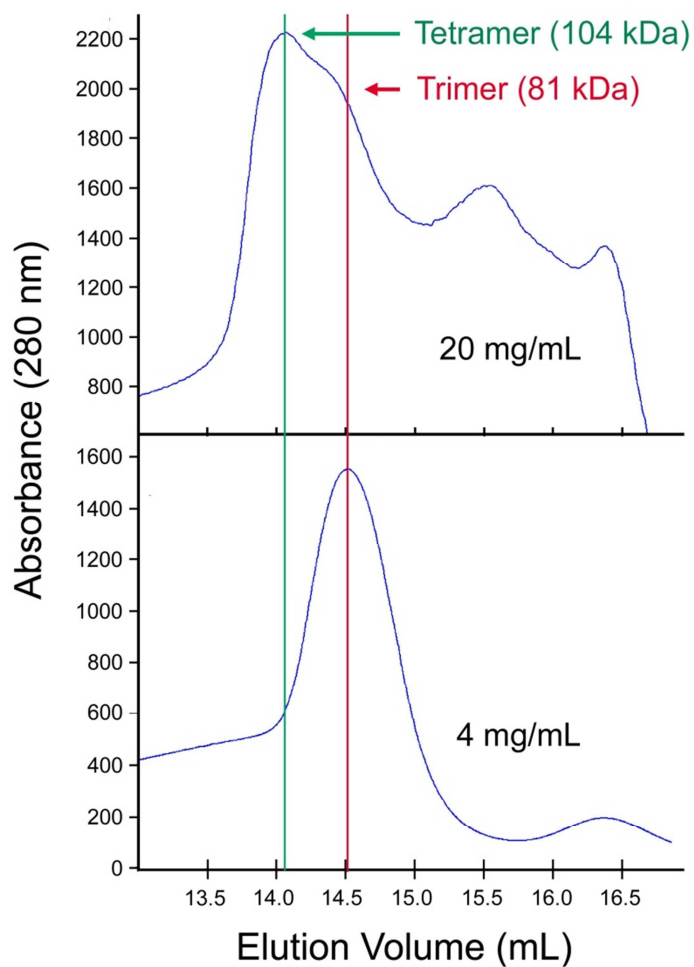
We wish to thank Leonid Flaks for his assistance in data collection at X8C, NSLS, Brookhaven, NY. This work was funded by a grant from the Canadian Institutes of Health Research to MSJ (MOP-53209) and an NSERC fellowship to SNA.

2.7 Supplemental Data

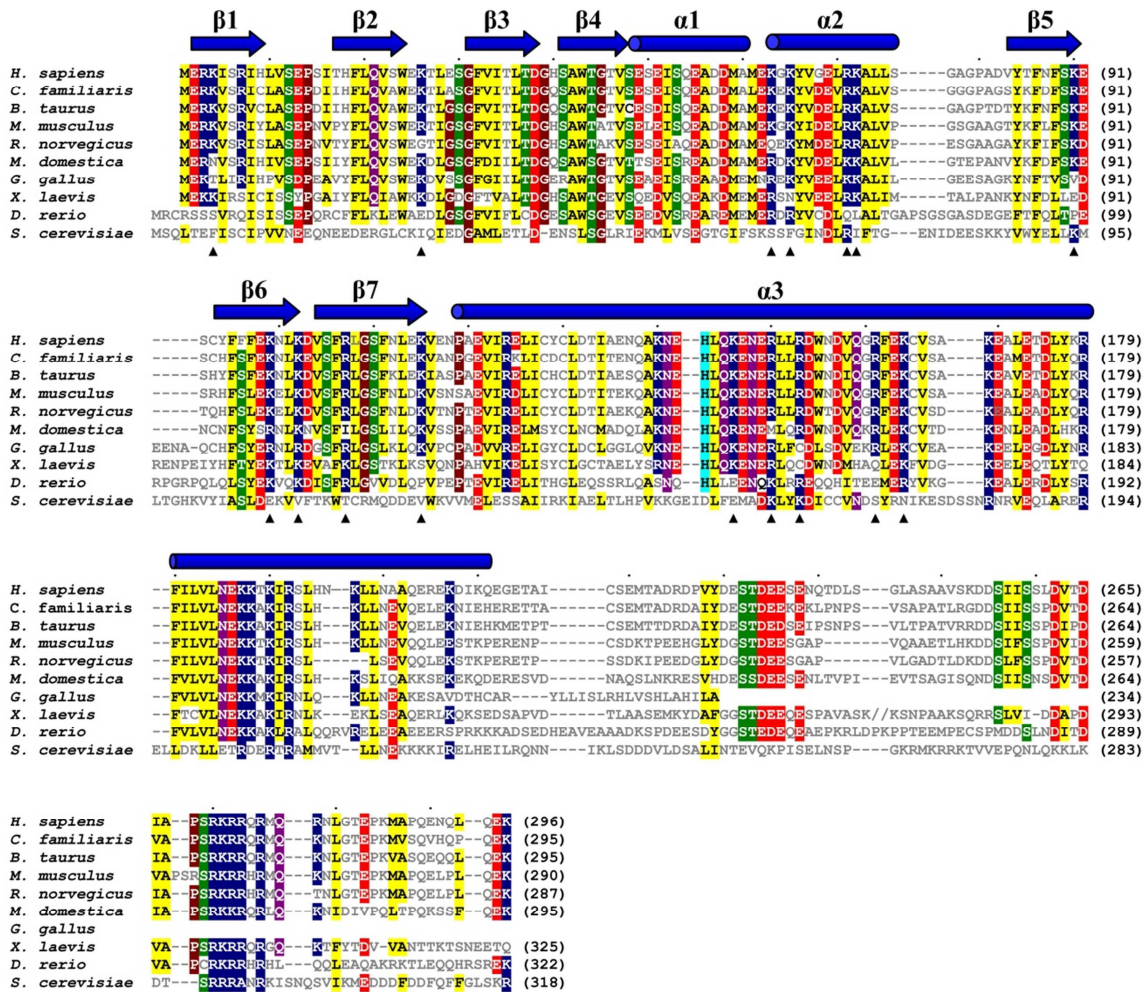
XLF¹⁻²²⁴



Supplemental Figure 2.1. Schematic of XLF residues. Fully exposed hydrophobic or charged/polar residues are colored yellow and blue, respectively. Residues mediating interactions of the XLF dimer interface are indicated in green. Light green: residues forming interface between opposing $\alpha 4$ helices of helical stalk/tail region. Dark green: residues mediating dimer interface formed by $\alpha 5$ and $\alpha 6$ interacting with $\alpha 4$ of the opposing subunit.



Supplemental Figure 2.2. Concentration-dependent oligomeric states of XLF. Gel filtration analysis of XLF¹⁻²²⁴ at ~20 and 4 mg/mL. A shift in the elution volume with different concentrations of XLF is indicated by green and red lines. Apparent molecular weights for the two highest peaks are provided.



Supplemental Figure 2.3. Structure and sequence conservation of human XRCC4. Sequence alignment of XRCC4 homologs. Conserved residues are hi-lighted as follows: hydrophobic, yellow; negative charge, red; positive charge, blue; proline and glycine, brown; threonine and serine, green; histidine, light blue; glutamine and asparagine, purple. Mutations made for functional analysis in XRCC4 are indicated by black triangles.

2.7.1 Supplemental Experimental Procedures

XLF expression vector construction

Human XLF cDNA was amplified by PCR using IMAGE clone 4026495 as template and primers CCCACGGCATATATACTATCTATGGAAGAACTGGAGCAAGGCCTG and GGGCTCGAGTAACTGAAGAGACCCCTTGGCTTCTTC. The PCR product was digested with BceAI and XhoI and ligated into pETDuet-1 (Novagen) using available NdeI and XhoI restriction sites to generate the expression vector encoding full length XLF, pMJ4489. The Gateway System (Invitrogen) was then used to create pMJ4441, a pDEST expression vector encoding XLF(amino acids 1-224) with a C-terminal hexahistidine tag immediately following residue 224. pMJ4441 was fully sequenced and found to contain a silent mutation in the XLF sequence at codon 183 of CCA to CCG. Mutants of XRCC4 and XLF were generated using the QuikChange Kit from Stratagene. All mutations were verified by DNA sequence analysis.

Mutant XLF purification.

Mutant XRCC4 was purified as previously described (Junop et al., 2000). Wildtype and mutant full length XLF were expressed in Rossetta(DE3)/pLysS cells induced with 1 mM IPTG for 4 hours at 37°C. Harvested cells were resuspended in buffer C (20 mM Hepes 8.0, 2 M NaCl, 10 mM imidazole, 10 mM β -mercaptoethanol), lysed, clarified, and applied to a Ni²⁺ charged IMAC column (Amersham), equilibrated with

buffer C. Bound XLF was washed successively with 25 ml buffer C supplemented with 40 and 90 mM imidazole. XLF was step-eluted with buffer C containing 240 mM imidazole and exchanged into buffer D (20 mM Hepes 8.0, 150 mM NaCl, 1 mM EDTA, 10 mM β -mercaptoethanol).

Chapter 3: Crystallization and preliminary X-ray diffraction analysis of the human XRCC4-XLF complex

Andres, S.N. and Junop, M.S. (2011). Crystallization and preliminary x-ray diffraction analysis of the human XRCC4-XLF complex. *Acta Cryst* **F67**, in press.

Reproduced with permission from the International Union of Crystallography (IUCr).

3.1 Author's Preface

The research presented in chapter 3 has been accepted to the peer-reviewed journal *Acta Crystallographica Section F: Structural Biology and Crystallization Communications*, and appears in its published format. This article details the methodology involved in crystallizing the XRCC4-XLF complex. S.N. Andres conducted all of the described experimental work. S.N. Andres and Dr. M.S. Junop were involved in writing of the manuscript.

3.2 Synopsis

A complex of human DNA repair proteins XRCC4 and XLF were co-crystallized under high salt and extreme dehydration conditions to produce diffraction to 3.9Å resolution. Initial phasing information was obtained from molecular replacement with single-wavelength anomalous diffraction using tantalum bromide clusters.

3.3 Abstract

XRCC4 and XLF are key proteins in the repair of DNA double-strand breaks through non-homologous end-joining. Together, they form a complex that stimulates ligation of double-strand breaks. Owing to the suggested filamentous nature of this complex, structural studies via X-ray crystallography have proven difficult. Multiple truncations of XLF and XRCC4 proteins were co-crystallized, but yielded low-resolution diffraction (~20Å). However, a combination of micro-seeding, dehydration and heavy metals improved diffraction of XRCC4^{Δ157}/XLF^{Δ224} crystals to 3.9Å. Although molecular replacement alone was unable to produce a solution, when combined with anomalous signal from tantalum bromide clusters initial phasing was successfully obtained.

3.4 Introduction

DNA double-strand breaks are a serious threat to chromosomal stability, and when left unrepaired cause genomic rearrangements or cell death. Mammals have two distinct pathways for repair of DNA double-strand breaks: homologous recombination and non-homologous end-joining (NHEJ). Of these, NHEJ is the primary repair method owing to

its unrestricted use throughout the cell cycle. NHEJ requires a core set of 7 proteins for binding, processing and ligating broken DNA ends (reviewed in Lieber *et al.*, 2010).

Two of these proteins, XRCC4 and XLF, have no known enzymatic function, yet are essential for repair, as evidenced by XLF^{-/-} and XRCC4^{-/-} mammalian cells, which display severe defects in double-strand break repair (Giaccia *et al.*, 1990; Zha *et al.*, 2007). XLF and XRCC4 have been shown to directly interact with one another and are responsible for stimulating LigaseIV to repair DNA ends (Ahnesorg *et al.*, 2006; Tsai *et al.*, 2007). However, the mechanism by which this occurs is unknown. XRCC4 and XLF are structural homologs, with each existing as a homodimer. Both proteins contain an N-terminal head and an extended C-terminal tail domain. Each head domain is comprised of a seven-stranded anti-parallel β -sheet, interrupted by a helix-turn-helix motif between strands 4 and 5. The C-terminal α -helical tail extends away from the base of the head domain and constitutes the primary dimerization interface. In XRCC4 this tail region remains fully extended while in XLF it wraps back up and around towards the head domain (Junop *et al.*, 2000; Andres *et al.*, 2007; Li *et al.*, 2008).

The head domains of XLF and XRCC4 are required for association with one another. Initial mutational studies that identified interacting surfaces between these proteins suggested that the oligomeric state of the XRCC4-XLF complex may consist of an extended filament (Andres *et al.*, 2007; Malivert *et al.*, 2010). This idea was further supported by recent data from small-angle x-ray scattering (Hammel *et al.*, 2010). Given the flexible and filamentous nature of the XRCC4-XLF complex, structural studies via X-

ray crystallography have proven challenging. Here we report the successful crystallization and initial diffraction of XRCC4-XLF crystals to 3.9Å resolution.

3.5. Materials and Methods

3.5.1 Molecular Cloning and Protein Expression

Full-length XRCC4, XRCC4^{Δ265}, XRCC4^{Δ202}, XRCC4^{Δ202 A60E}-BRCT, full-length XLF, and XLF^{Δ224} were cloned and expressed as previously described (Table 3.1; Modesti *et al.*, 1999; Junop *et al.*, 2000; Andres *et al.*, 2007; Wu *et al.*, 2009). XRCC4^{Δ157} was generated using the Gateway Cloning system (Invitrogen, Canada), and expressed identically to full-length XRCC4. XRCC4^{Δ136} was created by inserting a stop codon after residue 136 in XRCC4^{Δ157} using QuikChange mutagenesis (Stratagene, USA) and expressed using the M9 SeMet growth media kit (Medicilon Inc., People's Republic of China).

Table 3.1 Details of constructs used in crystallization trials

Protein	Mutation	Plasmid Backbone	Affinity Tags	Expression Cell Line
XRCC4	Full-length	pACYC184	C-term His6	BL21 (DE3)
XRCC4	$\Delta 265$	pET-28a	C-term His6	BL21 (DE3)
XRCC4	$\Delta 202$	pACYC184	None	BL21 (DE3)
XRCC4	$\Delta 157$	pDEST14	C-term His6	BL21 (DE3)
XRCC4	$\Delta 136$	pDEST14	None	BL21 (DE3)
XRCC4	$\Delta 202$, A60E	pACYC184	None	BL21 (DE3)
XLF	Full-length	pET-Duet1	C-term His6	Rosetta pLysS (DE3)
XLF	$\Delta 224$	pDEST14	C-term His6	Rosetta (DE3)
Ligase IV tandem BRCTs	654-911	pPRO-EXb	N-term His6	BL21 (DE3)

3.5.2 Purification

All constructs of XLF and XRCC4 were purified as previously described. His-tag fusions were not removed from any of the expressed proteins (Andres *et al.*, 2007; Junop *et al.*, 2000). XRCC4 ^{$\Delta 136$} purification differed from that of wild type XRCC4 as follows: after nickel affinity purification, XRCC4 ^{$\Delta 136$} was loaded onto a 5mL HiTrap Q HP, followed by a 5mL HiTrap SP HP (GE Healthcare, USA), both equilibrated with 20mM Tris (pH 8), 10mM DTT, 10% glycerol, 1mM EDTA and 150mM KCl. XRCC4 ^{$\Delta 136$} did not bind to either column and was collected in the unbound fraction. Following cation exchange, XRCC4 ^{$\Delta 136$} was further purified via gel filtration (HiLoad 16/60 Superdex 200; GE Healthcare, USA) using cation exchange buffer at 200mM KCl.

3.5.3 General Crystallization

XRCC4 and XLF were mixed in varying ratios: 25:50, 50:100, or 50:50 μM XLF:XRCC4. Crystallization was performed using hanging-drop vapour diffusion, combining 1 μL each of protein and crystallization solutions. Crystallization solutions from commercially available kits were used (Classics I and II, Ammonium Sulfate, PEGS, pH Clear I, Nucleix and JCSG I, II, III and IV Suites from Qiagen, Canada; Index Screen from Hampton Research, USA; Extension, Cryo, Membrane, and Low Ionic Screens from Sigma-Aldrich, Canada; Original Screen from Biogenova, USA). Crystallization trials were initially performed with well solutions of 2.5M ammonium sulfate (800 μL) and incubated at 277, 293 or 303 K. Other well solutions tested included 2.5-4M ammonium sulfate, 1.5-4M sodium chloride and 20% PEG3350. Crystals were further optimized by systematically varying each component of the primary crystallization condition. Additives were included during the optimization of initial crystallization conditions (Opti-Salts screen from Qiagen, USA; Silver Bullets and Additive screen from Hampton Research, USA).

3.5.4 Crystallization and Diffraction Collection

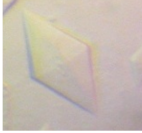




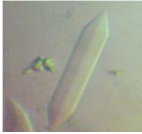
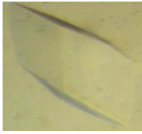

Crystals of XRCC4 ^{$\Delta 157$} -XLF ^{$\Delta 224$} grew from a combination of 1 μL XRCC4 ^{$\Delta 157$} (100 μM) and XLF ^{$\Delta 224$} (50 μM) in 20mM Tris pH 8, 200mM KCl, 1mM EDTA, 10mM DTT and 10% glycerol. The protein solution was combined with 0.8 μL of 1.8 M tri-ammonium citrate (pH 8) containing varying dilutions of crushed XRCC4 ^{$\Delta 157$} -XLF ^{$\Delta 224$} crystals, and 0.2 μL each of 0.1M barium chloride dihydrate and 2.0M sodium

thiocyanate. Hanging drops were initially dehydrated over 2.5M ammonium sulfate pH 7 at 303 K. After 24 hours, crystal trays were moved to 293 K, and after a further 24 hours, to 277 K. Four days later, crystals were further dehydrated over 4M ammonium sulfate. Five days later, crystals were soaked in a combination of 1 μ L of 0.5mM tantalum bromide and 0.5 μ L of 60% PEG 8000 for 3 hours prior to flash-cooling in liquid nitrogen. Diffraction data were collected on NSLS beamline X25 (Brookhaven, New York, USA), to a resolution of 3 \AA , using a wavelength of 1.2536 \AA , in a nitrogen stream at 100 K. Data were collected in 30 $^\circ$ wedges, with 0.5 $^\circ$ oscillation and 2s exposure per image. Initial scaling and space-group determination were performed using *HKL-2000* (Otwinowski and Minor, 2007).

3.6 Results and Discussion

XRCC4-XLF complexes comprised of varying protein lengths produced crystals with similar hexagonal rod morphologies under many crystallization conditions (Table 3.2). Additives were essential for reproducing the initial crystals. Diffraction was highly dependent on crystal size and only observed from crystals >0.4mm in length. Excess XRCC4 promoted initial crystal growth, while micro-seeding of the crystallization solution and incubation at 303 K controlled the extent of nucleation and increased growth in all 3 dimensions. Extended incubation at 303 K (>24 hours) produced larger crystals but with weaker diffraction (>20 \AA). Therefore, the incubation temperature was slowly decreased from 303 to 277 K, greatly improving resolution to 6-8 \AA .

Table 3.2 Crystals of XRCC4-XLF complexes and associated diffraction

XRCC4	XLF	Crystal of Complex	Diffraction
Full-length	Full-length		>20 Å
$\Delta 265$	Full-length		~8 Å
$\Delta 265$	$\Delta 224$		>20 Å
$\Delta 265$, A60E + BRCT	Full-length		>20 Å
$\Delta 202$	Full-length		>20 Å
$\Delta 202$	$\Delta 224$		>20 Å
$\Delta 157$	$\Delta 224$		~3 Å
$\Delta 136$	$\Delta 224$		~4 Å

Initial diffraction of XRCC4^{Δ157}-XLF^{Δ224} crystals was limited to 6-8Å and was further hampered by the presence of ice rings owing to insufficient cryoprotectant. As observed with other protein crystals, dehydration of XRCC4^{Δ157}-XLF^{Δ224} crystals improved cryoprotection (Heras *et al.*, 2003). The extent and duration of dehydration were optimized. Extreme dehydration was achieved by changing the well-solution to 4M ammonium sulfate and by soaking crystals in 60% PEG 8000. This not only improved cryoprotection, but also increased the diffraction resolution to ~3Å. Recently, Hammel *et al.* (2011) published a similar XRCC4-XLF structure to 3.9Å resolution, also using dehydration by PEG 3350 to achieve crystals, with a solvent content identical to that of the XRCC4^{Δ157}-XLF^{Δ224} crystals discussed here (~70%). Therefore, this effect may be more successful for crystals of high solvent content.

Data were collected on a microfocus beamline (X25 at NSLS) allowing multiple data sets to be collected from different regions of a single crystal. The crystals diffracted to 3Å resolution; however owing to anisotropic behaviour data were only processed to 3.9Å resolution (Figure 3.1). The crystals belonged to spacegroup C2, with unit cell parameters $a = 745.4$, $b = 149.6$, $c = 80.5$, $\beta = 94.7^\circ$ (Table 3.3). Alternately, Hammel *et al.* (2011) produced crystals that belonged to space group P6₅22, which may be the result of using a shorter XRCC4 truncation (1-140). However, these crystals also exhibited a long unit cell axis of 764 Å. The extremely long unit cell axis accounts for the anisotropy and is the result of limited lateral crystal contacts in the extended repeating unit (the structure will be discussed elsewhere; PDB entry 3rwr; Sheriff and Hendrickson, 1987).

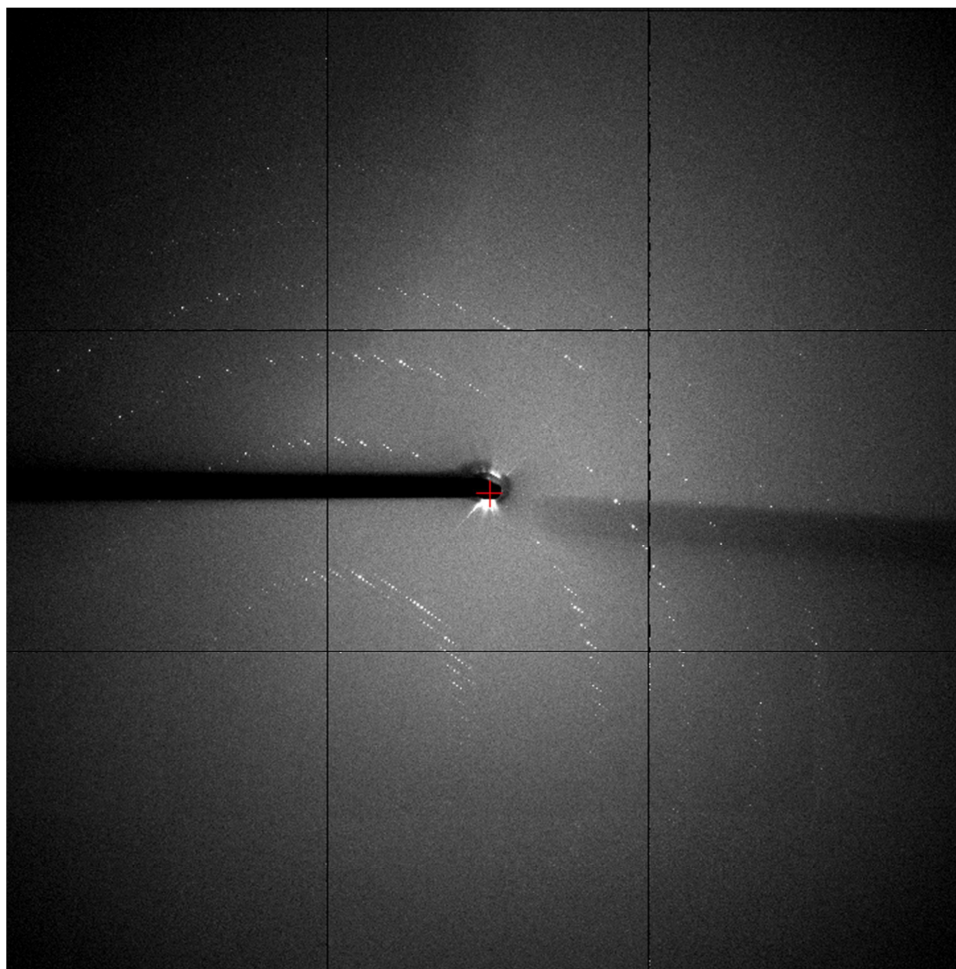


Figure 3.1 Diffraction image of XRCC4^{Δ157}-XLF^{Δ224} crystals illustrating anisotropy.

