STRUCTURE OF DNA LIGASE IV C-TERMINAL DOMAIN BOUND TO XRCC4

STRUCTURAL CHARACTERIZATION OF THE C-TERMINAL DOMAIN OF HUMAN DNA LIGASE IV BOUND TO XRCC4

By

SRILAKSHMI MEESALA, B.Sc., M.Sc.

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfilment of the Requirements

for the Degree

Master of Science

McMaster University

MASTER OF SCIENCE (2006) McMaster University

(Biochemistry and Biomedical Sciences) Hamilton, Ontario

TITLE:	Structural Characterization of the C-terminal Domain of Human DNA Ligase IV Bound to Xrcc4
AUTHOR:	Srilakshmi Meesala, M.Sc. (Hyderabad Central University)
SUPERVISOR:	Dr. Murray Junop
NUMBER OF PAGES:	xvii, 113

ABSTRACT

Non-homologous end joining (NHEJ) is the predominant mode of DNA double strand break (DSB) repair pathway in mammalian cells. At the heart of this repair pathway is Xrcc4-DNA ligase IV complex, which mediates ligation of the broken DNA strands. The C-terminal tandem BRCT repeats of human DNA ligase IV spanning residues 654-911 in complex with the functional fragment of Xrcc4 comprised of residues 1-203 were crystallized by the hanging drop vapour diffusion method at 20°C. Generation of single, well-packed, diffraction quality crystals suitable for structure determination involved usage of an Xrcc4 point mutant (A60E). Arriving at the crystallization condition included optimization of pH, variation of the precipitant concentration, investigation of the effects of small molecules, and alteration of the amount of crystal seed used as initial nuclei. A Crystal of selenomethionine-derived protein complex was grown using the above optimization steps and diffracted to 2.4 Å resolution. Data processing revealed that the crystal belonged to space group P1 with unit cell dimensions a = 67.33 b = 86.00 c = 111.52; α = 67.37 β = 83.00 γ = 74.56. The crystal structure of Xrcc4-DNA ligase IV complex was solved by single-wavelength anomalous diffraction using data collected at a wavelength of 0.9785Å corresponding to peak energy.

The structure maintains a 2:1 stoichiometry of Xrcc4 to the C-terminal domain of DNA ligase IV. The structure of the complex not only confirms the overall novel mode of interaction first observed in the 3.9 Å structure of the yeast ortholog lif1p-lig4p

complex, but it also discloses additional key features such as the DNA binding surface of the complex and the striking conformational changes occurring within Xrcc4 upon interaction with DNA ligase IV. Together, the structural information procured forms an important basis for a better understanding of the mechanism involved in the NHEJ repair pathway.

ACKNOWLEDGEMENTS

Firstly, I would like to thank my supervisor, Dr. Murray Junop, for providing me the opportunity to carry out research in his laboratory and be part of an exciting project. I am thankful for his valuable guidance and support in the past three academic years. His enthusiasm, inspiration and timely suggestions have been of tremendous help in the completion of this project. I am also appreciative for his patience whenever I asked questions concerning the X-ray crystallographic studies.

I would also like to thank my supervisory committee members, Dr. Alba Guarne and Dr. Daniel Yang for helpful discussions over the duration of these studies. In addition to a number of useful suggestions, Dr. Alba Guarne had also provided me with the server location information to check for merohedral crystal twinning. Dr. Daniel Yang provided important advice regarding space group determination of the new crystal form and assisted a number of times with issues associated with different software.

I would like to thank all the past and present members of Dr. Junop's laboratory for their assistance: Kun, Rachael, Andrew, Tracy, Margaret, Sean, Sara and Rishi. In particular, I would like to thank Kun Zhang for showing me how to purify Xrcc4-DNA ligase IV complex and grow crystals, Rachael Summerfield, for showing me how to mount crystals and collect data, Andrew Willems, for his advice at various stages and for demonstrating the usage of programs such as ClustalX 1.83 and Swiss-PDBViewer 3.7, Sean Jackson, for assisting me with the usage of WinCoot and PyMol 0.99 programs and Sara Andres, for proof-reading my thesis.

I would like to acknowledge the staff members of the National Synchrotron Light Source at beamline X8C at the Brookhaven National Laboratory for assistance with data collection.

I would like to express my final gratitude to my parents, my sister and my husband for being supportive and patient with me during my graduate studies.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

53BP1	p53 binding protein
ADSC	area detector system corporation
AMP	adenosine monophosphate
Amp	ampicillin
ATM	ataxia-telangiectasia mutated
ATP	adenosine triphosphate
B-NHEJ	back-up non-homologous end joining
BRCA1	breast cancer susceptibility gene1
BRCT	breast cancer carboxy-terminal domain
BSA	bovine serum albumin
CCD	charge-coupled device
CCP4	collaborative computational project number 4
Cm	chloramphenicol
CNS	crystallography and NMR systems
DNA	deoxyribonucleic acid
DNA-PK	DNA dependent protein kinase
DNA-PKcs	DNA dependent protein kinase catalytic subunit
D-NHEJ	DNA-PK dependent non-homologous end joining
Dnl4	Yeast DNA ligase 4
dNTP	2-deoxynucleotide-5'-triphosphate

DSB	double strand break
dsDNA	double stranded DNA
EDTA	ethylene diamine tetra acetic acid
FAT	FRAP, ATM, and TRRAP
FATC	C-terminal FRAP, ATM, and TRRAP
Fc	calculated structure factors
Fo	observed structure factors
HRR	homologous recombination repair
IPTG	isopropyl-thio- β-D-galactopyranoside
kDa	kilo Daltons
LB	Luria-Bertani broth
LDAO	N-lauryl-N, N-dimethylamine N-oxide
MAD	multiwavelength anomalous dispersion
MDC1	mediator of DNA damage checkpoint protein 1
MR	molecular replacement
MRE11	meiotic recombination 11
NCS	non-crystallographic symmetry
NDSB-195	non-detergent sulphobetaine 195
NHEJ	non-homologous end joining
NSLS	national synchrotron light source
PCR	polymerase chain reaction
PEG	polyethylene glycol

PI3KK	phosphatidylinositol 3-OH-kinase-related kinase
PNK	polynucleotide kinase
Pol x	polymerase x
Pol β	polymerase β
Pol λ	polymerase λ
Pol μ	polymerase μ
psi	pound-force per square inch
RAD50	radiation sensitive
rmsd	root mean square deviation
RP-A	replication protein A
SAD	single wavelength anomalous diffraction
SDS-PAGE	sodium dodecyl sulphate-polyacrylamide gel electrophoresis
siRNA	small interfering ribonucleic acid
TdT	terminal deoxy-nucleotidyl transferase
TEV	tobacco etch virus
V(D)J	variable diverse joining
Vm	Matthews coefficient
XLF	xrcc4 like factor
Xrcc4	X-ray cross complementation group 4
Xrs2	X-ray sensitive
Yku 70/Yku 80	yeast ku70/ku80

CHAPTER 1

INTRODUCTION

1.1 DNA Damage, Double Strand Breaks and Cellular Response

DNA contains crucial genetic information that forms the basis of life. Since a variety of DNA damaging agents pose a significant threat to genomic integrity by incurring a myriad of types of damage, guarding the genome is a key function of the cell. This guardianship is facilitated by rapid and efficient detection of DNA damage followed by a series of signal transduction pathways, which lead to specific repair mechanisms. The nature and severity of DNA damage plays an important role in cellular response to the damage. The multifaceted response evolved in eukaryotes to counteract the detrimental effects of DNA damage includes sensing the damage, activation of cell cycle checkpoints to arrest cell cycle progression, induction of transcriptional programs and enhancement of DNA repair pathways to repair and correct the damage (Fig. 1.1). However, when the extent of DNA damage is severe, the cell undergoes apoptosis thereby eliminating itself from a population that otherwise might suffer serious pathological consequences (Fig. 1.1). All these processes are meticulously coordinated to faithfully maintain, duplicate and segregate genetic material.

Endogenous DNA damage occurs at a high frequency. It has been estimated that base loss owing to spontaneous hydrolysis of DNA glycosyl bonds is in the order of 10^4 events per day for a mammalian cell (Dinner *et al.*, 2001). It is even more shocking to consider the effects of external damaging agents such as ionizing radiation. It is reported

that every cell exposed to 1 Gray of ionizing radiation will sustain 1000–2000 damaged bases, 800–1000 damaged sugars, 1000 single-strand breaks, approximately 40 double strand breaks (DSBs), 30 DNA–DNA cross-links and 150 DNA–protein cross-links (Kao *et al.,* 2005). Although all kinds of DNA lesions are deleterious to a cell, DSBs in DNA represent the most genotoxic lesions. While DNA DSBs can be generated by external sources such as ionizing radiation and radiomimetic chemicals, certain products released during normal cellular metabolic events such as reactive oxygen species can also trigger the formation of DSBs. Furthermore, DSBs are even observed when a DNA replication fork encounters a single stranded lesion. In addition to this, DSBs are also found in the context of normal physiology as intermediates in certain recombination pathways such as V(D)J recombination, immunoglobulin class-switching recombination, meiotic recombination and yeast mating-type switching. While V(D)J recombination is responsible for the production of a large repertoire of antigen receptors with different specificities, class-switching recombination facilitates conversion between the different isotypes of immunoglobulins. Together, both these processes are essential to the normal functioning of the immune system.

Given that DNA DSBs can arise by numerous routes, it becomes vital to analyze their consequences. Un-repaired or incorrectly repaired DSBs can cause cell death, chromosomal fragmentation, gross chromosomal rearrangements and chromosomal instability resulting in the activation of oncogenes and inactivation of tumour-suppressor genes, which in turn can lead to tumour formation.

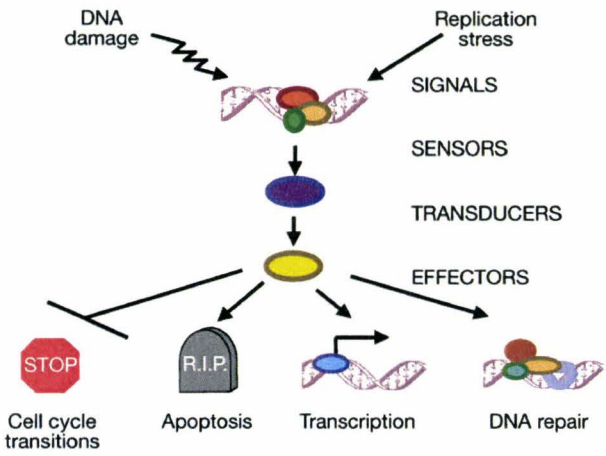

Figure 1.1 Cellular DNA damage response. Arrowheads represent activating events and perpendicular ends represent inhibitory events. Cell-cycle arrest is depicted with a stop sign, apoptosis with a tombstone. The DNA helix with an arrow represents damage-induced transcription, while the DNA helix with several oval-shaped subunits represents damage-induced repair. For the purpose of simplicity, the network of interacting pathways are depicted as a linear pathway consisting of signals, sensors, transducers and effectors. **Picture is adapted from Zhou *et al.*, 2000.**

To overcome the catastrophic effects of DNA DSBs, higher eukaryotes have evolved two major pathways for their repair namely the homologous recombination repair pathway (HRR) and non-homologous end-joining (NHEJ) repair pathway.

1.2 Various Types of Double Strand Break Repair

HRR utilizes chromosomal diploidy to restore the continuity of a broken DNA molecule and is based on sequence homology between the damaged DNA and an intact homologous sister chromatid, which is used as a template (Pâques *et al.,* 1999). Since the presence of a homologous sister chromatid is restricted to late S and G2 phases of cell cycle, HRR is confined to these phases. HRR is predominant in lower eukaryotes such as yeast, since they enable easy access and detection of the homologous chromosome due to the limited complexity of the genome in these organisms.

Although HRR is the major form of homology-directed repair and involves a conservative repair process, there also exists another minor non-conservative type of homology-directed repair known as single-strand annealing, in which DSBs are processed to single-stranded tails, and repair occurs through the annealing of a single-stranded tail at a nearby direct repeat sequence (Lin *et al.,* 1990).

In contrast to the above-mentioned repair pathways, NHEJ repair pathway, as the name suggests, does not require any sequence homology but joins the broken DNA ends end-to-end. Although the repair of DSBs by NHEJ occurs in organisms ranging from bacteria to mammals, it is the predominant contributor in higher eukaryotes such as mammals owing to the complexity of the genome, which hinders easy access to the homologous chromosome thus giving NHEJ precedence over HRR. This is also reflected by the fact that, despite the complexity, most of the major components involved in NHEJ have been initially identified in the mammalian system. Unlike HRR, which carries out accurate repair, NHEJ is imprecise and often results in the loss of a few nucleotides

providing a means for genetic evolution. NHEJ occurs throughout the cell cycle although it prevails during G0, G1 and early S phases. NHEJ is also the repair pathway chosen to carry out the ligation involved in V(D)J recombination responsible for generating immunoglobulin diversity and class switch recombination, which is employed to boost the immune response.

Recent studies have indicated the occurrence of two kinds of NHEJ pathways based on their prevalence, speed of repair and the proteins involved. D-NHEJ is the predominant faster pathway that depends on DNA-dependent protein kinase (DNA-PK) and XRCC4/DNA ligase IV, while B-NHEJ is a slower back-up pathway independent of DNA-PK but requiring DNA ligase III (Wang *et al.,* 2005). Since D-NHEJ is the most prevalent pathway, only that will be considered for further discussion and shall be referred to as NHEJ. Fig. 1.2 summarizes the steps involved in this repair pathway.

The key players involved in the NHEJ repair pathway have been identified by a series of elegant genetic and biochemical experiments. Current studies reveal that Ku70/Ku80 heterodimer, an abundantly available DNA end-binding factor in mammalian cells, is the first DNA repair protein involved in binding to broken DNA ends and thereby protecting the ends from being degraded by nucleolytic activity within the cell. A few reports have also demonstrated the role of Ku in bridging broken DNA ends ((Bliss *et al.,* 1997) (Cary *et al.,* 1997) (Ramsden *et al.,* 1998)). Furthermore, several predictions suggest that the Ku70/Ku80 heterodimer is the primary sensor in detecting DSBs *in vivo*. Structural analysis of this protein revealed a C-terminal asymmetric ring-like heterodimer formed by the interaction of the two monomers as shown in Figure 1.3 (Walker *et al.,*

2001). The DNA bound structure displays the duplex DNA passing through this ring, which forms a stable complex upon interaction with the positively charged amino acids lining the interior surface (Walker *et al.,* 2001) (Fig. 1.3). Both the subunits of the heterodimer Ku70/Ku80 fold into similar structures and are comprised of an N-terminal α/β domain and a β-barrel domain besides the already described C-terminal ring domain (Walker *et al.,* 2001).

It has been proposed that while the fixed size of the ring imposes a constraint on the nature of DNA, restricting the protein to exclusively bind dsDNA, the positively charged inner surface of the protein paves the way for generic sequence-independent interaction with the sugar-phosphate backbone of DNA, thus enabling the Ku70/Ku80 heterodimer to recognize and bind to all kinds of DNA DSBs. Binding of Ku70/Ku80 heterodimer to DNA ends increases the affinity of the other NHEJ components to DNA. Subsequent to the binding to DNA ends, the Ku:DNA complex at each end recruits DNA-PKcs to the site of damage (Hammarsten *et al.,* 1998), which brings about alignment and further bridging of the DNA ends and causes the Ku heterodimer to translocate inwards along the dsDNA (Yoo *et al.,* 1999) (Fig. 1.4).

Ku70/Ku80 heterodimer, the regulatory subunit of DNA-dependent protein kinase (DNA-PK), along with the massive catalytic subunit, DNA-PKcs of ~ 465 kDa constitute the DNA-PK complex, which functions as a serine-threonine kinase and belongs to the family of phosphatidylinositol-3-OH kinase-related kinases. Although several *in vitro* and *in vivo* studies have identified DNA-PK phosphorylation sites on most of the NHEJ components, only autophosphorylation of DNA-PK has been conclusively demonstrated

to specifically influence NHEJ by inducing conformational changes within the protein ((Chan *et al.,* 1996) (Douglas *et al.,* 2001) (Chan *et al.,* 2002) (Merkle *et al.,* 2002) (Ding *et al.,* 2003) (Block *et al.,* 2004) (Reddy *et al.,* 2004)). Among the varied phosphorylation targets of DNA-PK are also DNA damage response regulators such as p53 and chromatin structure stabilizers like Histone H1 ((Woo *et al.,* 1998) (Kysela *et al.,* 2005)). The massive DNA-PKcs molecule is composed of a C-terminal region containing conserved domains named as FAT, FATC and a motif homologous to the kinase domain of phosphatidyl inositol 3-OH kinases along with an N-terminal region that has been postulated to consist of helical repeats with a low degree of conservation ((Perry *et al.,* 2003) (Brewerton *et al.,* 2004)). The FAT motif has been named after its presence in three related proteins called FRAP, ATM and TRRAP. Since the second FAT motif is at the C-terminus of DNA-PK, it has been termed, FATC.

Electron microscopic studies of DNA-PK have revealed 3 distinct domains - palm, arm and head, which are structurally shared by a related kinase called ATM ((Boskovic *et al.,* 2003) (Rivera-Calzada *et al.,* 2005)). Furthermore, DNA-PK also shares the 50kDa C-terminal catalytic PI3KK motif within the head domain with related kinases ((Boskovic *et al.,* 2003) (Rivera-Calzada *et al.,* 2005)). While the remaining portion of the head domain is involved in interacting with Ku, the arm domain is thought to have a regulatory role while a section of the palm domain comprised of a leucine-rich region is responsible for interaction with DNA (Rivera-Calzada *et al.,* 2005). Combinatorial efforts of structural and modelling studies indicate that DNA-PKcs shows significant conformational changes upon binding to DNA, causing the palm domain to

reach out to the head domain while the arm moves inwards, thus creating a channel for accommodating dsDNA (Boskovic *et al.,* 2003) (Fig. 1.4). There is some interesting speculation that the activation of the kinase activity of DNA-PKcs in the presence of DNA is the result of these DNA-induced conformational changes. Emerging evidence suggests that DNA-PKcs might function as a scaffold assembling other components of NHEJ required to complete the end-joining process.

Artemis, another factor of NHEJ possesses a 5'-3' exonuclease activity on its own but acquires an endonuclease activity along with DNA-PKcs that is capable of cleaving a variety of DNA structures such as overhangs, hairpins, flaps, gaps and loops (Jeggo *et al.,* 2002). Artemis is thought to participate in processing the DNA ends during NHEJ and specifically be involved in opening the hairpin structures of DNA generated during V(D)J recombination.

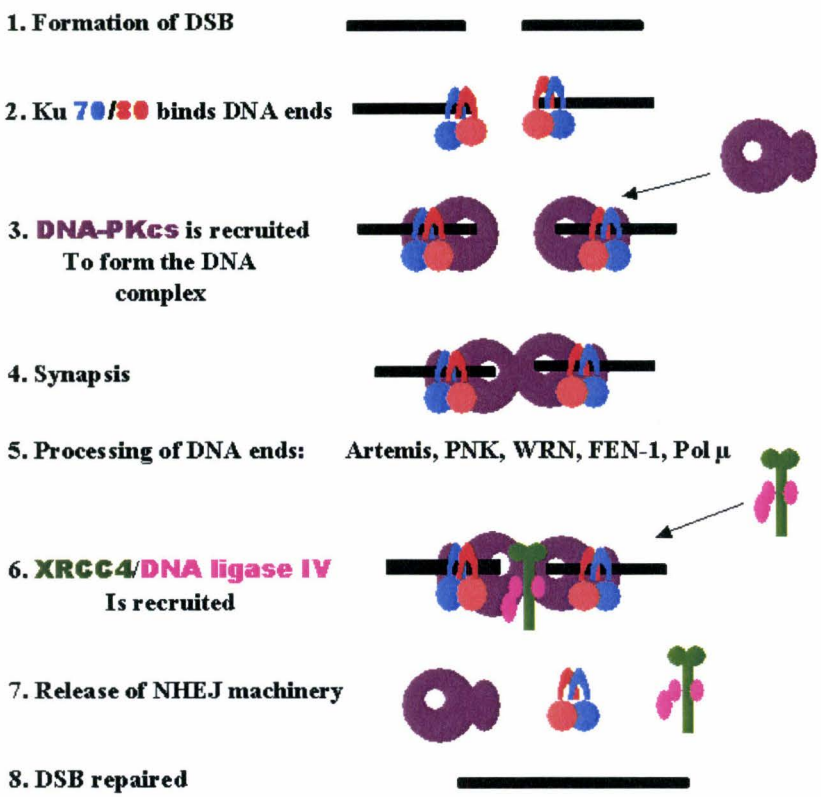

Figure 1.2 Mechanism of mammalian non-homologous end joining repair pathway depicting the steps involved and the key proteins of the pathway. The text on the left of the diagram indicates the particular step occurring at that point. Double stranded DNA is shown in black. Each protein in the diagram is coloured differently and the corresponding name of the protein is given in the same colour in the text for convenience. Arrow arising from a protein indicates recruitment of that protein during the step. **Picture is modified and adapted from Lees-Miller *et al.*, 2003.**

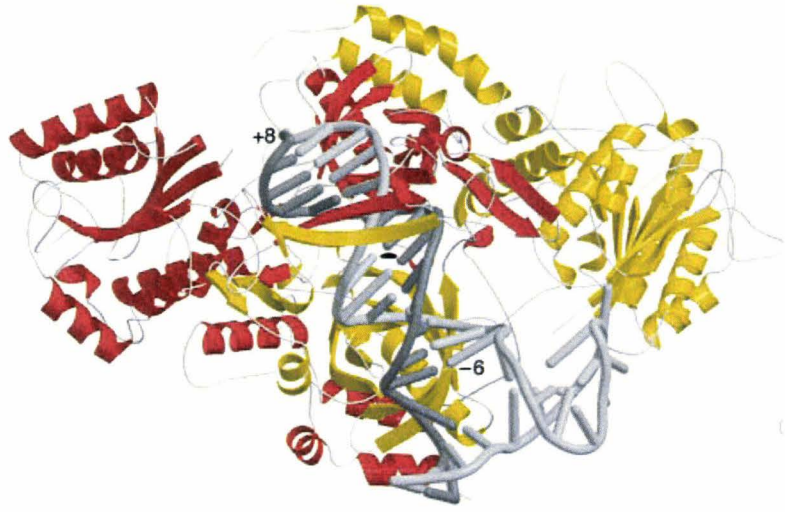

Figure 1.3 Ribbon diagram of Ku bound to DNA, viewed down the molecular dyad axis. The axis is shown as a black symbol. Ku70 is coloured red and Ku80 orange. The 34-residue oligonucleotide is light grey; the 21-residue oligonucleotide dark grey. The terminal base pairs of the central duplex are numbered +8 (broken DNA end) and –6. **Picture is adapted from Walker *et al.*, 2001.**

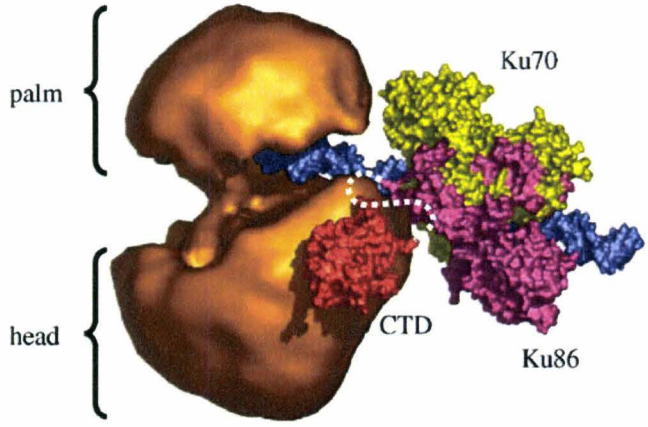

Figure 1.4 Hypothetical model for the DNA-PK holoenzyme assembled at a DNA end. The model combines data from the single-particle EM structure for the DNA-PKcs–DNA complex (Boskovic *et al.*, 2003) (coloured orange), the crystal structure of the Ku86:70–DNA complex (Walker *et al.*, 2001) (coloured yellow and pink), and the NMR structure of the Ku86-CTD (Harris *et al.*, 2004) (labelled as CTD and coloured red), which mediates assembly of the holoenzyme. DNA in the complex has been coloured blue. White dashed line represents the linker between the crystallized Ku86:70 and the C-terminal domain of Ku86. **Picture is adapted from Llorca *et al.*, 2004.**

Polynucleotide kinase (PNK) is also involved in NHEJ. It possesses both kinase and phosphatase functionalities and converts chemically modified DNA ends generated during radiation damage into ligatable termini (Chappell *et al.,* 2002). Several polymerases belonging to the POLX family such as polμ, polλ and polβ are also thought to be important for NHEJ, as is their counterpart TdT for V(D)J recombination (Daley *et al.,* 2005). Although the precise roles of these polymerases and nucleases in the DNA end-processing is not completely clear, it has been suggested that nucleases remove several nucleotides from single-stranded overhangs at the DSB termini, especially when a region of microhomology occurs, and DNA polymerases fill in gaps of several nucleotides. At the heart of NHEJ repair pathway resides Xrcc4-DNA ligase IV complex that is responsible for carrying out the central function of ligating the opposing broken DNA ends. Since this complex is the focus of my research, it will be described in detail later.

According to the current iterative processing model for NHEJ, the Ku:DNA complex does not follow a particular order to recruit the nuclease, polymerase and ligase (Ma *et al.,* 2005). Moreover, DNA ends not requiring one or more of the intermediate components proceed to the next or final ligation step without the involvement of other enzymes (Ma *et al.,* 2005). For example, DNA ends with ligatable termini and without any gaps do not require a nuclease and polymerase but can be directly ligated using a ligase. Moreover, the current model also implies that ligation of one strand of a DSB occurs independently of the other strand (Ma *et al.,* 2005). This model involves the processing, filling and ligation of one strand followed by processing, filling and ligation

of the other strand. These steps could be repeated a number of times depending on the DNA ends.

