Characterization of protein and DNA interactions of the human DNA repair protein XRCC1

Characterization of protein and DNA interactions of the human DNA repair protein XRCC1

By Mac Chun Yeung Mok, B.Sc.

A Thesis submitted to the School of Graduate Studies in partial fulfillment of the requirements for the Degree Doctor of Philosophy

McMaster University

© Copyright by Mac Chun Yeung Mok, June 22, 2016

Doctor of Philosophy (2016) McMaster University

Biochemistry and Biomedical Science Hamilton Ontario

Title: Characterization of protein and DNA interactions of the human DNA repair protein XRCC1

Author: Mac Chun Yeung Mok, B.Sc. (Lakehead University)

Supervisor: Dr. Murray S. Junop

Pages: xi, 151

Abstract

DNA single strand break repair and base excision repair are two repair pathways essential for life in humans. XRCC1 is required for repair in both pathways and is believed to act primarily as a scaffold protein for assembly of several other repair factors at the site of damage. In addition to orchestrating the repair event through protein-protein interactions, XRCC1 is thought to make direct contact with DNA; however, the importance of this interaction has not been demonstrated *in vivo*. Work described here localizes the *in vitro* DNA binding activity of XRCC1 to a minimal binding region (residues 219-415) encompassing the first BRCT domain (301-415) and an additional 80 residues N-terminal to the BRCT domain. Further analysis reveals that K243, K245, R246, K247, K271, R272, K274 act as key resides for mediating interaction with DNA. Interestingly, although the region N-terminal to the BRCT domain is predicted to lack structure, small angle X-ray scattering experiments demonstrate the presence of structure afforded by a series of proline residues. A XRCC1 DNA binding-deficient mutant is further shown to be deficient in foci formation at sites of single strand breaks demonstrating, for the first time, a direct role for XRCC1 DNA binding activity *in vivo*.

This work further identifies several compounds that inhibit interaction of XRCC1-Ligase3. Mutations in XRCC1 that disrupt association with Ligase3 have been shown to abolish DNA single strand break repair in G_1 phase of

the cell cycle. Hence, this interaction serves as a potential target for development of compounds able to sensitize cells toward chemotherapeutic agents that function by alkylating DNA. To this end, a ‘magnetic fishing’ assay was developed to monitor XRCC1-Ligase3 interaction and further used to identify compounds that disrupt the interacting complex. Six lead compounds were found to exhibit a dose-dependent response in disrupting XRCC1-Ligase3 interaction. Despite considerable effort to structurally characterize these compounds in complex with both XRCC1 and Ligase3, no compounds could be located in the crystal structures obtained.

Altogether, work presented here enhances understanding of XRCC1 function by establishing a role for DNA binding in repair, and provides useful leads that may be further developed as chemotherapeutic agents and/or probes for studying DNA repair in the single strand and base excision pathways.

Acknowledgments

I wish to thank my supervisor Dr. Murray Junop for his support, teaching, and guidance. He is a wonderful mentor, and a true scientist of which I inspire to become.

I wish to also thank my committee members Dr. Paul Berti and Dr. Giuseppe Melacini for their helpful discussion and patience over the course of the project.

I wish to also thank my lab mates that shown me how lucky I am to have been in this lab.

I wish to lastly thank my family for believing that I could do this.

Table of contents

List of Figures

List of Tables

Abbreviations

ADP – Adenosine diphosphate
BER – base excision repair
BME – beta-mercapto ethanol
BRCT – breast cancer gene 1 C-terminal
EMSA – electrophoretic mobility shift assay
GST – Glutathione S-transferase
IMAC – immobilized metal affinity chromatography
IPTG – isopropyl β-D-thiogalactopyranoside
IR – ionizing radiation
LB – Lennox broth
LIC – ligation independent cloning
MS – mass spectrometry
MWCO – molecular weight cut off
NLS – nuclear localization signal
NTD – N-terminal domain
OH – hydroxyl
PAR – poly adenosine diphosphate ribose
PCPD – per cell per day

PI – isoelectric point
PO4 – phosphate
PPI – protein-protein interaction
PPII – poly-proline type 2
PVDF – Polyvinylidene fluoride
ROS – reactive oxygen species
SAXS - small angle x-ray scattering
SSBR – single strand break repair
TBE – Tris/Borate/EDTA

1 Introduction

1.1 DNA single strand breaks and modified bases

Deoxyribonucleic acid (DNA) is the carrier of genetic information essential to sustain life. Genomic integrity is constantly threatened by DNA damaging agents that arise from both exogenous and endogenous sources. Different DNA damaging agents interact with DNA in specific ways and can affect the integrity of DNA in different ways. The sugar-phosphate backbone is particularly susceptible to damage by ionizing radiation (IR) and metabolic bi-products in the form of reactive oxygen species (ROS). The most common outcome of this damage is the formation of a single strand break, with up to 10,000 single strand breaks generated per cell per day (pcpd). Additionally, DNA bases are prone to other forms of damage that can also lead to formation of single strand breaks including spontaneous base loss (~10,000 pcpd), base oxidation (~100,000 pcpd), and methylation (~3,000 pcpd) [Madabhushi et al. 2014] (Figure 1.1). Exogenous sources of DNA damage also pose serious challenges to the integrity of the genome. For instance, the use of tobacco products introduces N-nitrosamines, which upon metabolic activation increase both single strand breaks and alkylated guanosine bases [Hecht et al. 1999]. The consequence of not repairing these damages include formation of mutations, DNA double strand breaks, which arise when DNA replication machinery encounters a single strand break, and chromosomal rearrangements. In humans the failure to repair such damage is associated with diseases such as cancer, premature aging,

immune deficiency and neurodegeneration [Maynard et al. 2016]. To prevent harmful levels of DNA damage, human cells have developed high fidelity and efficient pathways to repair these types of damage.