Table 3.3 Crystallographic data statistics

Data Collection	
Spacegroup	C2
Wavelength (Å)	1.25
Cell parameters (Å)	a = 745.4, b = 149.6, c = 80.5 $\alpha = \gamma = 90, \beta = 94.7$
Molecules in A.U.	24
Solvent Content (%)	71
Resolution range (Å) ^a	50.0–3.9 (4.2–3.9)
Unique reflections	78630
Data redundancy ^a	6.7 (6.6)
Completeness (%) ^a	97.6 (98.4)
$I/\sigma(I)$ ^a	11.7 (2.1)
R_{merge} (%) ^a	15.2 (96.3)
Mosaicity	0.35
Wilson scaling B factor (Å ²)	168.55

^aStatistics for the highest resolution data shell are shown in parentheses.

Even though individual structures of XRCC4 (Junop *et al.*, 2000; PDB entry 1fu1) and XLF (Andres *et al.*, 2007, PDB entry 2r9a; Li *et al.*, 2008, PDB 2qm4) have been solved, molecular replacement alone was unable to provide sufficient phasing. This may reflect the very large asymmetric unit and the high degree of structural similarity between XRCC4 and XLF. Tantalum bromide has been well-documented as a heavy-metal cluster suitable for determining phase of low-resolution structures (Knäblein *et al.*, 1997; Ban *et al.*, 1999; Banumathi *et al.*, 2002; Pomeranz Krummel *et al.*, 2009). Therefore, we

attempted to obtain tantalum bromide derivatives by soaking crystals at varying concentrations and for different lengths of time, with and without ‘back-soaking’. Soaking for less than 30 minutes did not produce a useful derivative, while soaking for >12 hours and/or back-soaking significantly decreased diffraction. Only direct soaking of crystals for ~3 hours generated sufficient phasing, without a significant loss in resolution.

The combination of low-resolution data, high solvent content and large unit cell required phasing using both molecular replacement and the anomalous signal from tantalum bromide. An initial search model of XRCC4-XLF was generated based upon mutational analysis and docking (Malivert *et al.*, 2010). Phases from molecular replacement were greatly improved by the additional phasing information from the tantalum bromide clusters (8 sites per asymmetric unit). Phasing from molecular replacement with single-wavelength anomalous diffraction in Phenix produced a FOM of 0.452 and LLG of -334048 (Adams *et al.*, 2010), and a structural solution was obtained (PDB entry 3rwr).

3.7 Acknowledgements We thank the helpful staff at X25, NSLS, Brookhaven National Laboratory. This work was supported by a Canadian Institute for Health Research grant to M.S.J. (MOP-89903) and an NSERC-CGS fellowship to S.N.A.

Chapter 4: Crystal structure of human XRCC4-XLF reveals an extended helical filament important for bridging DNA ends

Andres, S.N., Ristic, D., Wyman, C., Modesti, M., and Junop, M.S.

As prepared for submission to *Molecular Cell*.

4.1 Author's Preface

The research presented in chapter 4 has been prepared for submission to the peer-reviewed journal *Molecular Cell*, and appears in its submitted format. This article presents the structure of an XRCC4-XLF filament and identifies a bridging function for the filament. S.N. Andres conducted crystallization experiments and solved the structure of XRCC4-XLF. S.N. Andres carried out initial DNA binding assays that identified all XRCC4 and XLF mutants examined throughout this manuscript. Dr. M. Modesti performed DNA binding assays involving BRCT domains, as well as the DNA bridging assays. Dr. D. Ristic and Dr. C. Wyman were responsible for scanning force microscopy experiments. S.N. Andres and Dr. M.S. Junop wrote the manuscript, while the remaining authors were involved in editing the manuscript. Please note that references to Roy et al., refers to a paper that was co-submitted for publication with this manuscript of which S.N. Andres is also a co-author. It is from the laboratories of Dr. Katheryn Meek and Dr. Mauro Modesti, and is thus referenced as a personal communication.

4.2 Summary

DNA double-strand breaks pose a significant threat to cell survival and must be repaired. In higher eukaryotes such damage is repaired preferentially by non-homologous end-joining. Within this pathway, XLF and XRCC4 fulfill key roles required for efficient end-joining. Here we present the crystal structure of an extended protein filament of XRCC4-XLF at 3.94Å. DNA binding and bridging assays, combined with direct visualization, reveal how XRCC4-XLF filaments robustly bridge DNA molecules. This unanticipated, Ligase IV-independent bridging activity by XRCC4-XLF filaments suggests an early role for this complex during end-joining, in addition to its well-established later functions. Mutational analysis of the XRCC4-XLF C-terminal tail regions further identifies specialized functions in filament formation and interaction with DNA and LigaseIV. Based on this data, a model for XRCC4-XLF filament function in non-homologous end-joining is presented.

4.3 Introduction

DNA double-strand breaks pose a serious threat to genomic stability. They result from exposure to exogenous sources such as ionizing radiation, but are also the outcome of endogenous events including collapsed replication forks and intermediates of V(D)J recombination (Kuzminov, 2001; Roth et al., 1992). Mammalian cells have developed multiple mechanisms for repair, but predominantly employ non-homologous end-joining (NHEJ) (Jeggo et al., 1998). The core set of proteins required for NHEJ include Ku70/80, DNA-PKcs, XRCC4, XLF/Cernunnos, and LigaseIV (reviewed in Lieber et al., 2010). Ku70/80 initially binds DNA breaks preventing degradation and recruits DNA-PKcs and

XLF to the site of damage (Gottlieb and Jackson, 1993; Mimori and Hardin, 1986; Yano et al., 2008b). Synapsis of DNA ends is aided by DNA-PK (Ku70/80 and DNA-PKcs), which further promotes processing of damaged DNA ends by Artemis and other modifying enzymes (Ding et al., 2003; Ma, 2002; McElhinny et al., 2005).

XRCC4/LigaseIV is then free to seal DNA breaks. XLF directly interacts with XRCC4 and stimulates XRCC4/LigaseIV joining of incompatible DNA ends (Ahnesorg et al., 2006; Grawunder et al., 1997; Tsai et al., 2007).

XLF and XRCC4 are required for efficient NHEJ. Both XLF^{-/-} and XRCC4^{-/-} mammalian cells exhibit increased sensitivity to ionizing radiation coupled with serious defects in double-strand break repair (Giaccia et al., 1990; Zha et al., 2007). Furthermore, XRCC4 knockout mice are embryonic lethal due to increased neuronal apoptosis; but can be rescued by p53 deletion (Gao et al., 1998; Gao et al., 2000). Humans with XLF deficiencies display growth retardation, mental retardation, and immunodeficiency, reflecting key roles that NHEJ, and more specifically these two proteins, serve in cellular and mammalian development (Buck et al., 2006).

The mechanism by which XLF and XRCC4 function in NHEJ is not fully understood. Although they interact with one another at the protein-protein level, only XRCC4 interacts directly with LigaseIV (Ahnesorg et al., 2006; Critchlow et al., 1997; Deshpande and Wilson, 2007). XRCC4 binding stabilizes LigaseIV and augments LigaseIV adenylation (Grawunder et al., 1997). XLF, in turn, promotes re-adenylation and ligation of non-cohesive ends (Riballo et al., 2009; Tsai et al., 2007). Together these findings have led to the idea that XRCC4 and XLF contribute to a later role during NHEJ,

in particular helping to promote the final ligation step. More recently, however, live-cell imaging showed that recruitment of XLF and XRCC4 to damaged sites is rapid and dependent on the presence of Ku70/80 suggesting an earlier and additional role for XLF and XRCC4 in NHEJ (Yano and Chen, 2008; Yano et al., 2011). Furthermore Roy et al.,(personal communication) in an accompanying manuscript, report that perturbations of the XRCC4-XLF interaction result in double-strand break repair deficits including reduced frequency of coding-end joining during V(D)J recombination. Taken together, these data suggest that XRCC4 and XLF may have multiple functions during NHEJ, including their well-characterized late step in ligation and also an early step that we propose to be independent of XRCC4/LigaseIV.

How XLF and XRCC4 mediate their suggested roles throughout NHEJ is unclear. Individual crystal structures of XRCC4 and XLF reveal that they are related, with both forming homodimers having nearly identical N-terminal head domains and long alpha-helical tails. The tails differ in that XRCC4 maintains an extended conformation, while in XLF, they wrap back up towards the N-terminus forming a more compact structure (Andres et al., 2007; Junop et al., 2000; Li, et al., 2008). Mutational analysis within head domains identified residues R64, L65, and L115 of XLF, and K65 and K99 of XRCC4 as critical to XRCC4-XLF complex formation (Andres et al., 2007; Malivert et al., 2010). Given the location of these residues and the dimeric nature of the proteins, it has been proposed that the proteins could form a filament (Andres et al., 2007; Malivert et al., 2010). This idea has gained further support by recent small-angle x-ray scattering (SAXS) analysis that suggested a linear arrangement of XLF and XRCC4 proteins in

solution (Hammel et al., 2010).

Here we report the crystal structure of an extended XRCC4-XLF helical filament at 3.94Å. We determine a role for the XRCC4-XLF filament independent of LigaseIV in bridging double-strand breaks. XRCC4-XLF filaments are also shown to exist in two distinct states, one of which mediates DNA bridging. We further identify regions of XLF and XRCC4 important for DNA binding and demonstrate that these regions make independent contributions to filament function in NHEJ. In particular, the C-terminal tails of XLF are responsible for filament-DNA interactions, while the corresponding region of XRCC4 regulates formation of the filament required for DNA bridging. Sequestration of XRCC4 C-terminal tails through binding of LigaseIV tandem BRCT domains disrupts this species, ablating DNA bridging. A model for the mechanism of XRCC4-XLF filaments within NHEJ is presented.

4.4 Results

4.4.1 Structure of the XRCC4¹⁻¹⁵⁷-XLF¹⁻²²⁴ complex

Efforts to crystallize XRCC4-XLF involved various combinations of full-length and truncated proteins. Although crystals were obtained with all constructs tested, only XLF 1-224 (XLF²²⁴) and XRCC4 1-157 (XRCC4¹⁵⁷) produced diffraction quality crystals in our hands. XRCC4¹⁵⁷-XLF²²⁴ crystals diffracted to ~3Å, however anisotropy limited the data to 3.9Å. This anisotropy reflects the extremely high solvent content (~70%), and limited crystal contacts observed in the structure. Phase information was obtained using a combination of molecular replacement and single-wavelength anomalous diffraction

(MR-SAD). XRCC4¹⁵⁷-XLF²²⁴ crystals grew in spacegroup C2 with one extremely long axis (~750Å) and contained 6 dimers of each protein in the asymmetric unit. The overall structure was refined to a final R and R_{free} of 26.2 and 32.9 respectively (PDB 3RWR).

Table 4.1 lists a complete set of data collection and refinement statistics.

Table 4. 1 Crystallographic Data and Refinement Statistics

Data Collection	
Wavelength (Å)	1.2536
Space group	C ₂
Cell parameters (Å)	a = 745.4, b = 149.6, c = 80.5 α = γ = 90, β=94.7
Molecules in A.U.	24
Resolution range (Å) ^a	50.0–3.9 (4.2-3.9)
Unique reflections	78630
Data redundancy ^a	6.7 (6.6)
Completeness (%) ^a	97.6 (98.4)
I/σ(I) ^a	11.7 (2.1)
R _{merge} (%) ^a	15.2 (96.3)
Mosaicity	0.35
Wilson scaling B factor (Å ²)	168.55
Model and Refinement	
Resolution range (Å)	50.0–3.94
R _{work} (%)	26.2
R _{free} (%)	32.9
Reflections observed	73,366
Reflection test set	3,886
Number of protein atoms	3464
Rmsd bond lengths (Å)	0.0028
Rmsd bond angles (Å)	0.77
Average B factor (Å ²)	144.83

^a Statistics for the highest data resolution shell are shown in parentheses.

The basic repeating assembly of XRCC4¹⁵⁷-XLF²²⁴ observed within the asymmetric unit contained a tetramer of XRCC4¹⁵⁷ and XLF²²⁴ (Figure 4.1A). Both proteins remain essentially unaltered from previously determined individual structures (PDB 1FU1; PDB 2R9A). Average RMSD's of XLF²²⁴ and XRCC4¹⁵⁷ within the complex, compared to proteins on their own, were remarkably small at 1.38 Å and 1.47 Å respectively, indicating that complex formation does not require significant conformational changes. However, it does result in a ~30° offset between dimers (Figure 4.1A, black arrow). The standard deviation of RMSD's for each of the 12 monomers of XLF²²⁴ and XRCC4¹⁵⁷ within the asymmetric was 0.3 Å and 0.2Å respectively, and therefore further discussion of the XRCC4¹⁵⁷-XLF²²⁴ interface will not include all 24 observed interfaces.

Closer inspection of the XRCC4¹⁵⁷-XLF²²⁴ interface reveals extensive hydrogen-bonding through a continuous anti-parallel β -sheet formed between strands β 5-7 of XRCC4¹⁵⁷ and β 4' of XLF²²⁴ (Figure 4.1B). In the absence of XRCC4¹⁵⁷, β 4' exists as a loop between α 2 and α 3 of the helix-turn-helix motif in XLF²²⁴, indicating that small conformational changes result from complex formation (Andres et al., 2007; Li et al., 2008). The XRCC4¹⁵⁷-XLF²²⁴ interface is further strengthened through direct amino acid interactions (Figure 4.1C). A complete list of amino acids involved in this interface is provided in Supplementary Table S4.1. L65 and L115 of XLF²²⁴ make key contributions by sandwiching a highly conserved XRCC4¹⁵⁷ Phe residue (106) (Supplementary Figure S4.1, S4.2). In addition, K99 of XRCC4 makes hydrophobic interactions with L115 of

XLF²²⁴, through its aliphatic chain. Lysine 99 also forms a hydrogen bond with the carbonyl oxygen of S113 in XLF²²⁴. The interface is further stabilized through electrostatic interactions between R64 and E111 of XLF²²⁴ with E55 and K65 of XRCC4¹⁵⁷ (Figure 4.1C). These interactions are consistent with previously reported mutational analysis (Andres et al., 2007; Malivert et al., 2010).

4.4.2 XRCC4¹⁵⁷-XLF²²⁴ forms a protein filament

Each dimer of XRCC4¹⁵⁷ and XLF²²⁴ observed in the asymmetric unit contributes to a single continuous filament 840Å long. The filament is comprised of alternating XLF²²⁴ and XRCC4¹⁵⁷ dimers, with a total of 6 tetramers forming one complete revolution of a left-handed helix (Figure 4.1D). This filament confirms mutational data and SAXS analysis, suggesting XLF and XRCC4 might form a filament in solution (Andres et al., 2007; Hammel et al., 2010). The filament diameter is ~220Å, with an internal pore of ~70Å, towards which both proteins' N-terminal heads are directed. The C-termini differ; in the case of XRCC4¹⁵⁷, the tails continue outward in an extended conformation, while in XLF²²⁴, the tails wrap back, towards the inner pore. Though the overall filament structure is large, surface area buried between XLF²²⁴ and XRCC4¹⁵⁷ dimers is surprisingly small (750 Å²), when compared to the dimer interface of XRCC4¹⁵⁷ (1500 Å²) and XLF²²⁴ (3340 Å²). Stable complex formation typically requires buried surface areas greater than 1000 Å² (Janin and Chothia, 1990). Thus, in order for this filament to remain stable, additional interactions not observed in our crystal structure may be required. One obvious possibility for another binding partner is DNA.

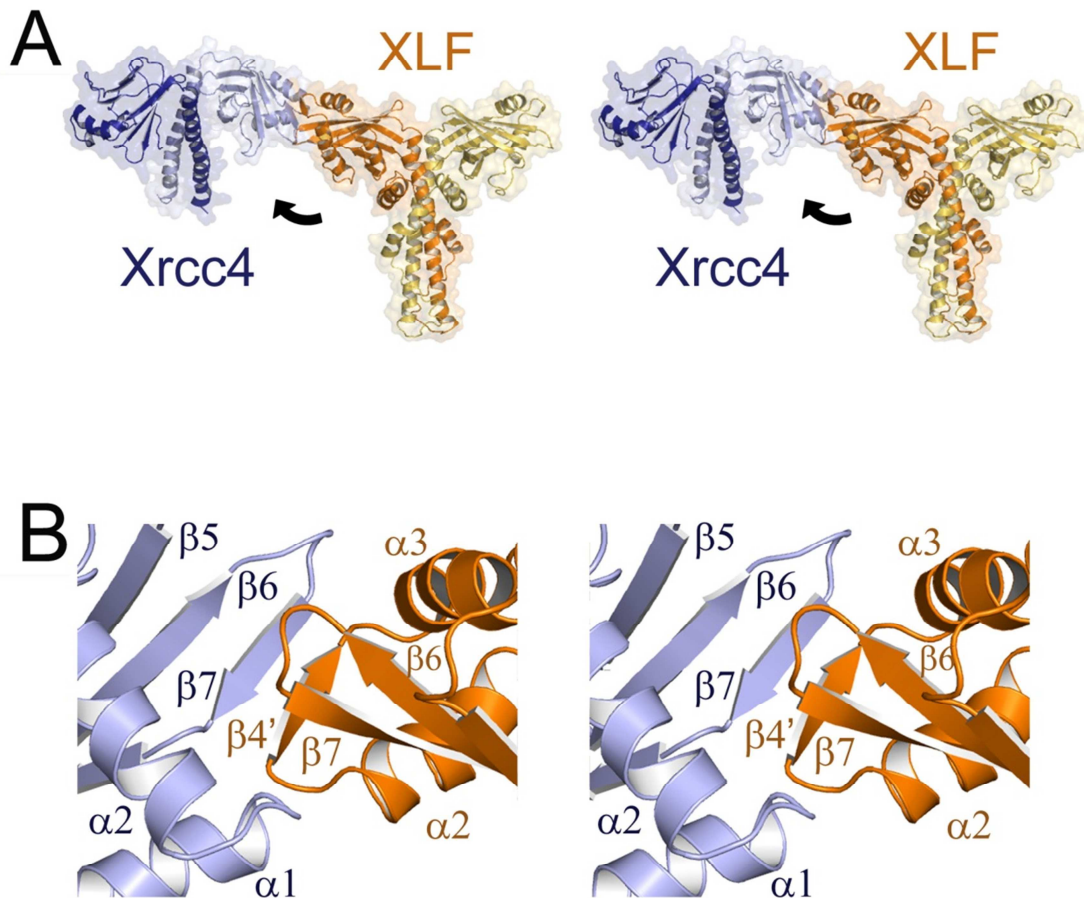


Figure 4.1 Crystal structure of the XLF¹⁻²²⁴ - XRCC4¹⁻¹⁵⁷ Complex
(A) Interaction of XLF²²⁴ homodimer (orange) with an adjacent XRCC4¹⁵⁷ homodimer (blue) as seen in the crystal structure. Black arrow illustrates ~30° offset between homodimers. (B) XRCC4¹⁵⁷-XLF²²⁴ head-to-head interface. *Continued on page 76.*

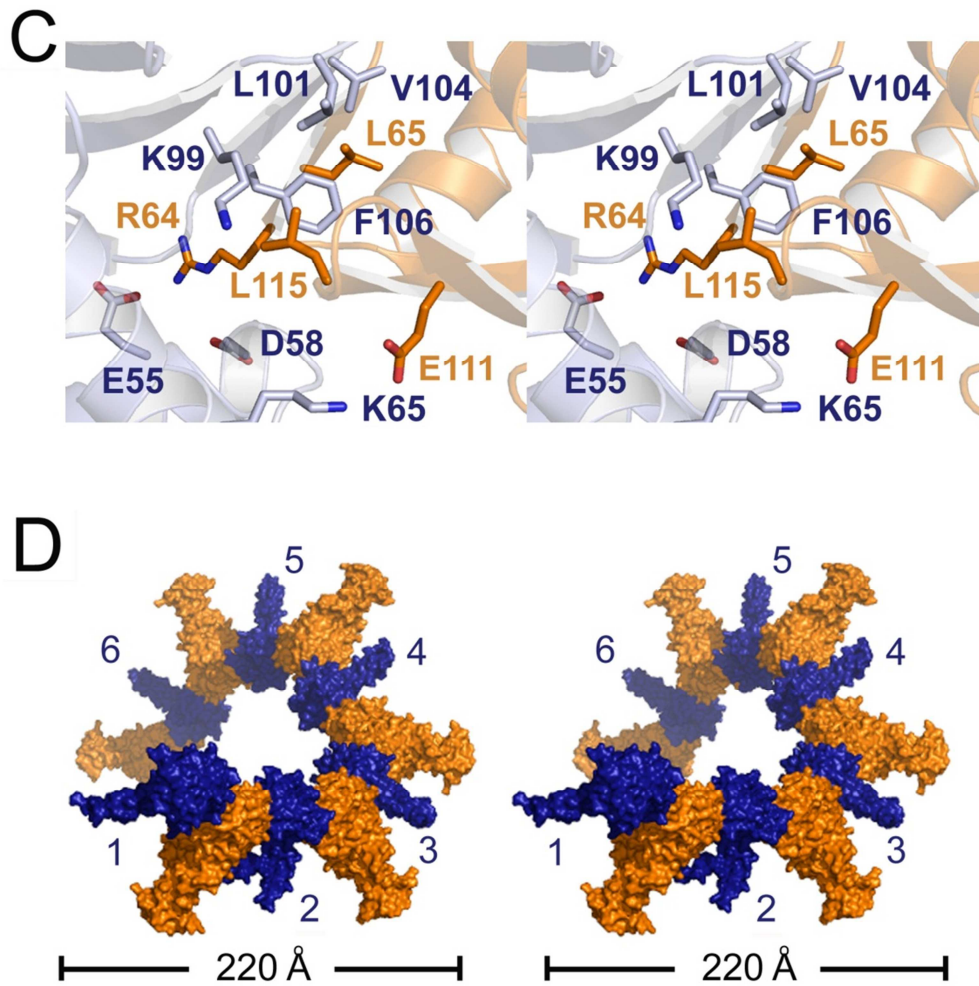


Figure 4.1 Crystal structure of the XLF¹⁻²²⁴ - XRCC4¹⁻¹⁵⁷ Complex
(C) Key amino acids directly involved in XRCC4¹⁵⁷-XLF²²⁴ interaction.
(D) XRCC4¹⁵⁷-XLF²²⁴ filament. XRCC4¹⁵⁷ homodimers are numbered.

4.4.3 Identification of DNA Binding regions in XLF and XRCC4

Both XRCC4 and XLF exhibit optimal DNA binding when DNA is greater than 100 bp, and this affinity increases as DNA length increases (Lu et al., 2007; Modesti et al., 1999). Furthermore, binding is reported to occur in a highly cooperative manner, suggesting a protein filament may form along DNA. In our structure, a single revolution of the XRCC4-XLF filament is 840Å long, covering ~ 250 bp of linear DNA. There are at least two models for how DNA and protein may interact with this filament; one involving wrapping of DNA around protein and another where the reverse occurs. Since the filament is already helical, the simplest, most energetically favourable way to model DNA binding is with DNA (18Å diameter) running through the internal pore (70Å) (Figure 4.2A).

To further investigate this nucleoprotein filament model, the DNA binding properties of XLF and XRCC4 were analyzed (Figure 4.2B, C). Both XRCC4 (lane 2, Figure 4.2B) and XLF (lane 2, Figure 4.2C) were able to independently bind a 1000 bp DNA fragment, confirming earlier reports that these proteins efficiently bind large DNA fragments (Lu et al., 2007; Modesti et al., 1999). Previous research has shown that the first 200 residues of XRCC4 retain full DNA binding activity (Modesti et al., 1999). In Figure 4.2B, we further localize this activity to amino acids 157 – 200 (lane 3). Mutational analysis within this region identified E170 and R192 as necessary for DNA binding (Figure 4.2B, lanes 4-5). These amino acids fall in a highly conserved segment of XRCC4 also responsible for tetramerization and interaction with the BRCT domains of LigaseIV (Supplementary Figure S4.1). A full list of amino acids in this region tested for

DNA binding is provided (Supplementary Table S4.2). Like XRCC4, DNA binding of XLF resides in its tail region (Andres et al., 2007). In Figure 4.2C, we confirmed this requirement and localized binding activity to a highly conserved lysine cluster at the C-terminus of XLF, similar to reports for XLF's yeast homolog (Sulek et al., 2007; Supplementary Figure S4.2). An Ala substitution at K293 partially decreased the overall binding capacity of XLF (Figure 4.2C, lanes 4). XLF K290A was also tested, but had no effect on XLF-DNA binding compared to wildtype (data not shown).