Adding to the growing list of NHEJ components, is a novel factor called Cernunnos/XLF ((Ahnesorg *et al.,* 2006) (Buck *et al.,* 2006)), recently identified by two separate groups through genetic and biochemical studies. The name of this protein XLF comes from its predicted structural similarity to Xrcc4 based on a set of complex database searches using advanced computer algorithms (Ahnesorg *et al.,* 2006). It was noted that inhibition of Cernunnos/XLF activity by the use of small interfering RNA (siRNA) in cells caused increased radiosensitivity and defective DNA DSB repair (Ahnesorg *et al.,* 2006). Given that Cernunnos/XLF has been shown to interact with Xrcc4 as well as with DNA ligase IV both *in vitro* and *in vivo* ((Ahnesorg *et al.,* 2006) (Callebaut *et al.,* 2006)), several interesting hypotheses predicting their mode of interaction and their functionality in NHEJ have been put forward. Latest developments have also shown that Cernunnos/XLF is the human homologue of the yeast NHEJ factor, NEJ1 (Callebaut *et al.,* 2006). Although NEJ1 has been shown to physically interact with lif1, the specific role of this protein in NHEJ has not been conclusively determined (Kegel *et al.,* 2001).

Most of the functional homologs of mammalian NHEJ repair pathway are even observed in yeast, suggesting that the overall mechanism is conserved in eukaryotes, at least to a certain extent. The yeast Yku70/Yku80 and lig4p/lif1p are the yeast equivalents of Ku70/Ku80 and DNA ligase IV/Xrcc4 respectively. Although yeast lacks a homologue for DNA-PKcs, there is strong evidence suggesting that the

Rad50/Mre11/Xrs2 complex functions as the end bridging factor in yeast NHEJ and also has a role in specifically stimulating intermolecular ligation by lig4p/lif1p complex ((Chen *et al.,* 2001) (Herrmann *et al.,* 1998)). Given that the Mre11 subunit of this complex has nuclease activity and may facilitate end-alignment (Paull *et al.,* 2000) and end processing, the Rad50/Mre11/Xrs2 complex can be considered to be functionally similar to mammalian DNA-PKcs/Artemis complex. Bioinformatics studies have also led to the identification of simpler NHEJ counterparts in bacteria (Aravind, L., 2001) (Doherty, A.J., 2001), such as the bacterial Ku homodimer and the ligase protein. Ligase, primase and nuclease domains together constitute the bacterial ligase protein, which unlike the homologous eukaryotic ligase lacks a C-terminal domain comprised of the tandem BRCT motifs.

1.3 Xrcc4-DNA Ligase IV Complex

Xrcc4 and DNA ligase IV are two indispensable components of the NHEJ repair pathway and cells defective in either of these proteins demonstrate radiosensitivity and impaired V(D)J recombination ((Li *et al.,* 1995) (Gao *et al.,* 1998)). Their significance is further highlighted by the fact that absence of any one of these crucial proteins in mice causes neuronal apoptosis and leads to embryonic lethality ((Barnes *et al.,* 1998) (Frank *et al.,* 2000) (Gao *et al.,* 2000)). On its own, Xrcc4 maintains a tetramer-dimer equilibrium, but upon association with DNA ligase IV, it forms a tight and stable 2:1 complex that is resistant to challenge with high salt concentrations ((Junop *et al.,* 2000) (Sibanda *et al.,* 2001)).

Although full-length human Xrcc4 is a 336 amino acid phosphoprotein, the N-terminal 200 residues (Fig. 1.5A) are sufficient to restore V(D)J recombination while the N-terminal 250 residues confer resistance to ionizing radiation in Xrcc4-defective XR-1 cells ((Critchlow *et al.,* 1997) (Mizuta *et al.,* 1997)). The crystal structure of the N-terminal 203 residues of Xrcc4 reveal that each protomer is comprised of a globular N-terminal head domain and an extended α-helical C-terminal stalk (Junop *et al.,* 2000) (Fig. 1.6A). The head domain is further comprised of a ß-sandwich, made up of two sets of antiparallel strands forming two ß-sheets and a helix-turn-helix motif (Junop *et al.,* 2000). Moreover, the crystal structure also depicts the formation of a tetramer through interactions mediated between the coiled coil regions of two individual dimers of Xrcc4 (Junop *et al.,* 2000) (Fig. 1.6B). Biochemical studies have demonstrated that this

tetramerization region of Xrcc4 dimer comprised of residues 165-201 extensively overlaps with the DNA ligase IV binding region, thus making each of the processes structurally independent of each other (Modesti *et al.*, 2003).

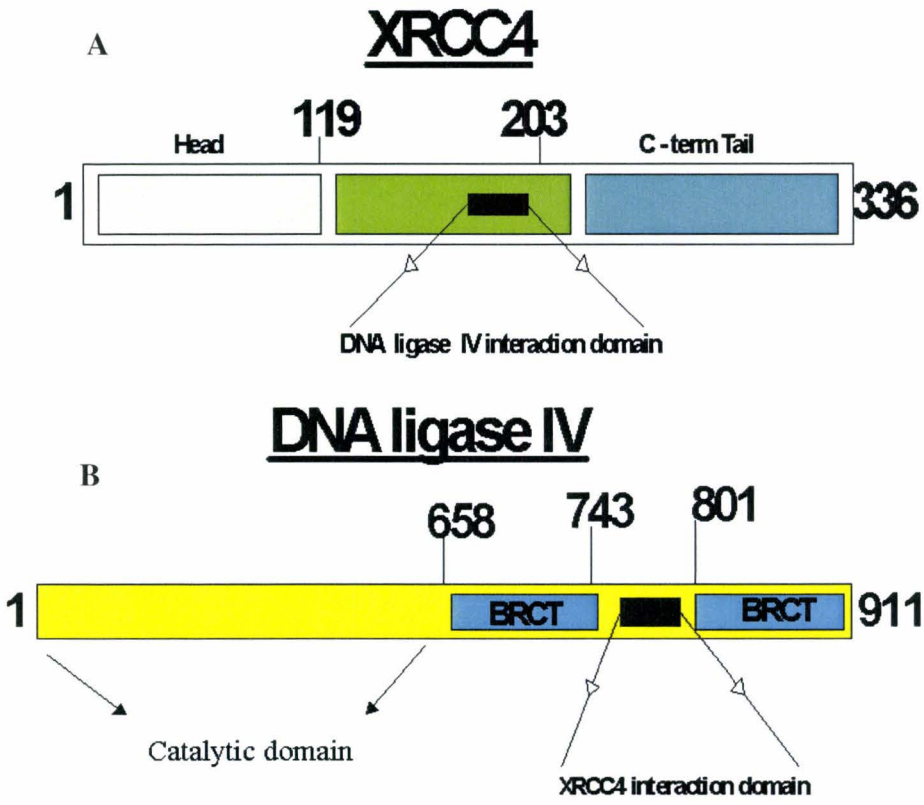

Figure 1.5 Domain boundaries of Xrcc4 and DNA ligase IV. (A) The three major domains of the 336 amino acid protein, Xrcc4 are exhibited in different colours. DNA ligase IV interaction domain is coloured in black. (B) Protein domains of DNA ligase IV. The catalytic domain resides at the N-terminal end. The two BRCT domains at the C-terminal end of the protein are coloured in cyan and they flank the Xrcc4 interaction domain coloured in black. **Picture is adapted from Modesti *et al.*, 2003.**

Figure 1.6 Crystal structure of Xrcc4 in different oligomeric states. (A) Ribbon diagram of Xrcc4 homodimer. One of the protomers is shown in light green while the other is in darker green. The N- and C-termini of an Xrcc4 protomer are indicated. (B) Ribbon diagram of Xrcc4 tetramer. The two dimers are coloured in green and red and each subunit of a dimer is differentiated with a lighter and a darker colour. **Picture is adapted from Junop *et al.*, 2000.**

DNA ligase IV is a 104kDa ATP-dependent ligase (Critchlow *et al.*, 1997) (Fig. 1.5B). The N-terminal catalytic domain of DNA ligase IV shares homology with DNA ligases I and III, while the C-terminal tandem repeat of BRCT domains are solely restricted to DNA ligase IV reflecting its unique ability to catalyze ligation during NHEJ which cannot be replaced by other mammalian ligases (Tomkinson *et al.*, 1997).

To date, several hypomorphic DNA ligase IV mutations causing LIG4 syndrome have been identified (Table. 1.1). Moreover, there are no cases cited of patients with DNA ligase IV null mutations suggesting that such a mutant might be non-viable in

humans. Hypomorphic mutations are mutations that cause reduced function of the gene product, or a negative change in expression of the gene. The key clinical features of DNA ligase IV functionally deficient patients mostly include microcephaly, radiosensitivity, developmental delay, cytopenia, predisposition to lymphoma and severe immunodeficiency in some cases (Enders *et al.,* 2006). The mutations R278H, Q280R and H282L lie close to the active site K273 of the conserved ligase motif and therefore lead to 5-10% residual adenylation and dsDNA ligation activity ((Badie *et al.,* 1997) (Riballo *et al.,* 1999) (Riballo *et al.,* 2001) (Wang *et al.,* 2001) (Enders *et al.,* 2006) (van der Burg *et al.,* 2006)). The R580X mutation is similar to the frameshift D423 that causes a premature termination after residue 442 giving rise to a protein that lacks both the nuclear localization signal as well as the Xrcc4 interaction domain, thereby preventing it from efficiently entering the nucleus (Girard *et al.,* 2004). Finally, the mutation R814X is missing only a part of the tandem BRCT domain but causes a phenotype characteristic of LIG4 syndrome suggesting that the presence of an intact second BRCT domain is important for activity. However, the immunodeficiency observed in these patients is mild or totally absent (Table 1.1), suggesting the presence of an almost normal V(D)J end-joining. This could possibly indicate that an intact second BRCT domain is crucial for proper functioning of other aspects leading to NHEJ, such as one of the DNA damage signalling pathways.

Table 1.1 Key clinical features of DNA ligase IV deficient patients reported to date[1].

	180BR	3703	411BR	2303	2304	99PO149	P1	P2	P-1	P-2
Age at diagnosis	14	4	9	46	48	9	2	4 wk	1.5	4 wk
Mutations	R278H	R814X	R278H	R580X	R469E	H282L	H282L	Q280R	Q280R	Q280R
	homoz.	homoz.	homoz.	R814X	R814X	R814X	K424fs	K424fs	K424fs	K424fs
Microcephaly	?	+	+	+	+	+	+	+	+	+
Radiosensitivity	+	+	+	?	?	+	-	+	-	?
Developmental Delay	-	+	+	+	?	+	+	-	-	-
Cytopenia	?	+	+	+	?	+	+	-	-	-
Leukemia	+	+	-	-	-	-	-	-	-	-
Immunodeficiency	-	-	+	+	+	+	+++	+++	+++	+++

[1]The code at the top of each column identifies a particular patient. The abbreviation "homoz." indicates homozygosity for the given mutation. The characteristics of R814X mutation are boxed to show the mild nature or lack of immunodeficiency observed in these patients. **The table is adapted from Enders *et al.*, 2006.**

It is postulated that Xrcc4 modulates the activity of DNA ligase IV through its association facilitated by the region between the two BRCT domains ((Grawunder *et al.*, 1998) (Sibanda *et al.*, 2001)). This strong interaction both stabilizes and stimulates the activity of DNA ligase IV ((Grawunder *et al.*, 1997) (Modesti *et al.*, 1999).

Early DNA binding studies have not only proven that Xrcc4 binds preferentially and cooperatively to linear or nicked circular DNA, but also that the functional fragment of Xrcc4 (1-200) is sufficient for binding and that deletion of residues 167-200 abolishes DNA binding (Modesti *et al.*, 1999). Furthermore, it has also been demonstrated that a supershift in the DNA binding gel occurs upon interaction of Xrcc4 with the tandem BRCT domains of DNA ligase IV, even though the domains themselves do not exhibit any DNA binding activity on their own (Modesti *et al.*, 1999). It is also important to note that upon association with the tandem BRCT fragment of DNA ligase IV, Xrcc4 displays an increased affinity for DNA (Dr. Mauro Modesti, personal communication).

1.4 Structural Analysis of Other Tandem BRCT Domains and the Previous Structure of Xrcc4 Bound to a Peptide of DNA Ligase IV

Although the structures of classical tandem BRCT repeats, characterized by short intervening linkers have been determined in proteins such as BRCA1, 53BP1 and MDC1 ((Williams *et al.,* 2001) (Derbyshire *et al.,* 2002) (Lee *et al.,* 2005)), the BRCT domains of DNA ligase IV are unique in possessing a somewhat more extended linker region. Tandem BRCT repeats with short linkers demonstrate a close head-to-tail interaction ((Williams *et al.,* 2001) (Derbyshire *et al.,* 20002) (Lee *et al.,* 2005)) and have been shown to bind phosphopeptides through the cleft generated by the interaction between the two domains. For instance, the crystal structure of BRCA1 tandem BRCT domains reveals an elongated structure with each BRCT fold comprised of a central, parallel four-stranded β-sheet and three α-helices, two of which pack against one face of the sheet while the third helix packs against the opposite face of the sheet (Williams *et al.,* 2001) (Fig. 1.7). As shown in Fig. 1.7, the BACH1 phosphopeptide bound structure of BRCA1 shows that a deep surface cleft is created by close associations between the two BRCT repeats facilitated through the extensive hydrophobic interface (Shiozaki *et al.,* 2004). The BACH1 phosphopeptide is shown to bind to this cleft, interacting with residues from both BRCT domains (Shiozaki *et al.,* 2004). Given that human DNA ligase IV associates with Xrcc4, it is intriguing to note that DNA ligase IV has also been shown to interact with phosphopeptides, exhibiting a distinct preference for phosphoserine containing phosphopeptides (Rodriguez *et al.,* 2003).

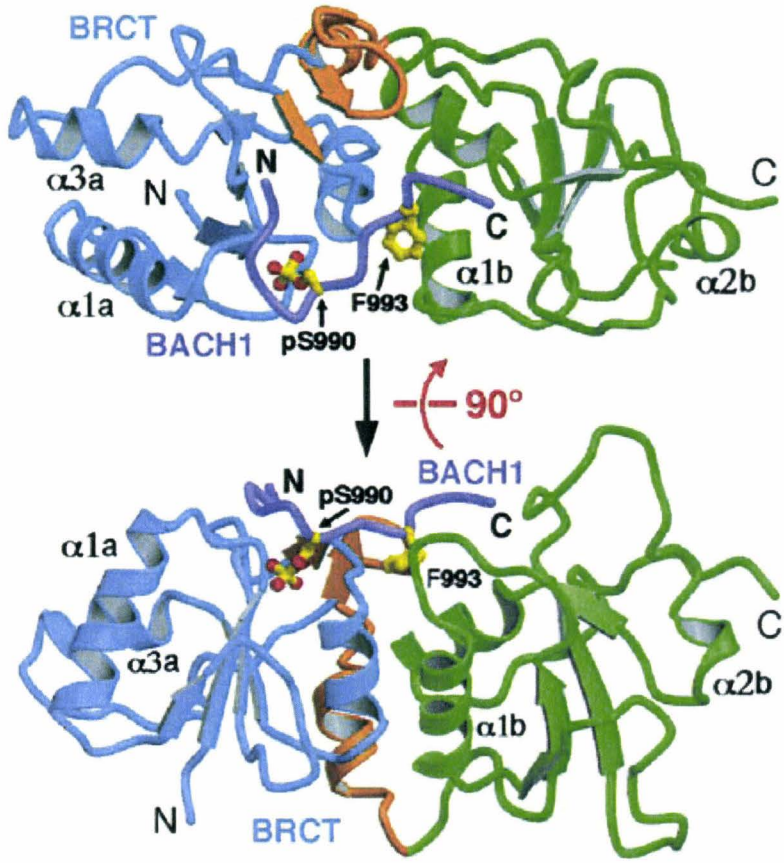

Figure 1.7 Overall structure of the BRCT repeats of human BRCA1 bound to a phosphorylated BACH1 peptide. Schematic representation of the BRCT-BACH1 complex. The two tandem BRCT repeats are coloured blue and green, respectively. The structural elements between the two BRCT repeats are shown in orange. The bound BACH1 peptide is coloured light purple. Phe 993 and the phosphorylated Ser 990 of the BACH1 peptide are highlighted. **Picture is adapted from Shiozaki *et al.*, 2004.**

The Xrcc4 interaction region of DNA ligase IV has been predicted to span residues 748-784 based on deletion analysis (Sibanda *et al.,* 2001). Furthermore, it has also been demonstrated that residues 180-213 of Xrcc4 are important for interaction with DNA ligase IV (Sibanda *et al.,* 2001). Based on this data, the structure of Xrcc4 bound to a polypeptide of DNA ligase IV corresponding to residues 748-784 was solved by X-ray crystallography. This structure shows that a single ligase chain interacts asymmetrically with the same region of both protomers of an Xrcc4 dimer (Sibanda *et al.,* 2001) (Fig. 1.8). Since this structure involves only a small peptide of DNA ligase IV of less than 40 residues, the structural and thereby the functional information gained from it is very limited. To further enhance the understanding of the role of this complex in NHEJ repair pathway, it is essential to gain additional structural knowledge using a larger fragment of DNA ligase IV.

In this research, the complex generated by the association of the tandem BRCT domains of DNA ligase IV with the N-terminal 203 residues of Xrcc4 is the subject for 3D structural analysis by X-ray crystallography, and the valuable information obtained from this structure will facilitate better understanding of NHEJ. Specifically, the structure will assist in comprehending the detrimental mutations arising in the tandem BRCT domains of DNA ligase IV leading to LIG4 syndrome. In addition to this, the structural information obtained might be of assistance in generating potential inhibitors of the full-length complex for blocking NHEJ repair and thereby radiosensitising cancerous zones, to make radiotherapy more effective.

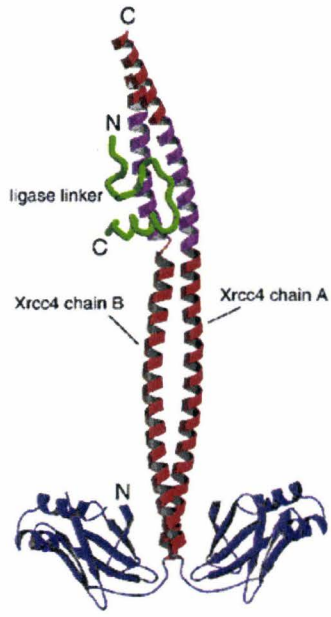

Figure 1.8 Crystal structure of Xrcc4 bound to a peptide of DNA ligase IV. The head domains are blue; the helical tails red and the DNA ligase IV binding region magenta. N- and C-termini are indicated for Xrcc4 protomer A and for the ligase chain. **Picture is adapted from Sibanda *et al.*, 2001.**

1.5 Thesis Objectives

The main goal of this investigation is to improve understanding of the several roles played by of Xrcc4-DNA ligase IV complex in NHEJ. Given that Xrcc4 forms a highly stable complex with DNA ligase IV and that the formation of the complex greatly stimulates the activity of DNA ligase IV, it becomes crucial to analyze the interactions that persist between the two proteins. Structural characterization of the complex would not only enable the identification of the underlying protein-protein interactions but also provide the information needed to detect any conformational changes taking place within the proteins upon the formation of the complex. Analysis of these structural changes could provide clues to comprehend the basis for enhanced ligation activity observed in DNA ligase IV within the complex.

The objective of this research is to determine the structure of the tandem BRCT domains of DNA ligase IV bound to the functional fragment of its stabilizing partner, Xrcc4 by X-ray crystallographic technique. The structural information obtained from this study will be used for further understanding the NHEJ repair mechanism and gaining significant functional insights. The structure will also aid in comprehending the mutations observed within the tandem BRCT domains of DNA ligase IV that are responsible for causing LIG4 syndrome characterized by immunodeficiency, microcephaly and developmental delay. Furthermore, the structural information procured would provide a basis for identifying the DNA binding surface of the complex. In addition this, the structure might also provide hints to characterize the functionalities of the tandem BRCT domains of DNA ligase IV.

A long-term benefit of this study might include the employment of the structural knowledge gained for assisting in the design of inhibitors to the full-length complex. Since most cancer therapies act primarily by causing extensive DNA damage including DNA DSBs, these inhibitors, which possibly function by preventing the association of Xrcc4 and DNA ligase IV may prove to be lead compounds for the development of novel anti-cancer therapeutic agents to be used in conjunction with radiation therapy or chemotherapy. The lack of a functionally active NHEJ repair pathway makes the cells more susceptible to treatment and forces the cells to undergo apoptosis.

CHAPTER 2

PURIFICATION AND CRYSTALLIZATION OF XRCC4-DNA LIGASE IV COMPLEX

2.1 Materials

The *E.coli* expression plasmids pWY1119 and pWY1190 along with the others listed in Tables 2.1 and 2.2 for co-expression of DNA ligase IV and Xrcc4 were kindly provided by my supervisor, Dr. Murray Junop. The chemicals for the preparation of the buffer solutions were purchased from Bioshop Canada, Ltd. (Burlington, ON). Ampicillin was obtained from Fisher Scientific (Ottawa, ON). Chloramphenicol was purchased from Calbiochem (San Diego, CA). TEV protease was purified in our lab (see Appendix I). Bradford Protein Assay reagents were purchased from Bio-Rad (Mississauga, ON). The HiTrap chelating nickel affinity column (~5mL), the Mono Q 10/10 anion exchange column and HiPrep 26/10 Desalting column were acquired from Amersham Biosciences/GE Healthcare (Baie d'Urfe, QC). Remaining chemicals including those required for preparing selenomethionine-containing media such as amino acids, vitamins and selenomethionine were purchased from Sigma-Aldrich (Oakville, ON). Pfu polymerase was purified in our lab and had activity equivalent to that of commercial turbo pfu polymerase (see Appendix II). Wizard plus SV miniprep kit was obtained from Promega (Madison, WI). This kit was used for performing all the minipreps to purify plasmid DNA. Ungreased VDX multi-well crystal plates, 22mm siliconized circle cover slides, Crystal Screens I and II, Crystal Screen Lite, Cryo Screen,

Natrix Screen, Additive Screens I, II and III and Detergent Screens I, II and III were purchased from Hampton Research (Aliso Viejo, CA). Wizard Screens I and II were supplied from Emerald Biostructures (Bainbridge Island, WA). Silicone lubricant high vacuum grease was obtained from Dow Corning (Midland, MI). Additives NDSB-195, 1,2,3-heptanetriol and 2,5-hexanediol were also obtained from Hampton Research (Aliso Viejo, CA). Different molecular weight polyethylene glycols (PEGs) were obtained from FLUKA (Ronkonkoma, NY). 30K Macrosep and 30K Microsep protein concentrators were obtained from Pall Corporation (Mississauga, ON).

Buffer solutions employed are as follows:

Column I: Immobilized Metal Affinity Chromatography using a Nickel column

Buffer A: 20mM Tris 8.0, 1mM BME, 0.03% LDAO, 500mM KCl, 10% glycerol
Buffer B: 20mM Tris 8.0, 1mM BME, 0.03% LDAO, 500mM KCl, 10% glycerol, 10mM imidazole
Buffer C: 20mM Tris 8.0, 1mM BME, 0.03% LDAO, 500mM KCl, 10% glycerol, 300mM imidazole

Column II: Mono Q anion-exchange column

Buffer D: 20mM Tris 8.0, 10mM DTT, 1mM EDTA, 10% glycerol, 0mM KCl
Buffer E: 20mM Tris 8.0, 10mM DTT, 1mM EDTA, 10% glycerol, 500mM KCl

Column III: Mono Q anion-exchange column

Buffer D: 20mM Tris 8.0, 10mM DTT, 1mM EDTA, 10% glycerol, 0mM KCl
Buffer E: 20mM Tris 8.0, 10mM DTT, 1mM EDTA, 10% glycerol, 500mM KCl

Column IV: HiPrep 26/10 Desalting Column (Buffer Exchange)

Buffer F: 20mM Tris 7.0, 10mM DTT, 1mM EDTA, 150mM KCl

Table 2.1 Plasmids with DNA ligase IV gene segments in pPROEX vector.

Plasmid Name	Expression Segment
pWY 1102	His-TEV-624-911
pWY 1104	His-TEV-630-911
pWY 1117	His-TEV-654-896
pWY 1119	His-TEV-654-911
pWY 1123	His-TEV-658-911
pWY 1124	His-TEV-624-896

Table 2.2 The two plasmids with Xrcc4 gene segments in pACYC vector.

Plasmid Name	Expression Segment
pWY 1190	1-203
pWY 1139	1-267

2.2 Methods

2.2.1 Overexpression of Native Deletion Mutant Complexes

The cloning experiments involved in generating the different plasmids for DNA ligase IV (in pPROEX vector) and Xrcc4 (in pACYC vector) listed in Tables 2.1 and 2.2 were previously performed by Dr. Murray Junop. These plasmids were overexpressed and purified as described below. Different combinations of plasmids containing constructs corresponding to the Xrcc4 functional fragment and DNA ligase IV tandem BRCT domains were co-transformed into *E.coli* BL21 DE3 cells by the heat shock method. A single colony of *E.coli* BL21 DE3 was picked from the LB-Amp (100μg/mL)-Cm (50μg/mL) agar plate to inoculate 45mL of LB-Amp (100μg/mL)-Cm (50μg/mL) for overnight culture at 37°C. 10mL of the overnight culture was then used to inoculate 1L of LB-Amp (100μg/mL)-Cm (50μg/mL), which was grown at 37°C. When the OD_{600} reached a value of approximately 0.5, the culture was induced with isopropyl-thio-β-D-galactopyranoside (IPTG) at a final concentration of 1mM and the cells were continued to grow at 37°C for 3 hours. Subsequently, the cultures were harvested by centrifugation at 7000Xg for 10 minutes, washed with 1X phosphate buffered saline and pelleted by centrifugation at 7000Xg for 10 minutes. The pellets were frozen in liquid nitrogen and stored at -80°C.