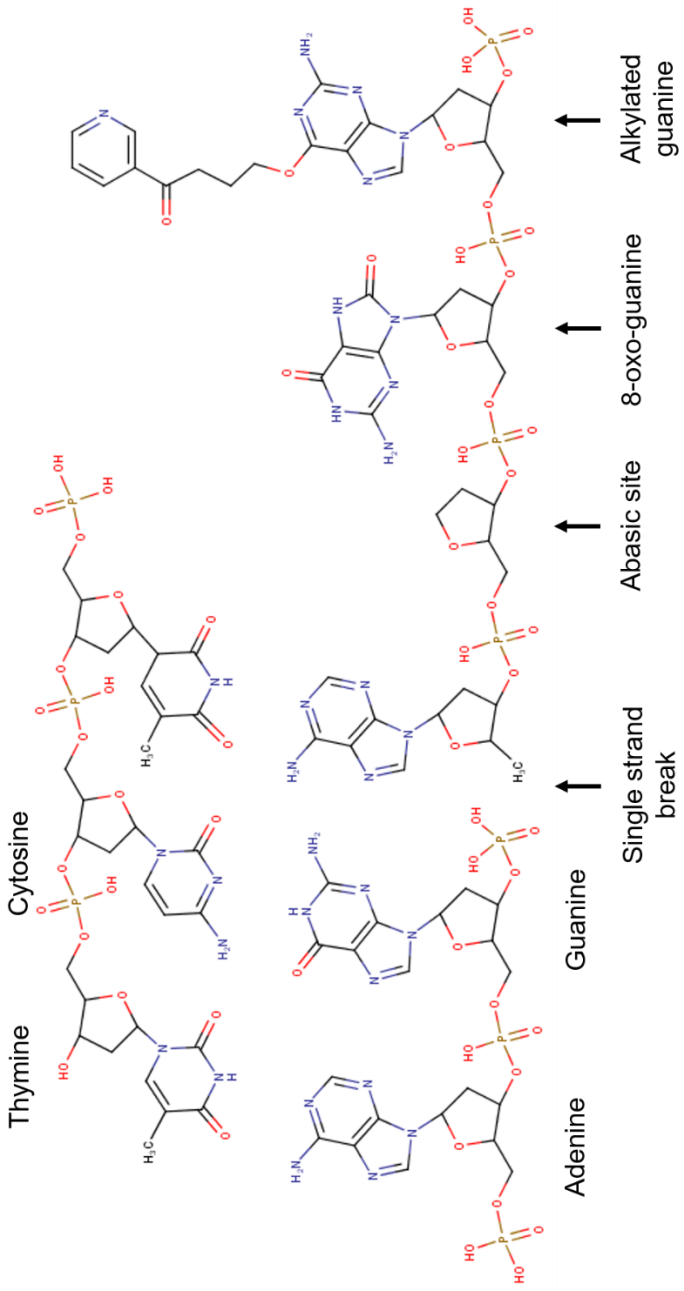


Figure 1.1 Commonly found DNA lesions induced by reactive oxygen species (ROS) and alkylating agents. The four DNA bases, thymine, cytosine, adenine and guanine are shown on the left for reference. Each arrow points to a different form of lesion including: single strand break, abasic site, oxidized base (8-oxo-guanine) and a guanine reacted with 4-(acetoxymethylnitrosamino)-1-(3pyridyl)-1-butanone, a known DNA adduct.

1.2 Base excision repair

In humans, an essential pathway responsible for repair of modified DNA bases is base excision repair (Figure 1.2). The pathway relies on the coordinated effort between several enzymes and scaffolding proteins to facilitate repair in a tightly regulated fashion. As the damage site is spatially restricted, enzymes must be recruited and displaced at the correct time and in the correct sequence for successful repair. The role of the glycosylase is to both detect and remove the damaged base, leaving behind an abasic site. To date, there are 11 known mammalian DNA glycosylases. Current understanding of this family of enzymes suggest that the enzyme initially binds DNA non-specifically and then scans the DNA for imperfect base pairing, whereupon it flips out the damaged base and cleaves the N-glycosidic bond to release the damaged base [Robertson et al. 2009]. In the case of the most commonly modified base, 8-oxo-guanine, repair begins with detection of the lesion by DNA glycosylase, OGG1. Certain glycosylases are bifunctional apurinic/apyrimidinic lyases, and may have the ability to induce a nick in the sugar phosphate backbone via a beta elimination reaction, generating a 3' sugar phosphate group [Jacobs et al. 2012]. In the presence of an abasic site or 3' sugar phosphate group, enzyme (apurinic/apyrimidinic endonuclease) APE1 is recruited [Fortini et al. 2007]. This enzyme is able to cleave the remaining DNA backbone and leave behind a 3' hydroxyl group, a process commonly referred to as 'end

tailoring'. If the starting substrate was an abasic site, APE1 cleaves 5' to the abasic site and leaves behind a 5' dRp group, which can be removed later by polymerase beta (PolB). APE1 also recruits a scaffolding protein, XRCC1, to the site of damage before APE1 is displaced. XRCC1 possess no known enzymatic activity, but is thought to function as a scaffold protein important for control of downstream repair process by timely recruitment and displacement of other repair enzymes. XRCC1 binds tightly to Ligase3 and recruits it to the damage site. XRCC1 and Ligase3 are believed to be constitutively bound in the cell and in the absence of XRCC1, Ligase3 is unstable and is depleted [Caldecott et al. 1995]. XRCC1 subsequently recruits PolB in order to replace the missing nucleotide and remove the 5' dRp group left behind from APE1 activity. Ligase3 completes the repair process by ligating the remaining nicked DNA [Caldecott 2003].

1.3 Short and long patch single strand break repair

DNA single strand breaks can be generated directly when ROS contacts the DNA backbone. This type of damage often generates a 'dirty break' where abnormal chemical groups such as phosphoglycolate remain bound to the damaged backbone [Fortini et al. 2007]. The repair of this type of damage is carried out by the single strand break repair pathway (Figure 1.2). This pathway can be further divided into two subpathways: short patch repair where only a single nucleotide needs to be replaced, or long patch repair if several nucleotides require removal. The short patch pathway is

initiated when the enzyme poly [ADP ribose] polymerase 1 (PARP1) detects a nick in the DNA backbone.

PARP1 association with DNA activates autoribosylation activity. Current understanding suggests that a pair of PARP1 enzymes interact at the damaged DNA and modify their partner by attaching a large poly(ADP)ribose chain at a specific residue [Masson et al. 1998]. This large negatively charged polymer acts as a signaling molecule to recruit XRCC1/Ligase3, the same complex also required for completion of base excision repair. If the correct chemical groups are not present at the damage site (namely a 3' hydroxyl and a 5' phosphate group), XRCC1 recruits polynucleotide kinase/phosphatase (PNKP) which can add either or both missing moieties [Whitehouse et al. 2001]. PolB is then recruited to fill in the missing nucleotide and allow Ligase3 to complete the repair process.

The long patch pathway follows a similar initiation process as the short patch (Figure 1.2). There are, however, several key differences in the processing of nucleotides. First, additional proteins are recruited to the damage site, in particular proliferating cell nuclear antigen (PCNA) which functions as a sliding clamp/processivity factor [Fan et al. 2004], and replication factor (RF)-C [Thompson et al. 2000]. Second, the repair process requires multiple polymerases (PolB, PolD and PolE) to fill in the missing nucleotides [Almeida et al 2007]. The addition of new nucleotides displaces the old, generating a '5' flap' structure. The flap is removed by the structure

specific 5’ flap-endonuclease (FEN)-1 [Hosfield et al. 1998]. In contrast to the short patch pathway, XRCC1/Ligase3 appear to be dispensable. Instead, Ligase1 is believed to be the ligase of choice to seal the DNA backbone [Levin et al. 2000]. Although the proteins involved in the long and short patch repair pathways have been identified, the decision within the cell to utilize one pathway over the other is still poorly understood.

In contrast to most enzymes that are present for only a single step of the repair process, XRCC1 remains at the damage site until the final step. Not only does it recruit and displace enzymes, it also enhances enzymatic activity of partners. XRCC1 is critically important in organisms, as loss is associated with loss of cerebellar interneurons in a developing neural system, increased brain damage and reduced recovery in stroke animals, and aggressive breast cancer phenotypes [Lee et al. 2009, Ghosh et al. 2015, Sultana et al. 2013].