4.4.4 DNA Binding and the XRCC4-XLF Filament

We tested DNA binding activity under conditions where XRCC4-XLF filaments are formed and observed a 'super-shifted' protein-DNA complex (Figure 4.2B, lane 7; 4.2C, lane 6). However, when XRCC4 DNA binding mutants were tested under these conditions, only an intermediate gel-shift was observed (Figure 4.2B, lanes 8-10). This 'intermediate' shift is less than the supershift, but greater than XLF binding alone. Since XRCC4 1-157 is the form present in the crystal structure, we tentatively suggest that this intermediate complex represents a simple filament as seen in Figure 1D, while the high molecular weight species (supershift) represents a more complex oligomer. Unlike XRCC4, XLF-DNA binding mutants in complex with wildtype XRCC4 produced either no shift, (XLF²²⁴), or an intermediate shift, (XLF K293A) (Figure 4.2C). The species observed in lane 8 is likely due to an incomplete DNA binding defect of the XLF K293A mutant (lane 3 vs. 4). What is clear is that DNA binding of XLF is more important for filament-DNA interaction than that of XRCC4 (Fig 4.2C, lane 7 vs. Fig 4.2B, lane 8).

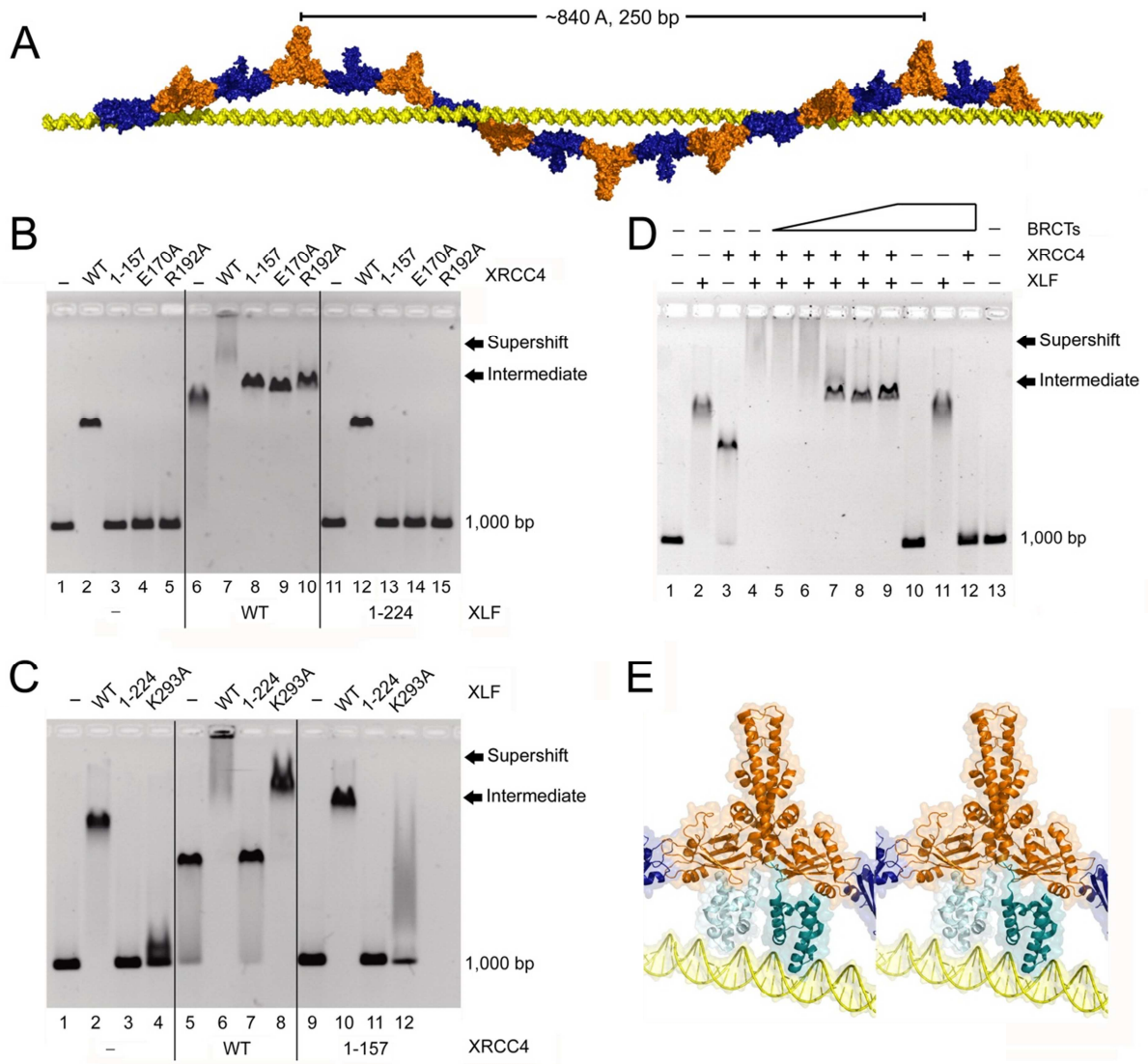


Figure 4.2 XRCC4-XLF DNA binding (A) Model of proposed filament interactions with DNA. Length of one filament revolution is indicated in Å and bp. DNA, yellow; XLF, orange; XRCC4, blue (B) XRCC4 wild-type (WT) and mutants (8 μM) were incubated with 100 ng of DNA with or without 2 μM XLF (WT and mutants), analyzed by EMSA. (C) WT XLF and its mutants (2 μM) were incubated with 100 ng of DNA with or without 8 μM XRCC4 (WT and mutants). (D) Effect of Ligase IV tandem BRCT domains on XRCC4-XLF-DNA complex formation. XRCC4 (8 μM) and XLF (2 μM) were incubated with 100 ng DNA fragment in the presence of increasing amount of Ligase IV tandem BRCTs domain (0, 0.5, 1, 2, 4 to 8 μM). (E) Model of DNA binding through XLF C-terminus, using PDB 2KV2 (green).

The absence of DNA binding in XLF fully disrupts the XRCC4-XLF filament's affinity for DNA, while XRCC4-DNA binding within the filament is dispensable to some degree, given that the intermediate species is still observed. Furthermore, previously identified mutations that prevent the XRCC4-XLF interaction are unable to bind DNA to levels beyond individual wildtype proteins, indicating that an XRCC4-XLF filament is necessary to produce either the super-shifted or intermediate species observed in Figure 4.2 (Supplementary Figure S4.3).

4.4.5 Ligase IV and DNA binding regions of XRCC4 modulate XRCC4-XLF filament-DNA interactions

The tandem BRCT domains of Ligase IV interact with residues 155-195 of XRCC4, including direct contact with R192 (Wu et al., 2009). XRCC4 R192 is also one of the residues implicated in DNA binding (Figure 4.2B). Similarly, other substitutions in this region of XRCC4 prevent interaction with both DNA and the BRCT domains of LigaseIV (Supplemental Table S4.2, section 4.8; Modesti et al., 2003). Not surprisingly, binding of tandem BRCT domains precludes XRCC4's interaction with DNA (Figure 4.2D, lane 12), whereas XLF's affinity for DNA is not affected by the tandem BRCT domains since XLF has little to no affinity for LigaseIV (Figure 4.2D lane 11) (Andres et al., 2007; Deshpande and Wilson, 2007). Thus, XRCC4-DNA binding and BRCT binding are mutually exclusive. As expected, the super-shifted species is maintained upon addition of sub-stoichiometric concentrations of BRCT domains (Figure 4.2D, lanes 5-6). Conversion to the intermediate species only occurs when equimolar concentrations of BRCT domains are added (Figure 4.2D, lanes 7-9). The smaller 'intermediate' species is

not disrupted, even at the highest concentrations of BRCT domains tested, and is reminiscent of the complex observed in EMSAs including wildtype XLF and XRCC4 DNA binding mutants. The formation of two species suggests changes in the composition of the XRCC4-XLF filament, including alterations to filament length, shape, or number of filaments and DNA simultaneously in complex.

4.4.6 Modeling DNA Binding in the XRCC4-XLF filament

XRCC4 binds both DNA and LigaseIV *in vitro*, yet the data presented here indicate that it cannot bind both simultaneously (Critchlow et al., 1997; Grawunder et al., 1997; Modesti et al., 1999). It was also noted earlier that disruption to XRCC4-DNA binding does not affect DNA-filament interactions as severely as when XLF-DNA binding is compromised. The model presented in Figure 4.2A directs the DNA binding tails of XRCC4 away from the DNA helix. Therefore, taken together, XRCC4's DNA binding appears to be dispensable for nucleoprotein filament formation. XLF's DNA binding domain, however, is required, and is directed towards the center of the filament. As illustrated in Figure 4.2E, we have modelled the DNA binding domain of XLF contacting DNA within the filament pore. As a size comparison, the missing 75 C-terminal residues of XLF have been modelled using an 85 amino acid DNA binding domain (PDB 2KV2). Of significance in this model is the fact that there is sufficient room between the DNA helix and the C-terminus of the current XLF structure for the remaining 75 residues, thus illustrating how XLF-DNA interactions may increase the overall filament stability by increasing buried surface area. Further structural analysis will be required to verify the validity of this model.

4.4.7 The XRCC4-XLF filament bridges DNA molecules

During DNA double-strand break repair, DNA ends are maintained in close proximity through assembly of a higher-order nucleoprotein complex (reviewed in Dobbs et al., 2010). The finding that XRCC4-XLF forms an extended filament, which binds DNA, suggests the filament may be involved in bridging DNA during repair. To test this hypothesis, we followed the experimental design outlined in Figure 4.3A. If XRCC4-XLF bridges the 1000 bp and 500 bp DNA, then the 500 bp DNA will be recovered from the streptavidin-coated beads. In the presence of XLF or XRCC4 alone, DNA found in the supernatant is bound to protein, but recovery of the 500 bp DNA from beads is not observed (Figure 4.3B, lanes 2-3). Therefore individual XLF or XRCC4 proteins are unable to bridge DNA. The XRCC4-XLF filament, however, produces a super-shifted species in the supernatant as expected, but also has recovery of the 500bp DNA from the bead, indicating the filament's ability to bridge DNA ends (Figure 4.3B, lane 4). As a control, XLF and XRCC4 mutants that abolish XRCC4-XLF complex formation (mutants studied by Roy et al., personal communication, in an accompanying manuscript) still bind DNA, but do not recover the 500bp DNA fragment, signifying that XRCC4-XLF must form a filament in order to bridge DNA (Figure 4.3C). In summary, these data suggest that the bridging activity of XRCC4-XLF filaments is functionally important in DNA repair.

Not unlike the effect seen in DNA binding, DNA bridging by XRCC4-XLF is also interrupted by the presence of BRCT domains (Figure 4.3B, lane 5). No DNA is recovered from the bead, and the supernatant fraction contains only the DNA binding

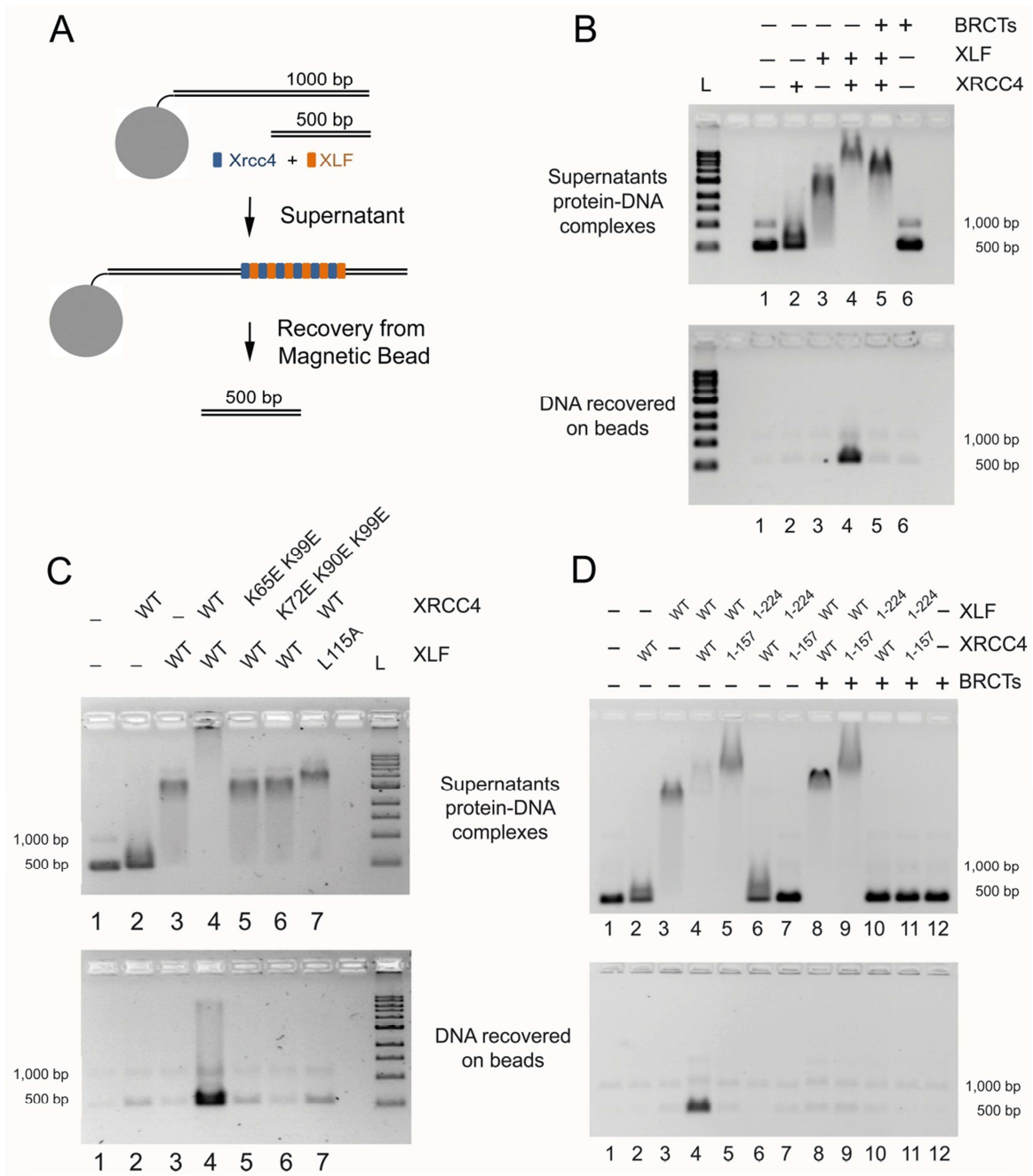


Figure 4.3 Bridging of DNA molecules by XRCC4-XLF (A) Schematic of DNA bridging assay. Proteins were incubated with magnetic beads linked to 1000 bp DNA and free 500 bp DNA. Beads were separated from supernatant and analyzed for presence of the 500 bp DNA. (B) 200 ng each of 1000 bp and 500 bp DNA fragments were incubated with XRCC4 (2 μ M), XLF (2 μ M) or LigaseIV tandem BRCT domains (BRCTs, 2 μ M). Top panel shows the analysis of the protein-DNA complexes in the supernatants. Bottom panel shows the recovery of DNA species on the beads. L = 1 kb DNA ladder (NEB). (C) Bridging assays performed as in (B) with mutants preventing XRCC4-XLF filament formation. (D) Bridging assays performed as in (B). XRCC4 1-157 and XLF 1-224 are truncated proteins lacking C-terminal tails and DNA binding activity.

intermediate species. Since the intermediate filament is observed when bridging is abolished, it is formation of the super-shifted filament that must be required for DNA bridging. Therefore, one would expect factors that alter the C-terminal tail of XRCC4, such as BRCT binding, to have negative effects on bridging activity.

As expected, similar results are observed when wildtype XLF and XRCC4 C-terminal DNA binding mutants are assayed for bridging capabilities. Figure 4.3D illustrates that in the presence or absence of the BRCT domains, any truncation of either XLF or XRCC4 that fully ablates DNA binding also prevents DNA bridging. In particular, reactions carried out with XRCC4¹⁵⁷ cannot bridge DNA, but are able to support filament-DNA binding, even though XRCC4¹⁵⁷ is unable to bind DNA alone (Figure 4.3D, lanes 5 and 9). These results further emphasize that XRCC4 tails are not necessary to bind DNA within the context of the filament, yet are still necessary for DNA bridging, which in itself requires the super-shifted filament species. Thus, the tails of XRCC4 are most likely involved in formation of the super-shifted nucleoprotein filament.

4.4.8 Formation of a higher-order XRCC4-XLF filament

Evidence so far suggests that the XRCC4-XLF filament observed within the crystal structure is only partially representative of the filament required for DNA bridging. Formation of two distinct nucleoprotein complexes further suggests that the simple filament observed in the crystal structure may oligomerize into higher-order complexes in the presence of DNA when BRCT domains are not present. To further investigate these possibilities we directly visualized complexes using scanning force microscopy (SFM), under conditions identical to DNA binding and bridging experiments.

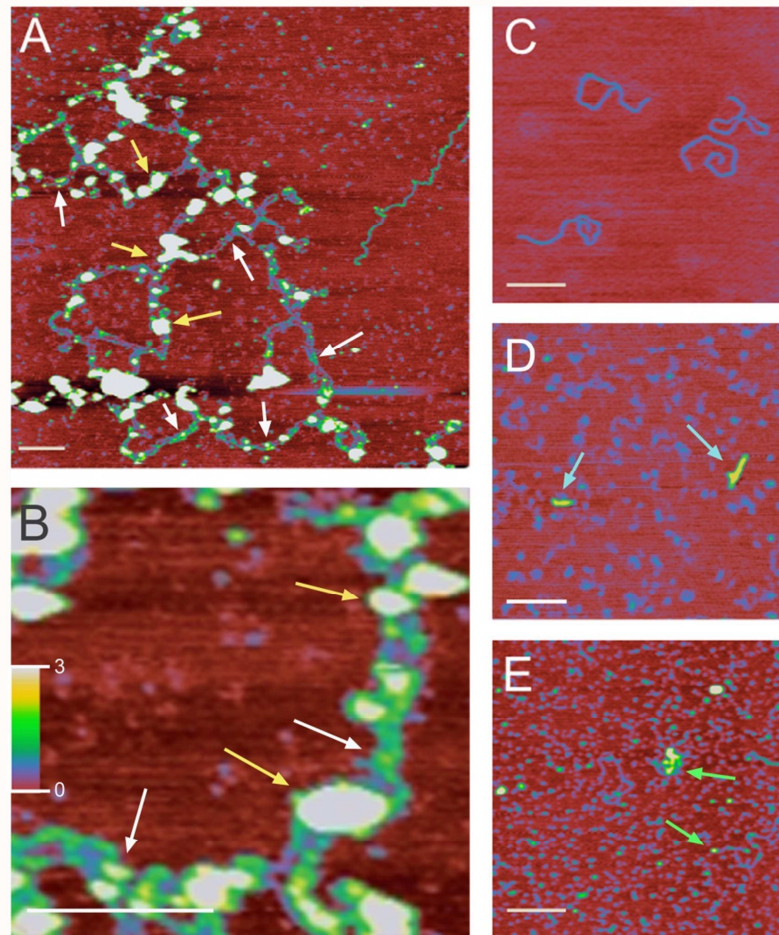


Figure 4.4 SFM analysis of protein-DNA networks. (A) Large protein-DNA networks observed in XRCC4-XLF + DNA binding reactions. Protein complexes attached to DNA appear as higher, wider objects that vary in size (yellow and white arrows). Protein-induced parallel bridging of DNA molecules is evident (white arrows). Length of the DNA networks indicates molecules are joined end-to-end. (B) Enlargement of (A) highlighting distance between parallel DNA molecules. Note that dimensions of biomolecules are distorted in SFM images. Relative size and separation between objects can, however, be used very accurately. (C) 1.8 kb linear DNA, contour length 6000Å, width 180Å, and height 3Å. (D) Image of XRCC4-XLF complexes. Protein complexes likely to be dimers, tetramers and small multimers appear as uniform objects distributed over the surface (blue objects). Protein filaments (blue arrow) measure $L \times H \times W = 3000 \times 20 \times 300 \text{Å}$. (E) Addition of BRCT domains disrupts DNA-protein networks. Protein complexes similar to (D) are uniformly distributed. Individual 1.8 kb DNA fragments were apparent, image center and near right edge. Some DNA molecules associated via protein complexes were observed, off center to right (green arrow). Image A is 2 X 2 microns. Image B is 500 X 500 nm. Images C-E are 1 X 1 microns. In all images the white bar is 200 nm long and height is indicated by color (0-3 nm red to yellow/white, scale bar in panel B).

In full agreement with earlier results, upon addition of both proteins to DNA, large protein-DNA networks were observed (Figure 4.4A, B). These networks included 1.8 kb DNA molecules, connected by protein and aligned end-to-end, and in parallel (separated by $\sim 200\text{\AA}$) (Figure 4.4A, B). Two different protein species are bound to DNA, a larger complex (yellow arrow) than the filaments observed in the absence of DNA (blue arrow, Figure 4.4D), and a smaller species (white arrow), found interspersed along parallel DNA molecules. Comparing dimensions of the larger protein complexes in the DNA networks (yellow arrow, Figure 4.4A, B) to those of objects identified as a filament in the protein alone control (blue arrow, Figure 4.4D) indicates that these complexes are composed of more than one filament. These striking observations are consistent with two XRCC4-XLF filaments and/or filament bundles bridging parallel DNA molecules through XRCC4 tetramerization. Consistent with this interpretation, addition of BRCT domains, which are known to prevent XRCC4 tetramerization, abolishes the networks and large filament bundle structures, but permits XRCC4-XLF association with DNA, similar to the lone DNA molecule coated with protein in Figure 4.4A (green arrows, Figure 4.4E).

4.5 Discussion

DNA double-strand breaks require rapid and efficient repair to circumvent apoptosis, and chromosomal rearrangements that lead to deregulation of normal cell function (Gao et al., 2000; Lieber et al., 1998). NHEJ exists for this purpose and choreographs several proteins in the repair of DNA double-strand breaks. Two of these proteins, XLF and XRCC4, are of great interest due to their central role in NHEJ, but have remained elusive with respect to their mechanism(s) of action. The structure of

XRCC4-XLF presented here identifies key features of this complex and sheds new light on their role in NHEJ.

XLF and XRCC4 form an extended, alternating filament. This filament binds DNA specifically within XLF's C-terminus, through a highly conserved, lysine-rich region located at the last 10 amino acids of the protein (Supplementary Figure 4.2). Formation of the XRCC4-XLF filament bundle is dependent on the availability of XRCC4's conserved C-terminal tails (157-200). This conserved region is also essential for DNA bridging activity (Figure 4.3B).

4.5.1 XRCC4-XLF Filament Bundles

Precedence for nucleoprotein filaments is seen in NHEJ's counterpart, homologous recombination (HR) (see Supplementary Discussion). A simple model of the XRCC4-XLF filament bound to DNA was presented in Figure 4.2A, yet results from DNA binding studies displayed two modes of protein-DNA interaction – an intermediate and super-shifted form, suggesting that the nucleoprotein filament exists in more than one structural state. Of these, only the super-shifting species was capable of stably bridging DNA. Nucleoprotein complexes were further characterized via SFM, and demonstrated to align DNA both end-to-end and in parallel. These observed nucleoprotein networks contained two distinctly sized complexes. If we extend the model presented in Figure 4.2A to include multiple filaments, several possibilities can be considered to explain these results.