2.2.2 Preparation of Selenomethionine-Containing Media

The medium used for growing selenomethionine-derived protein consisted of 2X M9 + 0.4% glucose minimal medium supplemented with 19 amino acids (-Met) at 40μg/mL, seleno-L-methionine at 40μg/mL and vitamins at 1μg/mL. It was made up as follows: all glassware were autoclaved and solutions autoclaved and/or filter-sterilized. 20X M9 stock solution was prepared by mixing 10g NH_4Cl, 30g KH_2PO_4, 68g Na_2HPO_4 anhydrous or 128 g $Na_2HPO_4.7H_2O$ and water to make up to a total volume of 500mL. Amino acid mix I was prepared by mixing all amino acids except methionine, tyrosine, tryptophan and phenyl alanine so that the concentration of each amino acid came to 4 mg/mL. This was filter sterilized and stored at -80 °C. Likewise, amino acid mix II was prepared by mixing tyrosine, tryptophan and phenyl alanine such that the concentration of each was 4 mg/mL. Aromatic amino acids were included in a separate mixture because of their lower solubility compared to the others. The pH of this mixture was adjusted to 8 with NaOH to improve solubility before bringing it up to volume. This solution was then filter sterilized and stored at -20°C for use on the same day. Vitamin mix was prepared by adding riboflavin, niacinamide, pyridoxine monohydrochloride, and thiamine at 1mg/mL of each in the final mixture. The mixture was then filter sterilized and stored at -20°C. 1L of final media was prepared by adding 2 mL 1M $MgSO_4$, 100 mL 20X M9, 2 mL $FeSO_4.7H_2O$ at 12.5 mg/mL, 10 mL 40% Glucose, 10 mL amino acid mix I at 4 mg/mL, 10 mL amino acid mix II at 4 mg/mL, 1 mL vitamin mix at 1 mg/mL, 4 mL seleno-L-methionine at 10 mg/mL and water to make up to a total volume of 1L with an adjusted pH of 7.4.

2.2.3 Overexpression of Selenomethionine-Derived Deletion Mutant Complexes

To produce selenomethionyl protein, the plasmid pWY 1190 (1-203 of Xrcc4) or pMJ4266 (A60E mutant of 1-203 of Xrcc4) was co-transformed into *E.coli* B834(DE3) cells along with the plasmid pWY1119 (654-911 of DNA ligase IV). Growth of these cells in selenomethionine ensured that all the methionine in the protein was in the form of selenomethionine. Expression of selenomethionyl protein was similar to regular protein expression with the exception of using selenomethionine-containing media. Since the doubling time of the cells in this media was about 3 hours, cells took much longer to grow and were induced at OD_{600} of 0.5 and harvested at OD_{600} of 2.0.

2.2.4 Generation of Surface Mutants of Xrcc4 and DNA Ligase IV Deletion Mutants

The plasmids pWY1190 (1-203 of Xrcc4) and pWY1119 (654-911 of DNA ligase IV) were chosen for being engineered to contain surface mutations. Six surface mutants of Xrcc4 (E25A/K26A, E49A, A60E, E91A, K102A, E114A/K115A/E117A) and five surface mutants of DNA ligase IV (K654A/S656A, D661A/E663A, S672A/Q673A, K731A/K733A, D761A/S672A) were generated by carrying out site-directed mutagenesis reactions. The complimentary primers were designed to include approximately 15 nucleotides on either side of the amino acid/amino acids to be mutated. The mutagenesis reaction was carried out by initially mixing 5μL 10X pfu turbo buffer, 1μL 12.5mM dNTP mix, 0.25μL template (100ng), 1μL primer # 1 (125ng/μL), 1μL primer # 2 (125ng/μL) and 41.75μL of sterile water. Once the reaction mixture was

mixed well and placed on ice, 1μL of pfu polymerase was added and the contents were thoroughly mixed before starting the program in the PCR machine as shown in Table 2.3.

Table 2.3 The program used for carrying out site-directed mutagenesis reactions.

LID	105°
NOWAIT	AUTO
T=95°	30 SEC
T=95°	30 SEC
T=55°	1 MIN
T=68°	18 MIN
GOTO 2 REP 18	
T=68°	40 MIN
HOLD 4°	ENTER
END	

After the completion of the reaction, the parental DNA was digested by adding 1μL of Dpn I and incubating at 37°C for 1hour. 1μL of the reaction mixture was then used to transform the plasmid into Top 10 chemically competent cells by the heat shock method and subsequently the cells were plated on LB-Amp (100μg/mL) agar and LB-Cm (50μg/mL) agar for DNA ligase IV and Xrcc4 mutants respectively. The next day, a single colony was picked from the plate to inoculate 10mL of media containing the appropriate drug and grown at 37°C for approximately 10-12 hours. Cells were then spun down and the pellet obtained was used to extract the plasmid by miniprep using Wizard plus SV miniprep kit. The plasmid obtained was then sequenced to validate the presence of the desired mutations(s).

2.2.5 Purification of Xrcc4-DNA Ligase IV Deletion Complexes

Purification of native and selenomethionyl protein complex followed the same protocol, which is described below. To purify the protein complex, each 2L cell culture pellet was re-suspended in 25mL of ice cold Buffer A. A protease inhibitor cocktail comprised of pepstatin, leupeptin, benzamidine and PMSF was added to prevent degradation of the desired protein complex and the cells were passed through a French Pressure Cell at 12000 psi. Lysing using the French Pressure Cell was repeated 4 times to obtain good lysis efficiency and the suspension was clarified by centrifugation at 30,000Xg for 40 minutes at 4°C after adding more of the protease inhibitor cocktail. Protease inhibitors leupeptin and benzamidine were added again and the filtered supernatant was applied to an equilibrated HiTrap chelating nickel affinity column (~5mL). The column was washed with Buffer B (10mM imidazole) at first and then with 5% (24.5mM imidazole) and 10% Buffer C (39mM imidazole) to eliminate unwanted impurities. The protein complex was eluted at 70% Buffer C (213mM imidazole) and fractions containing the protein were pooled and diluted with Buffer D to decrease the final salt concentration to 150mM KCl. The filtered sample was subsequently loaded onto a Mono Q column pre-equilibrated with 30% Buffer E (150mM KCl). The complex was then eluted with a linear gradient from 30% Buffer E (150mM KCl) to 80% Buffer E (400mM KCl). Pooled fractions were treated with 66μg of TEV protease per milligram of protein for 100 min at room temperature to remove the N-terminal tag comprised of six histidine residues, and the filtered sample was re-purified using a Mono Q column as described earlier. The progress of purification was monitored by analyzing the fractions

using 12% SDS-PAGE gels and staining with Coomassie Brilliant Blue. Finally, the selected fractions of the protein complex were exchanged against Buffer F using a HiPrep 26/10 Desalting column and concentrated to ~ 7 mgs/mL using both Macrosep and Microsep 30K protein concentrators. The concentration of the purified protein was determined using Bradford Assay (Bradford, 1976).

2.2.6 Crystallization of Different Deletion Mutant Complexes of Xrcc4 and DNA Ligase IV

2.2.6.1 General Crystallization Set Up

Prior to crystallization, the protein sample was centrifuged at 4°C to pellet large particles and/or precipitate that could interfere with crystallization. Crystallization was carried out using the hanging drop vapour diffusion method using VDX crystal plates and siliconized circle cover slides (McPherson, 1982). 0.8mL of the mother liquor was pipetted into each well of the 24-well VDX crystal plate. Subsequently, 1µL of mother liquor from the well was mixed with 1µL of protein solution on a siliconized cover slide and the slide was inverted and sealed with silicone lubricant high vacuum grease. Crystal trays were incubated at 20°C and 25°C in thermostable incubators or at 4°C in the cold room (Blow *et al.,* 1994). Initial screening involved 342 different conditions including Crystal Screens I and II, Crystal Screen Lite, Cryo, Natrix, Wizard I and Wizard II screens (Jancarik and Kim, 1991) (Cudney *et al.,* 1994).

2.2.6.2 Optimization Tool I: Fine Screening Technique

Fine screening optimization was carried out on the crystallization conditions that resulted in nucleation and/or crystal growth in the initial screening. Since there are several crystallization variables that could be altered in a given crystallization condition to affect size and quality of the crystals, a few significant variables were chosen. The variables specifically selected to optimize the crystals included pH of the buffer (McPherson, 1995), precipitant concentration (Jovine, 2000), salt concentration, temperature (Blow *et al.*, 1994), amount of mother liquor in the well and ratio of protein to precipitant concentration in the droplet. Unlike other parameters that were varied over a gradient, only two different temperatures, 20°C and 25°C were employed in optimizing the crystals. The fine screening technique of optimization was carried out using two-dimensional matrices with the two different variables to be altered on the two axes (Fig. 2.1). The process involved changing a particular set of two variables over a gradient while keeping the remaining variables constant and analyzing the effect of those variables on crystal growth. The best crystallization condition resulting from the first round of fine screening optimization was chosen to carry out the next round of fine screening, where the next two variables were altered. Each round of screening became part of an iterative process that started with a coarse grid over a wide range and ended with a fine grid over a narrow range.

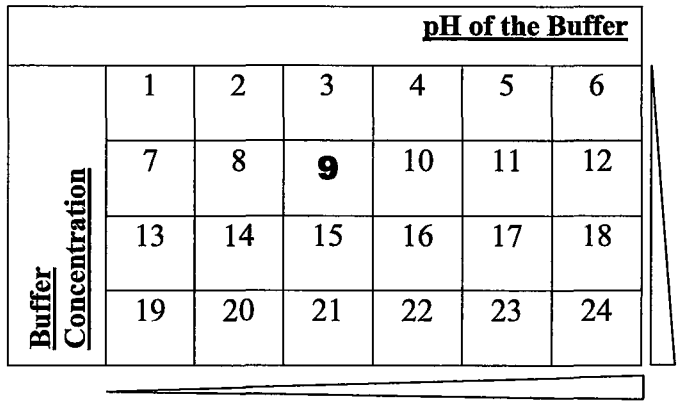

Figure 2.1 An example of a fine screening optimization grid comprised of a two dimensional matrix. The numbers 1-24 depict the 24 wells of the VDX crystal plate in 6 columns and 4 rows. The initial crystallization condition is positioned in well number 9 (column 3 and row 2). The X-axis represents a gradient of the buffer pH while the Y-axis represents a negative gradient of the buffer concentration.

2.2.6.3 Optimization Tool II: Crystallization Using Additives and Detergents

Optimization using additives and detergents involved the combination of the condition, which yielded crystals/nuclei individually, with a series of reagents such that the final concentration of the newly added reagent in the droplet was at 10%. Additive Screening kits I, II, III as well as Detergent Screening kits I, II and III from Hampton Research were employed as the optimization tools. Additives are usually comprised of small molecules including ions, salts and low molecular weight PEGs that could potentially improve the size and quality of macromolecular crystals by enhancing the stability of the protein, homogeneity of its conformation, crystal contacts and/or solvent interactions. Detergents, on the other hand mostly influence the solubility of the sample by perturbing solvent structure. Each optimization trial included testing of every unique crystallization condition with all the additives and detergents provided in the kits at 20°C

and 25°C. Since each of the kits consisted of 24 unique conditions, a total of 144 different conditions were tested at two different temperatures for each optimization trial.

2.2.6.4 Optimization Tool III: Crystallization Using Microseeding

Microseeding is an optimization technique employed to control excessive nucleation and enhance crystal growth to obtain fewer large crystals. This method is highly reproducible and is usually chosen when too many tiny nuclei/crystals are observed in the droplet as opposed a limited number. Prior to performing microseeding, 1mL of stabilization buffer was prepared for the crystallization condition. The stabilization buffer contained components of the mother liquor at concentrations similar to those in the hanging droplet after equilibration against the well solution. Using a preliminary initial crystallization condition, microseeding was carried out in the manner described below over a gradient of precipitant concentration to locate the condition with right amount of crystal seed and precipitant that enabled the growth of single large crystals.

1. A few wells where the droplets contained numerous tiny crystals without any precipitate were selected.
2. The crystals were harvested into a glass, crushing device with the aid of 500μL of stabilization buffer.
3. The crystals were then thoroughly crushed and the resulting suspension was transferred to an Eppendorf tube containing 100μL of stabilization buffer.
4. This was used as the crystal stock suspension.

5. The crystal stock was serially diluted to obtain different concentrations of crystal seed suspensions as shown in Fig. 2.2.
6. Mixing of the crystal stock by vortexing was carried out at every step to ensure a uniform distribution of the seed in the suspension.
7. The dilutions obtained included 1:100, 1:500, 1:2500, 1:12500, 1:62500 and 1:312500 of which only 1:2500, 1:12500, 1:62500 and 1:312500 were used to set up crystallization trials as the concentration of seed in the first few dilutions was predicted to be too high to obtain fewer, big crystals.
8. Finally, 1μL of the protein was mixed with 1μL of the crystal seed dilution on a siliconized cover slide and sealed against 800μL of mother liquor acting as the well solution.

Each optimization trial involved six different precipitant concentrations over four different concentrations of the crystal seed at 20°C and 25°C, totalling to 48 individual conditions.

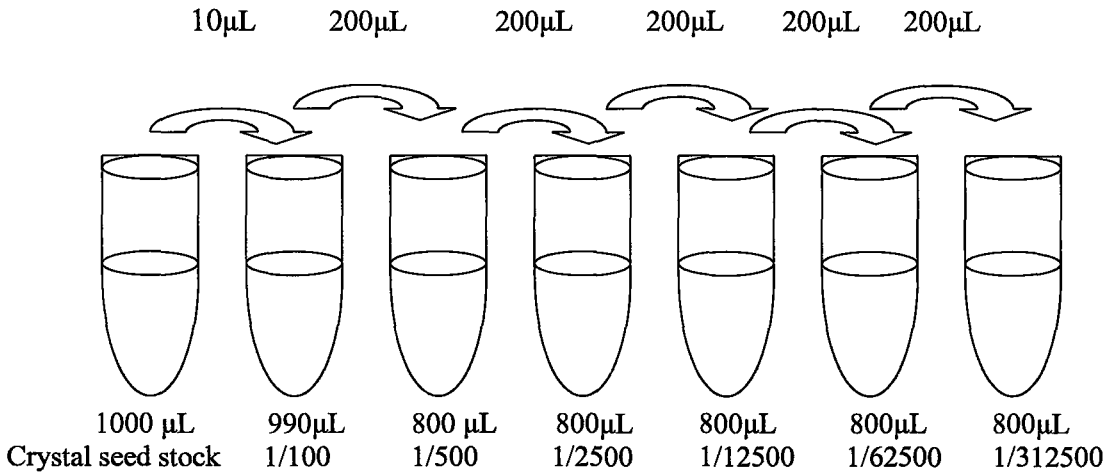

Figure 2.2 Preparation of serial dilutions used for microseeding. The first tube on the extreme left contains the crystal seed stock. The first dilution is achieved by transferring a given amount (shown on the top of the arrow) of crystal seed stock to the tube on the immediate right containing a certain quantity (displayed at the bottom of the tube) of stabilization buffer. Further dilutions are obtained by repeating this process of transferring crystal suspension from the current tube into the next tube containing a different amount of stabilization buffer solution. The final dilutions obtained are depicted at the very bottom of each tube.

2.3 Results and Discussion

2.3.1 Purification of Unmutated Xrcc4-DNA Ligase IV Deletion Complexes

A four-step purification strategy was employed to purify Xrcc4-DNA ligase IV deletion mutant complex to achieve the degree of purity required for crystallization. The purification scheme entailed a nickel affinity chromatography, followed by a Mono Q 10/10 anion exchange column, a TEV digestion to remove the N-terminal histidine tag and another Mono Q 10/10 anion exchange column as a final polishing step. A 4L culture of *E.coli* typically yielded ~ 10 mg of pure protein. Fig. 2.3 displays an SDS-PAGE gel of the purified fractions obtained from Mono Q column before and after TEV digestion.

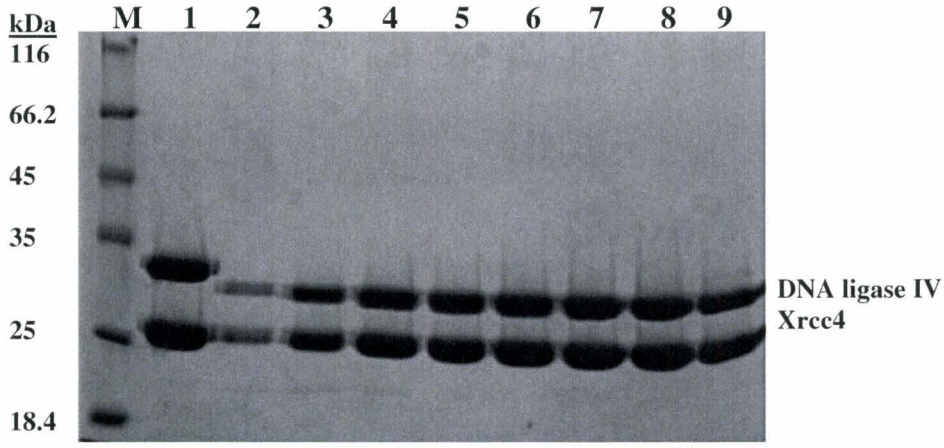

Figure 2.3 Purification of Xrcc4-DNA ligase IV deletion complex. The final purification was analyzed by Coomassie Brilliant Blue stained 12% SDS-PAGE gel loaded with 20μL protein solution per lane. Lane M represents the molecular weight markers in kDa; lane 1 depicts Xrcc4-DNA ligase IV eluate from Mono Q column before TEV digestion; lanes 2-9 depict Xrcc4-DNA ligase IV eluate from Mono Q column after TEV digestion. The positions of Xrcc4 and DNA ligase IV proteins on the gel are indicated on the right.

2.3.2 Crystallization of Xrcc4 - DNA Ligase IV Unmutated Deletion Complexes

The exact domain boundaries for Xrcc4 functional fragment and DNA ligase IV tandem BRCT domains could not be definitely determined. Thus, several different deletion mutants of both Xrcc4 and DNA ligase IV with slightly varying borders were employed in different combinations as shown in Table 2.4 for carrying out initial crystallization experiments. Before my arrival in the lab, crystals of complex 6 (Table 2.4) were obtained by our technician, Kun Zhang at 100mM MES pH 6.5, 30% PEG 5000 and 200mM ammonium sulphate with 20% NDSB-195. In order to screen for better quality crystals, preliminary crystallization trials with each of the deletion mutant complexes were carried out at 7mg/mL protein concentration using the 342 conditions from initial screens at both 20°C and 25°C and in the presence and absence of NDSB-195, an additive that was known to assist crystal growth with the deletion complex 6. Although several of the deletion mutant complexes readily crystallized at condition 26 of Crystal Screen II from Hampton Research ((100mM MES pH 6.5, 30% PEG 5000 and 200mM ammonium sulphate) with 20% NDSB-195) and a few other similar conditions, complexes 1 and 8 shown in Table 2.4 did not yield any crystals. The crystals produced from the remaining complexes morphologically resembled a cuboid with varying edges as shown in Fig. 2.4. Crystals of complexes 5 and 6 (Table 2.4) exhibited the best overall morphology amongst all the crystals and native crystals of complex 5 were chosen for further optimization and diffraction. However, since cells containing selenomethionine-derived protein of complex 6 already existed, selenomethionine-derived crystals were grown and optimized using this protein. Since initial screening resulted in several tiny,

multiple crystals, microseeding was carried out to improve crystal size after performing a few cycles of fine-screening optimization. The additive NDSB-195 was retained in the condition, as it proved to contribute significantly to crystal quality and size. The crystals thus obtained grew to dimensions of 0.6mm X 0.07mm X 0.025mm in 3-4 days. However, most of the crystals were visibly twinned or multiple in nature when observed under the microscope.

X-ray diffraction data sets of the crystals were obtained by retrieving the crystals from the droplet using Hampton Research cryoloops, briefly washing them in a cryoprotectant and placing them in a cryostream to facilitate the start of data collection. Macromolecular crystals are sensitive to X-rays at room temperature and frequently suffer from radiation damage, especially when X-ray experiments are carried out on highly intense synchrotron beamlines. Performing such experiments at cryogenic temperatures greatly reduce radiation damage and thus produce higher quality diffraction data. Cryoprotectants are added to modify the physiochemical properties of the solvent in a way that vitrified state can be reached at moderate cooling rates. This enables cooling of the crystal to cryogenic temperature without the crystallization of any water present either internally or externally in the crystal. The cryoprotectant, in this case, contained 25% ethylene glycol in addition to the components of the mother liquor, at concentrations similar to that in the droplet after equilibration with the well solution. Native crystals obtained in this manner diffracted to approximately 2.9 Å resolution while selenomethionine derived crystals diffracted to 3.0 Å resolution at the beamline X8C of National Synchrotron Light Source (NSLS). Unfortunately, the data collected

demonstrated partially overlapping intensity peaks in the diffraction pattern. In addition to this, difficulty was posed during indexing and merging, suggesting that the crystal was split or non-merohedrally twinned. In spite of making several attempts at solving the structure of the deletion mutant complex by molecular replacement (MR) and multiwavelength anomalous dispersion (MAD) using a number of different data sets collected during different times at the home source and the synchrotron, the structure could not be solved.

Table 2.4 Crystallization attempts with different deletion mutant complexes[1].

S. No	Xrcc4	DNA ligase IV	Results
1	1-203	658-911	No crystals
2	1-203	624-911	Medium crystals
3	1-203	630-911	Tiny crystals
4	1-203	654-896	Medium crystals
5	1-203	624-896	Good-sized crystals
6	1-203	654-911	Good-sized crystals
7	1-267	654-911	Tiny crystals
8	1-267	658-911	No crystals
9	1-267	624-911	Tiny crystals
10	1-267	630-911	Tiny crystals
11	1-267	624-896	Tiny crystals
12	1-267	654-896	Tiny crystals

[1]The first column indicates the serial number; the second and third columns, the deletion mutants used for Xrcc4 and DNA ligase IV in different combinations and the fourth column, the result obtained from crystallization trials of a particular deletion complex. The crystals with best morphology were obtained from the complexes with serial numbers 5 and 6. Complex 5 was chosen for optimization of native crystals while complex 6 was chosen for optimization of selenomethionine-derived crystals.

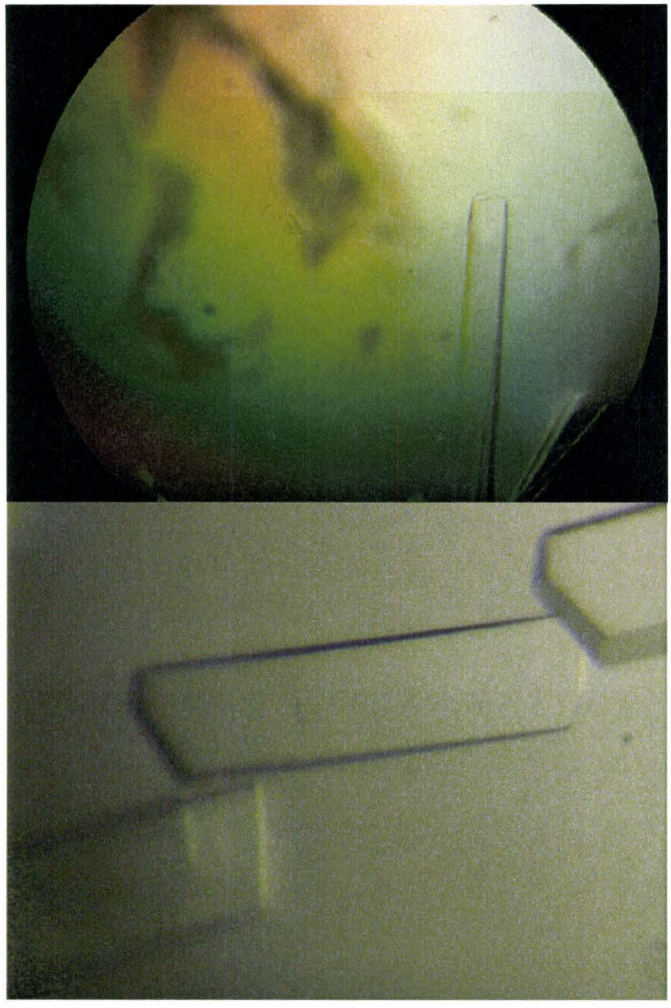

Figure 2.4 Twinned crystals from deletion mutant complex. The crystals obtained from deletion mutant complex 5 of Table 2.3 are displayed here. The crystals were grown in 100mM MES pH 6.5, 30% PEG 5000 and 200mM ammonium sulphate with 20% NDSB-195. The approximate crystal dimensions were 0.6mm X 0.07mm X 0.025mm. Most crystals were not single and exhibited twinned or multiple nature.

2.3.3 Generation of Surface Mutants of Xrcc4 and DNA Ligase IV Deletion Mutants

Since the crystals generated from the deletion mutant complex were twinned and the structure of the complex could not be solved using these crystals even after numerous trials, an alternative methodology was adopted. The goal of this methodology was to obtain a different crystal form of the complex that packed differently and behaved differently particularly without any inherent problems associated with crystal twinning. Since the packing of macromolecules in a crystal is partly contributed by the surface residues of the protein, it was chosen to alter certain charged surface residues of the complex, which could potentially generate new crystal contacts or alter existing ones without compromising the structural integrity of the complex and with minimal impact on the solubility of the protein complex. The surface mutations in Xrcc4 were chosen relying on the previously solved crystal structure (Junop *et al.,* 2000) (Fig. 2.5), whereas the surface mutations in DNA ligase IV were selected based on the solved structure of tandem BRCT domains of BRCA1 and their sequence homology to those of DNA ligase IV. The six mutants engineered for Xrcc4 and the five for DNA ligase IV are shown in Table 2.5. The products obtained from generating some of the mutant plasmids are shown on a 1% agarose gel stained with ethidium bromide in Fig. 2.6.