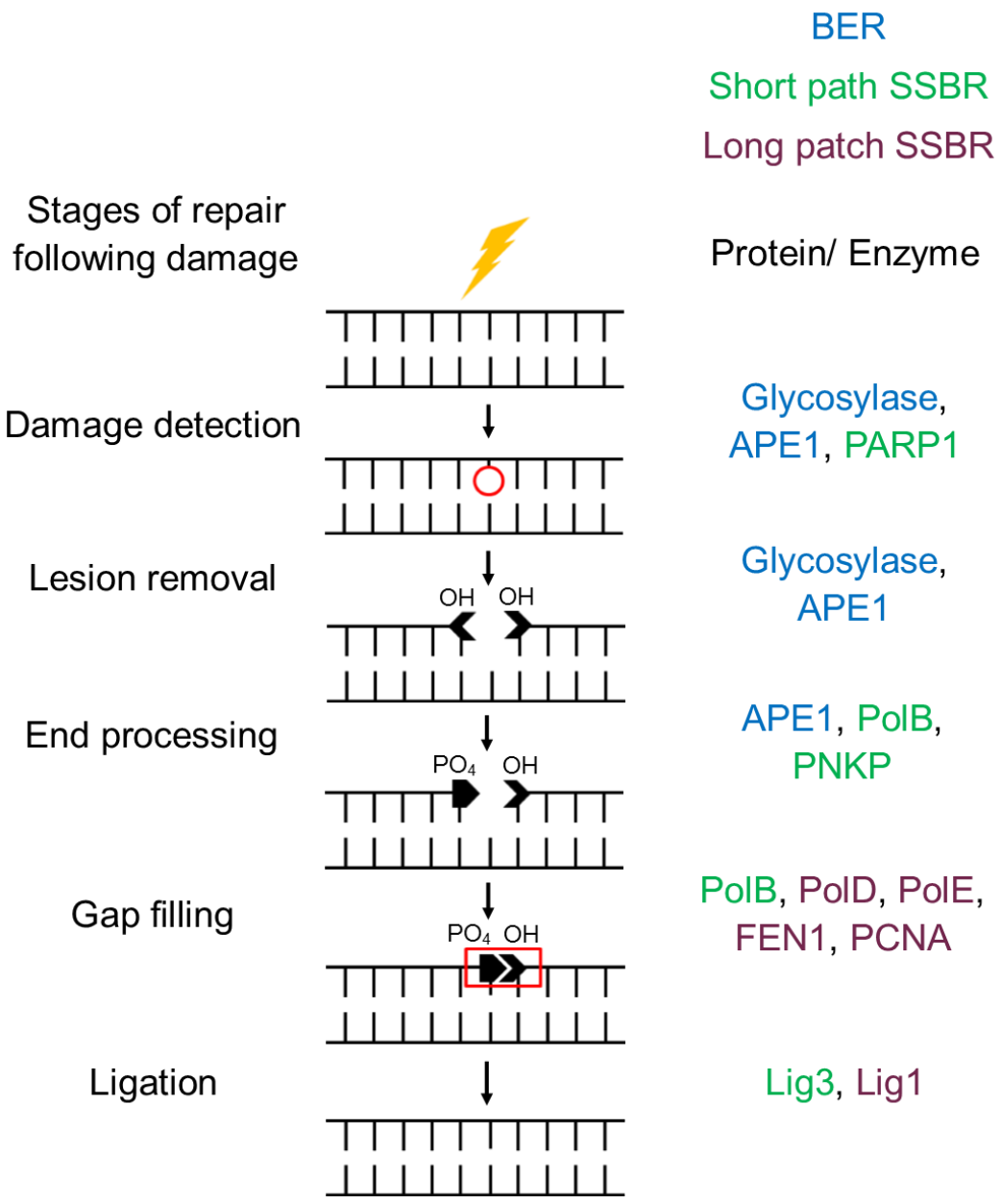


Figure 1.2 Depiction of the base excision and single strand break repair pathways. Enzymes involved in specific repair pathways are coloured in either blue for BER, green of short patch SSBR or purple for long patch SSBR. All pathways follow similar repair steps: damage detection, lesion removal, end processing, gap filling and ligation. An abasic site is depicted as a circle. PO4 and OH groups are depicted as arrows. For long patch SSBR, an additional flap removal step is carried out by FEN1 before ligation occurs.

1.4 Protein-protein and protein-DNA interactions in BER and SSBR

Protein-protein and protein-DNA interactions serve as mechanisms for tightly regulating DNA repair. The majority of these interactions are hetero-complexes where the complexes are not intended to be permanent, but instead are transient and depend on the environment and other factors such as post translational modifications [Jones et al. 1996].

1.4.1 XRCC1 as a scaffolding protein

Several staged enzyme-DNA interactions and reactions are required for repair following initial identification of the damage site. Access to the damage site must be controlled to ensure proper timing of reactions thereby preventing further damage to DNA intermediates undergoing repair. This scenario is exemplified by formation of abortive ligation products when ill-timed ligase activity acts on an inappropriate DNA substrate [Ahel et al. 2006]. Controlling these enzyme-DNA interactions is thought to be achieved in part by the scaffolding protein XRCC1, which orchestrates physical assembly of distinct repair complexes through protein-protein interactions [Good et al. 2011]. Common scaffolding mechanism that XRCC1 may employ include: tethering (to increase the local concentration of DNA substrate and enzyme), orientation (to correctly position the substrate and enzyme), and allosteric regulation (to alter conformation of either substrate or enzyme to favor reaction).

XRCC1 possesses a three-domain modular organization (Figure 1.3 A, B). The first structural domain is referred to as the N-terminal domain (NTD). This domain appears to be unique to XRCC1, in terms of both sequence identity and structural fold (protein Blast and Dali searches only return predicted XRCC1 proteins or homologues). X-ray structure of this domain consists of an anti-parallel beta sandwich, with loops connecting the beta strands, with very little alpha helical content.

XRCC1 also has two BRCA-1 C-terminal (BRCT) domains. According to the pFam server, the BRCT superfamily contains over 4000 proteins, most of which are involved in DNA metabolism. BRCT domains are approximately 100 residues in length and are named according to where they were first identified (**Br**east **C**ancer gene 1 C-**t**erminal). Sequence analysis shows that five regions are conserved within BRCT domains, most of which are hydrophobic residues that stabilize secondary structures within the domain [Bork et al. 1997, Taylor et al. 1998]. The structure of the first BRCT domain was solved using X-ray crystallography. The observed fold appears to be well conserved despite limited sequence similarity between BRCT domains from different proteins [Zhang et al. 1998] (Figure 1.3 C). The domain contains a four stranded, parallel beta sheet which is composed of several conserved hydrophobic residues that form the core of the domain. The beta sheet is flanked by three alpha helices, where helix-1and 3 form a helical bundle and flank one side while helix-2 flanks the other. Helix 1/3