One possibility to explain the large species observed in Figure 4.4 (black arrows), which greatly exceed the size of a single filament, involves packing multiple filaments

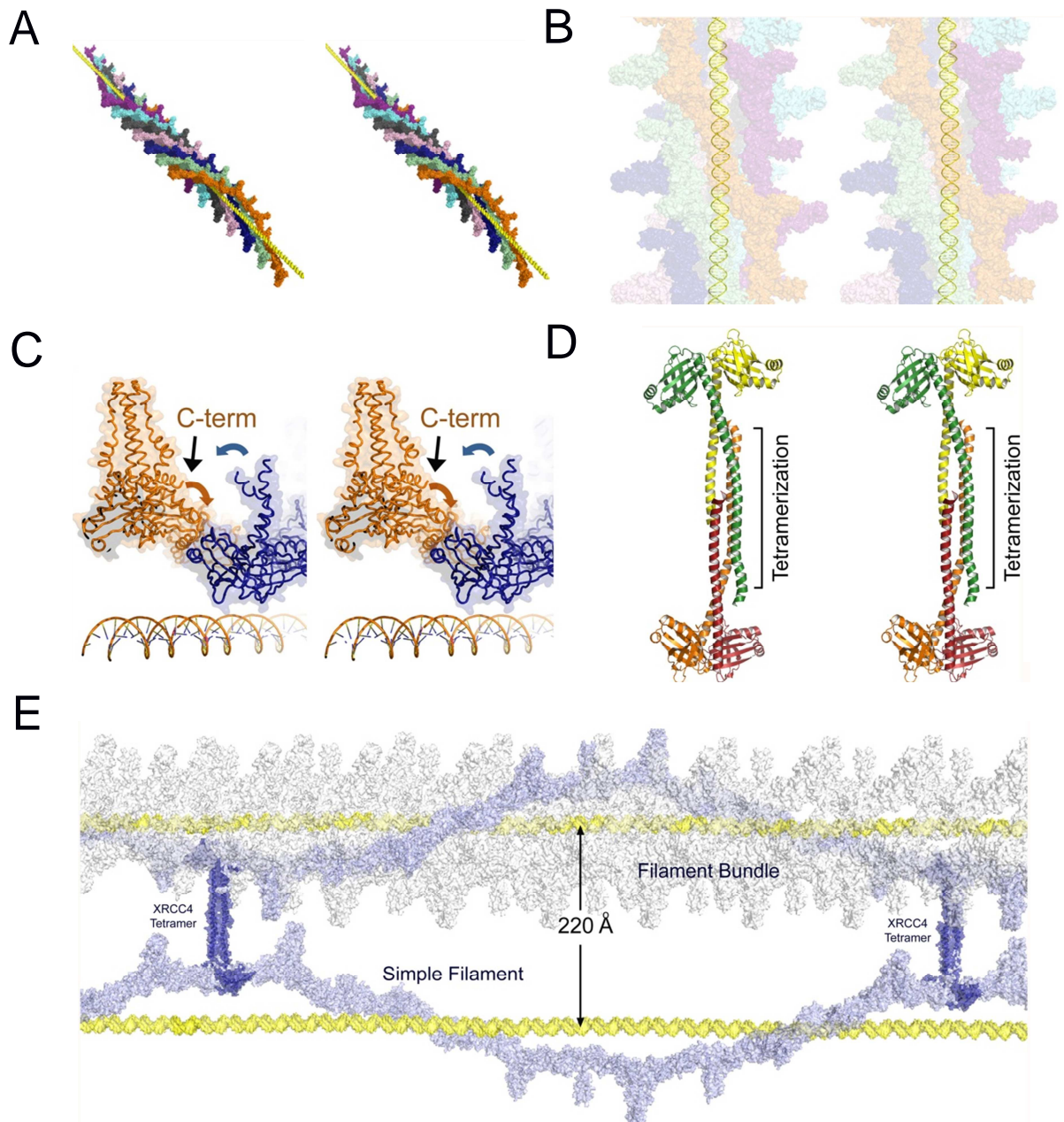


Figure 4.5 Model of XRCC4-XLF filaments bound to DNA (A) Multiple adjacent filaments bound to DNA (yellow). Each color is a separate filament. (B) DNA runs through the pore of the XRCC4-XLF filament. (C) Tails of an XRCC4¹⁵⁷ homodimer (blue) point towards the N-terminus of XLF²²⁴ (orange), in an adjacent filament. (D) XRCC4 dimers associate into a tetramer through C-terminal tails (PDB 1FU1). (E) Model of DNA bridging with single and complex filaments. Two DNA molecules coated in a simple or multi-filament bundle are bridged through XRCC4 C-terminal tails.

adjacent to one another. In this way, DNA can be fully coated by at least 7 successive single filaments (Figure 4.5A, B). We know the higher-order filament binds DNA through XLF, not XRCC4, yet the C-terminal tails of XRCC4 are absolutely required for bridging. Comparing the structure of XRCC4 to XLF, few differences are observed other than in the tail regions. XLF tails fold into a compact, globular form, while XRCC4 tails remain extended and flexible (Andres et al., 2007; Junop et al., 2000; Li et al., 2008). Secondary structure alignment of both proteins reveals that XLF's conserved hinge region (residues ~160-190), corresponds to a segment of XRCC4 where the dimerization domain ends (G160), suggesting that XRCC4 tails may also fold back, allowing interaction of its C-terminus with XLF from an adjacent filament within a bundle (Supplementary Figure S4.1, S4.2). Evidence for this cross-linking arrangement is present in the crystal structure. When modelled within the context of a filament bundle, the C-terminal tails of some XRCC4 dimers are directed towards the head domains of XLF in an adjacent filament (Figure 4.5C). This arrangement could explain why XRCC4 tails are required for XRCC4-XLF filament bundles to bridge DNA. Interestingly, this same region also corresponds to XRCC4's homo-tetramerization region (Figure 4.5D; Junop et al., 2000; Modesti et al., 2003). Results from SFM seem to support alignment of more than one bundle, situating DNA molecules both end-to-end and parallel to one another, separated by 200 Å. XRCC4 is able to homo-tetramerize, and through this interaction, filament bundles could be bridged (Figure 4.5E; Junop et al., 2000; Modesti et al., 2003). The spacing observed between adjacent DNA molecules in SFM (200 Å) correlates well with those predicted in this model (Figure 4.5E). Since tetramerization and BRCT binding are

mutually exclusive, addition of equimolar amounts of BRCT domains would disrupt all interactions dependent on XRCC4 tails, as was observed in Figure 4.4E.

In addition to the large species observed in SFM, smaller nucleoprotein complexes were also found juxtaposed on adjacent DNA molecules (white arrows, Figure 4.4B). The simplest explanation for these species is a similar arrangement to that suggested in Figure 4.5E, but with fewer or even a single filament coating each DNA. Both arrangements are consistent with the observation in Figure 4.4E, showing disruption of these complexes in the presence of BRCT domains. This model provides a mechanism to explain the observed correlation between filament state, DNA bridging, and availability of XRCC4 tails when either mutated or sequestered by Ligase IV binding.

Formation of filament bundles would greatly increase buried surface area within the overall filament structure (Figure 4.2A vs. Figure 4.5). This is particularly important as a single filament observed in the crystal structure only contains weak head-to-head interactions (750 \AA^2), hardly suitable for bridging large DNA molecules as is observed in our study. Thus, while XRCC4-XLF head interactions are sufficient to form a single filament, the corresponding tails of each protein may be involved in cross-linking adjacent protein filaments and filament bundles into a larger, more stable complex capable of bridging DNA.

4.5.2 Function of the XRCC4-XLF filament in NHEJ

The analysis presented here and elsewhere for XRCC4-XLF suggests a functional role in maintaining DNA ends in close proximity for appropriate repair (Akopiants et al., 2009). Recently, several factors have been implicated as co-factors in NHEJ that

facilitate appropriate end-bridging including ATM, 53BP1, Mre11/Rad50/Nbs1 (MRN), and H2AX (Shibata et al., 2010; Zha et al., 2011). Consistent with our conclusion that the XRCC4-XLF filament has a LigaseIV independent role in bridging DNA ends, data presented in the accompanying manuscript (Roy et al., personal communication) show that XRCC4 mutants, which cannot interact with XLF (and therefore cannot form filaments) have NHEJ deficits reminiscent of cells deficient in ATM (Zha et al., 2011). Even more, these data provide further insight into recent reports showing marked accentuation of NHEJ deficits in animals defective in both ATM and XLF (Zha et al., 2011). These data suggest that ATM's role in promoting bridging, perhaps by functioning upstream of MRN and 53BP1, are redundant with XRCC4-XLF filaments' function in bridging (Zha et al., 2011).

In the current model of NHEJ, XLF and XRCC4 are thought to function primarily at the ligation step, where XRCC4 stimulates LigaseIV, and XLF stimulates ligation of mismatched DNA ends (Grawunder et al., 1997; Tsai et al., 2007). However, our findings, in conjunction with those from an accompanying manuscript (Roy et al., personal communication), suggest that XRCC4 and XLF have an additional role, where XRCC4-XLF filaments bridge DNA double-strand breaks, an event likely to be important early on in NHEJ. The idea of an early role in NHEJ is further supported by rapid, DNA-PK-independent recruitment of XLF to damaged sites by Ku70/80 (Yano et al., 2008; Yano et al., 2008b; Yano et al., 2009). Recent work by Yano et al., (2011) identified the last 10 amino acids of XLF as necessary for interacting with Ku70/80. This is particularly interesting in light of the work presented here, which demonstrates that this

region is also required for DNA binding (XLF K293A). Together these results imply that Ku70/80 would initially recruit XLF to the damaged site and nucleate filament formation along DNA. This is in line with results showing that XLF recruited by Ku70/80 remains stably bound only when XRCC4 is also present at sites of DNA damage, presumably due to formation of the filament (Yano et al., 2008b).

As mentioned previously, XLF and XRCC4 are involved in ligation during NHEJ, but through a mechanism not fully understood. NHEJ is dependent on interactions between XRCC4 and the tandem BRCT domains of Ligase IV. When these BRCT domains were included in DNA binding and DNA bridging assays, the XRCC4-XLF filament formed an intermediate species incapable of stably bridging DNA. Therefore, it is possible that during NHEJ, XLF and XRCC4 initially form a filament bundle precipitated by Ku70/80, which is subsequently remodelled to a smaller bundle involved in ligation (perhaps by DNA-PK, as suggested in the accompanying manuscript of Roy et al., (personal communication). BRCT binding does not disrupt simple filament formation (Figure 4.2D), suggesting that these filaments can accommodate Ligase IV. It should be pointed out that depending on the amount of Ligase IV present, larger complexes might persist throughout NHEJ, with only small alterations where Ligase IV is bound to permit repair of DNA double-strand breaks. Further experimentation will be necessary to confirm this mechanism.

4.5.3 Conclusion

Here we have described the XRCC4-XLF crystal structure, composed of an extended filament of alternating XRCC4 and XLF homodimers. The XRCC4-XLF

filament forms two distinct species – a simple filament, as seen in the crystal structure, and a higher-order filament bundle, as observed by SFM. Filaments engage DNA through interaction with the C-terminus of XLF. Bridging of DNA ends, however, requires formation of filament bundles that are expected to provide stability to damaged DNA early and perhaps throughout NHEJ, while simple filaments that accommodate Ligase IV, would be expected to function later during ligation. Formation of filament bundles is regulated through availability of the C-terminus of XRCC4 by interaction with Ligase IV or itself, further implying multiple roles for XRCC4-XLF in NHEJ (Figure 4.6). Interestingly, in an accompanying paper (Roy et al., personal communication), DNA-PK hyper-phosphorylation within this region of XRCC4 suggests that post-translational modifications may also serve to further regulate filament formation.

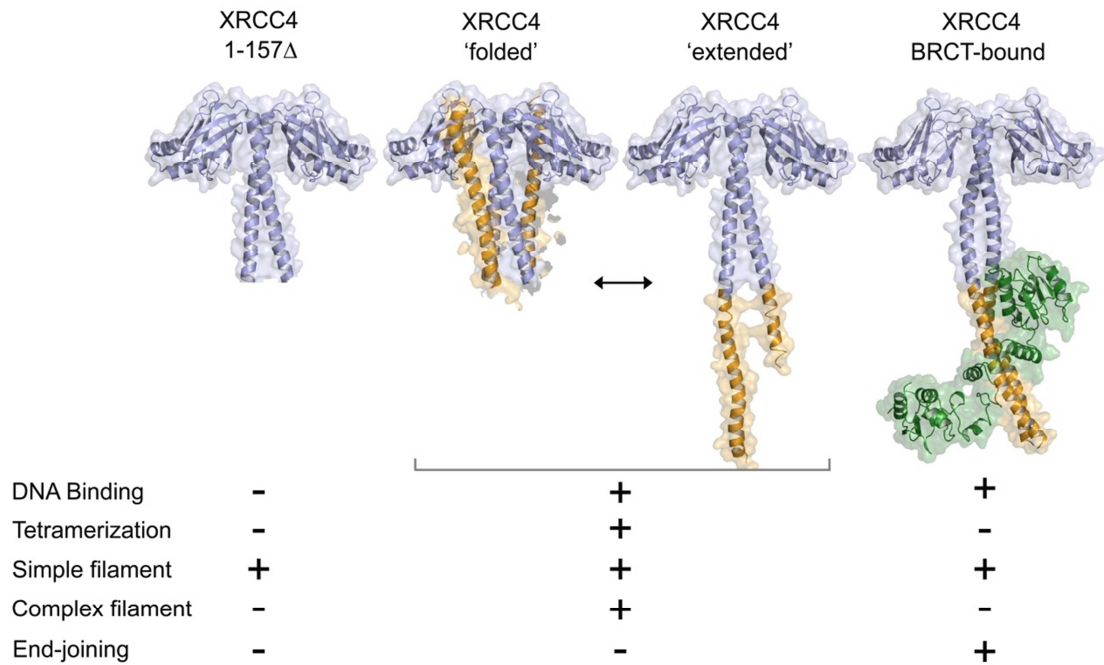


Figure 4.6 Summary of structural states of XRCC4. Structural states of XRCC4 are indicated with their associated function (PDB 1FU1 and PDB 3II6; Hammel et al., 2010; Junop et al., 2000; Modesti et al., 2003; Wu et al., 2009).

4.6 Experimental Procedures

4.6.1 XLF and XRCC4 expression vector construction

All XLF and XRCC4 expression plasmids, with the exception of XRCC4¹⁻¹⁵⁷, were from previously described work (Modesti et al., 1999; Junop et al., 2000; Andres et al., 2007). See Supplemental Data, section 4.8 for further details.

4.6.2 Protein Expression and Purification

XLF, XRCC4 and BRCT domains, wildtype and all mutations, were expressed and purified as previously described (Andres et al., 2007; Junop et al., 2000). XRCC4¹⁻¹⁵⁷ was expressed in M9 SeMET growth media kit (Medicilon Inc.).

4.6.3 Crystallization and Data Collection of XLF²²⁴-XRCC4¹⁵⁷

Crystals were grown using the hanging drop vapour diffusion method. A 1 μ L protein solution (50 μ M XLF²²⁴:100 μ M XRCC4¹⁵⁷ in 20 mM Tris, pH 8, 200 mM KCl, 1 mM EDTA, 10 mM DTT and 10% glycerol) was combined with microseeded crystallization solution (0.8 μ L of 1.8 M tri-ammonium citrate, pH 8) and additives (0.2 μ L each 0.1 M barium chloride dihydrate and 2.0 M sodium thiocyanate). Crystals were dehydrated over 4 M ammonium sulfate, pH 7, prior to a 3-hour soak in phasing solution before freezing (1 μ L, 0.5 mM tantalum bromide and 0.5 μ L of 60% PEG 8000). Diffraction data were collected at NSLS, X25 (Brookhaven, NY).

4.6.4 Structure Determination and Model Refinement

Data was processed with *HKL2000* (Otwinowski and Minor, 1997). Phases were determined using MR-SAD, followed by density modification with *Phenix* (Adams et al., 2010). An XRCC4-XLF docking model was used as a search model (Malivert et al.,

2010). 8 tantalum bromide sites were found. Iterative rounds of model building and refinement were carried out using *Phenix*, *Wincoot*, and *CNS v1.3* with DEN refinement (Adams et al., 2010; Brunger et al., 1998; Brunger, 2007; Emsley and Cowtan, 2004; Schroeder et al., 2010). Interacting regions were analyzed using *ZMM* (Zhorov 1981). All structural figures were generated using *PyMol* (DeLano, 2002). Buried surface area calculations were carried out through the *PISA* server, v.1.18 (Krissinel and Henrick, 2007).

4.6.5 Electrophoretic Mobility Shift Assays for DNA Binding

DNA preparation is described in Supplemental Information, section 4.8. Binding reactions (20 μ L) contained 100 ng of DNA, 20mM HEPES pH 8, 0.5 mM EDTA, 1 mM DTT, 5% glycerol, and 120 mM KCl after addition of proteins at the indicated concentrations. Reactions were incubated at room temperature for 30 min and resolved by electrophoresis in a 0.8 % agarose gel in Tris-borate buffer (89 mM Tris-base, 89 mM boric acid, 2mM EDTA pH 8.3) at 80 Volts/15 cm for 60 or 90 min. Gels were stained in Tris-borate buffer supplemented with 0.5 μ g/ml ethidium bromide (30 min.) and destained in deionized water (2 hrs).

4.6.6 DNA Bridging

DNA bridging was carried out as illustrated in Figure 3A. Full details are in Supplementary Information.

4.6.7 Scanning Force Microscopy

DNA substrates and XRCC4-XLF complexes were prepared as described in Supplementary Information. Images were obtained on a NanoScope IV SFM (Digital

Instruments; Santa Barbara, CA) operating in tapping mode in air with a type E scanner. Silicon Nanotips were from Digital Instruments (Nanoprobes). The length and height measurements were done with NanoScope software v 5.12 (Digital Instruments; Santa Barbara, CA). Although absolute dimensions in SFM are not accurate, relative size and separation between objects can be used very accurately.

4.7 Acknowledgements: The authors would like to thank Dr. M. Morar, McMaster University for structural discussion and X-25 beamline staff, NSLS, Brookhaven National Laboratory. This work was funded by grants from the USA National Cancer Institute (5PO1CA09584) and the Netherlands Organization of Scientific Research, Chemical Sciences Division (VICI grant 700.56.441) to C.W, the Association for International Cancer Research (grant 09-0633) and the Association pour la Recherche sur le Cancer (A09/2/5075) to M.M., the Canadian Institutes of Health Research to M.S.J. (MOP-89903) and an NSERC-CGS fellowship to S.N.A.

4.8 Supplemental Information

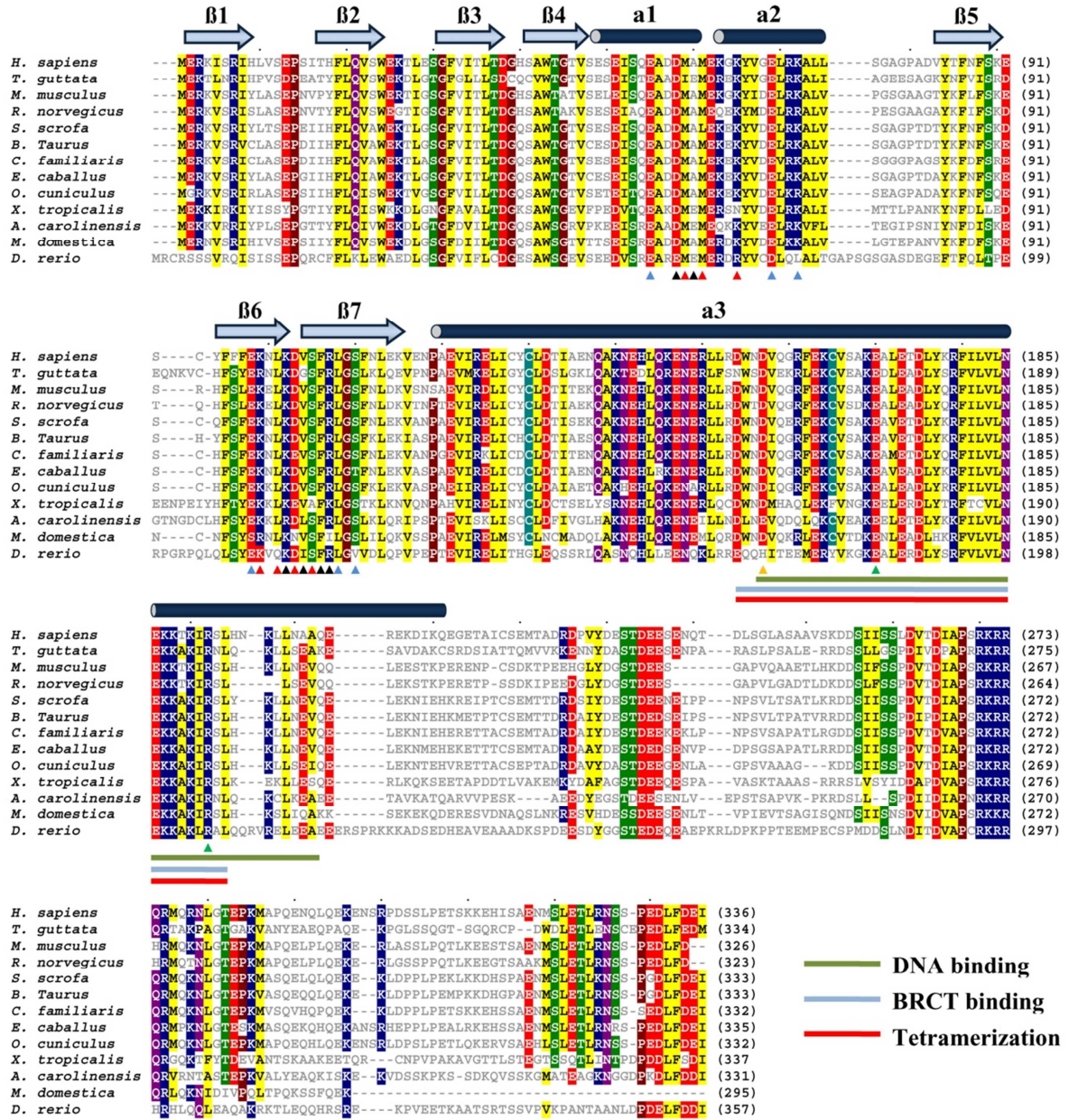


Figure S4.1 XRCC4 sequence alignment. Highly conserved residues are colored as follows: hydrophobic (yellow), positive (blue), negative (red), small polar (green), large polar (purple), cysteine (cyan), glycine and proline (brown). Triangles indicate: end of protein in crystal structure D157 (yellow), DNA binding mutants E170A and R192A (green), residues at XRCC4-XLF interface burying 0-20, 20-50, 50-170 Å² (light blue, red and black respectively). Regions of XRCC4 mediating DNA binding, Ligase IV interaction and homo-tetramerization are indicated in green, light blue and red lines, respectively

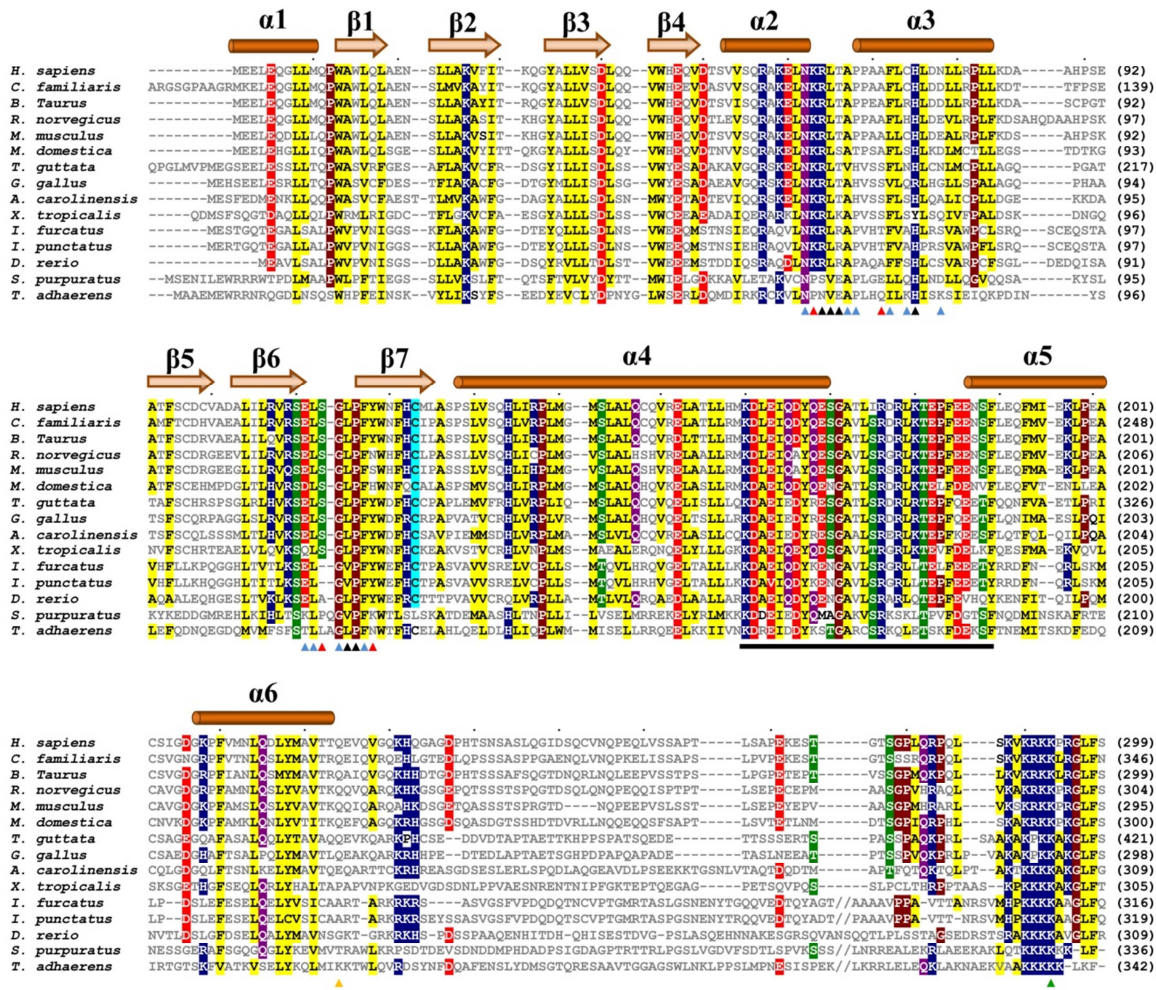


Figure S4.2 XLF sequence alignment. Highly conserved residues are colored as follows: hydrophobic (yellow), positive (blue), negative (red), small polar (green), large polar (purple), cysteine (cyan), glycine and proline (brown). Triangles indicate: end of protein in crystal structure Q224 (yellow), DNA binding mutant K293A (green), residues at XRCC4-XLF interface burying 0-20, 20-50, 50-170 Å² (light blue, red and black respectively). Hinge region, at which point the tails reverse direction, is underlined in black.