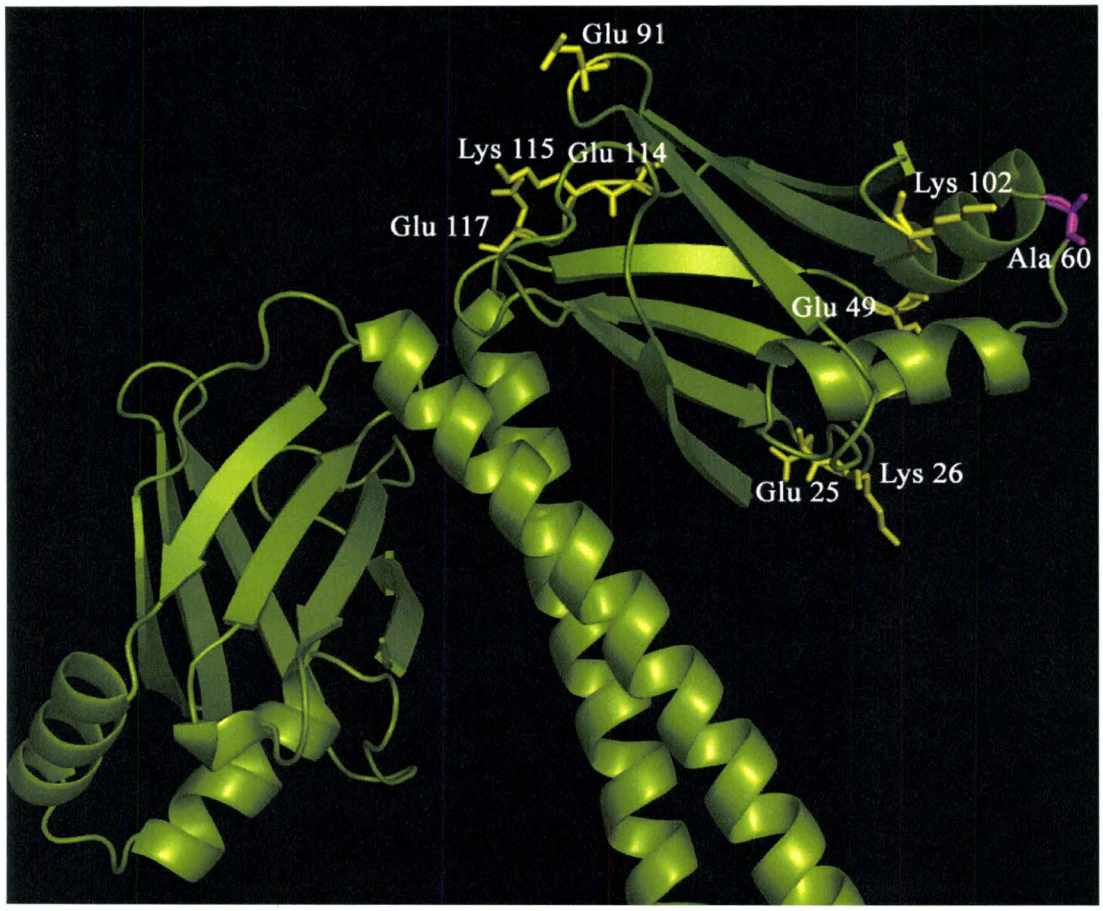

Figure 2.5 Various surface mutations engineered in Xrcc4. Based on the previously solved crystal structures and their packing interactions, the residues potentially involved in interacting with adjacent molecules were mutated. Position A60 where the mutation A60E has been engineered is coloured in magenta while the positions of the remaining mutations are coloured in yellow. **The structure in the picture is taken from PDB 1FU1 (Junop *et al.*, 2000). PyMol software was used to generate the picture (De Lano *et al.*, 2002).**

Table 2.5 Different surface mutants generated in Xrcc4 and DNA ligase IV[1].

Surface Mutations Engineered	
Xrcc4	**DNA ligase IV**
E25A/K26A	K654A/S656A
E49A	D661A/E663A
A60E	S672A/Q673A
E91A	K731A/K733A
K102A	D761A/S672A
E114A/K115A/E117A	

[1]The surface mutants coloured in black crystallized under similar conditions as the unmutated complex. The mutants shown in red did not produce sufficient soluble protein to carry out crystallization. The mutant A60E coloured in green crystallized under new conditions giving rise to two new crystal forms.

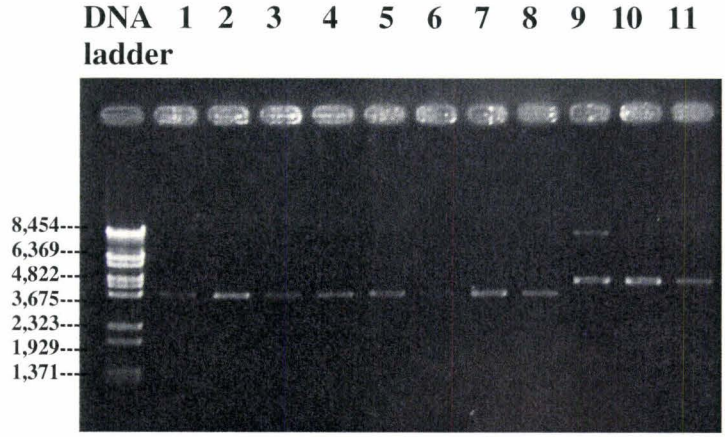

Fig. 2.6 Results of site-directed mutagenesis for making surface mutations in Xrcc4 and DNA ligase IV. A 1% agarose gel stained with ethidium bromide and visualized under UV light for observing the products obtained from mutagenesis reactions. The first lane in the left is the DNA ladder of BstE II digest whose molecular sizes are indicated on the left side of the figure in base pairs; lanes 1-8 are the products from reactions for generating mutants of Xrcc4; lanes 9-11 are the products from reactions for generating mutants of DNA ligase IV. The expected plasmid sizes for Xrcc4 and DNA ligase IV are ~ 3700 base pairs and ~ 4400 base pairs respectively, which are close to the products observed on the gel.

2.3.4 Purification of the Surface Mutant Complexes of Xrcc4-DNA Ligase IV Deletion Mutants

To test the ability of the mutant complexes listed in Table 2.5 to crystallize in a different form, each mutant complex was overexpressed, purified and tried for crystallization. Of the 11 mutant complexes that were purified, only 8 of them (shown in black in Table 2.5) yielded significant soluble protein to carry out crystallization trials and of those, only one mutant complex (shown in green in Table 2.5) crystallized under new conditions giving rise to a different crystal form. The mutants shown in red in Table 2.5 did not yield a reasonable amount of soluble protein to carry out crystallization attempts. The protein obtained after final purification of one of the surface mutant complexes that yielded ~ 10mg of soluble protein from a 4L culture of *E.coli* is shown in Fig. 2.7 by running it on a 12% SDS-PAGE gel. A four-step purification strategy was employed to purify these complexes to achieve the degree of purity required for crystallization. The purification scheme entailed a nickel affinity chromatography, followed by a Mono Q 10/10 anion exchange column, a TEV digestion to remove the N-terminal histidine tag and another Mono Q 10/10 anion exchange column as a final polishing step.

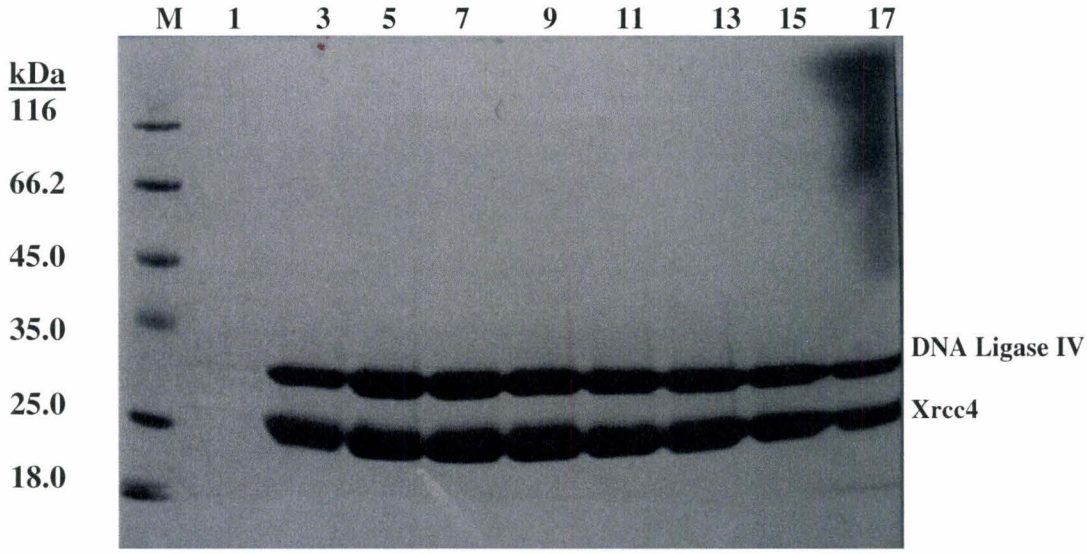

Figure 2.7 Purification of Xrcc4-DNA ligase IV surface mutant deletion complex. The final purification was analyzed by Coomassie Brilliant Blue stained 12% SDS-PAGE gel loaded with 20μL protein solution per lane. Lane M represents the molecular weight markers whose weights are depicted in kDa on the left; lanes 1-17 represent the alternate fractions obtained from Mono Q column after TEV digestion. The positions of Xrcc4 and DNA ligase IV proteins on the gel are indicated on the right.

2.3.5 Crystallization of the Mutants E91A, K102A, E49A, E25A/K26A of Xrcc4 and the Mutants K731A/K733A, D761A/S672A and K654A/S656A of DNA Ligase IV Complexed with Unmutated Binding Partners

The purified mutant complexes (mutants E91A, K102A, E49A, E25A/K26A of Xrcc4 complexed with unmutated DNA ligase IV and the mutants K731A/K733A, D761A/S672A and K654A/S656A of DNA ligase IV complexed with unmutated Xrcc4) were individually screened for crystallization at 7mg/mL protein concentration using the initial screening kits containing 342 conditions at 20°C in the absence and presence of NDSB-195, an additive that aided crystal growth with the unmutated protein complex. All the complexes only formed crystals/nuclei at the conditions earlier observed with the unmutated deletion mutant complex suggesting that the crystals belonged to the old crystal form and that the mutations introduced did not significantly alter the crystal packing. Furthermore, the crystals were also morphologically similar to the previous ones and demonstrated severe twinning. Therefore, additional screening for new crystals or further optimization of the observed crystals was not pursued with these mutants.

2.3.6 Crystallization of A60E Mutant of Xrcc4 (1-203) Complexed with Unmutated DNA Ligase IV (654-911)

Crystallization trials with the A60E mutant of Xrcc4 (pMJ4266) complexed with DNA ligase IV (pWY1119) were carried out in a manner similar to the other mutant complexes. First round of screening was performed with 7mg/mL of native protein using the initial screens consisting of 342 conditions at 20°C. The crystals that grew under

similar conditions as the unmutated deletion mutant complex were ignored, if also morphologically similar. New crystals were observed with the A60E mutant of Xrcc4 complexed with DNA ligase IV at two distinct conditions, which did not previously generate any crystals with the unmutated deletion mutant complex. These two new crystals forms obtained will now be discussed in detail.

2.3.6.1 New Crystal Form I

The New Crystal Form I was obtained at 20°C from condition 22 of Hampton Cryo Screen consisting of 85mM Tris 8.5, 25.5% PEG 4000, 170mM sodium acetate, and 15% glycerol (Fig. 2.8). The drop contained a single tiny diamond-shaped crystal in a relatively clearer part of the drop, whereas much precipitate was observed in the remaining portion of the drop. Due to its extremely small size, this crystal could not be used for X-ray diffraction analysis. In order to improve the crystal size and achieve reproducibility, optimization of the crystal condition was pursued. Varying the usual crystal growth parameters including temperature and the use of additives and detergents did not have any considerable positive impact on the nature of the crystals. Despite extensive efforts, the crystal size improved only marginally while still maintaining the original crystal quality. One of the crystals, a representative of the optimized crystals having dimensions of 0.1mm X 0.07mm X 0.04mm diffracted poorly at the synchrotron to approximately 7Å resolution.

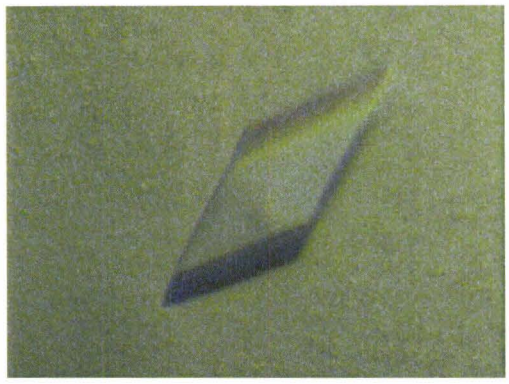

Figure 2.8 New Crystal Form I of A60E mutant complex. This crystal was obtained at new crystallization conditions (85mM Tris 8.5, 25.5% PEG 4000, 170mM sodium acetate, and 15% glycerol) different from the crystals generated from the unmutated complex. The crystal dimensions were 0.1mm X 0.07mm X 0.04mm.

2.3.6.2 New Crystal Form II

The second alternative crystal form was observed at 20°C with NDSB-195 in condition 8 of Wizard II screen containing 100mM Sodium-Potassium phosphate pH 6.2, 10% w/v PEG 8000 and 200mM sodium chloride (Fig. 2.9). Although the crystals had a cuboid shape with varying edges and looked somewhat similar to the crystals obtained from the unmutated complex, this condition was distinctly different from the original condition and the unmutated complex did not crystallize under this condition even after several attempts. This indicated that the crystals generated most likely belonged to a new crystal form with different packing interactions. Furthermore, these crystals exhibited morphologically superior quality compared to the old crystal form. The initial crystals observed were single and had reasonable dimensions to be used for X-ray diffraction analysis. However, the drop only contained two large crystals, which were a few months old. Previous X-ray diffraction studies of this protein complex had shown that the age of

the crystal exerts a monumental impact on the resolution and the diffraction quality of the data. Taking into consideration these factors, crystal optimization with fine screening was pursued to reproduce the original crystals and improve their quality. A considerable improvement in crystal quality was apparent from varying crystal growth parameters such as pH, precipitant concentration, salt concentration and drop size. The next optimization step, microseeding produced substantially larger crystals of approximately 0.6mm X 0.3mm X 0.2mm dimensions in 2-3 days. Both native and selenomethionine-containing crystals were obtained using the same optimization steps. The native crystals diffracted to 2.29 Å resolution and the selenomethionine-containing crystals diffracted to 2.4 Å resolution at the synchrotron.

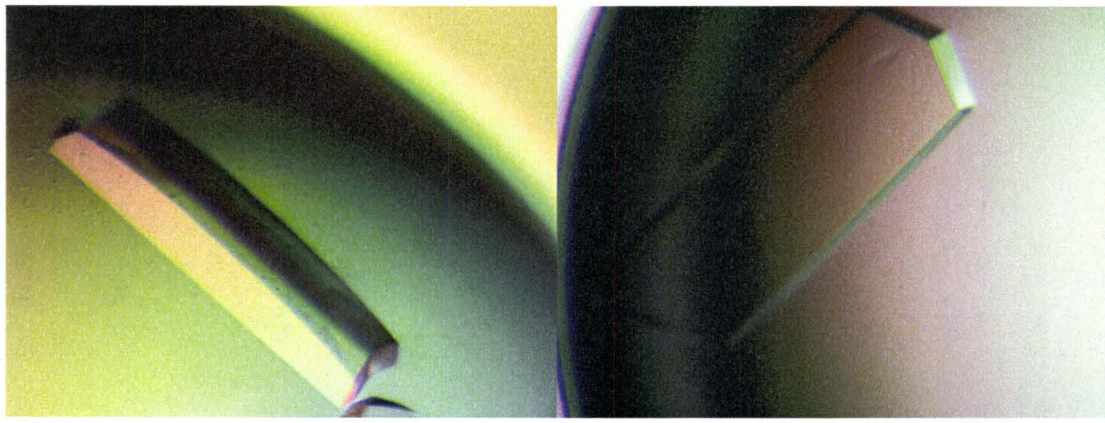

Figure 2.9 New crystal form II of A60E mutant complex. These crystals were obtained at a set of conditions that were different from the crystal growth conditions of the unmutated complex. The crystals grew to ~ 0.6mm X 0.3mm X 0.2mm dimensions in 100mM sodium-potassium phosphate pH 6.2, 10% w/v PEG 8000, 200mM sodium chloride and 20% NDSB-195. Although the overall crystal morphology was similar to the crystals obtained from the unmutated complex, the crystal quality was significantly better.

CHAPTER 3

STRUCTURE DETERMINATION OF THE C-TERMINAL DOMAIN OF DNA LIGASE IV IN COMPLEX WITH XRCC4

3.1 Materials and Methods

All data sets were collected at beamline X8C of NSLS using an ADSC Quantum-4 CCD detector. The software CBASS, which acts as the beamline and experimental control system was used to manipulate data collection steps. The program STRATEGY of HKL2000 suite (Otwinowski *et al.,* 1997) was employed to obtain information about the degrees of data to be collected to achieve completion. Upon the completion of data collection, all data sets were processed using Crystal Clear software package (Rigaku/MSC Ltd., La Jolla, CA) on a Dell Precision 330. CNS (Brunger *et al.,* 1998) and CCP4 (Collaborative Computational Project # 4, 1994) suite of programs were utilized to obtain phasing information. The programs encompassing the CCP4 suite were run with the aid of the graphical interface (Potterton *et al.,* 2003). The 2Fo-Fc and the Fo-Fc electron density maps generated from the model were calculated using the programs from CNS (Brunger *et al.,* 1998) and CCP4 packages. Initial stages of model building and refinement were carried out on an O2 SGI workstation (Helios-Oceana, Mississauga, ON). The program O (version 8) (Jones *et al.,* 1991) was used in model building whereas, structure refinement was accomplished using CNS software suite (Brunger *et al.,* 1998). WinCoot (Emsley *et al.,* 2004) and CCP4 suite were employed during the latter stages to model-build and refine the structure.

3.2 Results and Discussion

3.2.1 Data Collection and Processing

3.2.1.1 New Crystal Form I

Crystals belonging to the New Crystal Form I of A60E mutant of Xrcc4 complexed with DNA ligase IV diffracted to only about 7Å resolution at the synchrotron. Unlike the diffraction observed from crystals of the unmutated complex, this diffraction pattern did not demonstrate any overlapping intensities (Fig. 3.1) implying that the new crystal was devoid of any twinning problems. However, the resolution obtained was very low and therefore the data was not employed in solving the structure.

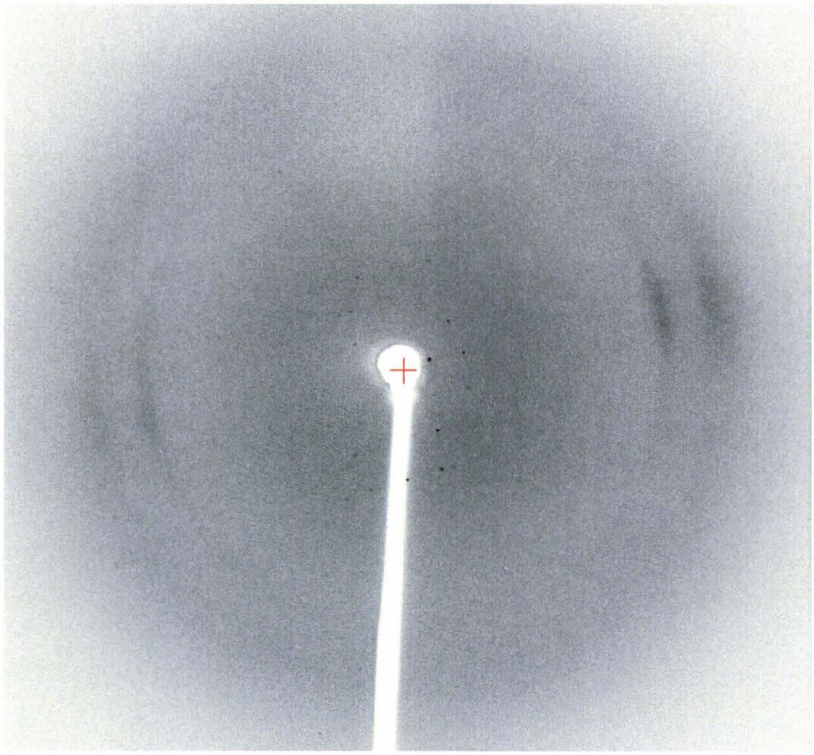

Figure 3.1 Diffraction pattern of New Crystal Form I at the synchrotron. The crystal diffracted weakly to 7.0 Å resolution due to its small size. The diffraction pattern does not demonstrate any overlapping intensities depicting the non-twinned nature of the crystal. **Crystal Clear software (Rigaku) was used to display this image.**

3.2.1.2 New Crystal Form II

The New Crystal Form II of A60E mutant of Xrcc4 bound to DNA ligase IV was crystallized using both native and selenomethionine-derived proteins.

3.2.1.2.1 Native Crystal Form II

The native crystals grown by microseeding were taken to NSLS, where they were briefly washed in a cryoprotectant buffer before mounting. One of the large crystals grown at 100mM sodium-potassium phosphate pH 6.0, 13% PEG 8000, 200mM sodium chloride and 20% NDSB-195 diffracted to 2.29Å resolution. The data set was collected with an oscillation range of 1.0° having a 45 second exposure time and a crystal to detector distance of 150mm using a wavelength of 1.100002 Å. After preliminary data analysis using the initial few diffraction images with the HKL2000 suite (Otwinowski *et al.,* 1997), the space group was suggested to be either C2 or P1. The space group C2 had been chosen for data collection in one of the earlier synchrotron trips with this crystal form and the data did not process well. The R-merge and the residual values were 0.55 0.24 respectively. Therefore the lower symmetry space group P1 was chosen this time for running the STRATEGY program of HKL2000 with the forethought that the data could be usefully employed for both the space groups.

After data collection, the crystallographic data of the native complex was then processed using Crystal Clear software (Rikagu). Although the processing displayed space groups C2 and P1, analysis of the processing statistics depicted in Table 3.1 suggested that P1 was the correct space group. The R-merge for C2 was found to be very

high suggesting that it was most likely not the correct space group (Table 3.1). On the contrary, the R-merge for P1 was 4.7%, which lies well within the normal range. Therefore, further analysis of the diffraction data was carried out in space group P1. Data processed with an R-merge of less than 10% is considered reasonable while anything greater is usually treated with suspicion. Table 3.2 summarizes the statistics for the native data processing of Xrcc4-DNA ligase IV complex in space group P1 and Table 3.3 gives the detailed processing statistics for each resolution shell.

Table 3.1 Comparison of crucial processing statistics of the same data in space groups C2 and P1[1].

	Unit Cell Parameters						Residual	R-merge
	a	**b**	**c**	**α**	**β**	**γ**		
C2	84.342	206.26	67.58	90°	104.63°	90°	0.23	0.52-0.6
P1	66.78	85.78	110.59	67.24°	82.47°	74.28°	0.0	0.047-0.093

[1]The R-merge and residual are the significant indicators to be analyzed for determining the correct space group. **Crystal Clear software (Rigaku) was used to generate this data.**

Table 3.2 Summary of significant processing statistics in space group P1 for a native crystal of Xrcc4-DNA ligase IV complex[1].

Space group	P1
Unit cell dimensions	67.51 86.11 111.73
	67.42 82.64 74.34
Resolution range	45.43 - 2.29 (2.37 - 2.29)
Total number of reflections	274017
Number of unique reflections	98177
Average redundancy	2.79 (2.75)
% Completeness	97.5 (96.0)
Rmerge	0.044 (0.370)
Reduced ChiSquared	1.00 (1.19)
Output <I/sigI>	12.0 (2.7)

[1]Values in () are for the last resolution shell. $R_{merge} = \Sigma| I_{obs} - I_{avg}| / \Sigma I_{avg}$. **Crystal Clear software (Rigaku) was used to generate this data.**

Table 3.3 Detailed processing statistics for each resolution shell in space group P1 for a native crystal of Xrcc4-DNA ligase IV complex[1].

Resolution range	I/sigI (unaveraged)	I/sigI (averaged)	Reduced ChiSquared	Rmerge (shell)	%Completeness (shell)
45.43 - 4.93	22.2	35.4	0.91	0.021	97.6
4.93 - 3.92	17.4	27.9	0.85	0.027	98.6
3.92 - 3.42	11.4	17.9	0.85	0.040	98.2
3.42 - 3.11	7.3	11.5	0.85	0.063	97.9
3.11 - 2.89	4.6	7.2	0.93	0.107	97.8
2.89 - 2.72	3.2	5.1	1.03	0.165	97.6
2.72 - 2.58	2.7	4.3	1.11	0.217	97.3
2.58 - 2.47	2.2	3.5	1.13	0.272	97.0
2.47 - 2.37	1.9	3.1	1.18	0.331	97.1
2.37 - 2.29	1.8	2.7	1.19	0.370	96.0

[1]$R_{merge} = \Sigma| I_{obs} - I_{avg}| / \Sigma I_{avg}$. **Crystal Clear software (Rigaku) was used to generate this data.**

The number of molecules in the unit cell and the solvent content of the crystal were estimated based on Matthews average solvent content analysis (Matthews, 1968). In short, Matthews co-efficient, Vm is defined as the ratio of the volume of the unit cell to the product of the molecular weight of the macromolecule and the number of

molecules within the unit cell. The most probable number of molecules in the unit cell would give a Matthews co-efficient within the empirically observed range of 1.6 and 3.6 $Å^3$/Da.

$$Vm = \frac{\text{Volume of the unit cell}}{\text{Molecular weight of macromolecule*Number of molecules per unit cell}}$$

Based on the crystal sensitivity to handling procedures, it was predicted to contain a high solvent content that lies at the far end of the spectrum for protein crystals. Table 3.4 illustrates the results obtained from Matthews solvent content analysis for the protein complex. Although there are two valid Matthews co-efficient values (3.8 $Å^3$/Da and 2.5 $Å^3$/Da) that correspond to somewhat high solvent content (67.1% and 50.6%) for the protein crystal, the statistics with the higher solvent content are more likely to be correct. Therefore, MR trials were first performed based on two copies of complex per unit cell.

Table 3.4 Matthews average solvent content analysis of Xrcc4-DNA ligase IV complex[1].