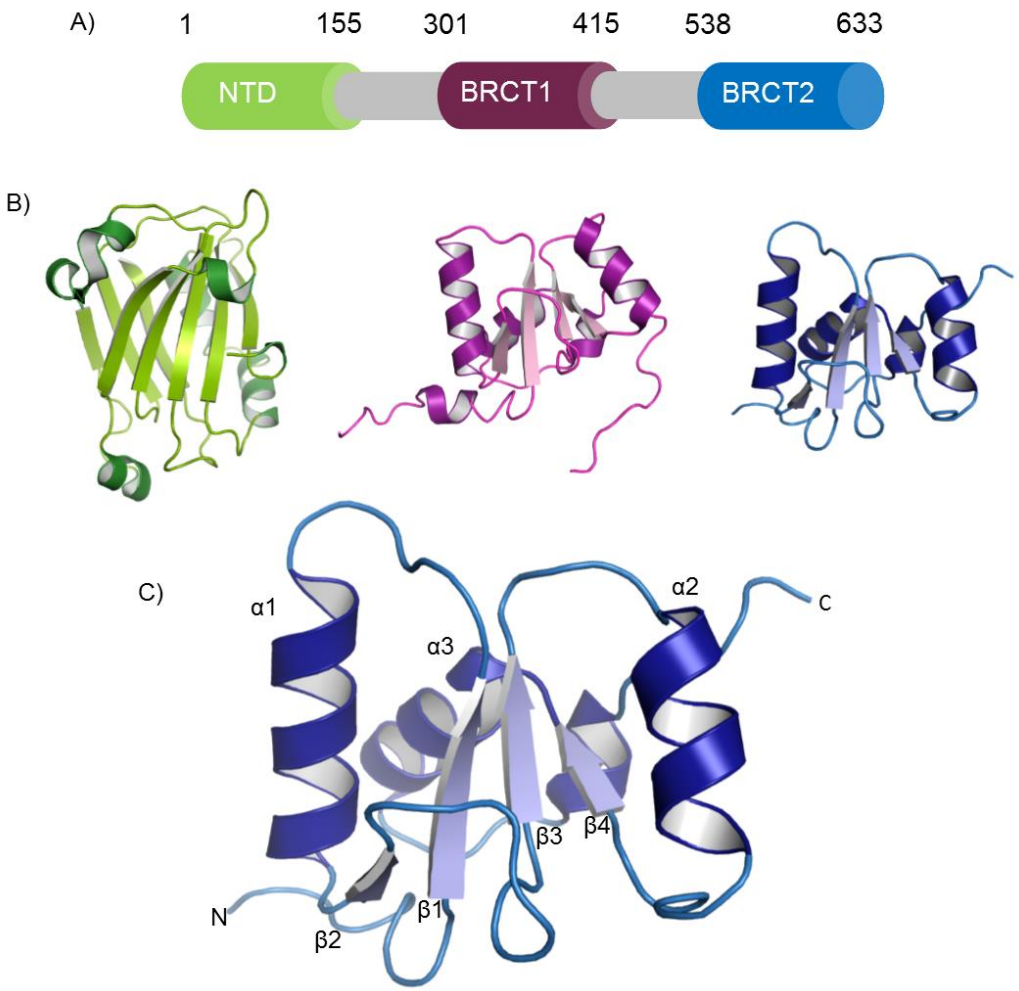


Figure 1.3 Domain organization of XRCC1 A) A schematic of the domain organization of XRCC1. Estimates of the domain boundaries are indicated by the residue number above. B) Structural domains of XRCC1 that has been determined by X-ray crystallography or NMR: NTD in green, BRCT1 in purple, BRCT2 in blue (PDBIDs 3LQC, 2D8M and 3PC6 respectively). C) A close examination of the BRCT fold: a four-stranded parallel beta sheet flanked by three alpha helices.

contain hydrophobic residues that are conserved across multiple proteins. Mutations that alter these conserved residues prevent proper folding of the domain, and certain residues in BRCA1, when mutated, result in a predisposition for cancer development. The second helix of this domain is the most variable region, where changes in both the length and residue composition have been observed [Glover et al. 2004].

BRCT domains are predominantly found to mediate protein-protein interactions. The domain architecture of proteins containing a BRCT is highly diverse, including single isolated domains (e.g. PARP1 and Ligase3), multiple isolated domains (e.g. XRCC1 and RFC), tandem repeat domains (e.g. BRCA1) and combinations of isolated and tandem repeats on the same protein (e.g. TopBP1) [Leung et al. 2011] (Figure 1.4 A). The most characterized architecture is the tandem BRCT where two domains are in close proximity within the primary amino acid sequence (Figure 1.4 B). The two BRCT domains dimerize to form a composite binding pocket for phosphorylated peptides where one BRCT domain provides charged residues that interact with the phosphorylated amino acid, and the other provides hydrophobic residues to interact with partner-specific sequences [Wardlaw et al. 2014] (Figure 1.4 C). Structural characterization of BRCA1, MDC1 and TopBP1 BRCT domains reveal that these tandem domains have high specificity for either phosphorylated serine or threonine residues. Further binding specificity is also attributed to slight alterations in the linker