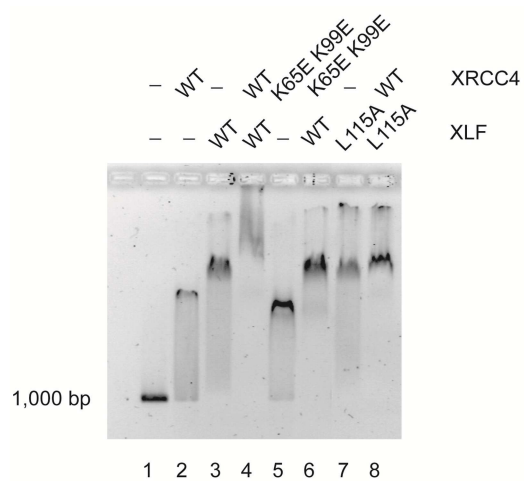


Figure S4.3 DNA Binding Analysis of XLF and XRCC4 filament mutants. Previously reported XLF and XRCC4 mutations that prevent filament formation were analyzed for DNA binding activity (Andres et al., 2007).

Table S4.1 XLF and XRCC4 interface residues with corresponding buried surface areas (BSA).

XLF	BSA(Å ²)	XRCC4	BSA(Å ²)
Asn - 62	3.0	Glu - 55	19.4
Lys - 63	29.5	Asp - 58	71.2
Arg - 64	167.5	Met - 59	19.8
Leu - 65	50.8	Ala - 60	56.4
Thr - 66	54.9	Met - 61	33.1
Ala - 67	17.2	Lys - 65	44.2
Pro - 68	22.4	Glu - 69	3.2
Ala - 71	37.7	Lys - 72	2.0
Phe - 72	1.7	Glu - 98	14.4
Cys - 74	21.09	Lys - 99	34.9
His - 75	70.6	Leu - 101	43.8
Asn - 78	2.3	Lys - 102	92.7
Glu - 111	16.3	Asp - 103	32.0
Leu - 112	14.9	Val - 104	63.3
Ser - 113	51.0	Ser - 105	31.1
Gly - 114	12.2	Phe - 106	86.4
Leu - 115	123.5	Arg - 107	56.1
Pro - 116	76.2	Leu - 108	0.8
Phe - 117	2.5	Ser - 110	1.34
Tyr - 118	24.9		

Table S4.2 DNA binding activity of XRCC4 mutants with mutations between amino acids 157-200.

XRCC4 Mutant	Binds DNA
C93A/C165A	+
R161Q/K164Q ^a	+
C165F	+
C165A/K169V/L172E	-
K169A	+/-
K169E	-
E170A	-
F180D/I181D	+/-
L184Q/K187D/I191S ^b	-
K187D/K188D	-
I191D/L194E/L198E	-
R192A	-
H195D	-

^a This XRCC4 mutant was previously identified as able to bind DNA (Andres et al., 2007).

^b This XRCC4 mutant was previously identified as unable to bind BRCT domains (Modesti et al., 2003)

4.8.1 Supplemental Experimental Procedures

XRCC4¹⁻¹⁵⁷ expression vector construction

XRCC4¹⁻¹⁵⁷ construction was constructed as follows: Forward and reverse primers (5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTTAGAAGGAGATAGAACCATGGAGAGAAAAATAAGCAG-3' and 5'-GGGGACCACTTTGTACAAGAAAGCTGGGTCTTAGTGATGGTGATGGTGATGATCATTCCAATCTCTCAG-3') were used in conjunction with pWY-1088, an XRCC4 expression plasmid (Junop et al., 2000) to create XRCC4¹⁻¹⁵⁷. This vector contained a C-terminal His₆-tag and was cloned into pDEST-14 using Gateway cloning technology (Invitrogen). Successful cloning was confirmed by DNA sequencing.

Electrophoretic Mobility Shift Assays for DNA Binding: DNA Substrate

Preparation

The 1000 bp DNA substrate was prepared by PCR using Phusion DNA polymerase (NEB), primers 5'-GAGTTTTATCGCTTCCATGAC and 5'-AATTTATCCTCAAGTAAGGGGC and PhiX174 DNA as template. The PCR product was purified by gel electrophoresis and QIAquick Gel Extraction Kit (QIAGEN) and stored in 10 mM Tris pH 8, 1 mM EDTA.

DNA Bridging Assay

The one-end biotinylated 1000 bp DNA substrate was prepared by PCR using Phusion DNA polymerase, primers 5'-Biotin-GAGTTTTATCGCTTCCATGAC and 5'-AATTTATCCTCAAGTAAGGGGC and PhiX174 DNA as template. The 500 bp DNA substrate was similarly prepared using primers 5'-GAGTTTTATCGCTTCCATGAC and

5'-CAGAAAATCGAAATCATCTTC. PCR products were purified and stored as described above.

Magnetic streptavidin-coated beads (Dynabeads M-280 Streptavidin, Invitrogen) were first passivated by washing the bead suspension three times with one volume of binding buffer (20 mM HEPES pH 8, 75 mM KCl, 0.5 mM EDTA, 1 mM DTT, 5 % glycerol, 400 µg/ml acetylated BSA) and finally resuspended in the same volume of binding buffer. For each reaction, 200 ng of end biotinylated 1,000 bp DNA were added to 10 µL of passivated bead suspension and incubated for 5 min at room temperature (> 90% attachment). Next, 200 ng of 500 bp DNA fragment were added before final addition of the XRCC4, XLF or BRCTs proteins (each at 2 µM) in a total volume of 40 µL in binding buffer. Reaction mixtures were incubated for 30 min at room temperature after which beads were collected with the magnet without any centrifugation step. The 40 µL supernatant fractions were analyzed by electromobility shift assay and reducing polyacrylamide gel electrophoresis in the presence of SDS. The beads were washed 2 times with one volume of binding buffer and finally resuspended in 40 µL of binding buffer without BSA. A 10 µL fraction of the bead suspension was analyzed by reducing polyacrylamide gel electrophoresis in the presence of SDS. Proteinase K (40 µg) and Sarkosyl (0.5 % final) were added to the remainder 30 µL bead suspensions, incubated for 30 min at 50 °C and resolved by electrophoresis in 0.8 % or 1 % agarose gels in Tris-borate buffer as described above.

Scanning Force Microscopy: DNA and Protein Complex Preparation

The double-stranded DNA used in the SFM experiments was made by linearization of pDERI1 (Ristic et al., 2001). Digestion of this plasmid with PvuI produced 1821 bp linear double-stranded DNA. The resulting linear DNA was purified by GFX™ column (Amersham) and checked for purity by gel electrophoresis.

XRCC4-XLF complexes were formed in 20 μ L reactions containing 2 μ M XRCC4, 2 μ M XLF, 20mM HEPES-KOH (pH8), 50mM KCl, 1mM DTT, 0.5mM EDTA, 5% glycerol. Complexes of XRCC4-XLF-DNA were produced by addition of 7.5 μ M DNA (concentration in bp) in reaction described above. Where present, DNA Ligase IV BRCT domains were added to binding reaction with final concentration of 2 μ M. Reactions were incubated at 19⁰C for 10 min and then placed on ice. Reactions were diluted 40-fold in reaction buffer just before deposition. Mica was freshly cleaved and treated with 90mM spermidine. After 20 s spermidine was removed by washing with reaction buffer. The excess buffer was removed and diluted reaction mixture was transferred to mica. After 15 s the mica was washed with water and dried in a stream of filtered air.

4.8.2 Supplemental Discussion

Precedence for XRCC4-XLF filaments in other repair pathways

HR is another important DNA double-strand break repair pathway which differs from NHEJ in its requirement for a homologous template during repair. HR makes use of a nucleoprotein filament conserved from humans (Rad51) to bacteria (RecA) and even archaea (RadA) (Di Capua et al., 1982; Galkin et al., 2006; Sandler et al., 1996). Rad51 (RadA) forms protein filaments through a β -strand polymerization motif, where the β -

strand of one Rad51 (RadA) protein binds to the β -sheet of an adjacent Rad51 (RadA) monomer (Chen et al., 2007; Conway et al., 2004). In a similar manner, RecA forms a continuous β -sheet across monomers within its growing filament. As an isolated protein, however, the interacting β -strand exists as a flexible loop (Datta et al., 2000). This is identical to what occurs in the loop region of XLF, when in contact with XRCC4 (Figure 1B, $\beta 4'$). Furthermore, RadA, like XRCC4-XLF, forms a left-handed protein helix to bind DNA (Chen et al., 2007). Interestingly, the XRCC4-XLF filament and RecA/Rad51/RadA filament appear to be examples of functionally convergent evolution in two related DNA repair pathways.

Chapter 5: The Effects of XRCC4-XLF Filaments on LigaseIV Activity

5.1 Author's Preface: The work described in this chapter consists of unpublished results, but is included as it pertains to the overall thesis objectives. An introduction for this chapter is omitted and replaced with a short abstract to avoid repetition with Chapter 1. All research was carried out by S.N. Andres.

5.2 Abstract

Final ligation of double-strand breaks in NHEJ requires XRCC4 and LigaseIV. Together, they form an extremely stable complex resistant to 1M NaCl (Critchlow et al., 1997; Grawunder et al., 1997). LigaseIV activity is stimulated by binding of XRCC4, but the mechanism by which this occurs is unknown (Grawunder et al., 1998a; Grawunder et al., 1998b). XLF is also known to stimulate ligation and interacts with XRCC4 (Tsai et al., 2007; Ahnesorg et al., 2006). Therefore the effects of XLF and XRCC4 on ligation, either alone or as part of an XRCC4-XLF filament, were investigated. To further understand the role of XRCC4 in ligation, structural studies have solved the crystal structure of XRCC4 in complex with Ligase IV (Δ 654-911) (Figure 1.8). This structure illustrates how XRCC4-LigaseIV interact, with tandem BRCT domains of LigaseIV wrapping around the C-terminal helical region of XRCC4 (Wu et al., 2009). The catalytic domain of LigaseIV, however, was not included in this structure, and therefore current structural information cannot fully explain how XRCC4 stimulates LigaseIV. Thus, to further understand XRCC4's role in ligation, structural studies of the full-length XRCC4-LigaseIV complex were also attempted.

5.3 Experimental Procedures

5.3.1 Expression of Human XRCC4-LigaseIV

Plasmid MJ-4052 is a pET-28a(+) co-expression vector, previously constructed by Dr. M.S. Junop. It contains a kanamycin resistance cassette and genes encoding the human proteins XRCC4 (amino acids 1-336) and full-length LigaseIV (amino acids 1-

911), with a C-terminal hexa-histidine fusion present on LigaseIV, all under the control of a T7 promoter.

pMJ-4052 was transformed into Rosetta (DE3) *E. coli* cells using a standard heat shock method (Cohen et al., 1972). Single colonies were isolated from agar plates containing 0.05 mg/mL kanamycin and 0.25 mg/mL chloramphenicol. These antibiotics were added to all media used in the growth and expression of pMJ-4052 discussed hereafter, at the same concentrations used in the agar plates. Bacterial colonies containing pMJ-4052, as identified by antibiotic resistance, were used to inoculate 40 mL of LB media and incubated with shaking at 37°C for 16 hours. This culture was used for sub-culture by placing 10mL of cultured media into 1L of LB (4L total). 1L cultures were incubated at 37°C, with shaking, until the optical density at 600 nm measured between 0.4 - 0.5. Cultures were then placed in an ice bath for 20 minutes, prior to induction with 1 mM IPTG for 16 hours, with shaking, at 16°C. Cultures were then centrifuged for 10 minutes at 3,315 g, the supernatant removed, and the resulting pellets flash-frozen in N₂(l) prior to storage at -80°C.

5.3.2 Purification of XRCC4-LigaseIV

Cell pellets from growth of 2L cultures of *E. coli* cells expressing XRCC4-LigaseIV were resuspended in NiA buffer (20mM Tris, pH8, 500mM KCl, 1 mM beta-mercaptoethanol, 0.03% LDAO and 10% glycerol), DNaseI (10u, Fermentas) and protease inhibitors (1µM Pepstatin-A, 1mM PMSF, 1mM Benzamidine HCl, 10µM Leupeptin). Lysis occurred via sequential passage through a French Press cell four times

at 1,000 psi. Further protease inhibitors and DNaseI were then added prior to and after lysate clarification by centrifugation at 48,384 g. Following centrifugation, the supernatant was removed and filtered before initial purification by nickel-affinity chromatography. A 5mL His-Trap HP column (GE Healthcare) connected to an ÄKTAFPLC (GE Healthcare) was equilibrated with NiA buffer containing 10mM imidazole and loaded with the filtered supernatant. Column washes with 25mL of NiA+10mM imidazole and 25mL of NiA+16mM imidazole was carried out prior to elution of XRCC4-LigaseIV with NiA+ 64mM imidazole. The eluted protein was diluted with 20mM Tris, pH8, 10mM dithiothreitol, 1mM EDTA, and 10% glycerol (QA buffer) to obtain a final concentration of 150mM KCl, prior to injection onto a MonoQ 10/100 GL column (GE Healthcare) equilibrated with QA+150mM KCl. After 32mL of washing with QA+150mM KCl, XRCC4-LigaseIV was eluted over a gradient of 150mM – 325mM KCl for 95mL and collected in 1.5mL fractions. Fractions containing the protein of interest were identified by SDS-PAGE analysis and exchanged into QA+25mM KCl via a HiPrep 26/10 Desalting column (GE Healthcare). The recovered protein was then loaded onto a MonoS 10/100 GL column (GE Healthcare) equilibrated with QA+25mM KCl. Following a 32 mL wash with the equilibrating buffer, XRCC4-LigaseIV was eluted over a gradient of 25mM – 200mM KCl for 100mL. Fractions containing pure protein were identified by SDS-PAGE, pooled, and simultaneously concentrated and buffer-exchanged into 10mM Tris, pH8, 150mM KCl, 1mM EDTA, and 5mM DTT using a Vivaspin 20 centrifugal concentrator (30 kDa MWCO, GE Healthcare).

5.3.3 Activity of XRCC4-LigaseIV

XRCC4-LigaseIV activity was tested by monitoring DNA ligation efficiency. 210 ng of a modified pUC-19 plasmid (pMJ-4293), linearized with FastDigest *NdeI* (Fermentas) was combined with 100 pmol each of co-expressed full-length XRCC4-LigaseIV, full-length XLF and/or full-length XLF L115A, in addition to 5 U of T4 DNA Ligase (Fermentas) and 2 μ L of 10X T4 DNA Ligase Buffer (Fermentas). Water was added to create a final reaction volume of 20 μ L prior to incubation for 1 hour at room temperature. Following incubation, 1u of Proteinase K (Fermentas) was added, and incubation continued at 50°C for 1 hour. 0.2 μ L of 10% SDS was then added to the reaction and incubated for 20 minutes at room temperature before being resolved by a 0.8% TBE agarose gel at a constant voltage of 50V for 16 hours. The resulting gel was stained in GelStar® Nucleic Acid Gel Stain (Lonza) for 3 hours and visualized using a UV transilluminator for qualitative analysis.

5.3.4 Crystallization of XRCC4-LigaseIV

XRCC4-LigaseIV was crystallized using the hanging drop vapour diffusion method (McPherson, 1976). Specific crystallization conditions will be reported in the results. However in general, 1 μ L of purified XRCC4-LigaseIV (2mg/mL) was combined with 1 μ L of crystallization precipitant and incubated over 800 μ L of 1.5M ammonium sulfate at 20°C. Crystallization precipitants were used from multiple commercially available kits, including Hampton Index (Hampton Research), Wizard I and II (Emerald Biosystems) and The Cryos Suite (Qiagen). Crystal growth parameters of conditions

yielding potential protein crystals were further optimized by repeating crystallization under conditions identical to that which produced the initial crystal, but with the addition of 0.2 μ L of additive from the Hampton Additive Screen (Hampton Research).

5.4 Results

5.4.1 Purification of XRCC4-LigaseIV

Full-length XRCC4-LigaseIV eluted at low concentrations of imidazole (16mM) from the His-Trap HP column and remained highly impure. Two additional purification steps, involving ion exchange chromatography, were therefore added. XRCC4-LigaseIV eluted from the MonoQ 10/100 GL column (GE Healthcare) column between 262mM and 288mM KCl, and from the MonoS 10/100 GL column (GE Healthcare) column between 150mM and 170mM KCl. A total of 3.7 mg of XRCC4-Ligase IV complex was purified from 4L of Rosetta (DE3) *E. coli* cells (Figure 5.1).

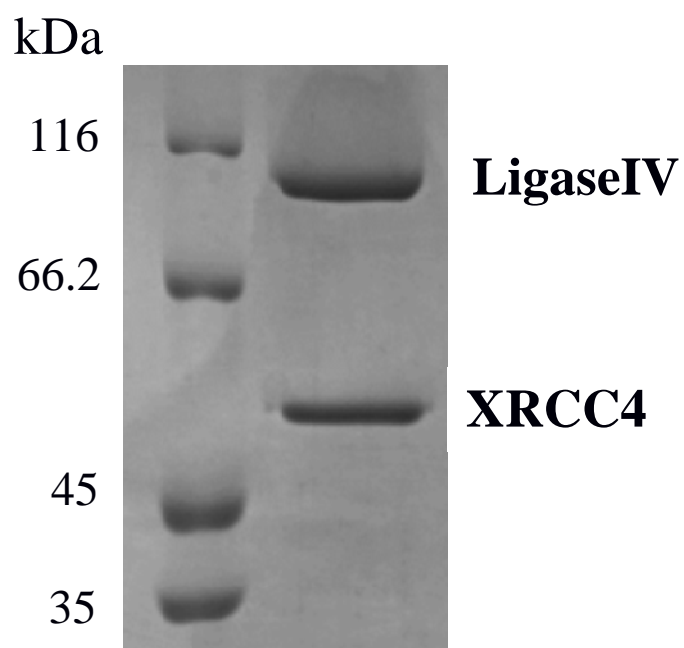


Figure 5.1 Purified Human XRCC4 (~53kDa) and Human LigaseIV (~104kDa) after nickel-affinity, anion and cation exchange purification from Rosetta (DE3) *E. coli* cells. 40 μ g of total protein was analysed by SDS-PAGE.

5.4.2 Activity of XRCC4-LigaseIV

XRCC4-LigaseIV was able to carry out both intra- and intermolecular ligation on a linearized plasmid with cohesive 3'-overhangs (lane 4, Figure 5.2A), but not to the same extent as commercial T4 DNA Ligase under the same conditions (Fermentas; lane 4, Figure 5.2B). However upon addition of XLF, intramolecular ligation was decreased, yet intermolecular ligation was highly stimulated for XRCC4-LigaseIV, when compared to XRCC4-LigaseIV alone. This effect was not observed for XLF with T4 DNA Ligase. Addition of extra XRCC4 to either T4 DNA Ligase or LigaseIV caused a decrease in intermolecular ligation (lane 7, Figure 5.2). However, addition of excess XRCC4 to XLF and XRCC4-LigaseIV did not change the overall intermolecular ligation of the DNA substrate.

It has been reported that XLF alone stimulates LigaseIV's intermolecular ligation, yet it has also now been observed that XLF can form a protein filament with XLF (Chapter 4; Hammel et al., 2011; Ropars et al., 2011; Tsai et al., 2007). To determine if the XRCC4-XLF filament is required to stimulate intermolecular ligation, XLF L115A replaced wildtype XLF in the activity assay. As reported earlier, XLF L115A prevents the interaction of XLF with XRCC4, thus preventing formation of the protein filament (Andres et al., 2007; Chapter 2). Therefore, when XLF L115A is combined with XRCC4, both LigaseIV and T4 DNA Ligase exhibit reduced levels of intra- and intermolecular ligation when compared to wildtype XLF (lane 4, Figure 5.3).

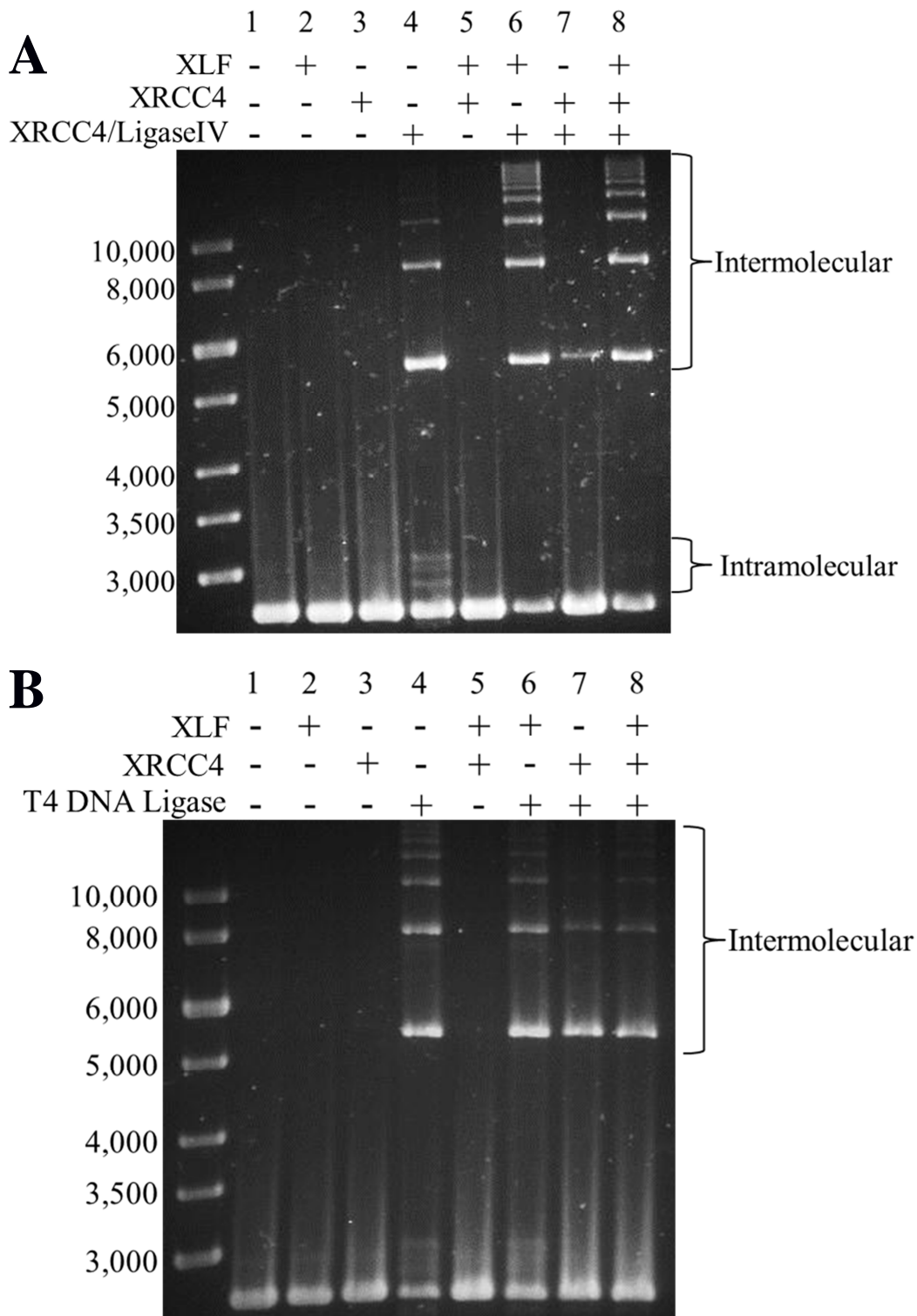


Figure 5.2 Effects of XLF and XRCC4 on Ligation (A) LigaseIV activity is stimulated by the addition of XLF, but decreased in the presence of excess XRCC4. (B) T4 DNA Ligase activity remains unaffected by addition of XLF, but is decreased in the presence of XRCC4. Inter- and intramolecular ligation are indicated by brackets. Molecular weight of the DNA ladder is in basepairs.