Estimated Molecules/Asymmetric Unit	Vm $Å^3$/Da	Solvent Content (%)
1	7.5	83.5
2	3.8	67.1
3	2.5	50.6
4	1.9	34.1
5	1.5	17.7

[1]The solvent content of the crystal and the Matthews coefficient are shown when a given number of molecules are present in the asymmetric unit.

3.2.1.2.2 Selenomethionine-Derived Crystal Form II

Crystals of selenomethionine-derived Xrcc4-DNA ligase IV complex grown in 100mM sodium-potassium phosphate pH 6.0, 13% PEG8000, 200mM sodium chloride and 10% NDSB-195 were taken to NSLS and mounted after a short wash in a cryoprotectant buffer solution. Notably, one half of a crystal that split exactly into two upon being dipped into the cryoprotectant buffer diffracted to 2.4 Å resolution. This crystal was not replaced with another as the crystal had diffracted to 2.4 Å resolution and the spontaneous splitting did not appear to damage the crystal.

It was proposed to collect MAD data from the selenomethionine-derived crystal. The MAD experiment is based on the principle that the magnitude of X-ray anomalous scattering from selenium present in the protein depends on the wavelength of the incident radiation. Since the maximum variations in X-ray scattering are observed around the absorption edge of selenium, the data collected at the absorption peak becomes crucial.

Since the main goal here was to obtain phase information from the selenium atoms of selenomethionine incorporated into the protein of the crystal, the methodology for data collection was designed to have minimum errors between the Bijvoet mates collected. Care was also taken to reduce the errors between different data sets collected. Bijvoet mates/Friedel mates are the pair of reflections arising from a set of lattice planes with Miller indices (h k l) and (-h -k -l) respectively. The differences between the Friedel paired reflections (ΔF) provides the anomalous signal for phasing whereas the dispersive signal is contributed by the difference between two data sets collected at different wavelengths. The appropriate data collection strategy to ensure completeness for the

specific crystal orientation in P1 space group was determined using the program STRATEGY of HKL2000. Before proceeding with the experiment, a selenium scan was performed between X-ray energies, 12594 eV and 12722 eV. Fig. 3.2 shows the final plot obtained which is a smoothed fit to the curve showing anomalous and dispersive peaks. The X-axis of the plot displays X-ray energy in electron volts (eV) and the Y-axis shows anomalous scattering factors f′ and f″.

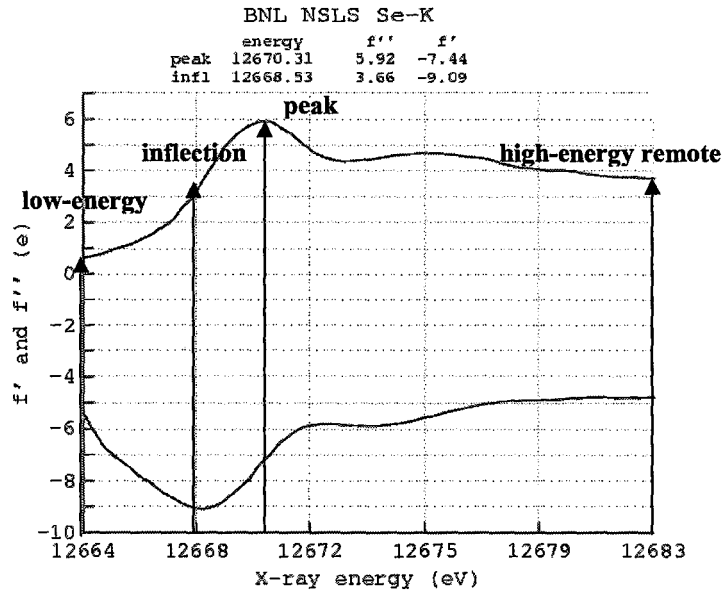

Figure 3.2 A smoothed fit to the curve of a selenium scan between energies 12664 and 12683 eV showing the anomalous and dispersive peaks. The peak and inflection energies along with the high-energy remote and low-energy are depicted. The f′ and f″ at peak energy of 12670.31 eV are –7.44 and 5.92 respectively. The f′ and f″ at inflection energy of 12668.53 eV are –9.09 and 3.66 respectively.

Based on the plot, the first data set was collected at the absorption peak of Selenium K edge (λ-peak = 0.9785; E = 12670.3 eV; f′= -7.44 and f″= 5.92) in pairs of small angular wedges in alternate forward and reverse directions to maximize accuracy of intensity difference measurements. Data collection in this mode is useful to obtain good anomalous differences and reduce the errors caused by crystal decay due to radiation damage. An oscillation width of 0.5° with 45 seconds of exposure was used to collect each 15° wedge of data and the next 15° wedge of data was collected 180° away. This process was repeated till 180° of data was collected at the absorption peak of selenium in both forward and reverse directions totalling to 360°. Next, a non-anomalous data set was collected at the low-energy wavelength following the regular native data collection protocol. Subsequently, data was collected at the inflection point of selenium K edge (λ-infl = 0.9789, E = 12666.3 eV, f′= -9.09; f″= 3.66) in a manner similar to data collection at the absorption peak. As a result of crystal decay, the diffraction patterns obtained from the data sets at absorption peak and inflection point were visibly different (Fig. 3.3). Moreover, significant ice formation on the crystal and surrounding area was noted toward the end of data collection, which was partly remedied by pouring liquid nitrogen over the crystal. The snow formation led to the observation of a few ice rings and high intensity spots not corresponding to data in some of the diffraction images. Therefore, it was decided to use the single data set collected at the absorption peak of selenium as a single anomalous diffraction (SAD) data set for the purposes of solving the structure. In contrast to MAD, SAD relies completely on the anomalous signal between the Bijvoet mates.

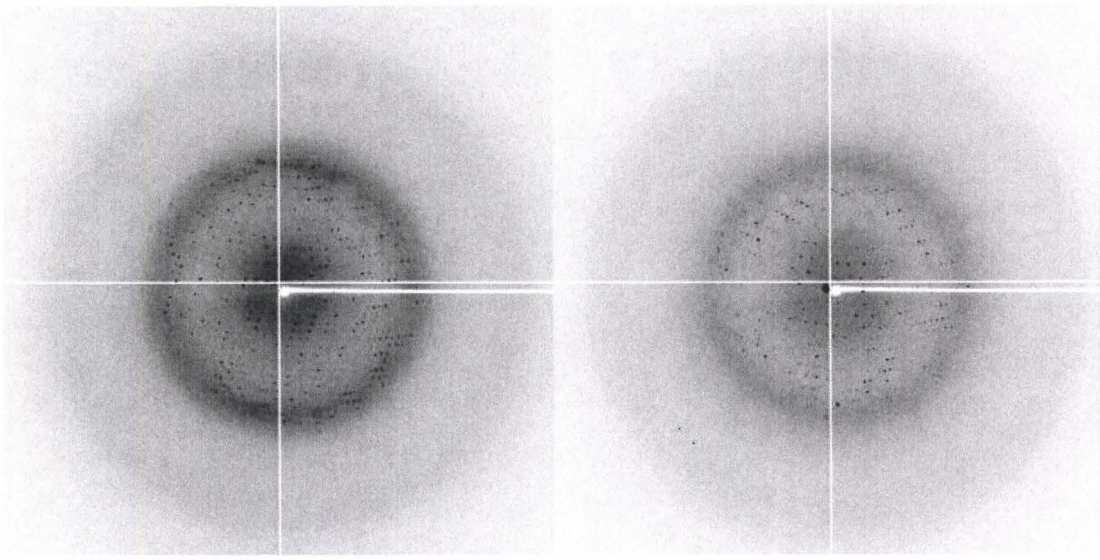

Figure 3.3 Comparison of the diffraction patterns obtained from the same crystal at the beginning (left) and at the end of data collection (right). Crystal decay leads to decreased intensity of the diffraction spots at the end of data collection (right). **Crystal Clear software (Rigaku) was used to generate these images.**

The crystallographic data obtained from the selenomethionine-derived crystal of Xrcc4-DNA ligase IV complex was processed using Crystal Clear software such that the Bijvoet mates were scaled together in order to enable the easy detection of an anomalous signal. P1 was chosen as the space group based on prior knowledge from the native data set. The processing statistics of the data set collected at the peak energy are summarized in Table 3.5 and the detailed processing statistics are provided in Table 3.6. Estimation of solvent content based on Matthews co-efficient calculation was identical to the native crystal predicting two molecules per unit cell with 67.1% solvent content.

Table 3.5 Summary of significant processing statistics in space group P1 for a selenomethionine-derived crystal of Xrcc4-DNA ligase IV complex collected at peak energy[1].

Space group	P1
Unit cell dimensions	67.33 86.00 111.52
	67.37 83.00 74.56
Data set	SeMet (peak)
Resolution range	39.89 - 2.40 (2.49 - 2.40)
Wavelength	0.9785
Total number of reflections	288584
Number of unique reflections	80165
Average redundancy	3.60 (3.52)
% Completeness	92.2 (94.3)
Rmerge	0.060 (0.336)
Reduced ChiSquared	0.97 (1.09)
Output <I/sigI>	11.0 (3.4)

[1]Values in () are for the last resolution shell. $R_{merge} = \Sigma| I_{obs} - I_{avg}| / \Sigma I_{avg}$. **Crystal Clear software (Rigaku) was used to generate this data.**

Table 3.6 Detailed processing statistics for each resolution shell in space group P1 for a selenomethionine-derived crystal of Xrcc4-DNA ligase IV complex collected at peak energy[1].

Resolution range	I/sig unaveraged	I/sig averaged	Reduced ChiSquared	Rmerge shell	%Completeness shell
39.89 - 5.17	11.6	20.4	0.97	0.043	98.0
5.17 - 4.10	11.9	21.2	0.88	0.041	98.9
4.10 - 3.58	10.4	18.8	0.88	0.046	98.7
3.58 - 3.26	7.50	13.5	0.89	0.064	98.5
3.26 - 3.02	4.60	8.10	0.95	0.109	85.5
3.02 - 2.85	3.40	6.10	1.00	0.156	97.7
2.85 - 2.70	2.70	4.60	1.01	0.207	74.6
2.70 - 2.59	2.40	4.20	1.04	0.246	96.2
2.59 - 2.49	2.10	3.60	1.05	0.307	79.4
2.49 - 2.40	1.90	3.40	1.09	0.336	94.3

[1]$R_{merge} = \Sigma| I_{obs} - I_{avg}| / \Sigma I_{avg}$. **Crystal Clear software (Rigaku) was used to generate this data.**

3.2.2 Phase Determination

3.2.2.1 Phasing by Molecular Replacement (MR) Using the Native Data Set

The program PHASER ((Read 2001) (Storoni *et al.,* 2004)), which uses the CCP4 graphical interface (Potterton *et al.,* 2003), was employed for phasing by molecular replacement (MR). The program encompasses MR ensembling, rotation and translation function algorithms that have been developed based on maximum likelihood and multivariate statistics. Moreover, the MR functionality in the program has been linked to packing and rigid body refinement functionalities to provide automated tree-searches, thus making it an extremely powerful MR phasing tool.

The previously solved crystal structures of Xrcc4 were both used as search models with and without the tandem BRCT domains of BRCA1. Since the solutions obtained from searches without the tandem BRCT domains packed better with enough room for accommodating the tandem BRCT domains, they were chosen for further analysis. Interestingly, the several solutions generated from a series of different runs exhibited varied translational coordinates although the rotational coordinates were identical (Fig. 3.4). Furthermore, analysis on the properties of space group P1 revealed that all of the translationally different solutions obtained were correct. Since the origin is not fixed in space group P1, the first copy of the complex can be positioned anywhere within the unit cell and subsequently the next copy of the complex can be located relative to the first by a series of rotation and translation functions. One of the solutions obtained was randomly chosen for performing model building and refinement. Given that this

initial model was incomplete and missing a significant portion of the expected complete structure comprised of the tandem BRCT domains, the phase information obtained was very weak. As observed in Fig. 3.5, the 2Fo-Fc and Fo-Fc electron density maps generated using the initial model showed very poor density for a portion of one of the tandem BRCT domains. The density for the remaining part of the tandem BRCT domains was completely absent. Furthermore, the difference (Fo-Fc) map did not provide significant information to carry out model building or refinement (Fig. 3.5). Since intensive efforts for model building did not demonstrate any improvement in the initial model or the R and free R values, which stayed at around 50%, the refinement was not pursued further.

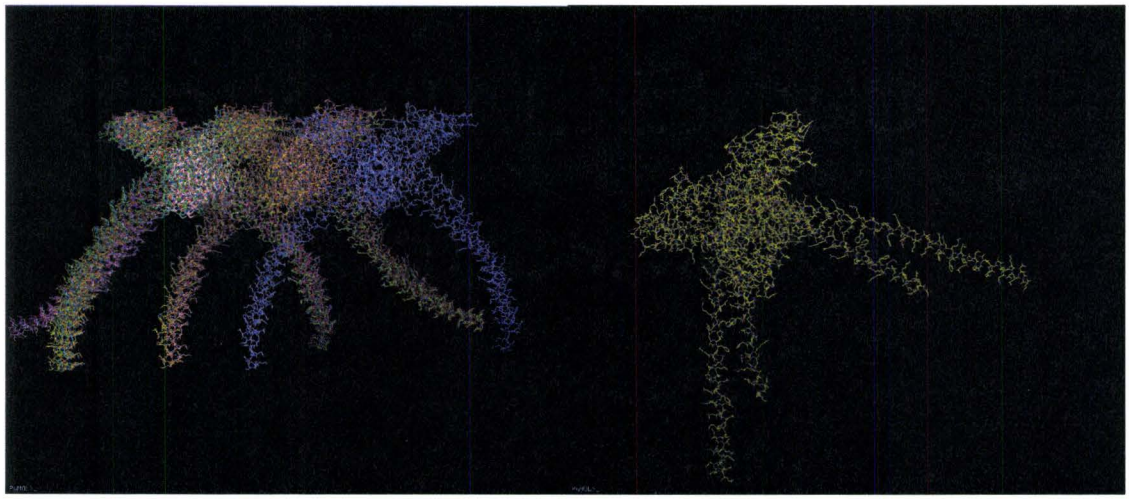

Figure 3.4 Molecular replacement solutions obtained from running the program PHASER ((Read, 2001) (Storoni *et al.*, 2004)). The various rotationally identical but translationally different solutions are shown on the left. Each solution is represented in a different colour. The picture on the right shows one of the solutions picked for being used as the initial model. The two Xrcc4 dimers in each solution differ by ~ 90° rotation. **PyMol software was used to generate the pictures (De Lano *et al.*, 2002).**

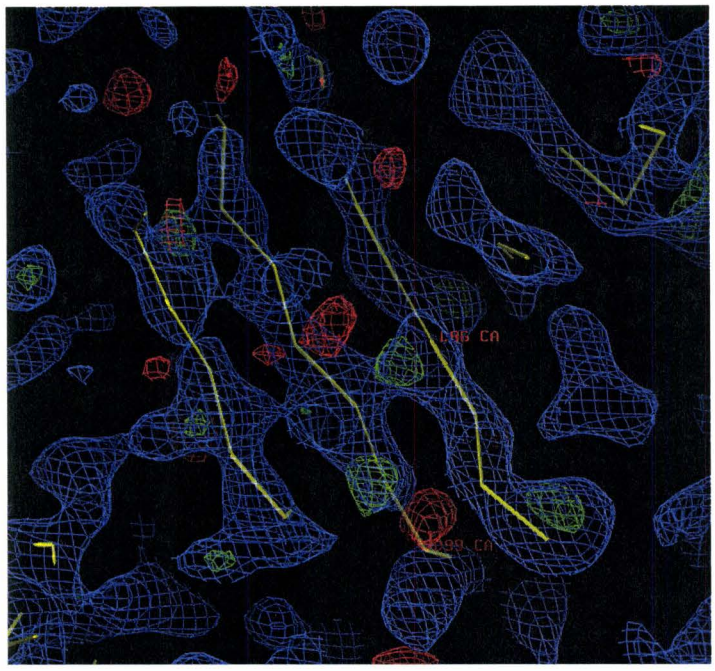

Figure 3.5 Electron density map generated using the initial phases from molecular replacement solution. The blue density represents the 2Fo-Fc map while the green and red represent the Fo-Fc maps contoured at +2.5 and –2.5 sigma values respectively. Although the map is partly interpretable, the information that can be gathered from it is very low due to the weak initial phase information. The possible location of the β-strands in one of the BRCT domains is shown in yellow. **The program O was used to display these electron density maps (Jones *et al.*, 1991).**

3.2.2.2 Phasing by Single Anomalous Diffraction (SAD) Using the Data Set Collected at Peak Energy

CNS version 1.1 software suite (Brunger *et al.,* 1998) was used with the peak data to solve the structure of Xrcc4-DNA ligase IV complex. A series of steps adhering to the SAD phasing protocol were followed to obtain the solution of the complex (Brunger *et al.,* 1998). SAD phasing relies on the anomalous signal included in the diffraction data collected at the absorption peak. In the first step, the reflection file produced from Crystal Clear software containing the anomalous peak data was converted to CNS reflection file format. Wilson scaling was employed to place the SAD data set on an absolute scale, and intensity statistics were analyzed to check the data quality. Next, a heavy atom search for 15 selenium sites (~ 2/3 of total (22)) was performed in the resolution range 15 Å to 4 Å. Each round of heavy atom search included 100 trials, all of which were assigned correlation coefficients at the end of the run. The trial numbers with the highest correlation coefficients in the list were chosen as they were expected to correspond to the correct site configuration. Table 3.7 illustrates the fourteen heavy atom sites located in the initial run, the top seven of which were selected for subsequent analysis based on their correlation coefficients and persistent presence in searches carried out using different parameters.

Table 3.7 The X, Y and Z-coordinates and the corresponding correlation coefficients of heavy atom sites identified during the initial run[1].

SITE #	X-AXIS COORDINATES	Y-AXIS COORDINATES	Z- AXIS COORDINATES	CORRELATION COEFFECIENT
1.	site.x_1=0;	site.y_1=0;	site.z_1=0;	**0.4524**
2.	site.x_2=20.6585;	site.y_2=38.4399;	site.z_2=6.25241;	**0.4523**
3.	site.x_3=19.557;	site.y_3=40.5376;	site.z_3=1.84095;	**0.4524**
4.	site.x_4=73.356;	site.y_4=104.64;	site.z_4=95.7159;	**0.4523**
5.	site.x_5=82.0947;	site.y_5=85.6024;	site.z_5=99.0652;	**0.4523**
6.	site.x_6=46.9121;	site.y_6=82.3685;	site.z_6=98.0514;	**0.4486**
7.	site.x_7=17.6487;	site.y_7=12.3585;	site.z_7=2.87481;	**0.4487**
8.	site.x_8=38.3;	site.y_8=14.1567;	site.z_8=1.00661;	0.4434
9.	site.x_9=55.9814;	site.y_9=34.4803;	site.z_9=62.8794;	0.4434
10.	site.x_10=82.6362;	site.y_10=84.1914;	site.z_10=98.2008;	0.4349
11.	site.x_11=40.7623;	site.y_11=78.1504;	site.z_11=98.8159;	0.4361
12.	site.x_12=63.133;	site.y_12=47.0736;	site.z_12=97.1571;	0.4363
13.	site.x_13=61.3083;	site.y_13=45.8355;	site.z_13=92.4417;	0.4302
14.	site.x_14=18.0621;	site.y_14=21.5709;	site.z_14=32.9145;	0.4375

[1]The top seven sites were chosen based on their higher correlation coefficients and repetitive occurrence in different search trials.

Using this list of possibly correct sites, heavy atom refinement and SAD phasing were carried out. SAD phasing was performed with the site positions obtained and also their inverse positions to overcome hand ambiguity. The resulting SAD phases were used along with anomalous difference structure factors to calculate anomalous difference Fourier maps, which aided in locating additional selenium sites. The newly generated sites were used to perform another heavy atom search and SAD phasing following an iterative pattern until most of the total sites were identified. The final list of heavy atoms obtained from this process included 16 of the total 22 selenium sites. The coordinates for this list of heavy atom sites are shown in Table 3.8. The phase ambiguity, which is integral to the SAD method of solving structure, was resolved by density modification by solvent flipping. The electron density maps produced from density modification with the original selenium sites and the flipped selenium sites were easily distinguishable (Fig.

3.6). The correct site configurations generated an electron density map with continuous stretches of density and showed clearly demarcated protein and solvent boundaries.

Table 3.8 Final list of heavy atom sites chosen for performing phasing[1].

SITE #	X-AXIS COORDINATES	Y-AXIS COORDINATES	Z- AXIS COORDINATES
1.	site.x_1=0;	site.y_1=0;	site.y_1=0;
2.	site.x_2=20.6585;	site.y_2=38.4399;	site.z_2=6.25241;
3.	site.x_3=19.557;	site.y_3=40.5376;	site.z_3=1.84095;
4.	site.x_4=73.356;	site.y_4=104.64;	site.z_4=95.7159;
5.	site.x_5=82.0947;	site.y_5=85.6024;	site.z_5=99.0652;
6.	site.x_6=46.9121;	site.y_6=82.3685;	site.z_6=98.0514;
7.	site.x_7=17.6487;	site.y_7=12.3585;	site.z_7=2.87481;
8.	site.x_8=38.2428;	site.y_8=14.1794;	site.z_8=0.92294;
9.	site.x_9=56.1274;	site.y_9=34.3525;	site.z_9=62.7315;
10.	site.x_10=61.156;	site.y_10=46.092;	site.z_10=92.676;
11.	site.x_11=63.239;	site.y_11=47.349;	site.z_11=97.226;
12.	site.x_12=18.294;	site.y_12=21.689;	site.z_12=33.200;
13.	site.x_13=19.187;	site.y_13=26.457;	site.z_13=36.320;
14.	site.x_14=54.476;	site.y_14=30.389;	site.z_14=66.015;
15.	site.x_15=24.571;	site.y_15=74.038;	site.z_15=85.277;
16.	site.x_16=56.691;	site.y_16=98.121;	site.z_16=102.706;

[1]The X, Y and Z coordinates of the sites are indicated as distances from the origin.

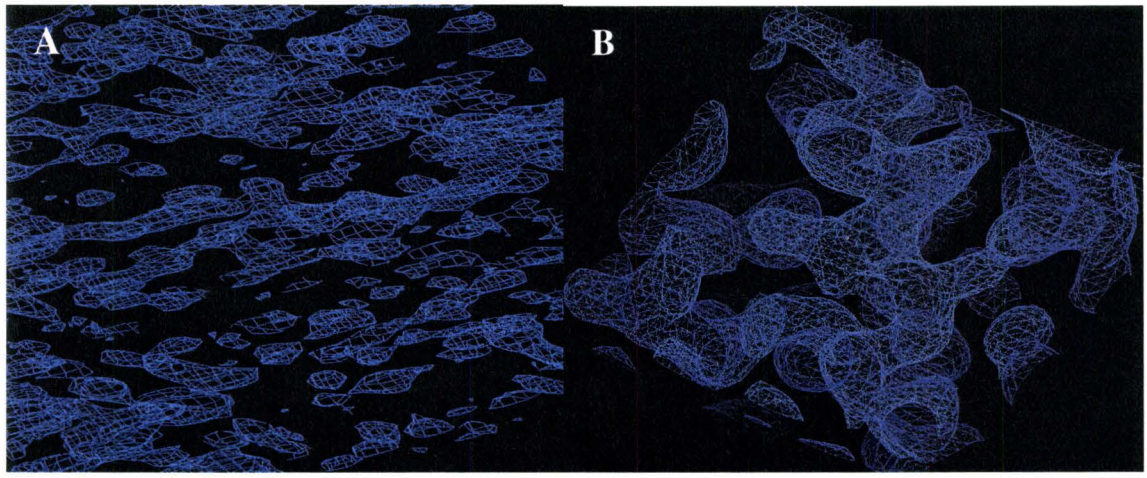

Figure 3.6 Electron density maps generated from SAD phasing and subsequent density modification. (A) Phasing was performed using the flipped heavy atom sites. (B) Phasing was performed using the un-flipped heavy atom sites to obtain an interpretable map showing part of an α-helical region of the tandem BRCT domains of DNA ligase IV. **The program O was used to display these electron density maps (Jones *et al.*, 1991).**

3.2.3 Model Building and Refinement

Prior to performing model building and refinement, a random subset of 5% of the data set was set aside and labelled the test set, using a CNS program (Brunger *et al.*, 1998). Since X-ray diffraction data are redundant to a certain extent, a small amount of data (5%) can be safely omitted from refinement procedures without seriously affecting the result. This test data set does not contribute or add bias to the refined model and therefore can be used for cross-validation purposes during model refinement.

The experimental phases determined from the selenium sites were used in constructing the initial model of Xrcc4-DNA ligase IV complex. The experimental electron density map generated after density modification displayed good density for most of the complex including the tandem BRCT domains. However, one of the head domains of Xrcc4 had significantly less and discontinuous density. The initial model for Xrcc4 was taken from the previously solved crystal structure of Xrcc4 bound to the peptide of DNA ligase IV (Sibanda *et al.*, 2001) and manually manipulated into the electron density on a silicon graphics computer using the program O (Jones *et al.*, 1991). Although the overall configuration of the model appeared to follow the density, several portions were distinctly different. The regions that did not fit the electron density were manually adjusted using the program O (Jones *et al.*, 1991). Also, it was ensured that all the residues of Xrcc4 stayed in register. The residues Ala 202 and Gln 203 at the C-terminal end of Xrcc4 had missing density and were therefore removed from the model.

Once the initial model of Xrcc4 was completed, the model of tandem BRCT domains was built step-by-step starting at the very N-terminal end of the domains. Polyalanine templates and helical templates were employed to trace the Cα atoms of the secondary structures. The Bone_skel command was used to generate a skeletonised electron density map, which enabled preliminary tracing of the structure. The sub-routine Baton_build, which is a part of program O (Jones *et al.,* 1991) was utilized in portions to maintain accurate distances between Cα atoms. In addition, a single BRCT domain from the previously solved crystal structure of BRCA1 tandem BRCT domains was used as a guide to determine the directionality and the overall orientations of each of the domains. Continuous density was initially lacking in the linker region connecting the two BRCT domains. Moreover, a number of loops did not initially exhibit electron density and so were not included in the initial model.