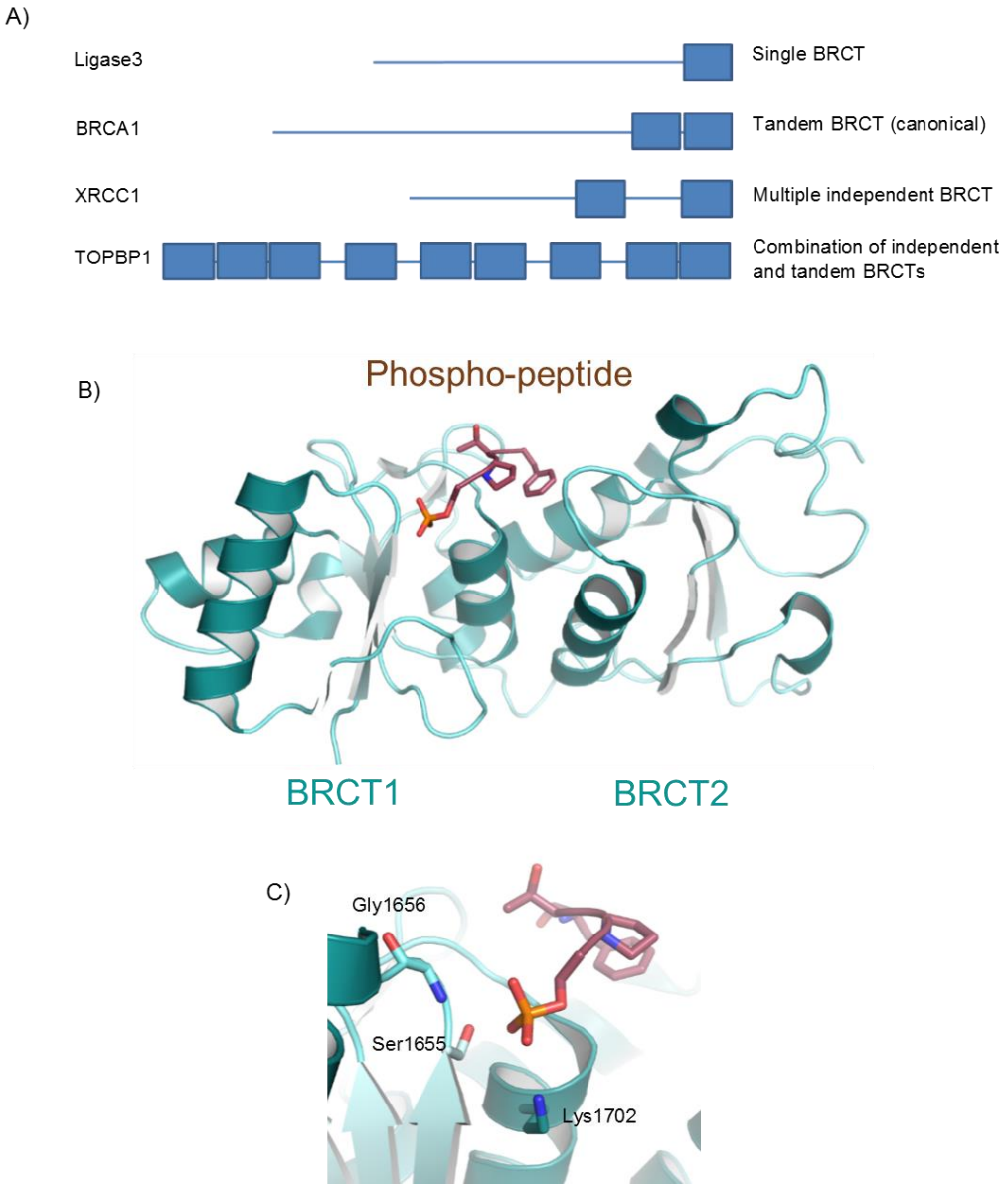


Figure 1.4 BRCT domain arrangement in proteins. A) Depiction of possible BRCT arrangements in Ligase3, BRCA1, XRCC1 and TOPBP1. Each blue square is a BRCT domain. B) Structure of the tandem BRCT of BRCA1 interacting with a phosphorylated peptide (PDBID: 3K0H). Helix 1 and 3 from BRCT2 forms a 4-helical bundle with helix 2 of BRCT1 and a helical linker. This interaction creates a hydrophobic pocket that mediates protein-protein interactions, as well as positions key residues to interact with a phosphate group show in C).

length, differential packing of adjacent BRCT domains, and the presence of additional domains and/or motifs [Leung et al. 2011].

Proteins that possess multiple BRCT domains may use each to interact with a partner with a high degree of specificity. For example, while the C-terminal BRCT domain of XRCC1 is responsible for interacting with Ligase3, the central BRCT domain shows no affinity for Ligase3 [Nash et al. 1997, Beernink et al. 2005]. Studies of TopBP1 BRCT domains show that multiple tandem repeats also interact with specific partners. Such studies demonstrate that interactions mediated by BRCT domains are highly specific and may serve as ideal targets for modulating protein-protein interactions critical for DNA repair.

1.4.2 Biological significance of XRCC1-protein interactions

In human XRCC1, the NTD domain has been shown to interact directly with polB through biochemical and structural characterization [Caldecott et al. 1996, Cuneo et al. 2010]. PolB is the main polymerase responsible for nucleotide addition in BER and SSBR. This interaction buries a surface area of approximately 1300Å^2 and is mediated by extensive hydrophobic interactions as well as several salt bridges (Figure 1.5). *In vitro* experiment shows that the XRCC1-polB interaction regulates PolB activity to ensure only a single nucleotide is added to the substrate [Kubota et al. 1996], and *in vivo* experiments show that disruption of the interaction

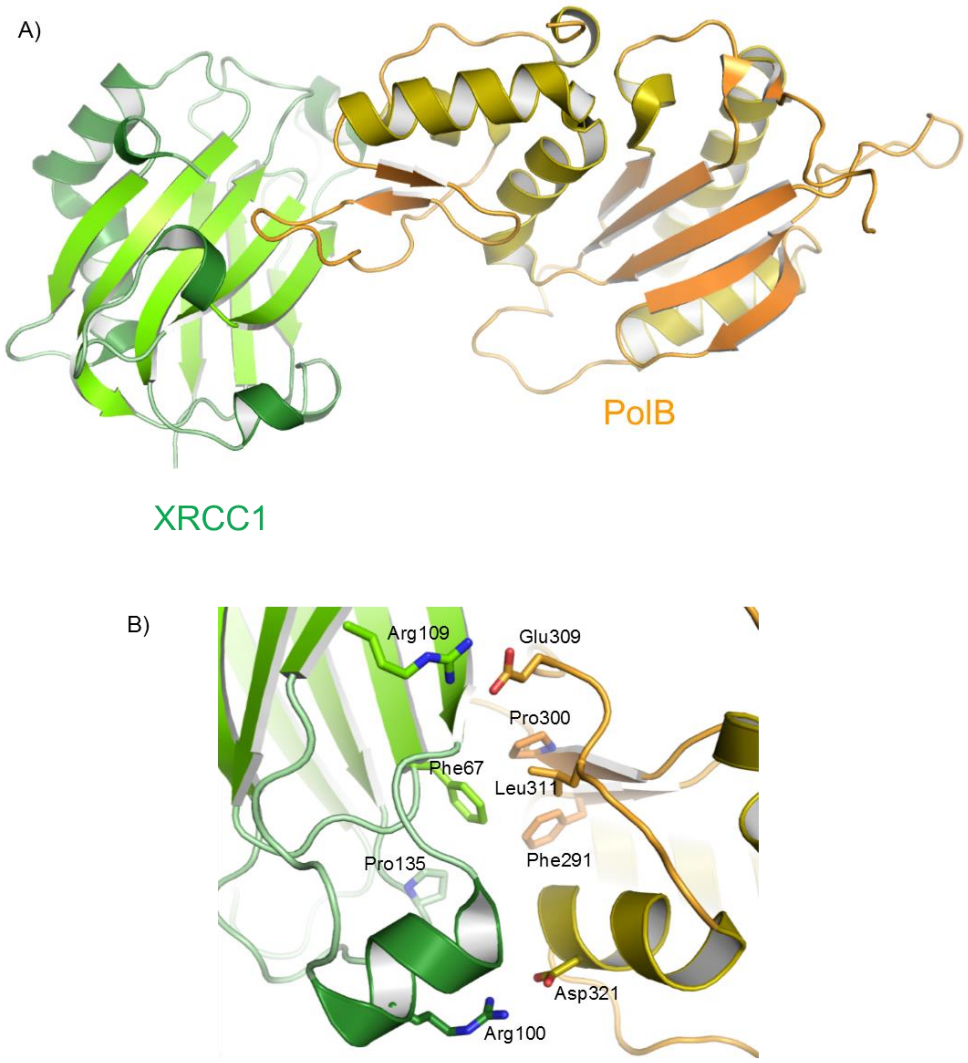


Figure 1.5 A) Crystal structure of XRCC1 in complex with PolB. The NTD of XRCC1 (green) contacts the thumb domain of PolB (orange) (PDBID 3LQC). B) A highlight of important residues mediating XRCC1-polB interaction. The interaction buries approximately 1300 $Å^2$ of surface area, and is mediated by 9 hydrogen bonds (not all shown) and hydrophobic interactions.

reduces repair efficiency and increases sensitivity to hydrogen peroxide [Dianova et al. 2004] and methanesulfonate [Wong et al. 2005].

Single BRCT domains are capable of serving as interacting platforms for protein-protein interactions. The first BRCT domain of XRCC1 is known to interact with early DNA repair responders, such as PARP1, APE1 and many DNA glycosylases. The region preceding the N-terminus of this BRCT domain also appears to contribute to the APE1 and glycosylase interactions. These interactions serve to recruit XRCC1 to the site of DNA damage [Masson et al. 1998, Vidal et al. 2000, Marsin et al. 2003], and disruption of these interactions prevents XRCC1 localization to DNA damage sites *in vivo* [Campalans et al. 2013, Campalans et al. 2015]. In most cases, the interaction with XRCC1 also regulates the activity of the enzyme, whether it be downregulating PARP1 activity, or stimulating APE1 and OGG1 activity. Despite the importance of these interaction complexes, no structural information is available.