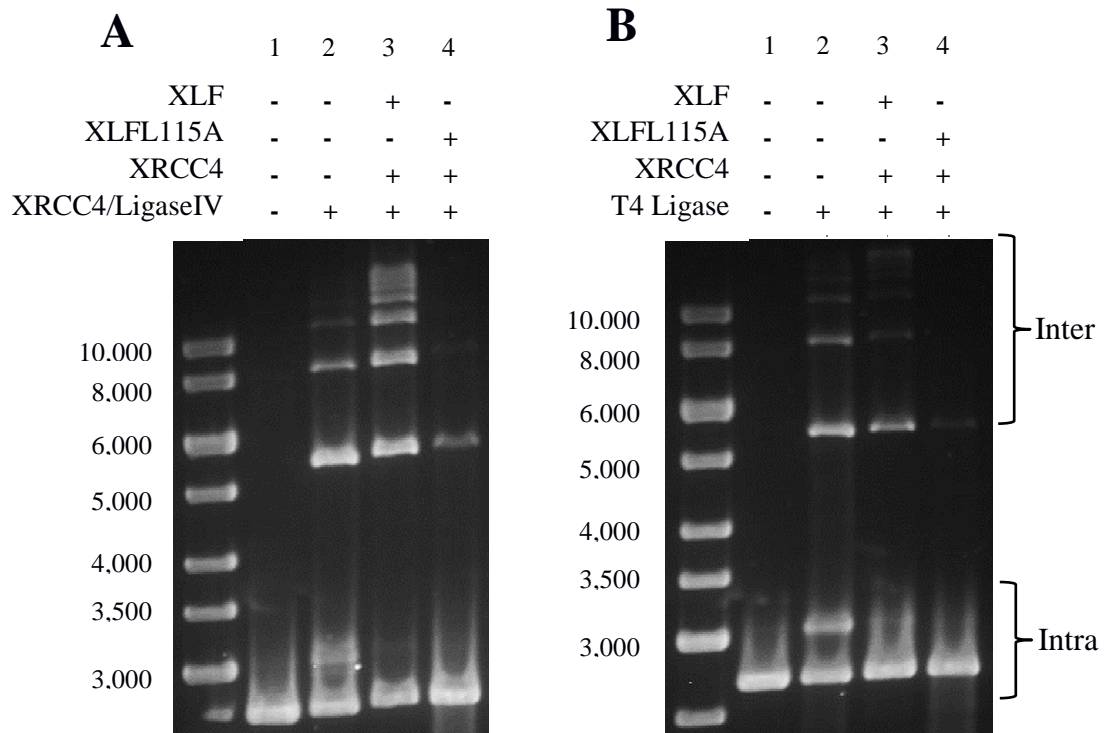
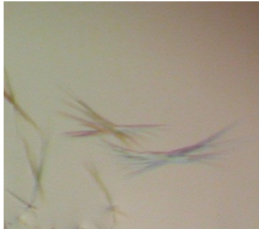
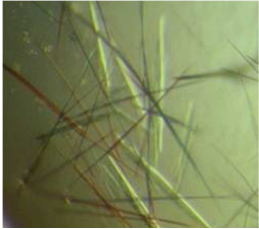
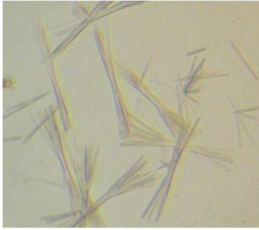



Figure 5.3 Effects of XLF L115A on Ligation (A) Effects of XLF L115A on ligation activity of LigaseIV. Ligase IV represents XRCC4 in complex with LigaseIV, as LigaseIV requires co-expression with XRCC4 for stability in bacterial cells. **(B)** Effects of XLF L115A on T4 DNA Ligase. Inter- and intramolecular ligation are indicated by brackets. Molecular weight of DNA ladder is in basepairs. XLF L115A abolishes XRCC4-XLF interactions (Andres et al., 2007).

5.4.3 Crystallization of XRCC4-LigaseIV

Four potential crystals of the XRCC4-LigaseIV complex were obtained using the hanging drop vapour diffusion method and are outlined in Table 5.1.

Table 5.1 Putative protein crystals of XRCC4-LigaseIV

Protein:Precipitant ratio (μL)	Precipitant	Crystals
1:1	1.1M di-ammonium tartrate, pH 7 (Hampton Index #26)	
1:1	60% v/v Tacsimate, pH 7 (Hampton Index #29)	
1:1	0.17M MgCl ₂ , 0.085M Tris, pH 8.5, 2.89M 1,6-hexanediol, 15% glycerol (The Cryos #3)	
1:0.5	0.04M Potassium phosphate, 16% PEG 8000, 20% Glycerol (The Cryos #69)	

5.5 Discussion

5.5.1 Crystallization of XRCC4-LigaseIV

While crystals were obtained of the full-length XRCC4-LigaseIV complex, they remained small and needle-like with no x-ray diffraction observed. Further optimization of these crystals is currently in progress but will be challenging. To obtain diffraction data suitable for determining a crystal structure, there is one primary objective for these crystals: decrease protein flexibility. Protein crystallography requires that every repeating unit that makes up the crystal is identical with respect to one another. This ensures that when exposed to x-rays, all resulting diffraction is constructive, so that a significant signal is produced. If any part of the protein(s) is (are) flexible, then destructive interference of the diffracting x-rays will occur, resulting in poor or even no diffraction signal. XRCC4-LigaseIV is a large complex of ~300 kDa (Lee et al., 2000). Large complexes typically pose difficulty in crystallization as the larger the complex, the greater potential for flexibility and a loss in diffraction signal; XRCC4-LigaseIV is no exception. Only the N-terminal 203 residues of XRCC4 have been crystallized, owing in large part to residues beyond 203 displaying flexibility. This has been confirmed by partial proteolysis experiments and more recently by SAXS analysis (Hammel et al., 2010). Furthermore, interaction with LigaseIV is unlikely to decrease this flexibility as the minimal interacting unit of LigaseIV (654-911) has already been crystallized with XRCC4, suggesting that the remaining 653 residues of LigaseIV are not in direct contact with XRCC4 (Wu et al., 2009). The only known feature of XRCC4's C-terminal tail is that it is required for bridging DNA ends in an XRCC4-XLF filament (Chapter 4).

Therefore, to decrease flexibility of XRCC4, it would be beneficial to either add XLF and DNA to XRCC4-LigaseIV, or truncate the C-terminus of XRCC4. Since adding more proteins will likely increase flexibility, truncation of XRCC4's C-terminus is the best option. As for LigaseIV, electron microscopy images of XRCC4-LigaseIV suggest some mobility of the N-terminus of LigaseIV; however the catalytic core appears to be limited in its overall flexibility (Recuero-Checa, 2009). Furthermore, the catalytic domain of LigaseI bound to DNA and portions of DNA ligase II have been solved, indicating that while crystallization of this complex will be difficult, it is not impossible (Cotner-Gohara, 2010; Pascal, 2004).

5.5.2 Stimulation of Ligation Activity

XRCC4-LigaseIV ligated a linearized plasmid to a closed circular form (intramolecular ligation) (Figure 5.3). Multiple bands were observed, representing various topoisomers, a feature characteristic of ligating DNA into a closed, circular state (Bjornsti et al., 1999). However, in the presence of XLF, these bands were no longer observed. For molecular biology purposes, this would be a beneficial tool for cloning purposes. However, within a cell, the site of a DNA double-strand break presents ligations of an intramolecular nature. The primary difference between this *in vitro* assay compared to *in vivo* activity is likely the availability and local concentration of DNA ends. *In vitro*, intramolecular ligation is likely preferred due to higher local concentrations of DNA ends, as they are on the same plasmid. This results in circular closed forms of DNA ligation products. However, in the presence of XLF, intermolecular ligation is increased, suggesting that a local concentration of DNA ends between different

plasmids is greater than the concentration of DNA ends on the same plasmid. This implies that in the presence of XLF, DNA ends from different plasmids are being sequestered in a conformation that prevents intramolecular ligation, perhaps due to steric constrictions imposed by circular plasmid DNA, therefore intermolecular ligation is favoured. *In vivo* the only available DNA ends would be the double-strand break, and accessibility would be determined by remodelling of the chromatin. Nevertheless, XLF should still be able to stimulate ligation in both situations likely by limiting DNA end diffusion and keeping the DNA ends at the site of damage.

XLF binds to XRCC4, and as a complex, both proteins are able to bridge DNA ends, but not in the presence of LigaseIV. However, they still form a nucleoprotein complex consisting of one or a few filaments (Chapter 4). Therefore, in the case of the assay presented here, what is the effect when the XRCC4-XLF interaction is abolished? As seen in Figure 5.3A, lane 4, abrogation of the XRCC4-XLF interaction via a point mutation in XLF (L115A) reduced ligation, suggesting that an XRCC4-XLF filament is required for stimulating ligation. One potential explanation for this observation relates to the XRCC4-XLF protein filaments' ability to bind DNA (Chapter 4). XRCC4-XLF DNA binding activity would limit DNA end diffusion by forming a potentially rigid nucleoprotein filament, thus *in vitro*, more intermolecular ligation is observed. However, if XLF is removed, a filament no longer forms and the DNA ends are more mobile and diffuse. Therefore, an XRCC4-XLF filament *in vivo* could maintain DNA ends of a double-strand break in close contact so that LigaseIV can efficiently access them for ligation.

A second possibility for increased ligation in the presence of XLF also relates to adenylation of LigaseIV and XRCC4-XLF filaments. As mentioned previously, ligation requires generation of a LigaseIV-adenylate complex. Once a single ligation event occurs however, ATP is removed from Ligase IV (Wei et al., 1995). Recent work by Riballo et al., (2008) found that in the presence of XLF and ATP, XRCC4-LigaseIV re-adenylation was highly stimulated compared to addition of ATP alone *in vitro* (Riballo et al., 2008). In terms of the ligation assay presented here, it further suggests that addition of XLF facilitates re-adenylation of LigaseIV using the ATP present within the buffer, allowing for more ligation events to occur, and therefore stimulated ligation is observed. Given that XLF and XRCC4 form protein filaments, it is possible that an XRCC4-XLF filament, and not XLF alone, is required for re-adenylation. It is possible that an XRCC4-XLF filament could add increased stability to XRCC4-LigaseIV on DNA such that once repair is finished, rather than being removed from the DNA, XRCC4-LigaseIV is bound longer, allowing time for re-adenylation. These results also suggest that the XRCC4-XLF filament may position LigaseIV in such a way that the catalytic lysine is more accessible for ATP to bind. Further experimentation would be needed to confirm this hypothesis, and could include experiments similar to Yano et al., (2011). Yano et al., (2011) examined recruitment of XLF and XRCC4 to a double-strand break and the length of time it remained at the damage *in vivo*. Applying this same experiment, but tracking LigaseIV and determining the length of time it remains at the site of a dsDNA break in normal, XLF^{-/-}, and XLF^{-/-} transfected with XLF L115A cell lines could be a starting point to understanding the mechanism behind XLF's ability to stimulate ligation and re-adenylate

LigaseIV. These results would be highly informative, and complementary to what is currently known about XRCC4-XLF's ability to stimulate LigaseIV efficiency *in vitro*.

Two other interesting observations were also made with regards to the ligation assay. The first observation was that addition of XLF did not stimulate T4 DNA Ligase activity, indicating that the XRCC4-XLF stimulation effect is specific to LigaseIV and NHEJ. The second observation was the inhibition of ligation by adding excess XRCC4 (lane 7, Figure 5.2). Since this was observed for both LigaseIV and T4 DNA Ligase, it suggests that XRCC4 is not binding to the proteins, but rather sequestering the DNA ends from the ligases. Also, XRCC4's DNA binding domain overlaps with XRCC4-LigaseIV binding (Chapter 5). These data suggest that the addition of excess XRCC4, which is not bound to LigaseIV or T4 DNA Ligase, is then free to bind DNA, blocking ligation activity. Therefore, it is possible that XRCC4 may also act as a ligation regulator to prevent LigaseIV from indiscriminately joining DNA ends in NHEJ.

Chapter 6: Summary and Conclusions

6.1 Summary

Work conducted during completion of this thesis focused on the structural and biochemical characteristics of human DNA repair proteins XRCC4 and XLF in order to better understand the mechanism by which these proteins function in NHEJ. The structure of XLF¹⁻²²⁴ was determined, and in conjunction with mutagenic and functional studies, the key residues involved in its interaction with XRCC4 were also elucidated. Similarly, amino acids in XRCC4 required for association with XLF were identified. A specific mutation to XLF (L115A) was unable to stimulate ligation of mismatched DNA ends and was also unable to interact with XRCC4. Therefore, the XRCC4-XLF interaction, and not XLF alone, appears to be necessary for stimulating DNA ligation of mismatched DNA ends during NHEJ. Structural and functional analysis of XLF and its interaction with XRCC4 suggested the potential formation of an extended XRCC4-XLF structure. To further examine the XRCC4-XLF interaction, and the ability to form extended oligomers, the XRCC4-XLF complex was probed through structural determination via x-ray crystallography. From this analysis, it was determined that the XRCC4-XLF complex could not readily pack into well-ordered crystals and that dehydration is key to improving data from crystals of large unit cells and high solvent content. Most significantly, though, the crystallization provided sufficient data to solve the structure of XRCC4¹⁻¹⁵⁷ in complex with XLF¹⁻²²⁴. The resulting structure revealed an XRCC4-XLF filament, confirming the model put forth from the initial XLF structure and mutagenic studies. Further studies suggested that the role of this filament in NHEJ appears to be essential for bridging a DNA double-strand break. Although the filament can exist in multiple states, both observed through biochemical assays and scanning force microscopy, only the larger multifilament complex is competent for bridging function. The multimeric state of the filament depends on the availability of a free XRCC4 tail

(157-200). Since LigaseIV interacts with this region of XRCC4, binding of LigaseIV appears to regulate the oligomeric state of the XRCC4-XLF filament. Additionally, DNA binding sites on both XLF and more notably, XRCC4 were established to expand the model of an XRCC4-XLF filament bridging DNA ends. Importantly, this work uncovered an essential XRCC4 regulatory region responsible for binding LigaseIV, DNA, or forming homotetramers. The bridging function of XRCC4-XLF multifilament bundles suggests an earlier role and mechanism for XRCC4-XLF in NHEJ, where the filament bridges DNA ends to stabilize them and maintain them in close proximity for efficient repair of the DNA double-strand break.

6.1.1 Challenges and Solutions in Crystallizing Multi-Protein Complexes

The use of structural biology to understand a protein's function has become common practice. In particular, x-ray crystallography can provide high-resolution models of protein structure that capture different states of protein activity, guiding biochemical work to allow more advanced understanding of a protein's mechanisms of action. This concept was applied to the research presented here, but not without some difficulty in crystallizing XRCC4-XLF and attempts at crystallizing XRCC4-LigaseIV. The purpose of crystallizing a protein is to increase the signal from x-ray diffraction. If a single protein was exposed to x-ray irradiation, the resulting diffraction pattern would be too weak to detect (low signal-to-noise ratio), and the sample would be destroyed long before sufficient data could be collected. In a crystal, however, there are multiple copies of the same repeating unit that are identical to one another and aligned in precisely the same orientation. Therefore diffraction, which behaves as a wave, is amplified when produced from multiple repeating units (constructive interference), and sample decay is spread out over a larger area. However, if there are differences between the individual repeating

units in the crystal or their alignment, the resulting diffraction pattern is significantly curtailed by destructive interference, and hence weakening of the signal (Glusker and Trueblood, 1985; McPherson, 2003). Challenges in crystallizing a single protein are compounded when crystallizing a multi-protein complex, as the probability of increasing flexibility and disorder in crystal lattice packing rises. This was the primary cause of difficulties when crystallizing XRCC4-XLF, especially given that the XRCC4 C-terminus is predicted to be mobile. Any region of a protein that is flexible has a tendency to reduce crystal contacts, thereby making stable crystal packing difficult. This effect was seen in the structure of XLF¹⁻²²⁴, where amino acids 80-92 in the head domains could not be modelled, due to a lack of electron density. This region was most likely a loop, as it fell between an α -helix and β -strand (Andres et al., 2007; Chapter 2). Loops are flexible secondary structure elements, and therefore would not be uniform in the crystal, resulting in destructive interference, and creating a lack of electron density for structure modelling.

Prior knowledge or predictions of the protein structure are useful tools when attempting to decrease protein mobility in a crystal. For XRCC4 and XLF, structures of each protein had been determined prior to crystallizing the complex (Junop et al., 2000; Chapter 2, Andres et al., 2007). From these structures it was observed that each had truncated C-terminal tails, due to predicted flexibility in these regions. Since the structure of XRCC4 and XLF's C-termini was not a priority to solve, these regions were deleted from the proteins, a technique commonly used for crystallizing flexible proteins in order to decrease the potential for poor crystal packing (Geiger et al., 2008; Huxford et al., 2000). Mutagenic and functional studies were also informative for crystallization, indicating that XRCC4 and XLF formed a complex through their individual head domains, and given that both proteins are homodimers, permitted generation of a model that XRCC4 and XLF would create a protein filament (Chapter 2, Andres et al., 2007).

The length of this filament in a protein crystal could extend infinitely and would only be limited by the amount of protein present. This would create a potentially elongated structure that could also put negative restrictions on protein packing within crystals. The best option for this case, then, was decreasing the protein complex to the minimal interacting region (XRCC4¹⁻¹⁵⁷, XLF¹⁻²²⁴), thus decreasing possible disorder in the protein crystal, yet still retaining enough of the proteins' structures to gain information on the XRCC4-XLF complex.

Initial structural information was obtained from the size of the crystal's unit cell. The dimensions of this unit cell contained one extremely long axis of 745 Å, suggesting the XRCC4-XLF structure existed as an extended filament. However, the initial crystals produced weak diffraction (>20 Å) again suggesting poor crystal packing. Nonetheless, one other important piece of information determined from initial data was the solvent content of the crystal. Typical macromolecular crystals have ~50% solvent content, while the predicted solvent content for XRCC4-XLF was ~70%. Such a large solvent content indicates there is a significant amount of volume unoccupied by protein, leaving room for mobility, which is damaging to crystal quality. To remove this impediment, the solvent content needed to be decreased, and to do this, dehydration proved very useful. Crystal dehydration was beneficial for several reasons. It removed solvent from the crystal, forcing the protein to pack more tightly, reducing mobility (Salunke et al., 1985; Frey, 1994). Also, less water in the crystal decreased the amount of radiation damage, as free radicals are formed in the water when ionized with x-ray radiation (Garman, 2010). Finally, a lack of water removed the danger of ice crystal formation, which can physically damage a crystal, thus, dehydration also served to enhance cryoprotection (Heras, B., and Martin, J.L., 2005). Overall, if a crystal can withstand dehydration, it is extremely beneficial to improving the diffraction quality of a crystal, as evidenced by the change

from 20Å to 3.94Å resolution data with the XRCC4-XLF complex, and ultimately the ability to obtain a structural solution (Chapter 3).

6.1.2 Protein Filaments are Not Unique to XRCC4-XLF

As discussed in the supplementary discussion of chapter 4, protein filaments are not unique to XRCC4-XLF and NHEJ. Filaments of Rad51 are formed during repair of DNA double-strand breaks via homologous recombination (Chen et al., 2007; Conway et al., 2004). Thus protein filaments, per se, are not a new theme in DNA repair. Work here, however, suggested that the similar structure of XRCC4 and XLF head domains and their dimeric relationships might make them uniquely suitable for generating a repeating, ‘filament’-like structure. This idea has gained recent validation through structural characterization of the protein spindle-assembly-6 (SAS-6). SAS-6 is involved in forming centrioles, which are the major microtubule organizing center for DNA replication in mammals (Leidel et al., 2005). The SAS-6 structure, recently discovered by Kitagawa et al., and van Breugel et al., (2011) bears an uncanny resemblance to XRCC4. Like XRCC4 and XLF, it has an N-terminal head domain containing a β -sandwich, as well as a long C-terminal α -helical tail extending away from the head, through which it dimerizes (Figure 6.1) (Kitigawa et al., 2011; van Breugel et al., 2011). SAS-6 also self-associates through its N-terminal head domain, much like XRCC4 interacts with XLF. While XRCC4 and XLF are different proteins, they are structural homologs, and therefore the XRCC4-XLF filament is also a form of self-association, such as observed with SAS-6. One main difference, though, is that XRCC4-XLF forms an

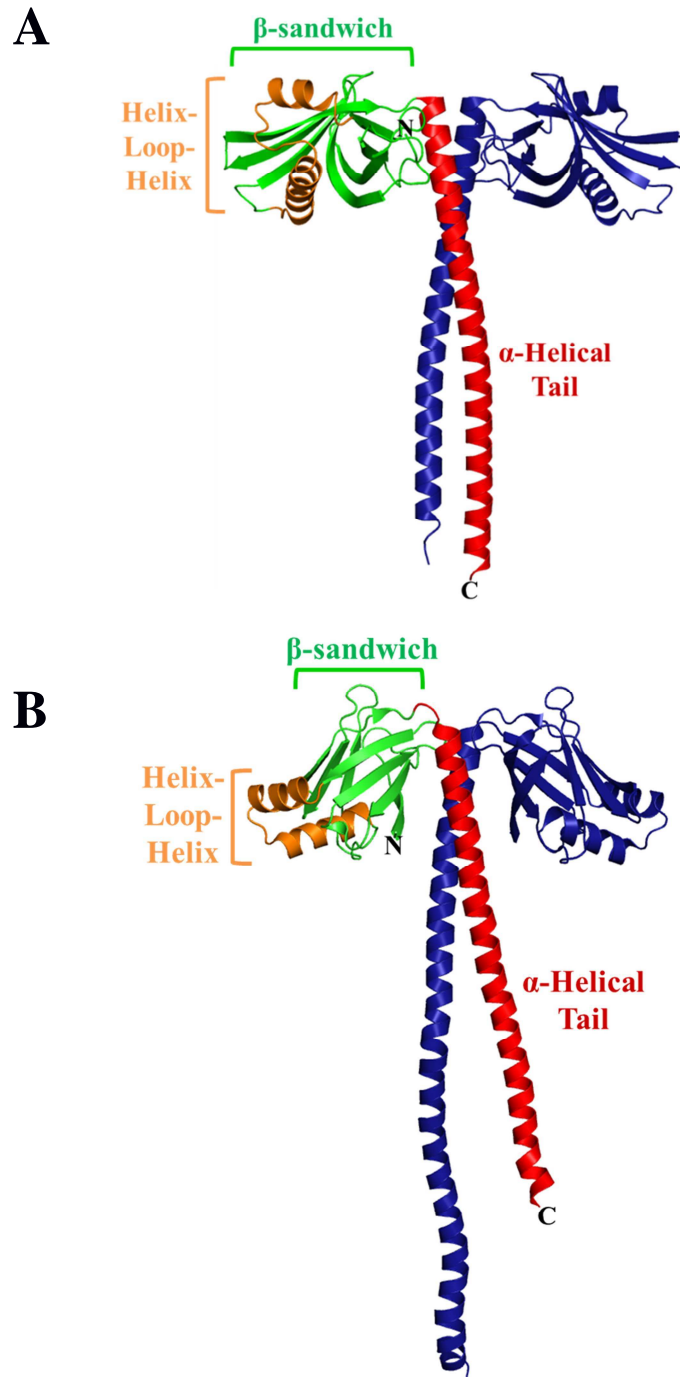


Figure 6.1 Comparison of XRCC4 and SAS-6 homodimers. (A) The crystal structure of SAS-6 from *Chlamydomonas reinhardtii* was determined by Kitigawa et al., and van Breugel et al., (2011). One monomer consists of a head domain encompassing a β -sandwich (green) and helix-loop-helix motif (orange), followed by an α -helical tail (red). The dimerizing monomer is in blue. N- and C-termini are in black (PDB 3QOX). (B) The crystal structure of human XRCC4, with one monomer in blue, and the other monomer illustrating the β -sandwich (green), helix-loop-helix motif (orange), and α -helical tail (red). N- and C-termini are in black (Junop et al., 2000; PDB 1FU1).

extended filament, whereas SAS-6 forms a 9-fold symmetric ring, with the C-terminal tails pointing out. The SAS-6 ring is still a protein filament, but based on the angle between each head domain interaction, has a fixed endpoint (Kitigawa et al., 2011; van Breugel et al., 2011). Again, this is comparable to XRCC4-XLF, which takes on a circular twist, but instead of closing, forms an extended left-handed helix, also with the C-terminal tails of XRCC4 and XLF pointing away from the centre (Chapter 4). Finally, the three examples of protein filaments examined thus far (XRCC4-XLF, Rad51, SAS-6) are all involved in activities required for the maintenance of DNA (double-strand break repair, replication), suggesting that protein filaments may be a specialized form of protein structure more widely involved in protecting genomic integrity than previously realized.