Subsequent to positioning of all the α-helices and β-strands of the BRCT domains, 2Fo-Fc and Fo-Fc electron density maps were generated in CNS (Brunger *et al.,* 1998) based on the initial model. The native data set to 2.4 Å resolution was employed in calculating these electron density maps. Given that majority of the main chain model showed reliable agreement with the observed map and that the maps displayed additional information, the building of loops was undertaken. Following that, some of the side-chains were incorporated into the model and rigid body refinement and simulated annealing were employed to refine the position and orientation of the model. At this stage, another program WinCoot (Emsley *et al.,* 2004) was employed to further position the side chains accurately using a series of 2Fo-Fc and Fo-Fc maps generated from the

CCP4 with the latest model. Automatic refinement was performed by executing a few cycles of maximum likelihood and restrained refinement using the program Refmac5 (Murshudov *et al.,* 1996, Murshudov *et al.,* 1997, Pannu *et al.,* 1998, Murshudov *et al.,* 1999, Winn *et al.,* 2001) of CCP4. Since the initially calculated structure factors usually tend to have poor agreement with the observed structure factors, the process of refinement involves improvement of the discrepancy between them. The refinement imposed constraints on the geometric parameters such as bond lengths, bond angles, torsion angles, chirality and Van der Waal contacts. Non-crystallographic symmetry (NCS) that existed between the two complexes in the unit cell was used to improve refinement, as it increased the ratio of the number of experimentally observed parameters to the components to be refined. NCS is defined as the symmetry present in the asymmetric unit between biologically identical molecules. Tight restraints were used for Cα atoms and regions exhibiting high similarity whereas, loose restraints were used with the side chains and dissimilar portions of the complex. The refinement cycles were interspersed with manual interventions in which residues were placed appropriately within the electron density and the geometry and torsion angles checked to conform to the Ramachandran plot. The final model obtained from this was subjected to B-factor refinement using a program from CNS suite of programs (Brunger *et al.,* 1998). Unfortunately, the previously discussed head domain of Xrcc4 continued to lack reliable electron density even to the end of refinement suggesting a partial disorder in that region. The final structure was complete and well defined including all the loop regions except the loop in one of the head domains of Xrcc4, which was eliminated. A total of 283

water molecules were added during the last step of refinement as early addition could potentially lead to bias and improper positioning of side chains. The entire process of model building and refinement was monitored by calculating the R and free R values. The initial model had R and free R values of 0.48 and 0.50 respectively. After intensive manual and automatic refinement, the final refined structure had an R value of 0.24 and a free R value of 0.28. These slightly high values could be attributed to the presence of partial disorder in one of the head domains of Xrcc4.

To minimize the risk of over-fitting the data, the refinement process was terminated when the free R value started to rise and the R and free R values began to diverge. The Ramachandran plot of the Xrcc4-DNA ligase IV model (Fig. 3.7) was computed using the program PROCHECK (Laskowski R.A., 1993). The validation subroutine included in WinCoot (Emsley *et al.,* 2004) was employed to validate the proper stereochemistry and torsion angles of the protein components in the model. Table 3.9 summarizes the statistics of the final refined crystal structure of Xrcc4-DNA ligase IV complex at 2.4 Å resolution.

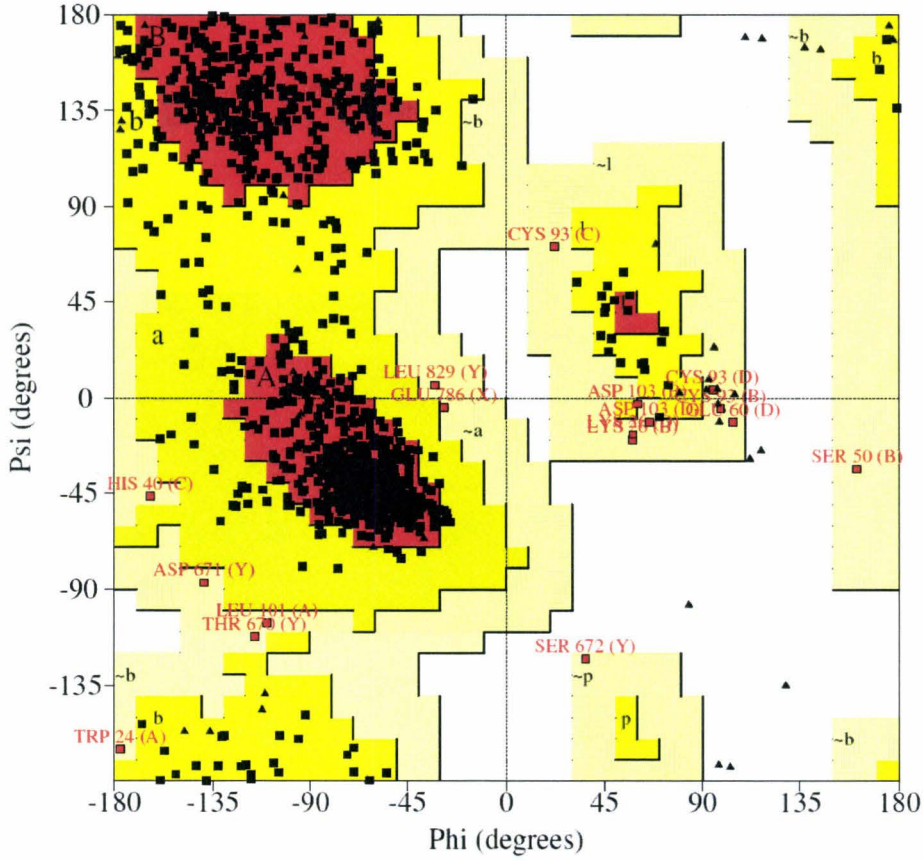

Figure 3.7 Ramachandran plot of Xrcc4-DNA ligase IV complex. Ramachandran plot of Xrcc4-DNA ligase IV complex showing the sterically reasonable values of the angles φ and ψ. While the regions coloured in red represent the most favoured regions, the areas in dark yellow represent additional allowed regions and the portions in pale yellow represent generously allowed regions. The portion in white indicates the disallowed region. Triangles represent glycine residues and squares represent all other residues.

Table 3.9 Statistics of final refined structure of Xrcc4-DNA ligase IV at 2.4 A°.

Refinement Statistics	
Number of reflections used for R factor calculations	72697
Number of reflections used for free R factor calculations	2667
Resolution Range (A°)	20-2.4
Rfree (%)[1]	0.28
Rworking (%)[2]	0.24
Stereochemistry statistics	
Number of glycine residues in protein	54
Number of proline residues in protein	30
Number of residues in protein	1307
Number of atoms with occupancy set to 0[3]	0
Number of atoms in protein	10929
Residues in Ramachandran plot (%)[4]	
Most favourable	83.1
Additionally allowed	15.5
Generously allowed	1.4
Disallowed	0
Deviation from Ideal Values	
Rmsd in bond lengths (A°)	0.046
Rmsd in bond angles (°)	3.627
Rmsd in general Planes (A°)	0.017
Estimated Overall Coordinate Error [5]	
ESU based on R value (A°)	0.318
ESU based on free R value (A°)	0.259
Thermal Factor[6]	
Overall B mean value (A°2)	62.653
Overall anisotropic B value[7]	
B11 (A°2)	2.44
B22 (A°2)	-1.83
B33 (A°2)	-0.6
B12 (A°2)	-0.55
B13 (A°2)	-1.53
B23 (A°2)	0.85

[1,2]R, Rfree = $\Sigma| F_{obs} - F_{calc}| / |\Sigma F_{obs}|$.

[3]The occupancy of certain atoms can be set to zero if the electron density is absent.

[4]Values were analyzed using PROCHECK (Laskowski, 1993).

[5]These values represent the estimated coordinate errors calculated by the Sigma A method.

[6]An overall thermal factor is used to describe the extent of movement generated for the whole model during the temperature changes.

[7]The anisotropic B-values describe the directional dependency of the overall average B-factor.

CHAPTER 4

STRUCTURAL DISCUSSION OF XRCC4-DNA LIGASE IV COMPLEX

4.1 Overall Structural Analysis

The structure of human Xrcc4-DNA ligase IV complex maintains a 2:1 stoichiometry and adopts an overall morphology of a tree with two additional branches (tandem BRCT domains) arising from the main trunk (Fig. 4.1A). There are two 2:1 protein complexes per crystallographic asymmetric unit, and they superimpose with a root mean square deviation (rmsd) of 0.81 Å and 0.82 Å for (Xrcc4 and DNA ligase IV) 652 Cα atoms and for all atoms respectively. In line with the presence of an extended linker region that is unique to the tandem BRCT domains of DNA ligase IV, the structure of Xrcc4-DNA ligase IV complex depicts an unusual model of interaction that was first observed in the structure of lif1p-lig4p complex, the yeast ortholog of Xrcc4-DNA ligase IV.

DNA ligase IV is found to bind tightly to the coiled coil tails of Xrcc4 wrapping itself around to form a clamp-shaped helix-loop-helix domain (Fig. 4.1B). Structural analysis reveals that the helix-loop-helix clamp interacts symmetrically with the A and B chains of Xrcc4, wherein a single amphipathic α-helix lies on either side of the tails of Xrcc4 and the connecting loop makes a uniform turn around the B chain of Xrcc4. The two individual BRCT domains maintain the conserved fold characteristic of their superfamily. Each BRCT domain is comprised of four parallel β-strands forming a β-

sheet and three α-helices flanking them (Williams *et al.,* 2001). Comparison of this structure with the previously solved crystal structure of Xrcc4 bound to the peptide of DNA ligase IV (Sibanda *et al.,* 2001) sheds light on the dramatic conformational changes occurring within Xrcc4 (Fig. 4.2A). The coiled coil region of Xrcc4 undergoes a bend causing a shift in the orientation of Xrcc4 homodimer by approximately 25°. This angle between the residues of different helices was measured with the superimposed structures using COOT (Emsley *et al.,* 2004). Interestingly, the kink created by binding DNA ligase IV is oriented exactly opposite to the kink observed in the earlier structure. In contrast, the relative conformation of chains spanning residues 170-200 remains more or less unaltered and their Cα atoms superimpose with a rmsd of 0.74 Å. Comparative analysis of the structure with its yeast ortholog lif1p-lig4p structure (Dore *et al.,* 2006) depicts that only residues 180-200 (Cα atoms) of Xrcc4 superimpose on its equivalent lif1p with a root mean square deviation of 4.44 Å with the remaining portion of Xrcc4 taking an altogether different pattern leaning close towards the second BRCT domain (Fig. 4.2B).

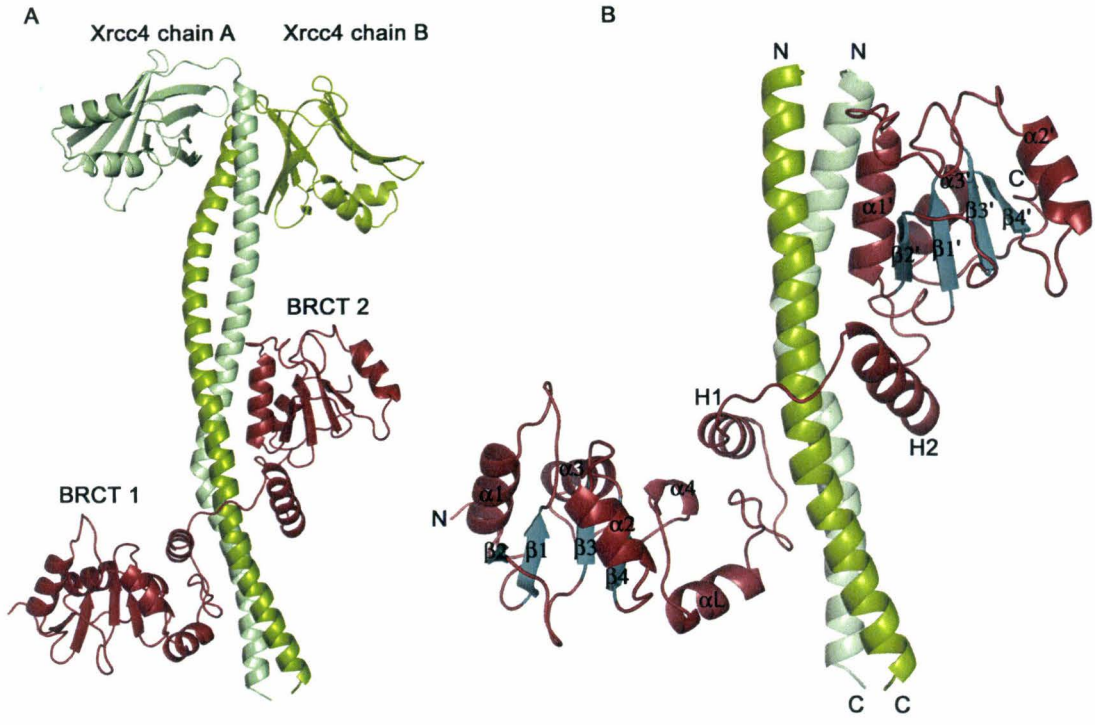

Figure 4.1 Ribbon diagram of Xrcc4-DNA ligase IV complex. The two Xrcc4 chains A and B are coloured in pale green and split pea green respectively while the tandem BRCT domains of DNA ligase IV are shown in red. (A) Overall structure of C-terminal domain of DNA ligase IV bound to Xrcc4 homodimer. (B) Characteristic fold of tandem BRCT domains of DNA ligase IV in complex with Xrcc4. The N-terminal and C-terminal ends of the tandem BRCT domains are labelled as also the various helices and strands. The four parallel β-strands in each of the domains are coloured in deep teal. **PyMol software was used to generate the pictures (De Lano *et al.*, 2002).**

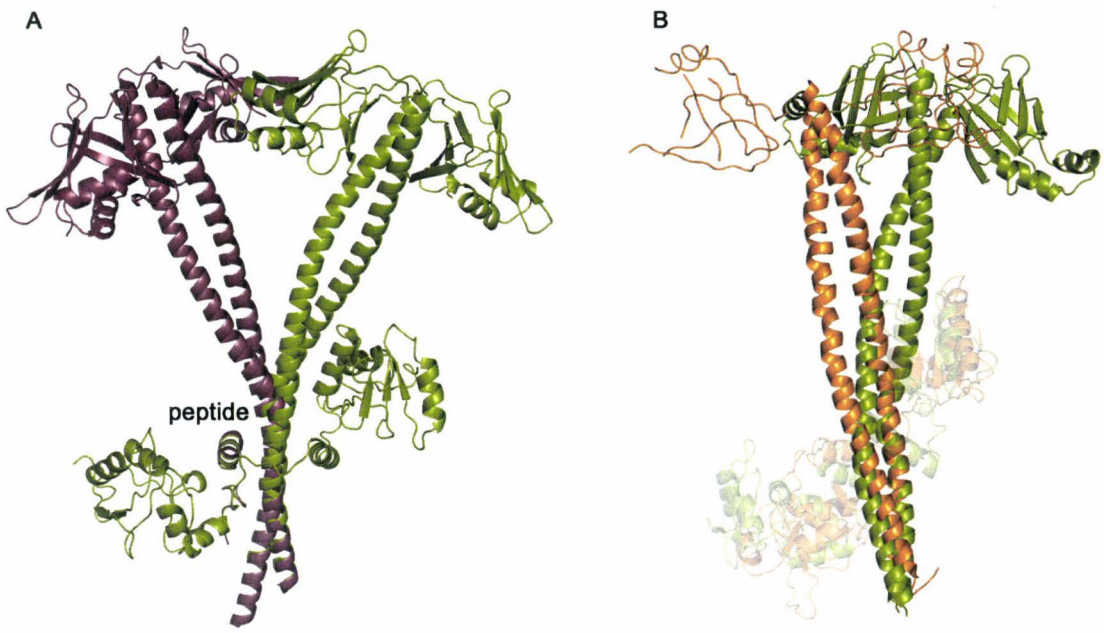

Figure 4.2 Structural alignment of the complex with previously solved similar structures. (A) Structural alignment of the current structure of the complex (human Xrcc4-DNA ligase IV shown in green) with the previously solved structure of Xrcc4 bound to a peptide of DNA ligase IV (shown in raspberry red). (B) Structural alignment of the current structure of the complex (shown in green) with the yeast lif1p-lig4p complex (shown in orange). **The structures with PDB codes 1IK9 (Sibanda *et al.*, 2001) and 1Z56 (Dore *et al.*, 2006) were taken for performing structural alignments. PyMol software was used to generate the pictures (De Lano *et al.*, 2002).**

4.2 Xrcc4-DNA Ligase IV Interface

The Xrcc4-DNA ligase IV interaction interface is extensive and buries 2128.6 $Å^2$ of total surface area, which is slightly more than the buried surface area observed in the yeast lif1p-lig4p complex (Dore *et al.,* 2006) (1867.7 $Å^2$) based on calculations using POPS server[1]. The extensive interaction interface can be broadly divided into four major sections each of which contributes significantly to stabilize the complex. These sections include the 3:5 beta hairpin immediately following the first BRCT, the central helix-loop-helix clamp, the loop preceding the second BRCT and helices α1' and α3' of the second BRCT (Fig. 4.3). The principal fraction of buried surface area is bestowed by the helix-loop-helix clamp, which encompasses 903.7 $Å^2$ of surface area constituted by 33 residues (Fig. 4.4). It is readily apparent that the beta hairpin and the short amphipathic alpha helix (H1) of DNA ligase IV are in total agreement with the previously reported structure of the peptide of DNA ligase IV bound to Xrcc4 (Sibanda *et al.,* 2001) (Fig. 4.2A): the C alpha atoms superimpose with an rmsd of 1.28 Å. Residues 780-788 constitute the loop region of the clamp and are responsible for making polar contacts with the B chain of Xrcc4 (Fig. 4.4). Asn 783 and Ser 784 are engaged in establishing hydrogen bond contacts with Lys 178B and Arg 179B. In contrast, the corresponding loop in the yeast homologue structure (Dore *et al.,* 2006) is comprised of a proline-glycine rich region. C-terminal to this, is a long amphipathic alpha helix H2 spanning residues 789 to 805 (Fig. 4.3 and Fig. 4.4). Five hydrophobic residues, all on one side of the helix are pointing toward Xrcc4 surface and are involved in making contacts, thereby creating a flat

[1] **http://ibivu.cs.vu.nl/programs/popscompwww/**

hydrophobic scaffold (Fig. 4.4). Interestingly, this longer helix H2 is absent in lif1p-lig4p complex (Dore *et al.,* 2006) and instead is replaced with an extended coil. Although the amphipathic alpha helix of DNA ligase IV interacts with both chains of Xrcc4 burying 318 $Å^2$ of hydrophobic surface area, pronounced interactions are observed with the B chain of Xrcc4 where the helix positions itself at a 45° angle relative to the chain. The loop following the clamp is comprised of another stretch of hydrophobic residues forming an additional binding surface for Xrcc4. It is interesting to note that Arg 814 located further down the loop participates in a network of hydrogen bonds (Fig. 4.5) that facilitate interaction with Xrcc4 as well as help maintain the conformation of that portion of the second BRCT motif. The buried Arg814 residue engages in charged hydrogen bonding contacts with Glu 173B, Glu 800 and Ser 811 (Fig. 4.5). Substantial interactions of DNA ligase IV with the A chain of Xrcc4 are facilitated by helices α1' and α3' of the second BRCT domain (Fig. 4.5). While helix α1' interacts with both chains of Xrcc4 through its entire length forming a parallel triple helix bundle, only residue Asp 899 at the C-terminal end of helix α3' makes contacts with Arg 161 of the A chain. Altogether, the combined contribution of all these significant interactions make Xrcc4-DNA ligase IV complex extremely stable.

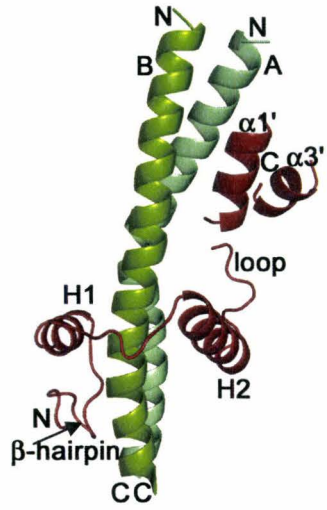

Figure 4.3 Xrcc4-DNA ligase IV - Major interactions. The interface is classified into four major sections based on the structural elements of DNA ligase IV. Starting at the N-terminus, the sections include the β-hairpin, helix-loop-helix clamp, loop region and the helices α1' and α3'. The N- and C-termini of Xrcc4 and DNA ligase IV have been depicted. **PyMol was used to generate the picture (DeLano *et al.*, 2002).**

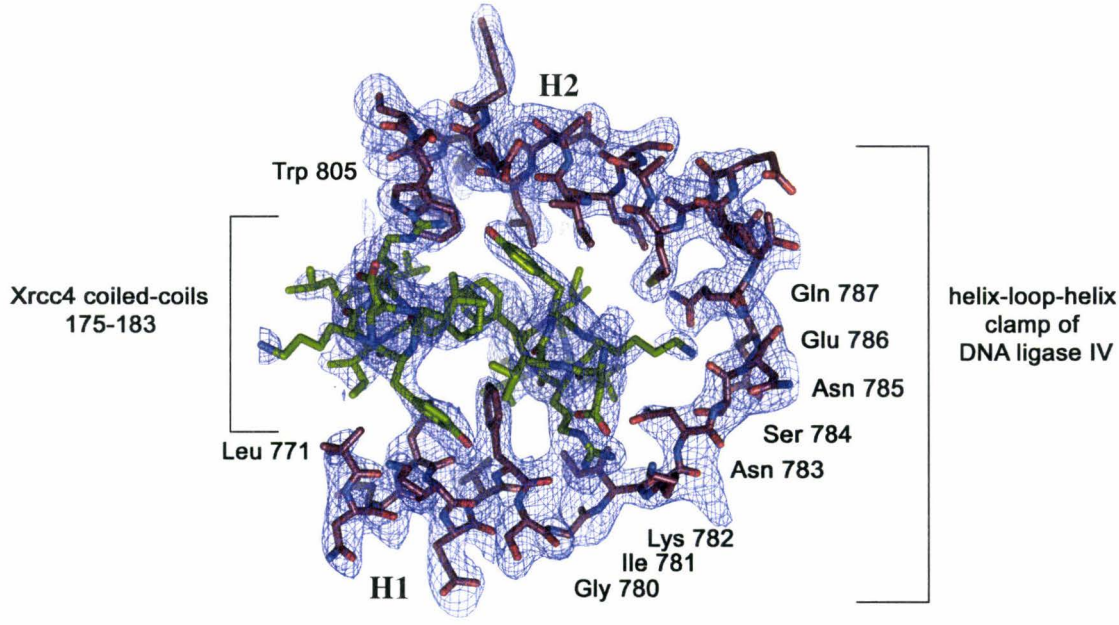

Figure 4.4 A top-view looking down the helices of Xrcc4 at the interface between the helix-loop-helix clamp of DNA ligase IV and the coiled-coil chains of Xrcc4 dimer and its surrounding 2Fo-Fc electron density map (3 σ contour level). The DNA ligase IV clamp spans residues 771 to 805 and the Xrcc4 coils span residues 175-183. **The program COOT was used to generate the electron density map (Emsley *et al.*, 2004).**

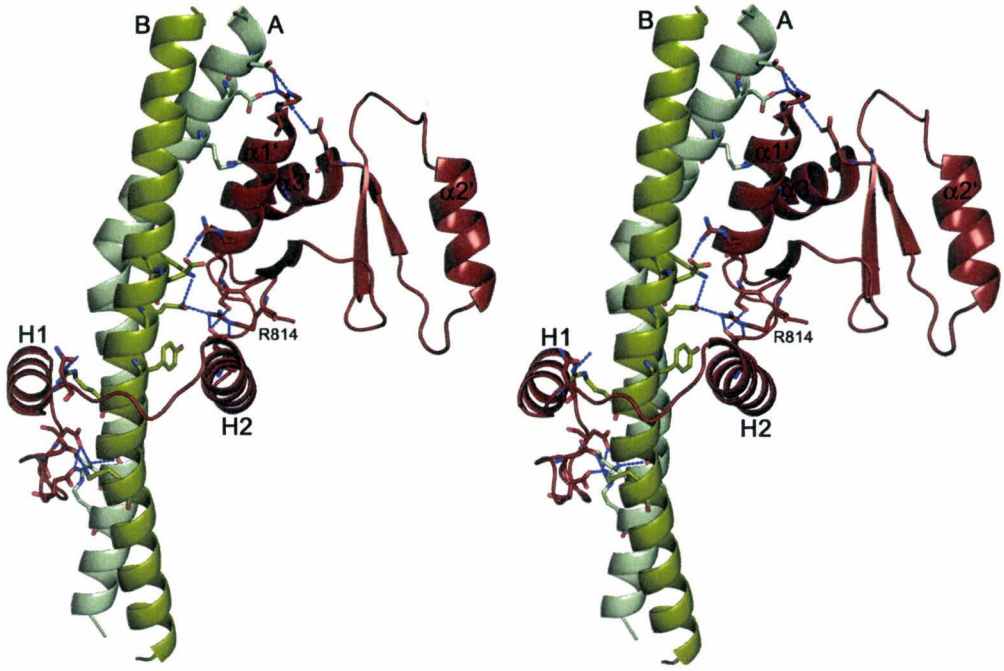

Figure 4.5 A cross-eyed stereo view of the key hydrogen bonding pattern of the complex. The residue Arg 814, which is involved in a network of hydrogen bonds, is labelled as also are the helices of DNA ligase IV. **PyMol software was used to generate the pictures (De Lano *et al.*, 2002).**

4.3 Comparison of the Surface Electrostatics of Xrcc4-DNA ligase IV Complex with Lif1p-Lig4p Complex; Characterization of the DNA Binding Site of Xrcc4-DNA ligase IV Complex and its Relevance to R814X Mutation

From examination of the amino acid sequences of human and yeast DNA ligase IV C-terminal domains (Fig. 4.7), it is apparent that there are considerable differences in the positions of charged amino acids that do not necessarily alter the structure significantly but nonetheless have a monumental impact on the electrostatic surface potential. These residues may in turn influence the interactions with DNA and other proteins involved in NHEJ. Although both human and yeast structures (Dore *et al.*, 2006) demonstrate a positively charged cleft in Xrcc4 below the second BRCT domain, the electrostatic features of the BRCT domains themselves are quite different especially in particular regions (Fig. 4.6 A, B, C and D). Interestingly, there is a substantial positively charged zone at the very C-terminal end of the second BRCT domain of human DNA ligase IV (Fig. 4.6 A), which remarkably lies on the same face as the positively charged cleft of Xrcc4 lying below the second BRCT. In addition, the base of each of the head domains of Xrcc4 carries a positively charged area, which apparently aligns with the above-described positive surface of the complex. Given that the positively charged zone of Xrcc4-DNA ligase IV complex is confined along one surface of the complex, it is conceivable to suggest the presence of a preferential DNA-binding site corresponding to that region.