The C-terminal BRCT domain of XRCC1 is believed to be constitutively bound to Ligase3, since this interaction is important for *in vivo* stabilization of Ligase3 [Caldecott et al. 1995]. This interaction was the first BRCT-BRCT heterodimer structure solved, involving the C-terminal BRCT domain of XRCC1 and the C-terminal domain of Ligase3 [Cuneo et al. 2011]. The interaction of these BRCT domains buries a surface area of 1210 $Å^2$ and involves both polar and non-polar interactions (Figure 1.6).

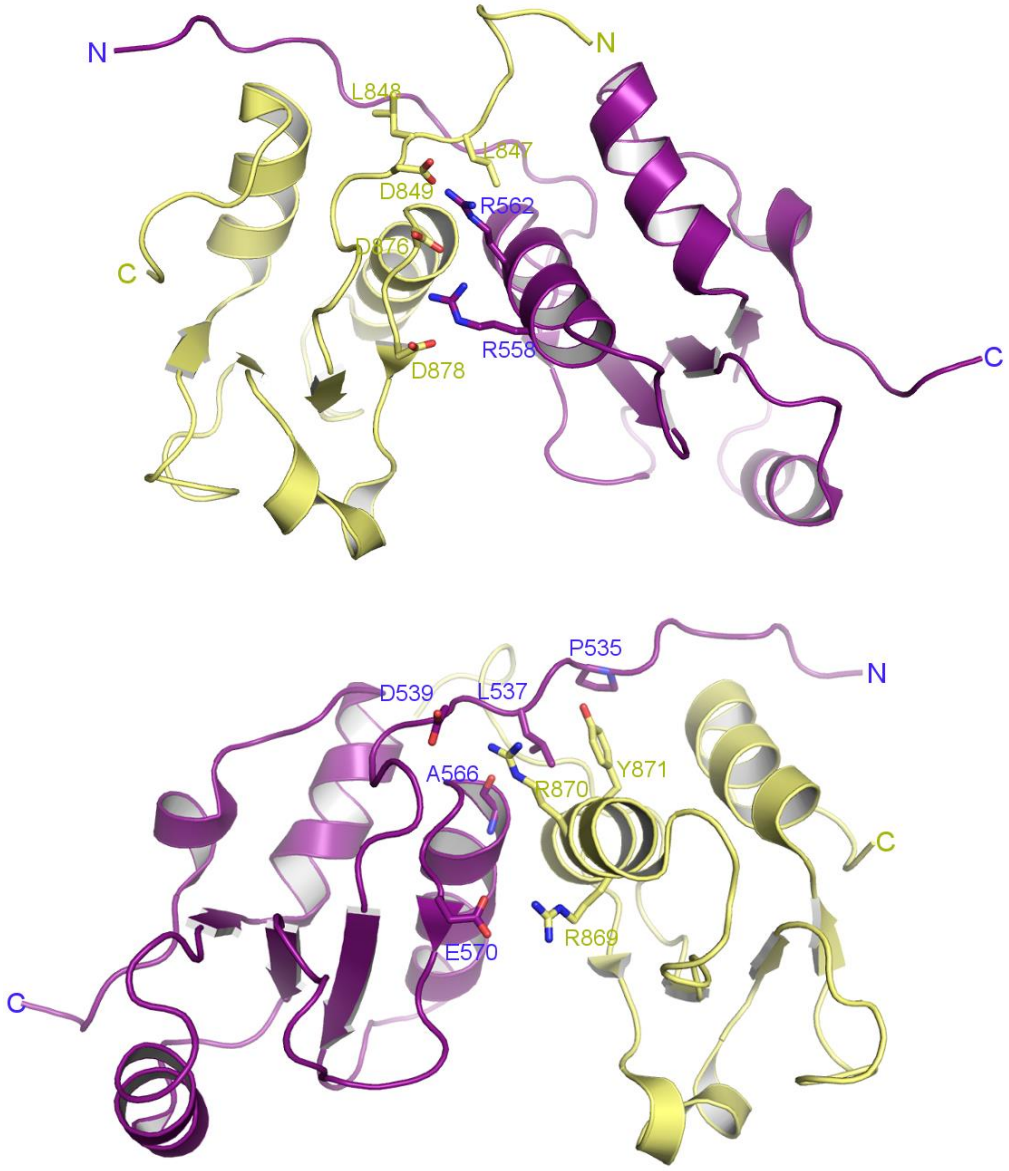


Figure 1.6 Structure of the XRCC1 and Ligase3 interaction complex (PDBID 3QVG). XRCC1 is coloured purple and Ligase3 in yellow. Important residues are highlighted for XRCC1 (in blue) and Ligase3 (in yellow).

Interestingly, the structure clearly demonstrates that the BRCT-BRCT interaction is not limited to just the BRCT domains themselves, but also several residues N-terminal to the BRCT domain of XRCC1. Disruption of this interaction *in vivo* has been shown to decrease SSBR efficiency in a cell cycle dependent manner [Moore et al. 2000], and a knockout of XRCC1 also reduces Ligase3 levels *in vivo* [Caldecott et al. 1995].

Two large linkers (between 100-150 residues each) connect the individual domains of XRCC1. These linkers, once phosphorylated by the kinase CK2, can also mediate protein-protein interactions [Parsons et al. 2010]. Namely, PNKP and APTX possess FHA domains that interact with phosphorylated XRCC1 [Ali et al. 2009, Cherry et al. 2015]. Interactions between XRCC1, PNKP and APTX are important for stimulating PNKP reactivity *in vitro* [Lu et al. 2010] and for protection against methanesulfonate treatments *in vivo* [Della-Maria et al. 2012, Luo et al. 2004]. *In vivo* experiments also demonstrate that if phosphorylation is eliminated via mutation of phosphorylated residues, PNKP and APTX cannot be recruited to sites of DNA damage [Wei et al. 2013].

It is clear that XRCC1 mediated interactions are important for recruiting enzymes, as well as regulating the enzymatic activity of partner proteins. The controlling mechanisms responsible for driving recruiting events, as well as the mechanisms that XRCC1 employs to regulate enzymatic activity remain a hot topic of study.

1.4.3 XRCC1-DNA interactions

Although each repair protein involved in SSBR interacts with DNA at various stages of repair, their interaction is transient [Almeida et al. 2007]. In contrast, XRCC1 functions by mediating interactions with different repair proteins, but stays associated at the site of DNA damage until repair is complete. Full length XRCC1 protein is able to interact with many forms of DNA, including nicked, gapped, double stranded and single stranded DNA, albeit with slight variations in affinity [Mani et al. 2004]. The full length protein slightly favors DNA containing small lesions (i.e. 1 nucleotide gaps, nicks) over undamaged ssDNA, dsDNA or DNA harbouring a large 5-nucleotide gap. The NTD of XRCC1 has been shown to interact with damaged forms of DNA, specifically nicked and gapped DNA, which could potentially provide a damage sensing role [Marintchev et al. 1999] and explain why full length protein prefers damaged DNA with small lesions. Additionally, *in vitro* crosslinking experiments and electron microscopy analysis shows both the central and C-terminal BRCT domain of XRCC1 interacting with DNA substrates [Nazarkina et al. 2007, Yamane et al. 2000]. However, it is not clear whether DNA binding of XRCC1 observed *in vitro* is necessary for DNA repair *in vivo*.