6.1.3 The Importance of the C-terminal Tails of XRCC4 and XLF

Even though the C-terminal tails of XRCC4 and XLF were not included in the XRCC4-XLF filament structure, their importance in NHEJ is significant. The first indication of the XRCC4 C-terminal tails' importance was revealed by their involvement in the formation of multi-filament bundles of the XRCC4-XLF filament (discussed in Chapter 4). Briefly, the crystal structure displayed evidence of XRCC4 tails folding over and 'cross-linking' neighbouring filaments. Consistent with this interpretation, sequestration of the C-terminal tails of XRCC4 through LigaseIV binding abolished formation of the XRCC4-XLF multi-filament in bridging assays and SFM studies. This further suggested that XRCC4 tails are needed for formation of multi-filament bundles associated with DNA bridging activity. Other evidence outside this body of work also suggests that the XRCC4 C-terminus is important in NHEJ. Research by Dr. Kathy Meek and Dr. Mauro Modesti (personal communication), examined the role of XLF and XRCC4 in V(D)J recombination. Their *in vivo* plasmid-based repair assay showed high

levels of recombination of coding ends in the presence of XLF, XRCC4, Ku70/80, and LigaseIV. Disruption of the XRCC4-XLF interaction via XRCC4 mutation caused a significant decrease in recombination efficiency. Interestingly, disrupting the XRCC4-XLF interaction interface by mutating XLF (L115A) did not have a noticeable effect on recombination efficiency. It is known that XLF L115 disrupts the head-to-head interaction between XRCC4 and XLF; therefore, the filament must be held together by another region of XRCC4, with the most likely candidate being the C-terminus. Similarly, the XLF C-terminus is also of interest. The research presented in Chapter 2 and 4 defined the XLF C-terminus as a DNA binding region, but the XLF C-terminal tail may also be of importance for stabilizing interactions with XRCC4. Yano et al., (2011) removed XLF's C-terminus, and *in vivo*, observed a decreased interaction, even though the head domains of each protein remained intact. Furthermore, studies of the yeast XLF homolog have shown that the N-terminus of XRCC4 associates with the C-terminus of *S. cerevisiae* XLF, suggesting that similar interaction may occur for the human homologs (Deshpande et al., 2007). A more extensive interaction interface between XRCC4 and XLF involving their C-terminal tails is realistic, since the buried surface area between the XRCC4 and XLF head domains is only 750 \AA^2 . A stable protein-protein interaction typically buries at least 1000 \AA^2 (Janin et al., 1990). The XRCC4 and XLF dimers each bury $\sim 2300 \text{ \AA}^2$ and $\sim 6500 \text{ \AA}^2$ respectively (Andres et al., 2007; Junop et al., 2000). Therefore, an increase in buried surface area by additional C-terminal tail interactions would make the XRCC4-XLF filament more stable than the current known state. Finally, the XRCC4 and XLF C-termini are the sites of phosphorylation by DNA-PKcs and ATM, and while these sites are not required for DNA double-strand break repair, it has been proposed that they are regulatory sites for disassembly of the repair complex (Yu et al.,

2003; Yu et al., 2008). Therefore, the C-terminal tails of both XRCC4 and XLF are likely significant for stable filament assembly and disassembly in NHEJ.

6.2 Future Directions

As more information becomes available about XRCC4 and XLF in NHEJ, more questions arise surrounding their role in DNA double-strand break repair. Many individual pieces of information are known about each protein, yet the details of how the data fits together into a cohesive NHEJ mechanism are still unclear. For example, XRCC4 and XLF bind and bridge DNA (Chapter 4) yet how these activities function in relation to the other DNA repair proteins, in particular LigaseIV, is unknown. Future work on XRCC4 and XLF should aim to examine two main targets – structures of larger protein-DNA complexes, and how these proteins impact ligation both biochemically and *in vivo*.

6.2.1 Structural Studies

Further structural information of two key complexes would be extremely useful in understanding how XLF and XRCC4 function in NHEJ. The first complex of interest would be the XRCC4-XLF filament with the C-terminal tails of each protein and in the presence of DNA. Based on the results in Chapter 4, this would determine the XLF-XRCC4-DNA interaction and whether or not the C-terminus of XRCC4 is involved in DNA binding when an XRCC4-XLF filament is formed, as opposed to solely aiding in formation of a multi-filament bundle. It would also indicate whether or not the XRCC4-XLF-DNA model in Chapter 4 is correct, with DNA running through the centre of the protein helix. Crystallizing XRCC4 or XLF separately with DNA would not be beneficial, since XLF and XRCC4 require long DNA substrates in order to bind

efficiently, and much research now points to an XRCC4-XLF filament binding DNA (Chapter 4). The second complex to solve would be that of XRCC4, XLF, and LigaseIV in the presence of DNA. Such a structure would show how an XRCC4-XLF filament is modified in the presence of LigaseIV, and would also illustrate how the catalytic domain of LigaseIV is positioned and regulated via XRCC4-XLF. Unlike the first complex, ‘breaking-down’ this large multi-protein structure into smaller parts would be of use. Determining only full-length XRCC4 and LigaseIV in the presence and/or absence of DNA, or solving the structure of an XRCC4-LigaseIV-XLF complex would also be informative, albeit not as complete an answer as the proposed XRCC4-XLF-LigaseIV-DNA structure.

Crystallography would be the ideal method to achieve a detailed answer to these mechanistic questions, as structures are typically of high enough resolution to determine individual amino acid interactions. However, the size of the suggested complexes is large, especially if XRCC4-XLF retains a filament structure. Therefore it is likely that the ability to generate a high quality crystal would be extremely challenging. On the other hand, it is also possible that by introducing DNA or LigaseIV, the proteins may adopt a ‘locked’ conformation, reducing protein flexibility. However, crystallography is not the only way to examine these structures. Electron microscopy has already proven informative for full-length XRCC4 bound to LigaseIV, and would also be useful for gaining insight of these larger structures (Recuero-Checa et al., 2009). Using results from electron microscopy and modelling in the known crystal structures of each protein could provide informative results, providing frozen ‘snapshots’ of the proteins when associated into larger protein-DNA complexes. To avoid a static image, though, and to obtain a structure of more biological significance, SAXS could also be used to complement structural analysis of these complexes. SAXS has already provided initial data on

XRCC4-XLF, and established that XRCC4-XLF is capable of forming a filament in solution (Hammel et al., 2010). The primary benefit of SAXS analysis is that data is acquired in an aqueous environment, which is arguably of more biological significance, and may in turn provide more biologically relevant states of protein interaction. Also, like electron microscopy, known crystal structures could be modelled into envelopes of protein complexes generated via SAXS analysis. Despite their obvious strengths, analysis by SAXS and electron microscopy only yield low resolution data. If any major changes occur between the individual protein conformations in the crystal structure compared to the SAXS or electron microscopy data, modelling atomic resolution structures into lower resolution data may become very difficult. Therefore, while crystallography will likely be challenging, a combination of a high-resolution structure (crystallography), with a lower-resolution structure (SAXS or electron microscopy), is the best approach for determining further mechanistic insight of these important NHEJ complexes.

6.2.2 Effect of XRCC4-XLF on Ligation Activity of LigaseIV

Initial studies suggested that XLF stimulates ligation of mismatched DNA ends (Tsai et al., 2007). Research presented in Chapter 5 suggests that it is not XLF alone that causes this, but the XRCC4-XLF filament. Chapter 4 also indicated that XRCC4-XLF filament bundles are able to bridge a DNA double-strand break, which would also be expected to impact ligation efficiency. The precise mechanism behind XRCC4-XLF stimulation of ligation is still unclear. Future work on how XRCC4-XLF affects ligation should address this question. Simple experiments to begin examining this topic could include taking the ligation assay outlined in Chapter 5 and expanding it to include not just the XLF L115A mutant, but also XRCC4 mutants that abolish the XLF interaction, as

well as DNA binding mutants in both XLF and XRCC4. This would first confirm that the XRCC4-XLF interaction is required to stimulate ligation, as shown by the initial data in Chapter 5, and it would also discriminate whether XLF or XRCC4's DNA binding activities are required for efficient ligation. In conjunction with biochemical studies, similar experiments need to be transferred to an *in vivo* ligation assay, such as that carried out by Smith et al., (2003), using cell lines deficient in XLF, XRCC4, or LigaseIV and complementing with each protein and its mutants that lack DNA- or protein-binding. These results would help determine the biological relevance of *in vitro* observations.

One additional key player in NHEJ that should not be overlooked in these assays is Ku70/80. Ku70/80 recruits XLF and XRCC4 to DNA double-strand breaks, and interacts with LigaseIV. *In vitro*, Ku70/80 stimulates ligation 20-fold, compared to XRCC4 alone (McElhinny et al., 2000; Yano et al., 2008a, b). Extending the *in vitro* assays described above to include Ku70/80 would be informative with regards to how XLF, and the XLF-XRCC4 filament affects ligation rate in the presence of Ku70/80. Also, defining binding interfaces between Ku70/80 and other DNA repair proteins would add more information that could help to probe various interactions and their importance at all stages of NHEJ. Currently, it is already known that Ku70/80 interacts with XLF through a DNA/XLF-C-terminal reaction, and with LigaseIV BRCT domains through its N-terminus (Yano et al., 2011; Costantini et al., 2007). Therefore, extending the above assays to include Ku70/80 would establish a more realistic NHEJ reaction, and inform on XRCC4-XLF filaments' overall function in NHEJ.

6.3 Significance and Conclusions

The research presented within this dissertation provides new insight into NHEJ, with particular regard to the roles of XRCC4 and XLF. Within the field it is currently

thought that XRCC4 and XLF function solely at the final stage of ligation, yet no evidence exists to explain the mechanism by which each protein stimulates LigaseIV activity. The structure of XLF provided the first look at this protein, confirming its predicted structural homology to XRCC4, with a unique divergence in the C-terminal tail, yet it did not explain LigaseIV stimulation by XLF (Chapter 2; Andres et al., 2007). The identification of amino acids necessary for XRCC4 and XLF interaction, and further determining that this interaction was required to stimulate ligation, indicated a role for XLF and supported the idea that XLF and XRCC4 are only active towards the end of NHEJ. More recently, however, an earlier role for XRCC4 has been suggested through various reports in the literature. Budman et al., (2007) carried out an *in vitro* end joining assay, illustrating that processing of DNA ends required XRCC4, placing XRCC4 at an earlier step in ligation. Similarly, the presence of XLF is required to stimulate gap filling by Pol μ and λ , again indicating a role in processing, rather than simply ligation (Akopiants et al., 2009). The results, then, of Chapter 4 would suggest by extension that it is an XRCC4-XLF filament that is involved in DNA end processing, and not XLF or XRCC4 alone. The observed effects for Pol μ and λ are likely a result of XRCC4-XLF filaments bridging DNA ends, as polymerization is often more efficient when DNA ends are aligned, although this requires experimental confirmation (Zhang et al., 2001). More recently, studies have also suggested that XRCC4 and XLF are involved in initial binding of a DNA double-strand break. Ku70/80 has been shown to recruit and interact with XLF in a DNA-mediated interaction after formation of a double-strand break. Furthermore, addition of XRCC4 to double-strand breaks stabilizes XLF at the break site (Yano et al., 2008b). This suggests that XRCC4 and XLF are part of NHEJ right from the start, and most likely, throughout all of NHEJ presumably as an XRCC4-XLF filament.

XRCC4-XLF filaments were observed bridging DNA both in parallel and end-to-end according to SFM images (Chapter 4). This suggests that bridging may occur in multiple ways during NHEJ, and would likely be subject to the conformation of DNA ends. DNA double-strand breaks are generally illustrated in a linear fashion, or end-to-end. However, within a cell, these breaks may also be parallel, with the break at the ‘tip’ of the nucleoprotein complex (Figure 6.2). When a double-strand break occurs, chromatin is remodelled to make room for DNA repair proteins, creating perturbations up to several megabasepair lengths away from the site of damage in eukaryotic cells, and 40 basepairs away in *Saccharomyces cerevisiae* (Rogakou et al., 1999; Tsukuda et al., 2005). This leaves the DNA ends exposed and mobile, able to take on either conformation

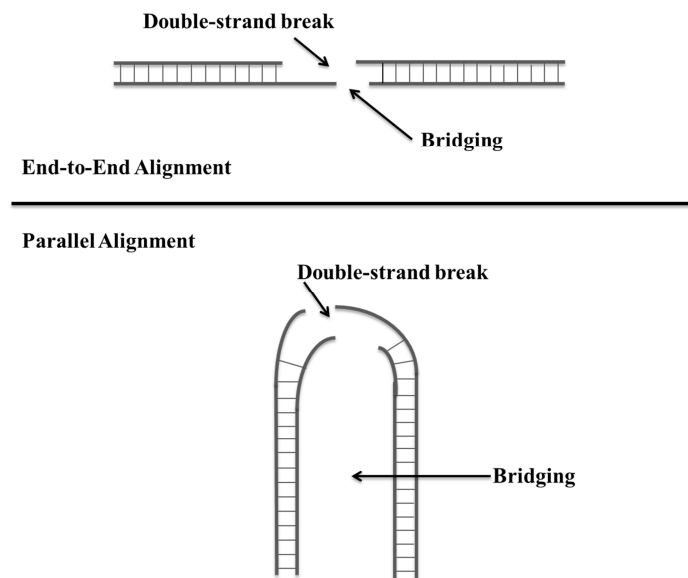


Figure 6.2 Illustration of the potential conformation of DNA ends at a double-strand break. DNA ends may align end-to-end, with XRCC4-XLF filaments bridging across the break. DNA ends may also align in parallel, with the double-strand break at the ‘tip’, and XRCC4-XLF filaments bridging between the parallel DNA strands.

observed via SFM. If DNA bridging occurs by formation of multiple filaments wrapping around DNA, stabilized by XRCC4 C-terminal tails binding adjacent filaments (Chapter 4, Figure 4.5A-C), then this would allow for the end-to-end alignment observed in SFM. However, if DNA bridging occurs by forming XRCC4 homotetramers between XRCC4-XLF filaments on adjacent DNA strands (Chapter 4, Figure 4.5D, E), then this would allow for the parallel bridging observed. These two modes of bridging can also account for the disruption observed when BRCT domains of LigaseIV are added to the bridging assay. BRCT domains bind to the same region of XRCC4 that is responsible for homotetramerization, and also encompasses the C-terminal tails of XRCC4, thus blocking XRCC4 C-terminal tails from other interactions required for bridging. Therefore, the different modes of DNA bridging observed in SFM are likely biologically relevant, as XRCC4-XLF filaments are adaptable for different modes of bridging DNA double-strand breaks.

SFM also illustrated that XRCC4-XLF filaments existed both as single and multi-filament bundles, and that binding of LigaseIV BRCT domains, or removing XRCC4 C-terminal tails only allowed formation of a simple filament (Chapter 4). Combined with the theory presented above that XRCC4-XLF filaments persist throughout NHEJ as either a single or multi-filament bundle, the current model for NHEJ can be adapted as seen in Figure 6.3. Upon double-strand break formation, Ku70/80 still binds the initial break, recruiting XLF and XRCC4 (Mimori et al., 1986; Yano et al., 2008a, b). Once recruited, XRCC4 and XLF would form multi-filament bundles (consisting of 2 or more filaments) that bridge the DNA double-strand break. At this point, Ku70/80 would then recruit DNA-PKcs, which would potentially disrupt the multi-filament bundle to a smaller species of fewer or even a single filament(s), at the site of the double-strand break. The presence of this smaller XRCC4-XLF filament(s) may no longer bridge the DNA break,

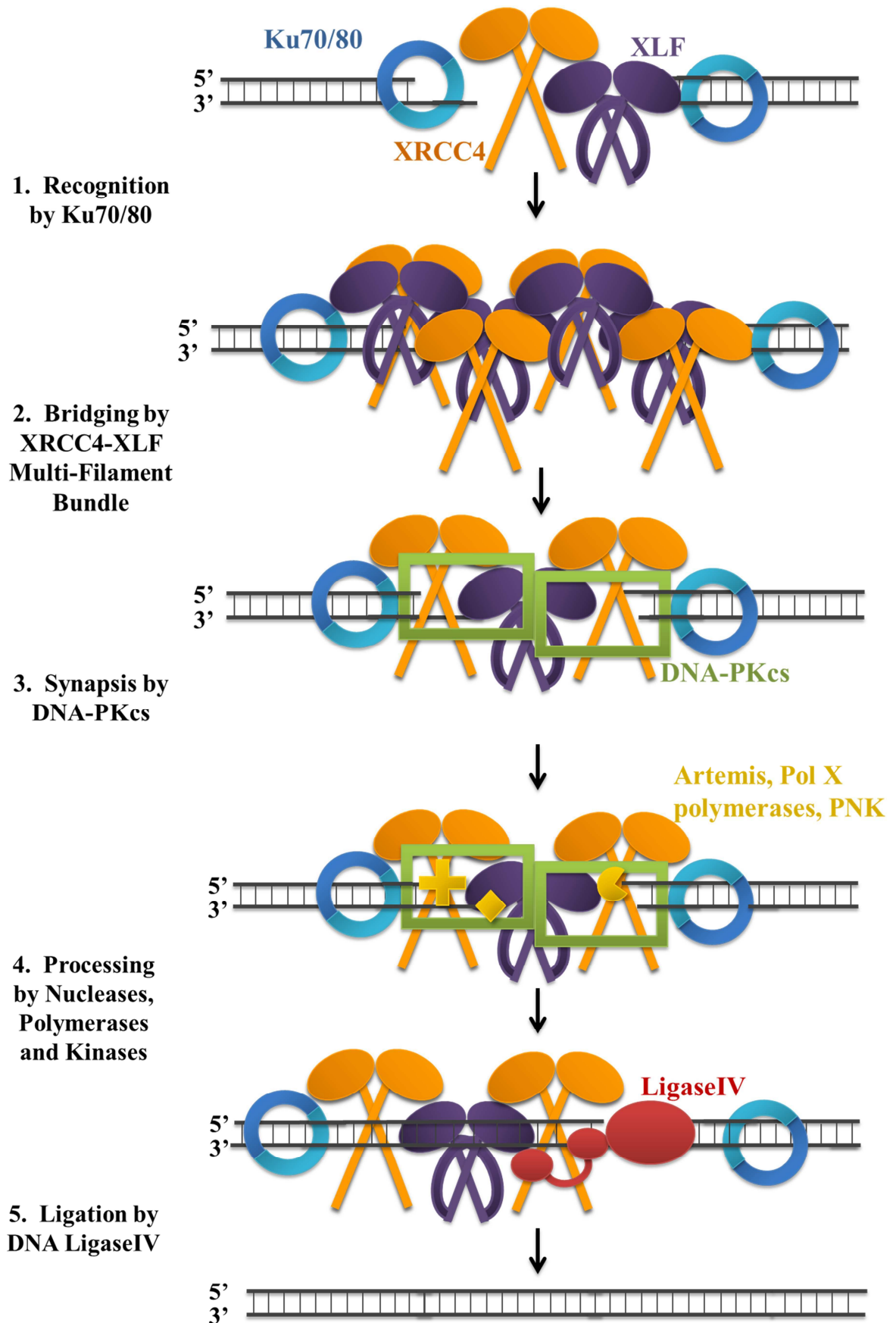


Figure 6.3 Proposed model for non-homologous end-joining. The initial model in Chapter 1 (Figure 1.2) is now modified to include the multi-filament bundle of XRCC4-XLF bridging the initial DNA double-strand break, and its subsequent remodelling to a smaller filament upon recruitment of the remaining non-homologous end-joining repair proteins. The XRCC4-XLF filament is depicted as a straight line for clarity. For three-dimensional representation of the filament encircling the DNA, please refer to Chapter 4.

but still aid in the alignment and stability of the DNA ends. This would explain the increased activity of the PolX family of polymerases in the presence of XRCC4 and XLF, since alignment of the break by XRCC4-XLF would prevent the Pol X polymerases from having to align the DNA ends (Akopiants et al., 2009). During ligation, the XRCC4-XLF filament will be a smaller species than the large multi-filament bundle, as SFM illustrated that the addition of multiple BRCT domains of LigaseIV resulted in smaller filament species, however the presence of the filament will still add stability and alignment of the DNA break, accounting for the increased ligation efficiency observed when XLF is present (Tsai et al., 2007; Chapter 5). Additionally, while a smaller filament bundle or single filament is present at the site of active repair, the multi-filament bundle may persist outside this area, as chromatin remodelling when double-strand breaks occur leaves DNA free of histones over a large area (Rogakou, 1999; Tsukuda, 2005). An XRCC4-XLF multi-filament bundle outside the active repair zone could protect the DNA from nucleases, but also add an overall stability to the repair complex, which would also contribute to the increased polymerization and ligation effects observed. Parts of this adapted NHEJ model are still highly speculative and require further research, but the addition of an XRCC4-XLF filament that persists throughout NHEJ fits in with recent data and challenges the current understanding of the non-homologous end-joining pathway.

References

- Abramoff, M.D., Magelhaes, P.J., Ram, S.J. (2004). Image Processing with ImageJ. *Biophotonics International*. **11**: 36-42.
- Adams, P.D., Grosse-Kunstleve, R.W., Hung, L.W., Ioerger, T.R., McCoy, A.J., Moriarty, N.W., Read, R.J., Sacchettini, J.C., Sauter, N.K., and Terwilliger, T.C. (2002). PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr*. **58**: 1948-1954.
- Ahnesorg, P., Smith, P., and Jackson, S.P. (2006). XLF interacts with the XRCC4-DNA ligase IV complex to promote DNA nonhomologous end-joining. *Cell* **2**: 301-313.
- Akopiants, K., Zhou, R.Z., Mohapatra, S., Valerie, K., Lees-Miller, S.P., Lee, K.J., Chen, D.J., Revy, P., de Villartay, J.P., and Povirk, L.F. (2009). Requirement for XLF/Cernunnos in alignment-based gap filling by DNA polymerases lambda and mu for nonhomologous end joining in human whole-cell extracts. *Nucleic Acids Res* **37**: 4055-4062.
- Alt, F.W., and Baltimore, D. (1982). Joining of immunoglobulin heavy chain gene segments: implications from a chromosome with evidence of three D-JH fusions. *Proc Natl Acad Sci USA* **79**: 4118-4122.
- Andres, S.N., Modesti, M., Tsai, C.J., Chu, G., and Junop, M.S. (2007). Crystal structure of XLF: a twist in non-homologous end-joining. *Mol Cell* **28**: 1093-1101.
- Aoufouchi, S., Flatter, E., Dahan, A., Faili, A., Bertocci, B., Storck, S., Delbos, F., Cocea, L., Gupta, N., Weill, J.C., and Reynaud, C.A. (2000). Two novel human and mouse DNA polymerases of the PolX family. *Nucl Acids Res* **28**: 3684-3693.
- Ban, N., Nissen, P., Hansen, J., Capel, M., Moore, P.B. and Steitz, T.A. (1999). Placement of protein and RNA structures into a 5 Å-resolution map of the 50S ribosomal subunit. *Nature* **400**: 841-847
- Banumathi, S., Dauter, M., & Dauter, Z. (2002). SAD/MAD phasing at high resolution using tantalum bromide cluster *Acta Cryst* **A58**: c81.
- Bennardo, N., Cheng, A., Huang, N., and Stark, J.M. (2008). Alternative-NHEJ is a mechanistically distinct pathway of mammalian chromosome break repair. *PLoS Genet* **4**: e1000110.
- Bennet, R.A., Swerdlow, P.S., and Povirk, L.F. (1993). Spontaneous cleavage of bleomycin-induced abasic sites in chromatin and their mutagenicity in mammalian shuttle vectors. *Biochemistry* **32**: 3188-3195.