This extensive positive zone at the C-terminal end of human DNA ligase IV is constituted by residues, Lys 876, Arg 879, Arg880, Lys 883, Arg 884, Lys 885 and Lys 887, all of which are in close proximity to each other and together generate a potential

binding surface for double stranded DNA. With the exception of K883, all the residues are conserved to a good extent through the course of evolution although only Lys 887 and Arg 879 are conserved to yeast (Fig. 4.7). In contrast to human DNA ligase IV, the C-terminal domain of yeast lig4p does not display extensive localized positive charge in any one particular region (Fig. 4.6 C and D). While a small amount of positive charge is noticed at the region corresponding to human DNA ligase IV C-terminal end, it is too distributed and interspersed with large negative charges making it a formidable DNA binding site. Notably, the positive residues in this region of lig4p, namely Arg 890 and K891 are quite apart from neighbouring positive residues such as Lys 902 and Arg 916. This could possibly indicate the presence of significantly lower interactions between the tandem BRCT domains of lig4p with DNA or a complete lack of involvement of the C-terminal domain of lig4p in DNA binding. Nevertheless, since the structure of lif1p-lig4p complex is missing an entire loop region corresponding to residues 840-858, the electrostatic properties observed are not conclusive and could alter considerably in the complete structure. Furthermore, there is also the possibility of the involvement of other, not so obvious regions within the C-terminal domain of lig4p that could potentially play a role in interacting with DNA.

From earlier DNA binding studies, we know that the DNA binding pattern of Xrcc4 follows a cooperative mode of binding requiring the polypeptide to encompass residues 1-28 as well as 168-200 for generating a stable protein-double stranded DNA complex (Modesti *et al.,* 1999). Furthermore, it is also known that the Xrcc4-double stranded DNA complex undergoes a supershift forming a ternary complex by the C-

terminal domain of DNA ligase IV (Modesti *et al.,* 1999). Based on this information coupled with the surface electrostatic properties of the complex, the structure was carefully analyzed to identify residues potentially significant for binding DNA. Although several positively charged conserved residues constitute the positive cleft of Xrcc4, all of them demonstrate electrostatic interactions with DNA ligase IV in either one of the two chains of the homodimer excluding K190 and K197, thus strongly indicating their role in DNA binding. Likewise, the positive residues K876, K879, K880 and K887 that form an integral part of the localized positively charged zone at the C-terminal end of DNA ligase IV are solvent exposed and available for interaction, indicating their possible involvement in binding DNA. Importantly, multiple sequence alignments of these regions in Xrcc4 and DNA ligase IV (Fig. 4.7 and Fig. 4.8) performed using ClustalX (Thompson *et al.,* 1997) program based on the results obtained from PSI-BLAST[2], demonstrated that all of the above listed amino acids are all highly conserved surface residues not engaged in interactions associated with tertiary and quaternary structure formation

Given that R814X is one of the mutations that gives rise to LIG4 syndrome, it is quite interesting to note that this deletion in human DNA ligase IV causes loss of the entire DNA binding surface of the C-terminal domain of DNA ligase IV. The elimination of this DNA binding surface of ligase IV possibly causes the protein complex to associate with DNA in an improper conformation that does not facilitate efficient ligation. Furthermore, since the deletion also causes loss of helices α1' and α3', which

[2]**http://spiral.genes.nig.ac.jp/homology/psi_blast-e.shtml**

interact with Xrcc4, the stability of the protein complex could also be partly compromised. The absence of a DNA binding surface on the tandem BRCT domains of DNA ligase IV along with the partial loss of Xrcc4 interaction interface could together be responsible for the substantial reduction in the adenylation/ligation activity of the enzyme causing LIG4 syndrome in patients homozygous for this mutation.

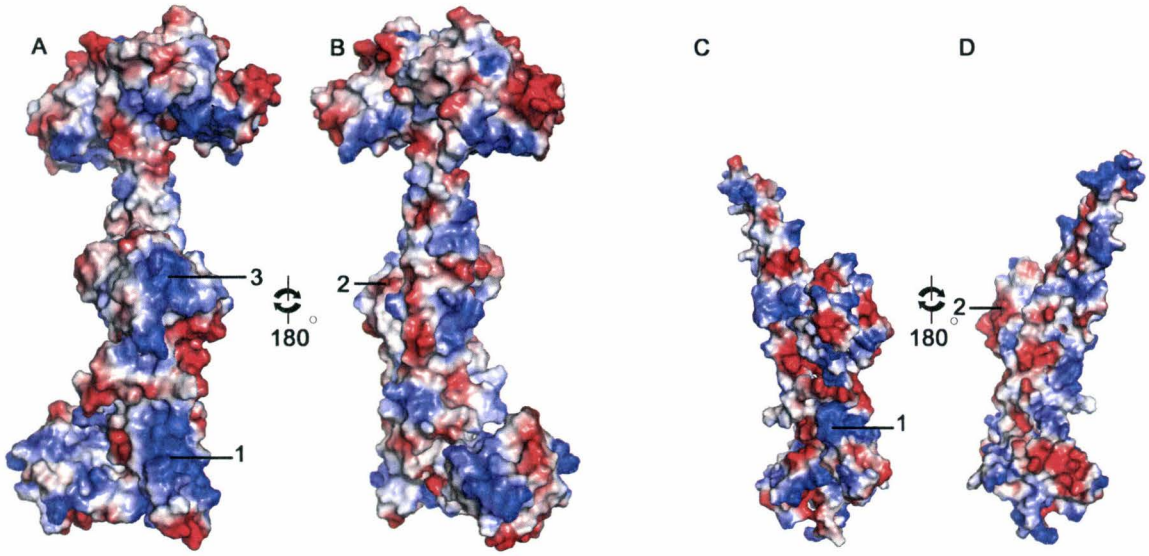

Figure 4.6 Surface electrostatic potential distribution. (A) and (B) represent the electrostatic potential distribution on the surface of Xrcc4-DNA ligase IV complex in two views that differ by 180° rotation along Y-axis. (C) and (D) represent the electrostatic potential distribution on the surface of lif1p-lig4p complex (excluding the heads of lif1p) in two corresponding views following the same orientation for tandem BRCT domains as Xrcc4-DNA ligase IV complex in (A) and (B) respectively. 1 corresponds to the positively charged cleft of Xrcc4/lif1p and 2 corresponds to the common negative region in tandem BRCT domains whereas 3 corresponds to the distinct positive charge at the C-terminal end of tandem BRCT domains of DNA ligase IV. Figures were obtained using the default contour levels of PyMol. **PyMol software was used to generate the pictures (De Lano *et al.*, 2002).**

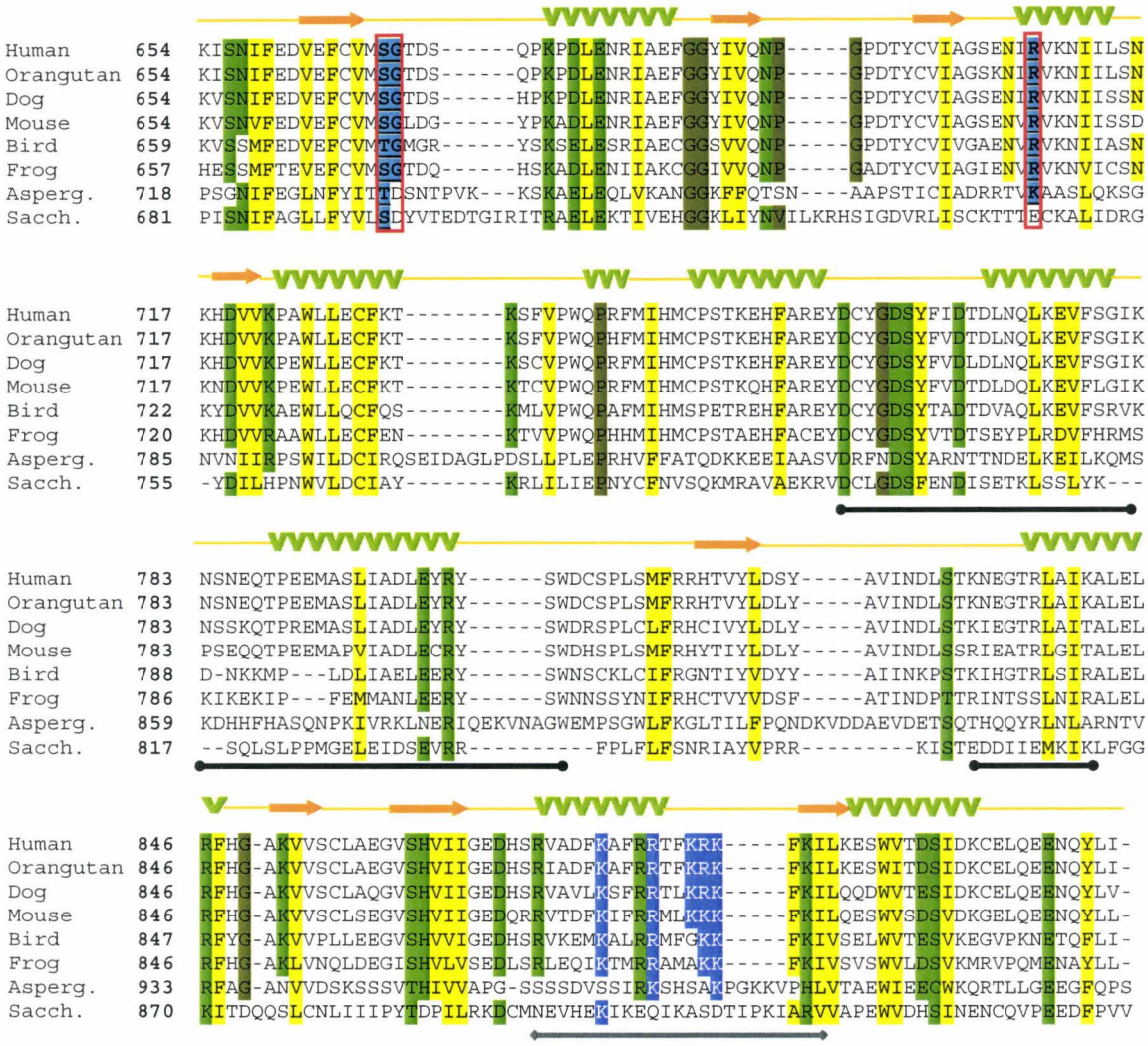

Figure 4.7 Evolutionary conservation of amino acids in DNA ligase IV C-terminal domain. Multiple sequence alignment of DNA ligase IV homologs of human, orangutan, dog, mouse, bird, frog and Saccharomyces cerevisiae (Sacch.). The related gene in Aspergillus (Asperg.) is based on sequence similarity. Secondary structures in human DNA ligase IV are indicated in orange arrows (β-strands) and green ribbons (α-helices). Regions involved in interactions with Xrcc4 are underlined with black lines ending in circles, those important for DNA binding with a grey line ending in diamonds. Conserved sequences are highlighted in yellow for hydrophobic residues significant for tertiary and quaternary structure formation, green for polar and charged residues making structurally critical hydrogen bonds and salt bridges, dark yellow for structurally essential glycines and prolines, blue for positively charged residues that might interact with DNA and cyan for residues potentially involved in phosphopeptide binding. Residues responsible for interaction with phosphopeptide are also underlined, shown in bold font and boxed in red. **The alignment was performed using ClustalX (Thompson *et al.*, 1997).**

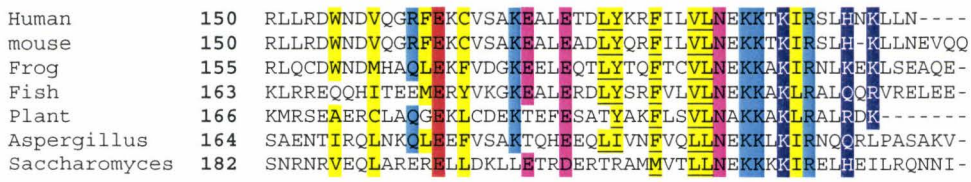

Figure 4.8 Multiple sequence alignment of Xrcc4 in homologs of human, mouse, frog, fish, plant, Aspergillus and Saccharomyces. Conserved hydrophobic residues interacting with DNA ligase IV in any one of the chains of Xrcc4 homodimer are highlighted in yellow whereas those interacting in both chains are shown in yellow and underlined. Positively charged conserved residues not interacting with DNA ligase IV are indicated in dark blue and those interacting with DNA ligase IV in one chain of Xrcc4 are indicated in cyan. Negatively charged residues interacting with DNA ligase IV in one of the chains of Xrcc4 are shown in magenta and those not interacting are shown in red. **The alignment was performed using ClustalX (Thompson *et al.*, 1997).**

4.4 Heterodimerization of Xrcc4 with XLF Retains the DNA Ligase IV Binding Interface

Several recent genetic and biochemical studies have identified a novel component of the NHEJ repair pathway called Cernunnos/XLF ((Ahnesorg *et al.,* 2006) (Buck *et al.,* 2006)). Based on sequence-structure analysis, it has been suggested that Cernunnos/XLF exhibits significant structural similarities to Xrcc4. Furthermore, a series of elegant experiments have demonstrated interactions between XLF and Xrcc4-DNA ligase IV complex ((Ahnesorg *et al.,* 2006) (Callebaut *et al.,* 2006)). Based on this premise, a model was proposed, which involved heterodimerization of Xrcc4 and XLF. With our available structure of Xrcc4-DNA ligase IV complex, in our attempt to verify this hypothesis, we modelled the corresponding region of XLF (178 - 220) homologous to the DNA ligase IV-binding-region of Xrcc4 (159-201) by replacing one of the protomers of Xrcc4 in our structure (Fig. 4.9). This modelling was based on recently published bioinformatics analysis and sequence alignment data of XLF with Xrcc4 (Callebaut *et al.,* 2006). Strikingly, we noticed that in spite of the low level of sequence similarity in this region, the newly constructed heterodimeric model still retained significant interactions with DNA ligase IV. It is interesting to note that there is asymmetry introduced in the Xrcc4-XLF heterodimeric model and only chain B of Xrcc4 can be replaced to model XLF while still continuing to maintain interactions with DNA ligase IV. The BRCT domains remain primarily anchored to XLF through a set of interactions, which are both hydrophilic and hydrophobic in nature. Beginning at the N-terminal end of XLF in the model, it is apparent that Q159 of XLF could potentially interact with E834 of DNA

ligase IV. Moving toward the C-terminus, A163, I166 of XLF and L843, F847 of DNA ligase IV are directed into the central interface engendering strong hydrophobic associations (Fig. 4.9). A second-set of potential electrostatic interactions is displayed between R169, E173 of XLF and L810 and R814 of DNA ligase IV. The side-chain of R169 electrostatically interacts with the carbonyl group of L810, which is oriented to position its carbonyl group towards R169, whereas A173 and R814 exhibit compelling interactions between their side-chains. Finally, the C-terminal end of XLF in the model presents a substantial hydrophobic patch comprised of residues F180, L181, F184 and M185 (Fig. 4.9). This non-polar surface facilitates key interactions with Xrcc4 as well as the second alpha helix of the helix-loop-helix clamp of DNA ligase IV.

Although the buried contacts within Xrcc4 homodimer interface are quite strong, it is still possible to envision an Xrcc4-XLF heterodimeric complex *in vivo* based on the interactions observed in the model. The above-described monumental interactions suggest a high propensity for the existence of such a complex *in vivo*. Furthermore, the positively charged residues at the C-terminal end (190-201) of XLF are freely available for interaction and assist in maintaining the DNA binding pocket of the complex. Taken together, this analysis strongly supports the previously postulated model involving combinatorial associations between Xrcc4 and XLF and their interaction with DNA ligase IV.

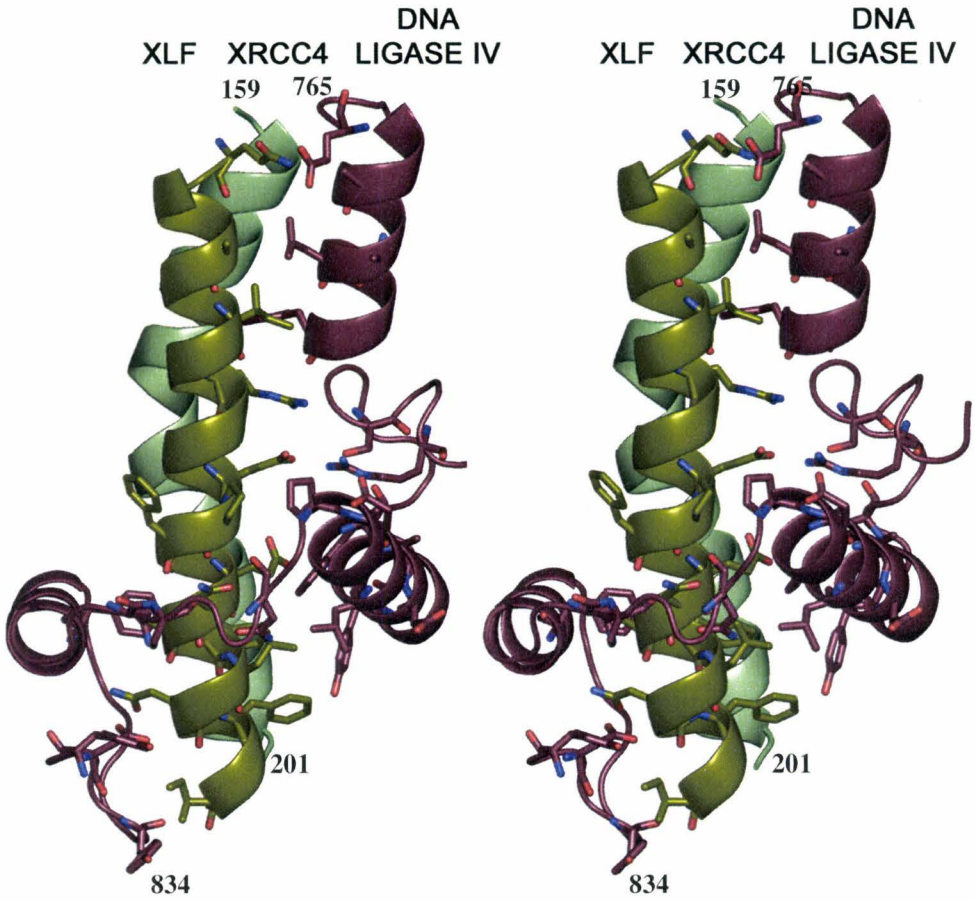

Figure 4.9 A cross-eyed stereo image showing interactions observed in Xrcc4-XLF heterodimeric model with the C-terminal domain of DNA ligase IV. Xrcc4 is coloured in pale green. XLF is coloured in olive green. DNA ligase IV is shown in raspberry red. The figure depicts the interactions between modelled XLF chain and portions of C-terminal domain of DNA ligase IV interacting with Xrcc4-XLF heterodimer. The Xrcc4/XLF chains span residues 159-201 while DNA ligase IV spans residues 765-834. **PyMol software was used to generate the pictures (De Lano *et al.*, 2002).**

4.5 Tandem BRCT Domains of DNA Ligase IV

To date, crystal structures of several tandem BRCT domains with short intervening linkers have been reported ((Williams *et al.,* 2001) (Derbyshire *et al.,* 2002) (Lee *et al.,* 2005)). All these structures are characterized by a close head-to-tail interaction between the two domains resulting in the formation of a pocket capable of binding phosphopeptides. Interestingly, the structure of Xrcc4-DNA ligase IV complex discloses a different mode of interaction where the long linker between the two BRCT domains almost encircles the coiled-coil tails of Xrcc4 and positions the two BRCT domains far apart on either side and at a 45° angle relative to Xrcc4 tails. Although tandem BRCT repeats with long linkers could represent a separate group that interact differently with their partners, structural superimposition of these domains individually on BRCA1 tandem BRCT domains (Williams *et al.,* 2001) (which were placed by superimposing their C-terminal BRCT domain on the corresponding domain of DNA ligase IV in the structure of the complex) after removing a few loops (comprised of residues 654-661, 671-672, 704, 745-747, 755-815, 822-836, 858-860, 867-870, 879 and 885 of DNA ligase IV and the corresponding residues in BRCA1) reveals significant similarity: the Cα atoms superimpose with an rmsd of 1.38A° (Fig. 4.10) and the Cα atoms of only the residues involved in phosphopeptide binding superimpose with an rmsd of 1.43 Å. Strikingly, structural superimposition also highlights the importance of helix α2 and demonstrates its displacement by helix A upon interaction with Xrcc4. Furthermore, the conserved residues of tandem BRCT domains that are responsible for interacting with phosphopeptides remain conserved in human DNA ligase IV (Fig. 4.7)

and it has been shown that human DNA ligase IV demonstrates a preference and specificity to bind to phosphoserine-containing phosphopeptides (Rodriguez *et al.,* 2003). Given that phosphopeptide binding by tandem BRCT domains requires the domains to stay in close proximity, it is conceivable to propose a model involving two different forms of DNA ligase IV that exhibit two different modes of interaction depending on the binding partner. Consistent with the proposed model, DNA ligase IV has been found to remain associated with Xrcc4 only during interphase but not throughout the cell cycle (Przewloka *et al.,* 2003).

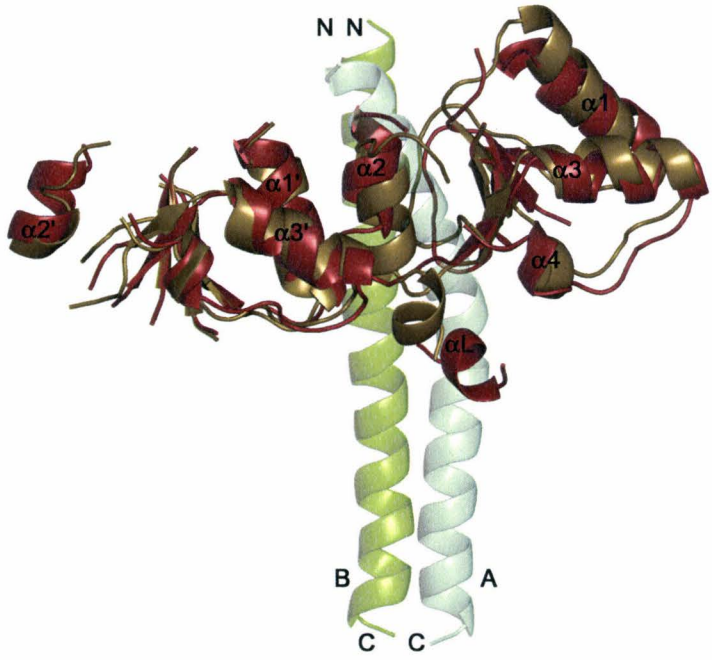

Figure 4.10 Structural superimposition of tandem BRCT domains of DNA ligase IV individually on BRCA1 tandem BRCT domains. The BRCA1 BRCT domains were positioned by superimposing their C-terminal BRCT domain on the corresponding domain of DNA ligase IV in the structure of the complex. Xrcc4 chains A and B are shown in pale green and split pea green respectively. Tandem BRCT domains of DNA ligase IV are coloured in brick red while those of BRCA1 are coloured in sand brown. **PyMol software was used to generate the pictures (De Lano *et al.,* 2002).**

4.6 Summary and Future Work

4.6.1 Summary

Although the unmutated complex of Xrcc4 (1-203) with DNA ligase IV (654-911) crystallized readily, the crystals obtained demonstrated issues associated with twinning. Generation of a new crystal form of the complex required methodical surface mutagenesis coupled with subsequent purification and screening for crystals. The mutant A60E of Xrcc4 bound to DNA ligase IV produced two different new crystal forms, of which, the second crystal form was selected for further optimization and structure determination owing to the poor diffraction of the first crystal form. Determination of the best crystallization condition involved a combination of three optimization techniques. Single crystals of the complex were grown by microseeding in 100mM sodium/potassium phosphate, 15% PEG 8000, 200mM sodium chloride and 20% NDSB-195 at 20°C.

Native crystals thus obtained diffracted X-rays to 2.29 Å resolution at the X8C beamline at BNL-NSLS facility where a complete data set was collected. The data set was processed in Rigaku Crystal Clear software. The data processing statistics, mainly the R-merge and residual, were used as a guide to determine the correct space group of the complex as P1 with unit cell dimensions of a=67.51; b=86.11; c=111.73 and α=67.42; β=82.64; γ=74.34. Although a real molecular replacement solution was obtained with the native data set, the phase information derived from it was very weak and could not be employed for model building and structure refinement.

Subsequently, selenomethionine-derived crystals were grown under the same conditions, which diffracted to 2.4 Å resolution at the X8C beamline at BNL-NSLS

facility. The MAD data set collected at peak, inflection and low energies was processed using Crystal Clear software in space group P1 with the same unit cell dimensions. The data set collected at peak energy was utilized for locating the selenium sites and performing subsequent SAD phasing and density modification. The experimental electron density map generated from density modification by solvent flipping was continuous and showed distinct α-helices. The structure of the complex was built through a series of steps using one of the solved structures of Xrcc4 as the initial model positioned approximately into the electron density.