1.5 Modulating DNA repair as a therapeutic approach for cancer treatment

1.5.1 Current therapeutic pitfalls

Many cancers are treated using radiation therapy and chemical compounds such as topisomerase inhibitors to generate lethal amounts of DNA double strand breaks [Jekimovs et al. 2014]. Alternatively, chemicals such as cisplatin are effectively used to generate DNA interstrand crosslinks that are highly toxic to cells [Madhusudan et al. 2005]. These therapeutic approaches are most toxic to rapidly dividing cells and are therefore well suited for combating many tumors, as damaged DNA disrupts replication and commonly leads to cell cycle arrest or apoptosis within cancer cells. These treatment options are initially effective, but farily unselective, often damaging normal dividing cells in tissues such as bone marrow and the intestinal lining of the digestive track [Luqmani et al. 2005]. Additionally, cancer cells frequently develop resistance to DNA damaging chemotherapeutics making relapse in tumors very challenging to treat. An estimated 1 in 10^6 cells of a tumor is naturally resistant to a given drug, hence depending on the size of the initial tumor, some cancerous cells will survive the initial treatment and reform a treatment resistant tumor [Luqmani et al. 2005]. Typical resistance mechanisms involve upregulation of DNA repair factors and/or loss or mutation of the initial drug target, causing the tumor to stop responding to treatment.

1.5.2 Exploiting synthetic lethality in cancer treatment

One of the fundamental hallmarks of cancer is genomic instability. Most cancer cells have defects in some aspect of their DNA repair system. For example, some breast cancer cells are characterized by dysfunctional BRCA1 and BRCA2 proteins, both of which are required for repair of DNA double strand breaks via the homologous recombination pathway. Cells that possess a dysfunctional DNA repair pathway often become dependent on a compensatory pathway. In the presence of dysfunctional double strand break repair, the cells become reliant on the single strand break repair pathway, of which PARP1 serves a critical role. Several groups have shown synthetic lethality (where inhibiting the function of two proteins by small molecules or genetic mutation causes cell death) by inhibiting PARP1 in BRCA deficient cells. Importantly, this approach can be highly selective for tumor cells, leaving normal cells with functional double strand break repair activity unscathed [Farmer et al. 2005, Bryant et al. 2005]. The ability to exploit synthetic lethal relationships involving DNA repair shows promise of being broadly applicable. In particular, recent studies have shown effective targeting of phenotypically aggressive XRCC1-deficient tumors by inhibiting DNA double strand break repair factors such as ATM and DNA PKcs, suggesting that synthetic lethality approaches may be applicable to any redundant two-pathway system for genomic stability [Sultana et al. 2013]. Despite the obvious potential for exploiting XRCC1 and the many protein-

protein interactions it mediates, there are currently no reports in the literature targeting XRCC1 for chemotherapeutic intervention.

1.5.3 Targeting protein-protein interactions

Each DNA repair pathway consists of multi-step processes and is carefully controlled via transient protein-protein interactions (PPIs) [Madhusudan et al. 2005, Almeida et al. 2007]. Typically, PPIs are highly specific and disruption leads to pathway dysfunction [Campalans et al. 2015]. Thus, in principle, PPIs serve as ideal targets for small molecule modulation. PPIs have traditionally been viewed as 'undruggable' due to the difficulty associated with disrupting large protein-protein interaction surfaces that frequently lack 'distinct' structural features. More recently, PPIs have gained renewed interest as useful targets for small molecule intervention with the emergence of new understanding in the field, such as the 'hot-spot residue' theory which focuses on disrupting PPIs by preventing crucial 'hot-spot' residues from contacting their perspective partners [Milroy et al. 2014, Nero et al. 2014]. Such compounds are referred to as orthosteric modulators. Additionally, allosteric modulators that disrupt PPI interfaces by inducing conformational change within the target protein have also been identified. Currently there are 12 anti-cancer small-molecule modulators targeting PPIs that have progressed to clinical trials, underscoring the promise of PPIs for targeted therapeutic intervention such as those mediated by XRCC1 [Nero et al. 2014].

1.6 Research outline

XRCC1 is essential for both DNA base excision repair and single strand break repair. Although the role of XRCC1 as a scaffolding protein in mediating interactions is critical for regulating DNA repair, little is known about how XRCC1 coordinates protein interactions with respect to DNA complex assembly.

The first goal of this thesis is to identify residues of XRCC1 that mediate interactions with DNA and determine whether DNA binding activities observed for XRCC1 *in vitro* have any impact on function *in vivo*. Several reports have suggested that XRCC1 interacts with DNA, yet no consensus of where the interaction may be occurring and how it occurs have been reached. In chapter 2, the minimal DNA binding region within XRCC1 is localized to residues 219-415. Further analysis shows that the key residues mediating this interaction are K243, K245, R246, K247, K271, R272 and K274. Amino acid substitution of these residues is also shown to abolish DNA binding *in vitro* and prevent DNA repair foci formation *in vivo* by elevating rates of XRCC1 dissociation from sites of DNA damage. In chapter 3, an XRCC1-DNA complex is characterized using small angle x-ray scattering. Efforts to determine the crystal structure of this complex are also presented.

A second aim of this thesis is focused on disrupting the BRCT-BRCT mediated complex of XRCC1-Ligase3 using small molecules. Due to its

critical role in two major DNA repair pathways, XRCC1 represents an ideal therapeutic target as inhibiting these repair pathways has been shown to induce synthetic lethality in various types of cancer. The final step of BER and SSBR converge at the ligation step, which is performed by the XRCC1/ligase3 complex. Inhibiting this repair step would effectively disrupt repair by both pathways. In chapter 4, an assay platform is developed which allows detection of XRCC1/ligase3 complex formation, and is further used to identify small molecule compounds with ability to modulate XRCC1/ligase3 interaction.

The final chapter of the thesis summarizes the results of previous chapters and discusses their impact and significance on the field of DNA repair.

2 The DNA binding region of XRCC1 encompasses the central BRCT domain and preceding 80 residues

2.1 Preface to Chapter 2

Dr. Anna Campalans and Dr. Pablo Radicella performed the *in vivo* cell imaging experiments. I performed the all other experiments.

2.2 Abstract

Successful single strand break and base excision DNA repair requires the scaffolding protein XRCC1. XRCC1 orchestrates the repair process by interacting with protein partners, and presumably DNA. Here we show that there exists a DNA binding module independent of the previously reported N-terminal domain (NTD). In contrast to the NTD, which was suggested to be capable of damage sensing, we demonstrate that the newly identified DNA binding module shows no significant preference for binding DNA containing nicks or gaps. We localize this DNA binding module to residues 219-415 of XRCC1, and show that amino acid substitution of positively charged residues (K243, K245, R246, K247, K271, R272, K274) to alanine abolishes DNA binding. Furthermore, the biological importance of DNA binding by XRCC1 is established using a cell-based to monitor XRCC1 foci formation in response to DNA damage.