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

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

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

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

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

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

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

Buck, D., Malivert, L., de Chasseval, R., Barraud, A., Fondanèche, M.C., Sanal, "O., Plebani, A., Stéphan, J.L., Hufnagel, M., le Deist, F., Fischer, A., Durandy, A, de Villartay, J.P., and Revy, P. (2006). Cernunnos, a novel nonhomologous end-joining factor, is mutated in human immunodeficiency with microcephaly. *Cell* **124**: 287-299.

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

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

Calsou, P., Delteil, C., Frit, P., Drouet, J., and Salles, B. (2003). Coordinated assembly of Ku and p46 subunits of the DNA-dependent protein kinase on DNA ends is necessary for XRCC4-ligase IV recruitment. *J Mol Biol* **326**: 93-103.

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

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

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

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

Chan, D.W., Ye, R., Veillette, C.J., and Lees-Miller, S.P. (1999). DNA –dependent protein kinase phosphorylation sites in Ku70/80 heterodimer. *Biochemistry* **38**: 1819-1828.

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

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

Chen, L.T., Ko, T.P., Chang, Y.C., Lin, K.A., Chang, C.S., Wang, A.H., and Wang, T.F. (2007). Crystal structure of the left-handed archaeal RadA helical filament: identification of a functional motif for controlling quaternary structures and enzymatic functions of RecA family proteins. *Nucleic Acids Res* **35**: 1781-1801.

Chen, X., and Tomkinson, A.E. (2011). Yeast Nej1 is a key participant in the initial end binding and final ligation steps of nonhomologous end joining. *J Biol Chem* **286**: 4931-4940.

Cohen, S.N., Chang, A.C.Y., and Hsu, L. (1972). Nonchromosomal antibiotic resistance in bacteria: genetic transformation of *Escherichia coli* by R-factor DNA. *Proc Natl Acad Sci USA* **69**: 2110-2114.

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

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

- Cotner-Gohara, E., Kim, I.K., Hammel, M., Tainer, J.A., Tomkinson, A.E., and Ellenberger, T. (2010). Human DNA ligase III recognizes DNA ends by dynamic switching between two DNA-bound states. *Biochemistry* **49**: 6165-6176.
- Critchlow, S.E., Bowater, R.P., and Jackson, S.P. (1997). Mammalian DNA double-strand break repair protein XRCC4 interacts with DNA ligase IV. *Curr Biol* **7**: 588-598.
- Datta, S., Prabu, M.M., Vaze, M.B., Ganesh, N., Chandra, N.R., Muniyappa, K., and Vijayan, M. (2000). Crystal structures of Mycobacterium tuberculosis RecA and its complex with ADP-AIF (4): implications for decreased ATPase activity and molecular aggregation. *Nucleic Acids Res* **28**: 4964-4973.
- DeFazio, L.G., Stansel, R.M., Griffith, J.D., and Chu, G. (2002). Synapsis of DNA ends by DNA-dependent protein kinase. *EMBO J* **21**:3192-3200.
- DeLano, W.L. (2002). The PyMOL User's Manual (Palo Alto, CA, USA.: DeLano Scientific).
- DeRose, E.F., Clarkson, M.W., Gilmore, S.A., Galban, C.J., Tripathy, A., Havener, J.M., Mueller, G.A., Ramsden, D.A., London, R.E., and Lee, A.L. (2007). Solution structure of polymerase mu's BRCT domain reveals an element essential for its role in nonhomologous end joining. *Biochemistry* **46**: 12100-12110.
- Deshpande, R.A. and Wilson, T.E. (2007). Modes of interaction among yeast Nej1, Lif1 and Dnl4 proteins and comparison to human XLF, XRCC4 and Lig4. *DNA Repair (Amst)* **6**: 1507-1516.
- Desiderio, S.V., Yancopoulos, G.D., Paskind, M., Thomas, E., Boss, M.A., Landau, N., Alt, F.W., and Baltimore, D. (1984). Insertion of N regions into heavy-chain genes is correlated with expression of terminal deoxytransferase in B cells. *Nature* **311**: 752-755.
- deVillartay, J.P., Shimazaki, N., Charbonnier, J.B., Fischer, A., Mornon, J.P., Lieber, M.R., and Callebaut, I. (2009). A histidine in the beta-CASP domain of Artemis is critical for its full in vitro and in vivo functions. *DNA Repair (Amst)* **8**: 202-208.
- deVriest, E., van Driel, W., Bergsma, W.G., Arnberg, A.C., and van der Vliet, P.C. (1989). HeLa nuclear protein recognizing DNA termini and translocating on DNA forming a regular DNA-multimeric protein complex. *J Mol Biol* **208**: 65-78.
- Di Capua, E., Engel, A., Stasiak, A., and Koller, T. (1982). Characterization of complexes between RecA protein and duplex DNA by electron microscopy. *J Mol Biol* **157**: 87-103.
- Ding, Q., Reddy, Y.V., Wang, W., Woods, T., Douglas, P., Ramsden, D.A., Lees-Miller, S.P., and Meek, K. (2003). Autophosphorylation of the catalytic subunit of the DNA-dependent protein kinase is required for efficient end processing during DNA double-strand break repair. *Mol Cell Biol* **23**: 5836-5848.

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

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

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

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

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

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

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

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

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

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

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

Fukumura, R., Araki, R., Fujimori, A., Mori, M., Saito, T., Watanabe, F., Sarashi, M., Itsukaichi, H., Eguchi-Kasai, K., Sato, K., Tatsumi, K., and Abe, M. (1998). Murine cell line SX9 bearing a mutation in the dna-pkcs gene exhibits aberrant V(D)J recombination not only in the coding joint but also in the signal joint. *J Biol Chem* **273**: 13058-13064.

Galkin, V.E., Wu, Y., Zhang, X.P., Qian, X., He, Y., Heyer, W.D., Luo, Y., and Egelman,

E.H. (2006). The Rad51/RadA N-terminal domain activates nucleoprotein filament ATPase activity. *Structure* **14**: 983-992.

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

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

Garcia-Diaz, M., Bebenek, K., Sabariego, R., Dominguez, O., Rodriguex, J., Kirchhoff, T., Garcia-Palmero, E., Picher, A.J., Juarez, R., Ruiz, J.F., Kunkel, T.A., and Blanco, L. (2002). DNA polymerase lambda, a novel DNA repair enzyme in human cells. *J Biol Chem* **277**: 13184-13191.

Garman, E.F. (2010). Radiation damage in macromolecular crystallography: what is it and why should we care? *Acta Cryst* **D66**: 339-351.

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

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

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

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

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

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

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

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

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

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

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

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

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

Gu, Y., Jin, S., Gao, Y., Weaver, D.T., and Alt, F.W. (1997). Ku70-deficient embryonic stem cells have increased ionizing radiosensitivity, defective DNA end-binding activity, and inability to support V(D)J recombination. *Proc Natl Acad Sci USA* **94**: 8076-8081.

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

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

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

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

Hartley, K.O., Gell, D., Smith, G.C.M., Zhang, H., Divecha, N., Connelly, M.A., Admon, A., Lees-Miller, S.P., Anderson, C.W., and Jackson, S.P. (1995). DNA-dependent

protein kinase catalytic subunit: A relative of phosphatidylinositol 3-kinase and the ataxia telangiectasia gene product. *Cell* **82**: 849-856.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Kadyk, L.C., and Hartwell, H. (1992). Sister chromatids are preferred over homologs as substrates for recombinational repair in *Saccharomyces cerevisiae*. *Genetics* **132**: 387-402.

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

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

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

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

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

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

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

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

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

Lee, J.W., Blanco, L., Zhou, T., Garcia-Diaz, M., Bebenek, K., Kunkel, T.A., Wang, Z., and Povirk, L.F. (2003). Implication of DNA polymerase lambda in alignment-based gap

filling for nonhomologous DNA end joining in human nuclear extracts. *J Biol Chem* **279**: 805-811.

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

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

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

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

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

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

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

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

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

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

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

- Lieber, M.R., Hesse, J.E., Mizuuchi, K., and Gellert, M. (1988b). Lymphoid V(D)J recombination: nucleotide insertion at signal joints as well as coding joints. *Proc Natl Acad Sci USA* **85**: 8588-8592.
- Lim, D.S., Kim, S.T., Xu, B., Maser, R.S., Lin, J., Petrini, J.H., and Kastan, M.B. (2000). ATM phosphorylates p95/nbs1 in an S-phase checkpoint pathway. *Nature* **404**: 613-617.
- Lu, H., Pannicke, U., Schwarz, K., and Lieber, M.R. (2007). Length-dependent binding of human XLF to DNA and stimulation of XRCC4.DNA Ligase IV activity. *J Biol Chem* **282**: 11155-11162.
- Ma, Y., Pannicke, U., Schwarz, K., and Lieber, M.R. (2002). Hairpin opening and overhang processing by an Artemis/DNA-dependent protein kinase complex in nonhomologous end joining and V(D)J recombination. *Cell* **6**: 781-794.
- Ma, Y., Pannicke, U., Lu, H., Niewolik, D., Schwarz, K., and Lieber, M.R. (2005). The DNA-dependent protein kinase catalytic subunit phosphorylation sites in human Artemis. *J Biol Chem* **280**: 33839-33846.
- Mahajan, K.N., McElhinny, N., Mitchell, B.S., and Ramsden, D.A. (2002). Association of DNA polymerase mu (pol mu) with Ku and ligase IV: role for pol mu in end-joining double-strand break repair. *Mol Cell Biol* **22**: 5194-5202.
- Maiti, R., Van Domselaar, G. H., Zhang, H., and Wishart, D.S. (2004). SuperPose: a simple server for sophisticated structural superposition. *Nucleic Acids Res*, **32**: 590-594.
- Malivert, L., Ropars, V., Nunez, M., Drevet, P., Miron, S., Faure, G., Guerois, R., Mornon, J.P., Revy, P., Charbonnier, J.B., Callebaut, I., & de Villartay, J.P. (2010). Delineation of the Xrcc4-interacting region in the globular head domain of cernunnos/XLF. *J Biol Chem* **285**: 26475-26483.
- Mandel, C.R., Kaneko, S., Zhang, H., Gebauer, D., Vethantham, V., Manley, J.L., and Tong, L. (2006). Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease. *Nature* **444**: 953-956.
- Martin, I.V., and MacNeill, S.A. (2002). ATP-dependent DNA ligases. *Genome Biol* **3**: reviews3005-3005.7.
- Maser, R.S., Monsen, K.J., Nelms, B.E., and Petrini, J.H. (1997). hMre11 and hRad50 nuclear foci are induced during the normal cellular response to DNA double-strand breaks. *Mol Cell Biol* **17**: 6087-6096.
- Matsuoka, S., Rotman, G., Ogawa, A., Shiloh, Y., Tamai, K., and Elledge, S.J. (2000). Ataxia telangiectasia-mutated phosphorylates Chk2 in vivo and in vitro. *Proc Natl Acad Sci USA* **97**: 10389-10394.

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

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

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

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

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

Meek, K., Douglas, P., Cui, X., Ding, Q. and Lees-Miller, S. P. (2007) *trans*-Autophosphorylation at DNA-dependent protein kinase's two major autophosphorylation site clusters facilitates end processing but not end joining. *Mol Cell Biol* **27**: 3881–3890.

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

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

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

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

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

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

- Moshous, D., Callebaut, I., de Chasseval, R., Corneo, B., Cavazzana-Calvo, M., Le Deist, F., Tezcan, I., Sanal, O., Bertrand, Y., Philippe, N., Fischer, A., and de Villartay, J.P. (2001). Artemis, a novel DNA double-strand break repair/V(D)J recombination protein, is mutated in human severe combined immune deficiency. *Cell* **105**: 177-186.
- Mueller, G.A., Moon, A.F., DeRose, E.F., Havener, J.M., Ramsden, D.A., Pedersen, L.C., and London, R.E. (2008). A comparison of BRCT domains involved in nonhomologous end-joining: Introducing the solution structure of the BRCT domain of polymerase lambda. *DNA Repair* **7**: 1340-1351.
- Murshudov, G. (2005). Refmac5 (CCP4: Supported Program). April 15, 2.
- Murzin, A.G. (1993). OB(oligonucleotide/oligosaccharide binding)-fold: common structural and functional solution for non-homologous sequences. *EMBO J* **12**: 861-867.
- Niewolik, D., Pannicke, U., Lu, H., Ma, Y., Wang L.C., Kulesza, P., Zandi, E., Lieber, M.R., and Schwarz, K. (2006). DNA-PKcs dependence of Artemis endonucleolytic activity, differences between hairpins and 5' or 3' overhangs. *J Biol Chem* **281**: 33900-33909.
- O'Driscoll, M., Cerosaletti, K.M., Girard, P.M., Dai, Y., Stumm, M., Kysela, B., Hirsch, B., Gennery, A., Palmer, S.E., Seidel, J., Gatti, R.A., Varon, R., Oettinger, M.A., Neitzel, H., Jeggo, P.A., and Concannon, P. (2001). DNA ligase IV mutations identified in patients exhibiting developmental delay and immunodeficiency. *Mol Cell* **8**: 1175-1185.
- Oettinger, M.A., Schatz, D.G., Gorka, C., and Baltimore, D. (1990). RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination. *Science* **248**: 1517-1523.
- Ooi, S.L., Shoemaker, D.D., and Boeke, J.D. (2001). A DNA microarray-based genetic screen for nonhomologous end-joining mutants in *Saccharomyces cerevisiae*. *Science* **551**: 2552-2556.
- Otwinowski, Z., and Minor, W. (1997). Processing of X-ray Diffraction Data Collected in Oscillation Mode. *Methods Enzymol.* **276**: 307-326.
- Paillard, S., and Strauss, F. (1991). Analysis of the mechanism of interaction of simian Ku protein with DNA. *Nucleic Acids Res* **19**: 5619-5624.
- Pascal, J.M., O'Brien, P.J., Tomkinson, A.E., and Ellenberger, T. (2004). Human DNA ligase I completely encircles and partially unwinds nicked DNA. *Nature* **432**: 473-478.
- Pavlicek, A., and Jurka, J. (2006). Positive selection on the nonhomologous end-joining factor Cernunnos-XLF in the human lineage. *Biol. Direct* **15**: doi:10.1186/1745-6150-1-15

- Pflugrath, J.W. (1999). The finer things in X-ray diffraction data collection. *Acta Crystallogr D Biol Crystallogr* **55**: 1718-1725.
- Pheiffer, B.H., and Zimmerman, S.B. (1982). 3'-phosphatase activity of the DNA kinase from rat liver. *Biochem Biophys Res Commun* **109**: 1297-1302.
- Pomeranz Krummel, D.A., Oubridge, C., Leung, A.K., Li, J., and Nagai, K. (2009). Crystal structure of human spliceosomal U1 snRNP at 5.5 Å resolution. *Nature* **458**: 475-480.
- Ramsden, D.A., and Gellert, M. (1995). Formation and resolution of double-strand break intermediates in V(D)J rearrangement. *Genes Dev* **9**: 2409-2420.
- Recuero-Checa, M.A., Doré, A.S., Arias-Palomo, E., Rivera-Calzada, A., Scheres, S.H., Maman, J.D., Pearl, L.H., and Llorca, O. (2009). Electron microscopy of Xrcc4 and the DNA ligase IV-Xrcc4 DNA repair complex. *DNA Repair Amst* **8**: 1380-1389.
- Resnick, M.A., and Martin, P. (1976). The repair of double-strand breaks in the nuclear DNA of *Saccharomyces cerevisiae* and its genetic control. *Mol Gen Genet* **143**: 119-129.
- Riballo, E., Critchlow, S.E., Teo, S.H., Doherty, A.J., Priestley, A., Broughton, B., Kysela, B., Beamish, H., Plowman, N., Arlett, C.F., Lehmann, A.R., Jackson, S.P., and Jeggo, P.A. (1999). Identification of a defect in DNA ligase IV in a radiosensitive leukaemia patient. *Curr Biol* **9**: 699-702.
- Riballo, E., Woodbine, L., Stiff, T., Walker, S.A., Goodarzi, A.A., and Jeggo, P.A. (2009). XLF-Cernunnos promotes DNA ligase IV-XRCC4 re-adenylation following ligation. *Nucleic Acids Res* **37**: 482-492.
- Rich, T., Allen, R.L., and Wyllie, A.H. (2000). Defying death after DNA damage. *Nature* **407**: 777-783.
- Richter, C., Park, J.W., and Ames, B.N. (1988). Normal oxidative damage to mitochondrial and nuclear DNA is extensive. *Proc Natl Acad Sci USA* **85**: 6465-6467.
- Ristic, D., Wyman, C., Paulusma, C., and Kanaar, R. (2001). The architecture of human Rad54-DNA complex provides evidence for protein translocation along DNA. *Proc Natl Acad Sc. USA* **98**: 8454-8460.
- Roberts, S.A., Strande, N., Burkhalter, M.D., Strom, C., Havener, J.M., Hasty P., and Ramsden D.A. (2010). Ku is a 5'-dRP/AP lyase that excises nucleotide damage near broken ends. *Nature* **464**: 1214-1217.
- Robins, P., and Lindahl, T. (1996). DNA ligase IV from HeLa cell nuclei. *J Biol Chem* **271**: 24257-24261.

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

Rooney, S., Alt, F.W., Lombard, D., Whitlow, S., Eckersdorff, M., Fleming, J., Fugmann, S., Ferguson, D.O., Schatz, D.G., and Sekiguchi, J. (2003). Defective DNA repair and increased genomic instability in Artemis-deficient murine cells. *J Exp Med* **197**: 553-565.

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

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

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

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

Sandler, S.J., Satin, L.H., Samra, H.S., and Clark, A.J. (1996). RecA-like genes from three archaean species with putative protein products similar to Rad51 and Dmc1 proteins of the yeast *Saccharomyces cerevisiae*. *Nucleic Acids Res* **24**: 2125-2132.

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

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

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

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

Shibata A., Barton O., Noon A.T., Dahm K., Deckbar D., Goodarzi A.A., Löbrich M., and Jeggo P.A. (2010). Role of ATM and the damage response mediator proteins 53BP1 and MDC1 in the maintenance of G(2)/M checkpoint arrest. *Mol Cell Biol* **30**: 3371-3383.

- Sibanda, B.L., Chirgadze, D.Y., and Blundell, T.L. (2010). Crystal structure of DNA-PKcs reveals a large open-ring cradle comprised of HEAT repeats. *Nature* **463**: 118-121.
- Sibanda, B.L., Critchlow, S.E., Begun, J., Pei, X.Y., Jackson, S.P., Blundell, T.L., and Pellegrini, L. (2001). Crystal structure of an Xrcc4-DNA LigaseIV complex. *Nat Struct Biol* **12**: 1015-1019.
- Siddiqi, M.A., and Bothe, E. (1987). Single- and double-strand break formation in DNA irradiated in aqueous solution: dependence on dose and radical scavenger concentration. *Radiat Res* **112**: 449-463.
- Singleton, B.K., Torres-Arzaus, M.I., Rottinghaus, S.T., Taccioli, G.E., and Jeggo, P.A. (1999). The C terminus of Ku80 activates the DNA-dependent protein kinase catalytic subunit. *Mol Cell Biol* **19**: 3267-3277.
- Smith, J., Riballo, E., Kysela, B., Baldeyron, C., Manolis, K., Masson, C., Lieber, M.R., Papadopoulo, D., and Jeggo, P. (2003). Impact of DNA ligase IV on the fidelity of end joining in human cells. *Nucleic Acids Res* **31**: 2157-2167.
- Spagnolo, L., Rivera-Calzada, A., Pearl, L.H., and Llorca, O. (2006). Three-dimensional structure of the human DNA-PKcs/Ku70/Ku80 complex assembled on DNA and its implications for DNA double-strand break repair. *Mol Cell* **22**: 511-519.
- Sulek, M., Yarrington, R., McGibbon, G., Boeke, J.D., and Junop, M. (2007). A critical role for the C-terminus of Nej1 protein in Lif1p association, DNA binding and non-homologous end-joining. *DNA Repair (Amst)* **6**: 1805-1818.
- Sun, H., Treco, D., Schultes, N.P., Szostak, J.W. (1989). Double-strand breaks at an initiation site for meiotic gene conversion. *Nature* **338**: 87-90.
- Suwa, A., Hirakata, M., Takeda, Y., Jesch, S., Mimori, T., and Hardin, J.A. (1994). DNA-dependent protein kinase and (Ku protein-p350 complex) assembles on double-stranded DNA. *Proc Natl Acad Sci USA* **91**: 6904-6908.
- Taccioli, G.E., Gottlieb, T.M., Blunt, T., Priestley, A., Demengeot, J., Mizuta, R., Lehmann, A.R., Alt, F.W., Jackson, S.P., and Jeggo, P.A. (1994). Ku80 is the product of the XRCC5 gene: a direct link with DNA repair and V(D)J recombination. *Science* **265**: 1442-1445.
- Teo, S.H., and Jackson, S.P. (2000). Lif1p targets the DNA ligase Lig4p to sites of DNA double-strand breaks. *Curr Biol* **10**: 165-168.
- Tomkinson, A.E., Totty, N.F., Ginsburg, M., and Lindahl, T. (1991). Location of the active site for enzyme-adenylate formation in DNA ligases. *Proc Natl Acad Sci USA* **88**: 400-404.

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

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

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

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

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

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

Ward, J.F., and Kuo, I. (1976). Strand breaks, base release, and postirradiation changes in DNA gamma-irradiated in dilute O₂-saturated aqueous solution. *Radiat Res* **66**: 485-498.

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

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

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

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

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

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

Wu, P.Y., Frit, P., Meesala, S., Dauvillier, S., Modesti, M., Andres, S.N., Huang, Y., Sekiguchi, J., Calsou, P., Salles, B., Junop, M.S. (2009). Structural and functional interaction between the human DNA repair proteins DNA ligase IV and XRCC4. *Mol Cell Biol* **29**: 3163-3172.

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

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

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

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

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

Yoo, S., and Dynan, W.S. (1999). Geometry of a complex formed by double strand break repair proteins at a single DNA end: recruitment of DNA-PKcs induces inward translocation of Ku protein. *Nucleic Acids Res* **27**: 4679-4686.

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

Yu, Y., Wang, W., Ding, Q., Ye, R., Chen, D., Merkle, D., Schriemer, D., Meek, K. and Lees-Miller, S. P. (2003) DNA-PK phosphorylation sites in XRCC4 are not required for survival after radiation or for V(D)J recombination. *DNA Repair* **2**: 1239–1252.

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

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

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

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

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

Appendix I

List of Publications by S.N. Andres

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

Cooper, C.A., Zhang, K., Andres, S.N., Fang, Y., Kaniuk, N.A., Hannemann, M., Brumell, J.H., Foster, L.J., Junop, M.S., and Coombes, B.K. (2010). Structural and biochemical characterization of SrcA, a multi-cargo type III secretion chaperone in *Salmonella* required for pathogenic association with a host. *PLoS Pathog* **5**: e1000751.

Wu, P.Y., Frit, P., Meesala, S., Dauvillier, S., Modesti, M., Andres, S.N., Huang, Y., Sekiguchi, J., Calsou, P., Salles, B., and Junop, M.S. (2009). Structural and functional interaction between the human DNA repair proteins DNA ligase IV and XRCC4. *Mol Cell Biol* **29**: 3163-3172.