The crystal structure of Xrcc4-DNA ligase IV complex is shown to contain two 2:1 complexes in the unit cell. The structure displays the extensive interaction interface between the proteins, which is significantly larger than the previously predicted interface based on deletion analysis ((Grawunder *et al.,* 1998) (Sibanda *et al.,* 2001)). Furthermore, the structure provides highly valuable information regarding the electrostatic surface potential distribution of the complex, which appears to be quite different from the electrostatic potential of the yeast lif1p-lig4p complex. Based on this critical analysis, a potential dsDNA binding surface of the complex was identified. The structure was used as a basis to study the associations of a recently identified factor of NHEJ, called Cernunnos/XLF with DNA ligase IV. Employing the current structure of the complex, an Xrcc4-XLF heterodimeric model was constructed, which strongly supports the recently proposed hypothesis involving heterodimerization of Xrcc4 and XLF. In addition to this, the structure forms an essential tool for understanding and analyzing various mutations in DNA ligase IV C-terminal domain that are responsible for

giving rise to LIG4 syndrome phenotype observed in these patients. Finally, the structure also provides a means of analyzing NHEJ through the course of evolution by comparing it with the structure of yeast lif1p-lig4p complex.

4.6.2 Future Work and Concluding Remarks

Given that DNA ligase IV demonstrates a preferential binding to phosphoserine-containing phosphopeptides, it would be quite interesting to gather structural information of the phosphopeptide-bound complex of the protein. Examination of the different domains present in this protein show that the phosphopeptide binding functionality can most likely be attributed to the tandem BRCT domains of the protein. From the current structure of Xrcc4-DNA ligase IV complex, we know that the linker between the tandem BRCT domains exhibits a morphology similar to a clamp and places the two BRCT domains far apart from each other. Since phosphopeptide binding requires the tandem BRCT domains to come into close proximity of each other, the structure of tandem BRCT domains of DNA ligase IV bound to a phosphopeptide would reveal any conformational changes taking place within the C-terminal domain of DNA ligase IV and its functional relevance to either NHEJ or other processes occurring within the cell.

Determining the crystal structure of the full-length Xrcc4-DNA ligase IV complex bound to DNA would be a major milestone in comprehending the mechanism of NHEJ repair pathway. The structural data gleaned from this complex would provide the framework for understanding the molecular mechanism by which the complex coordinates ligation of the broken DNA ends with other preceding events occurring within the pathway. In addition to providing better insights into the molecular mechanism of ligation, the structure would also demonstrate the significance of various motifs within the protein that are key for proper functioning of the enzyme. To date, a number of hypomorphic mutations in several motifs of DNA ligase IV have been

identified in patients and their understanding is currently based on the crystal structures of other DNA ligases, which are incapable of participating in NHEJ. Characterizing the structure of the full-length complex would provide extremely significant information for performing an in-depth analysis of these mutations. Since the activity of this complex is essential for NHEJ, inhibitors targeting the complex would also block the repair pathway and therefore can be used in conjunction with other cancer therapies for the treatment of cancer. The detailed structural knowledge gained would form the basis for rational design of key inhibitors to the complex.

Finally, determining the structure of Cernunnos/XLF and identifying its precise role in NHEJ would prove extremely useful in unravelling the key steps involved in NHEJ. As Cernunnos/XLF is known to interact with Xrcc4, the structure of the protein either on its own or along with Xrcc4 would furnish significant information on different aspects of its behaviour. Given that the specific function of this protein in NHEJ remains more or less ambiguous, the structural knowledge procured from it would enlighten the scientific community on the unknown aspects of NHEJ pathway.

APPENDICES

APPENDIX I

Purification of TEV Protease

AI.1 Overexpression Protocol

The TEV protease expression plasmid, pRK793 containing His-TEV(S219V)-Arg was transformed into BL21-RIL cells and the cells were grown at 37°C in LB containing 100μg/mL of Amp and 30 μg/mL of Cm. When the cells reached the mid-log phase (OD_{600} ~ 0.5), IPTG was added to a final concentration of 1mM and the temperature was reduced to 30°C. After 4 hours of induction, the cells were harvested and frozen in liquid nitrogen.

AI.2 Purification Protocol

Cells were re-suspended in a lysis buffer containing 50mM phosphate pH 8.0, 100mM sodium chloride, 10% glycerol and 25mM imidazole such that 10mL of buffer was used per 1gram of wet weight cells. Subsequently, protease inhibitors were added and cells were lysed by four passages through a French Pressure cell at 12,000 psi. The lysate volume was then determined and 5% of polyetheleneimine was added to a final concentration of 0.1%. The lysate was then mixed by inversion and centrifuged at 15,000Xg for 30 minutes. The supernatant was then applied on to a nickel column equilibrated with lysis buffer. The column was washed with seven column volumes of lysis buffer, and then the TEV protease protein was eluted using a 10 column volume

gradient to 50mM phosphate pH 8.0, 100mM sodium chloride, 10% glycerol and 200mM imidazole. Appropriate fractions were examined on an SDS-PAGE gel and pooled. EDTA and DTT were then added to the sample to a final concentration of 1mM. Subsequently, the sample was concentrated and loaded on to a SP-sepharose column in 25mM phosphate pH 8.0, 200mM sodium chloride, 10% glycerol, 2mM EDTA and 10mM DTT using 3% of the column volume. The eluted fractions were then examined on an SDS-PAGE gel and pooled appropriately. The protein sample was concentrated to 1.0mg/mL and frozen at -80°C in 50% glycerol.

APPENDIX II

Purification of Pfu Polymerase

AII.1 Overexpression Protocol

The recombinant pfu-pET15b expression plasmid was transformed into BL21(DE3) pLysS cells and the cells were grown at 37°C until an OD_{600} of 0.5, upon which they were induced with IPTG to a final concentration of 1mM. Following induction, the cells were continued to grow for 3 hours at the same temperature and then harvested and frozen in liquid nitrogen.

AII.2 Purification Protocol

A 2L cell pellet was re-suspended in 70mL of lysis buffer containing 20mM Tris 8.5, 0.03% LDAO, 500mM KCl, 10% glycerol and 10mM imidazole. Subsequently,

protease inhibitors were added and the cells were lysed by four passages through a French Pressure cell at 12,000 psi. The lysate was then centrifuged at 15,000Xg for 40 minutes. The supernatant obtained was then heated at 75°C for 15 minutes, mixed by inversion and placed on ice for 15 minutes. Following this, the sample was centrifuged at 15,000Xg for 20 minutes and the supernatant with added protease inhibitors was applied on to a Nickel column equilibrated with lysis buffer. The column was washed with four column volumes of lysis buffer, and the protein was eluted with a 18 column volume gradient to 20mM Tris 8.5, 0.03% LDAO, 500mM KCl, 10% glycerol and 300mM imidazole. The eluted sample was diluted to lower the salt concentration and loaded on to a SP-sepharose column in 50mM MES pH 5.6, 1mM EDTA and 10mM DTT. A gradient was applied from 100mM KCl to 500mM KCl over 150mL of volume to elute the protein sample. The eluted fractions were then examined on an SDS-PAGE gel, pooled appropriately, concentrated to ~ 0.175mg/mL and frozen at -20°C in glycerol.

REFERENCES:

Ahnesorg, P., Smith, P., Jackson, S.P. XLF interacts with the XRCC4-DNA ligase IV complex to promote DNA nonhomologous end-joining. *Cell.* 124, 301-313 (2006).

Aravind, L., Koonin, E.V. Prokaryotic homologous of the eukaryotic DNA-end binding protein Ku, novel domains in the Ku protein and prediction of a prokaryotic double-strand break repair system, *Genome Res.* 11, 1365-1374 (2001).

Badie, C., Goodhardt, M., Waugh, A., Doyen, N., Foray, N., Calsou, P., Singleton, B., Gell, D., Salles, B., Jeggo, P., Arlett, C.F., Malaise, E.P. A DNA double-strand break defective fibroblast cell line (180BR) derived from a radiosensitive patient represents a new mutant phenotype. *Cancer Res.* 57, 4600-4607 (1997).

Barnes D.E., Stamp, G., Rosewell, I., Denzel, A. & Lindhal, T. Targeted disruption of the gene encoding DNA ligase IV leads to lethality in embryonic mice. *Curr. Biol.* 8, 1395-1398 (1998).

Bliss, T.M., Lane, D.P. Ku selectively transfers between DNA molecules with homologous ends. *J. Biol. Chem.* 272, 5765-5773 (1997).

Block, W.D., Yu, Y., Merkle, D., Gifford, J.L., Ding, Q., Meek, K., Lees-Miller, S.P. Autophosphorylation-dependent remodeling of the DNA-dependent protein kinase catalytic subunit regulates ligation of DNA ends. *Nucl. Acids Res.* 14, 4351-4357 (2004).

Blow, D.M., Chayen, N.E., Lloyd, L.F., Saridaki, S.E. Control of nucleation of protein crystals. *Protein Science.* 3, 1638-1643 (1994).

Boskovic, J., Rivera-Calzada, A., Maman1, J.D., Chaco`n, P., Willison, K.R., Pearl, L.H., Llorca, O. Visualization of DNA-induced conformational changes in the DNA repair kinase DNA-PKcs. *The EMBO Journal.* 22, 5875-5882 (2003).

Bradford, M.M. A rapid and sensitive method for the quantitation of microgram quality of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* 72, 248-254 (1976).

Brewerton, S.C., Dore, A.S., Drake, A.C., Leuther, K.K., Blundell, T.L. Structural analysis of DNA-PKcs: modeling of the repeat units and insights into the detailed molecular architecture. *J Struct. Biol.* 145, 295-306 (2004).

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., Read, R.J., Rice, L.M., Simonson, T., Warren, G.L. Crystallography & NMR System: A New Software Suite for Macromolecular Structure Determination. *Acta Cryst.* D54, 905-921 (1998).

Buck, D., Malivert, L., de Chasseval, R., Barraud, A., Fondaneche, M.C., Sanal, O., Plebani, A., Stephan, J.L., Hufnagel, M., le Deist, F., Fischer, A., Durandy, A., de Villartay, J.P., Revy, P. Cernunnos, a novel nonhomologous end-joining factor, is mutated in human immunodeficiency with microcephaly. *Cell.* 124, 260-262 (2006).

Callebaut, I., Malivert, L., Fischer, A., Mornon, J.P., Revy, P., de Villartay, J.P. Cernunnos interacts with the XRCC4/DNA-ligase IV complex and is homologous to the yeast nonhomologous end-joining factor NEJ1. *J Biol Chem.* 281, 13857-13860 (2006).

Cary, R.B., Peterson, S.R., Wang, J., Bear, D.G., Bradbury, E.M., Chen, D.J. DNA looping by Ku and the DNA dependent protein kinase. *Proc. Natl. Acad. Sci.* 94, 4267-4272 (1997).

Chan, D.W., Chen, B.P., Prithivirajsingh, S., Kurimasa, A., Story, M.D., Qin, J., Chen, D.J. Autophosphorylation of the DNA-dependent protein kinase catalytic subunit is required for rejoining of DNA double-strand breaks. *Genes Dev.* 16, 2333-2338 (2002).

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

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

Chen, L., Trujillo, K., Ramos, W., Sung, P., Tomkinson, A.E. Promotion of Dnl4-catalyzed DNA end joining by the Rad50/Mre11/Xrs2 and Hdf1/Hdf2 complexes. *Mol. Cell.* 8, 1105-1115 (2001).

Collaborative Computational Project, Number 4. The CCP4 Suite: Programs for Protein Crystallography. *Acta Cryst.* D50, 760-763 (1994).

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

CrystalClear: An integrated program for the collection and processing of area detector data. Rigaku corporation. (1997-2002).

Cudney, R., Patel, S., Weisgraber, K., Newhouse, Y., McPherson, A. Screening and optimization of strategies for macromolecular crystal growth. *Acta Cryst.* D50, 414-423 (1994).

Daley, J.M., Laan, R.L., Suresh, A., Wilson, T.E. DNA joint dependence of pol X family polymerase action in nonhomologous end joining. *J Biol. Chem.* 280, 29030-29037 (2005).

DeLano, W.L. The PyMOL molecular graphics system. DeLano Scientific, San Carlos, CA, USA. (2002).

Derbyshire, D.J., Basu, B.P., Serpell, L.C., Joo, W.S., Date, T., Iwabuchi, K., Doherty, A.J. Crystal structure of human 53BP1 BRCT domains bound to p53 tumour suppressor. *EMBO J.* 21, 3863-3872 (2002).

Ding, Q., Reddy, Y.V., Wang, W., Woods, T., Douglas, P., Ramsden, D.A., Lees-Miller, S.P., Meek, K. 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 (2003).

Dinner, A.R., Blackburn, G.M., Karplus, M. Uracil-DNA glycosylase acts by substrate autocatalysis. *Nature.* 413, 752-755 (2001).

Doherty, A.J., Jackson, S.P., Weller, G.R. Identification of bacterial homologues of the Ku DNA repair proteins, *FEBS Lett.* 500, 186-188 (2001).

Dore, A.S., Furnham, N., Davies, O.R., Sibanda, B.L., Chirgadze, D.Y., Jackson, S.P., Pellegrini, L., Blundell, T.L. Structure of an Xrcc4-DNA ligase IV yeast ortholog complex reveals a novel BRCT interaction mode. *DNA Rep.* 3, 362-368 (2006).

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

Emsley, P., Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr.* D60, 2126-2132 (2004).

Enders, A., Fisch, P., Schwarz, K., Duffner, U., Pannicke, U., Nikolopoulos, E., Peters, A., Orlowska-Volk, M., Schindler, D., Friedrich, W., Selle, B., Niemeyer, C., Ehl, S., A severe form of human combined immunodeficiency due to mutations in DNA Ligase IV *The Journal of Immunology.* 176, 5060-5068 (2006).

Frank, K.M., Sharpless, N.E., Gao, Y., Sekiguchi, J.M., Ferguson, D.O., Zhu, C., Manis, J.P., Horner, J., DePinho, R. A., Alt, F.W. DNA ligase IV deficiency in mice leads to defective neurogenesis and embryonic lethality via the p53 pathway. *Mol. Cell.* 5, 993-1002 (2000).

Gao, Y., Ferguson, D.O., Xie, W., Manis, J.P., Sekiguchi, J., Frank, K.M., Vhaudhuri, J., Horner, J., DePinho, R.A., Alt, F.W. Interplay of p53 and DNA-repair protein XRCC4 in tumorigenesis, genomic stability and development. *Nature.* 404, 897-900 (2000).

Gao, Y., Sun, Y., Frank, K.M., Dikkes, P., Fujiwara, Y., Seidl, K.J., Sekiguchi, J.M., Rathbun, G.A., Swat, W., Wang, J., Bronson, R.T., Malynn, B.A., Bryans, M., Zhu, C., Chaudhuri, J., Davison, L., Ferrini, R., Stamato, T., Orkin, S.H., Greenberg, M.E., Alt, F.W. A critical role for DNA end-joining proteins in both lymphogenesis and neurogenesis. *Cell.* 95, 891-902 (1998).

Girard, P. M., Kysela, B., Harer C, J., Doherty A. J., Jeggo, P. A. Analysis of DNA ligase IV mutations found in LIG4 syndrome patients: the impact of two linked polymorphisms. *Hum Mol Genet.* 13, 2369-2376 (2004).

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

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

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

Herrmann, G., Lindahl, T., Schar, P. Saccharomyces cerevisiae LIF1: a function involved in DNA double-strand break repair related to mammalian XRCC4. *EMBO J.* 17, 4188-4198 (1998).

Jancarik, J., Kim, S.H. Sparse matrix sampling: a screening method for crystallization of proteins. *J. Appl. Cryst.* 24, 409-411 (1991).

Jeggo, P., O'Neill, P. The Greek Goddess, Artemis, reveals the secrets of her cleavage. *DNA Rep.* 1, 771-777 (2002).

Jones, T.A., Zou, J.Y., Cowan, S.W., Kjeldgaard, M. Improved methods for the building of protein models in electron density maps and the location of errors in these models. *Acta Cryst.* A47, 110-119 (1991).

Jovine, L. A simple technique to control macromolecular crystal nucleation efficiently using a standard vapor-diffusion setup. *J. Appl. Cryst.* 33, 988-989 (2000).

Junop, M.S., Modesti, M., Guarne, A., Ghirlando, R., Gellert, M., Yang, W. Crystal structure of the Xrcc4 DNA repair protein and implications for end joining. *EMBO J.* 19, 5962-5970 (2000).

Kao, J., Rosenstein, B.S., Peters, S., Milano, M.T., Kron, S.J. Cellular Response to DNA Damage. *Ann. N.Y. Acad. Sci.* 1066, 243-258 (2005).

Kegel, A., Sjostrand, J.O., Astrom, S.U. Nej1p, a cell type-specific regulator of nonhomologous end joining in yeast. *Curr. Biol.* 11, 1611-1617 (2001).

Kysela, B., Chovanec, M., Jeggo, P.A. Phosphorylation of linker histones by DNA-dependent protein kinase is required for DNA ligase IV-dependent ligation in the presence of histone H1. *Proc. Natl. Acad. Sci.* 102, 1877-1882 (2005).

Laskowski, R.A., MacArthur, M.W., Moss, D.S., Thornton, J.M. PROCHECK: a program to check the stereochemistry of protein structures. *J. Appl. Cryst.* 26, 283-291 (1993).

Lee, M.S., Edwards, R.A., Thede, G.L., Glover, J.N. Structure of the BRCT repeat domain of MDC1 and its specificity for the free COOH-terminal end of the gamma-H2AX histone tail. *J. Biol. Chem.* 280, 32053-32056 (2005)

Lees-Miller, S.P., Meek, K. Repair of DNA double strand breaks by non-homologous end joining. *Biochimie.* 85, 1161-1173 (2003).

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

Lin, F.L., Sperle, K., Sternberg, N. Intermolecular recombination between DNAs introduced into mouse L cells is mediated by a nonconservative pathway that leads to crossover products. *Mol. Cell Biol.* 10, 103-112 (1990).

Llorca, O., Pearl, L.H. Electron microscopy studies on DNA recognition by DNA-PK. *Micron.* 35, 625-633 (2004).

Ma, Y., Lu, H., Schwarz, K., Lieber, M.R. Repair of double-strand DNA breaks by the human nonhomologous DNA end joining pathway: the iterative processing model. *Cell Cycle.* 4, 1193-1200 (2005).

Matthews, B.W. The Solvent Content of Protein Crystals. *J. Mol. Biol.* 33, 491-497 (1968).

McPherson, A. Increasing the size of microcrystals by fine sampling of pH limits. *J. Appl. Cryst.* 28, 362-365 (1995).

McPherson, A. Preparation and analysis of protein crystals. J.Willey and Sons. New York (1982).

Merkle, D., Douglas, P., Moorhead, G.B., Leonenko, Z., Yu, Y., Cramb, D., Bazett-Jones, D.P., Lees-Miller, S.P. The DNA-dependent protein kinase interacts with DNA to form a protein-DNA complex that is disrupted by phosphorylation. *Biochemistry* 41, 12706-12714 (2002).

Mizuta, R., Cheng, H.L., Gao, Y., Alt, F.W. Molecular genetic characterization of XRCC4 function. *Int. Immunol.* 9, 1607-1613 (1997).

Modesti, M., Hesse, J.E., Gellert, M. DNA binding of Xrcc4 protein is associated with V(D)J recombination but not with stimulation of DNA ligase IV activity. *EMBO J.* 18, 2208-2018 (1999).

Modesti, M., Junop, M.S., Ghirlando, R., van de Rakt., M., Gellert, M., Yang, W., Kanaar, R. Tetramerization and DNA ligase IV interaction of the DNA double-strand break repair protein XRCC4 are mutually exclusive. *Journal of Mol. Biol.* 334, 215-228 (2003).

Murshudov, G.N., Lebedev, A., Vagin, A.A., Wilson, K.S., Dodson, E.J. Efficient anisotropic refinement of Macromolecular structures using FFT. *Acta Cryst.* D55, 247-255 (1999).

Murshudov, G.N., Vagin, A., Dodson, E. Application of maximum likelihood refinement in the refinement of protein structures. *Proceedings of Daresbury study weekend* (1996).

Murshudov, G.N., Vagin, A.A., Dodson, E.J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Cryst.* D53, 240-255 (1997).

Otwinowski, Z., Minor, W. Processing of x-ray diffraction data collected in oscillation mode. *Methods in Enzymology.* 276, 307-326, (1997).

Pannu, N.J., Murshudov, G.N., Dodson, E.J., Read, R.J. A incorporation of prior phase information strengthen maximum-likelihood structure refinement. *Acta Cryst.* D54, 1285-1294 (1998).

Pâques, F., Haber, J.E. Multiple pathways of recombination induced by double-strand breaks in Saccharomyces cerevisiae. *Microbiol. Mol. Biol. Rev.* 63, 349-404 (1999).

Paul Emsley and Kevin Cowtan; Coot: Model-building tools for molecular graphics. *Acta Cryst* D60, 2126-2132 (2004).

Paull, T.T., Gellert, M. A mechanistic basis for Mre11-directed DNA joining at microhomologies. *Proc. Natl. Acad. Sci.* 97, 6409-6414 (2000).

Perry, J., Kleckner, N. The ATRs, ATMs, and TORs are giant HEAT repeat proteins. *Cell.* 112, 151-155 (2003).

Potterton, E., Briggs, P., Turkenburg, M., Dodson, E. A graphical user interface to the CCP4 program suite *Acta. Cryst.* D59 1131-1137 (2003)

Przewloka, M.R., Pardington, P.E., Yannone, S.M., Chen, D.J., Cary, R.B. In vitro and in vivo interactions of DNA ligase IV with a subunit of the condensin complex. *Mol Biol Cell.* 2, 685-697 (2003).

Ramsden, D.A., Gellert, M. Ku protein stimulates DNA end joining by mammalian DNA ligases: a direct role for Ku in repair of DNA double-strand breaks. *EMBO J.* 17, 609-614 (1998).

Read, R.J., Pushing the boundaries of molecular replacement with maximum likelihood. *Acta Cryst.* D57, 1373-1382 (2001).

Reddy, Y.V., Ding, Q., Lees-Miller, S.P., Meek, K., Ramsden, D.A. Non-homologous end joining requires that the DNA-PK complex undergo an autophosphorylation dependent rearrangement at DNA ends. *J. Biol. Chem.* 279, 39408–39413 (2004).

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., Jeggo, P.A. Identification of a defect in DNA ligase IV in a radiosensitive leukaemia patient. *Curr. Biol.* 9, 699-702 (1999).

Riballo, E., Doherty, A.J., Dai, Y., Stiff, T., Oettinger, M.A., Jeggo, P.A., Kysela, B. Cellular and biochemical impact of a mutation in DNA ligase IV conferring clinical radiosensitivity. *J. Biol. Chem.* 276, 31124-31132 (2001).

Rivera-Calzada, A., Maman, J.D., Spagnolo, L., Pearl, L.H., Llorca, O. Three-dimensional structure and regulation of the DNA-dependent protein kinase catalytic subunit (DNA-PKcs). *Structure.* 13, 243-255 (2005).

Rodriguez, M., Yu, X., Chen, J., Songyang, Z. Phosphopeptide binding specificities of BRCA1 COOH-terminal (BRCT) domains. *J Biol. Chem.* 278, 52914-52918 (2003).

Shiozaki, E.N., Gu, L., Yan, N., Shi, Y. Structure of the BRCT repeats of BRCA1 bound to a BACH1 phosphopeptide: implications for signaling. *Mol. Cell.* 14, 405-412 (2004).

Sibanda, B., Critchlow, S., Begun, Jake., Pei, X., Jackson, S.P., Blundell, T., Pellegrini, L. Crystal structure of an Xrcc4-DNA ligase IV complex. *Nature Struct. Biol.* 8, 1015-1019 (2001).

Storoni, L.C., McCoy, A.J., Read, R.J. Likelihood-enhanced fast rotation functions. *Acta Cryst.* D60, 432-438 (2004).

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., Higgins, D. G. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 24 ,4876-4882 (1997).

Tomkinson, A.E., Mackey, Z.B. Structure and function of mammalian DNA ligases. *Mutat. Res.* 407, 1-9 (1998).

Tomkinson, A.E., Vijayakumar, S., Pascal, J.M., Ellenberger, T. DNA Ligases: structure, reaction mechanism, and function *Chem. Rev.* 106, 687-699 (2006).

van der Burg, M., van Veelen, L.R., Verkaik, N.S., Wiegant, W.W., Hartwig, N.G., Barendregt, B.H., Brugmans, L., Raams, A., Jaspers, N.G., Zdzienicka, M.Z., van Dongen, J.J., van Gent, D.C. A new type of radiosensitive T-B-NK severe combined immunodeficiency caused by a LIG4 mutation. *J. Clin. Invest.* 116, 137-145 (2006).

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

Wang, H., Rosidi, B., Perrault, R., Wang, M., Zhang, L., Windhofer, F., Iliakis, G. DNA ligase III as a candidate component of backup pathways of nonhomologous end joining. *Caner Res.* 65, 4020-4030 (2005).

Wang, H., Zeng, Z.C., Perrault, A.R., Cheng, X., Qin, W., Iliakis, G. Genetic evidence for the involvement of DNA ligase IV in the DNA-PK-dependent pathway of non-homologous end joining in mammalian cells. *Nucleic Acids Res.* 29, 1653–1660 (2001).

Williams, R.S., Green, R., Glover, J.N. Crystal structure of the BRCT repeat region from the breast cancer-associated protein BRCA1. *Nature Struct. Biol.* 8, 838-842 (2001).

Winn, M., Isupov, M., Murshudov, G.N. Use of TLS parameters to model anisotropic displacements in macromolecular refinement *Acta Cryst.* D57, 122-133 (2001).

Woo, R.A., McLure, K.G., Lees-Miller, S.P., Rancourt, D.E., Lee, P.W. DNA-dependent protein kinase acts upstream of p53 in response to DNA damage. *Nature.* 394, 700-704 (1998).

Yoo, S., Dynan, W.S. 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 (1999).

Zhou, B.B., Elledge, S.J. The DNA damage response: putting checkpoints in perspective. *Nature.* 408, 433-439 (2000).