2.3 Introduction

The successful repair of modified DNA bases and single stranded breaks via the base excision repair (BER) and single strand break repair (SSBR) pathways requires highly coordinated repair events. In humans, the X-ray cross complementing group 1 (XRCC1) protein is thought to act as a scaffold where it actively recruits repair enzymes to the site of damage [Thompson et al. 2000, Almeida et al. 2007]. These recruitment events are facilitated through direct protein-protein interactions between XRCC1 and

its partners. XRCC1 is 633 residues in length and has a modular design in which three distinct domains (the N-terminal domain and two independent BRCT domains) are separated by two intervening regions predicted to function as flexible, unstructured linkers [Caldecott et al. 2003]. Each of the three domains functions as a protein-protein interaction platform, where the N-terminal domain (NTD, residues 1-183) interacts with PolB [Caldecott et al. 1996, Marintchev et al. 2000]; the first BRCT domain (BRCT1; residues 301-415) interacts with PARP1, APE1, MPG, hNTH1 and hNEIL2 [Masson et al. 1998, Vidal et al. 2001, Campalans et al. 2005]; and the second BRCT domain (BRCT2; 534-633) interacts with Ligase3 [Nash et al. 1997]. The structures of all three domains have been determined by either X-ray crystallography or NMR [Marintchev et al. 1999, Zhang et al. 1998, PDB # 2D8M, 1CDZ, 1XNA]. In addition, the flexible linker regions have also been implicated in protein-protein interactions, where the first linker interacts with REV1 [Gabel et al. 2013] and hOGG1 [Marsin et al. 2003], and the second linker interacts with APTX, APLF and PNKP [Date et al. 2004, Iles et al. 2007, Whitehouse et al. 2001].

XRCC1 has been shown to bind DNA *in vitro* with a preference for interaction with damaged forms of DNA, specifically those harbouring nicks and gaps [Mani et al. 2004]. This DNA binding specificity has been attributed to interactions mediated by the N-terminal domain [Marintchev et al. 1999]. XRCC1 has also been demonstrated to interact with DNA repair

intermediates that contain chemical groups such as a 3-phospho-α,β-unsaturated aldehyde [Nazarkina et al. 2007]. This type of DNA binding does not require the NTD, instead it requires the central BRCT domain (BRCT1) and the activities of a preceding enzyme acting on the DNA, namely a glycosylase (i.e. OGG1). At this time the meaning of these XRCC1-DNA intermediate complexes remains somewhat unclear as they have only been observed under cross linking conditions. Yet another report indicated, under different cross linking conditions, the second BRCT domain (BRCT2) can also form a complex with DNA [Yamane et al. 2000]. Again, this BRCT2-DNA complex has only been seen under cross-linking conditions, presumably due to the native complex not being stable enough to survive electrophoretic separation. Since XRCC1 selectively interacts with damaged forms of DNA, but is also able to interact with all forms of DNA in general, it is possible that these two modes of binding serve distinct roles in DNA metabolism.

In this study we examined whether a XRCC1-DNA complex can form independently of the NTD. We hypothesized that while damage-sensing might occur within the NTD, another region may function as the major contributor to binding observed with non-damaged forms of DNA. Here we report a robust DNA binding module that encompasses the BRCT1 domain, as well as the preceding N-terminal linker region. We term this DNA binding region the **C**entral **D**NA **B**inding domain (CDB). Our results also identify key

residues that contribute to this interaction and mediate XRCC1 foci formation at sites of DNA damage *in vivo*.

2.4 Materials and methods

2.4.1 Preparation of protein expression vectors

The human XRCC1 gene was acquired from Open Biosystems (clone ID 4646806, accession BC023593). The Gateway cloning system (Invitrogen) was used to generate constructs of full length XRCC1, $XRCC1^{1-183}$, $XRCC1^{219-415}$, $XRCC1^{301-415}$, and $XRCC1^{219-300}$. Genes were PCR amplified by mixing 2 x i-pfu mix (10 µL, Froggabio) with water (7 µL), gene plasmid (1 µL), forward primer (1 µL, 1 µM) and reverse primer (1 µL, 1 µM). An additional C-terminal hexa-repeat histidine tag was added to full length XRCC1. Primers used PCR reactions are listed in Figure 2.1. PCR amplified products (3.5 µL) was reacted with pDONR201 (0.5 µL, Invitrogen) in the presence of BP clonaseII (0.5 µL, Invitrogen) and incubated at room temperature overnight. The reaction was then treated with proteinase K (1 µL, Thermo Scientific) at 37 °C for 30 min. The reaction was then transformed into TOP10 cells and plated on LB agar plates containing kanamycin (0.05 mg/mL). Plates were incubated at 37 °C overnight. Single colonies were selected for overnight cultures, which were grown in LB containing kanamycin (0.05 mg/mL) at 37 °C overnight. Overnight cultures were then harvested by centrifugation at 3005 x g for 10 min and the plasmid was extracted by using a miniprep kit (Geneaid) according to

manufacturer's instructions. The resulting BP entry clones (3.5 µL) were reacted with 0.5 µL of either pDEST15 (Invitrogen) for full length XRCC1, pDEST544 (Addgene # 11519) for XRCC1$^{219-415}$ and XRCC1$^{219-300}$ or pDEST17 (Invitrogen) for XRCC1$^{1-183}$, XRCC1$^{301-415}$ in the presence of LR clonaseII (1 µL) at room temperature overnight. The mixture was treated with proteinase K (1 µL, Thermo Scientific) at 37 °C for 30 min. The reaction was then transformed into TOP10 cells and plated on LB agar plates containing ampicillin (0.1 mg/mL). Single colonies were selected for overnight cultures, which were grown in LB containing ampicillin (0.1 mg/mL) at 37 °C overnight. Overnight cultures were then harvested by centrifugation at 3005 x g for 10 min. The plasmid was extracted by using a miniprep kit (Geneaid) according to manufacturer's instructions.

An XRCC1$^{219-633}$ truncation was cloned into the pLic-His vector using ligation independent cloning (LIC) [Cabrita 2006]. pLic-His plasmid (10 µL) was mixed with NEB4 buffer (2 µL), water (6.5 µL) and SacII (1.5 µL) to digest the plasmid for cloning. The reaction was incubated at 37 °C for 1 hour. The digested plasmid was ran on a 0.7% TAE agarose gel and purified by gel extraction using a gel extraction kit (Qiagen) according to manufacturer's instructions. The gene sequence was amplified using PCR. The PCR product (3 µL) was mixed with the digested plasmid (1 µL) and CloneEZ enzyme (1 µL) (Genescript). The reaction was incubated at room temperature for 30 min. The reaction was then transformed into TOP10 cells

and plated on LB agar plates containing ampicillin (0.1 mg/mL). Single colonies were selected for overnight cultures, which were grown in LB containing ampicillin (0.1 mg/mL) at 37 °C overnight. Overnight cultures were then harvested by centrifugation at 3005 x g for 10 min and the plasmid was extracted by using a miniprep kit (Geneaid) according to manufacturer's instructions.

For cell-based functional studies, XRCC1 was fused to YFP in pEYFP-N1 (Clontech). This vector was kindly provided by Dr. Anna Campalans and Dr. Pablo Radicella. Overlapping PCR [Liu et al. 2008] was used for site-directed mutagenesis and NLS insertion (see below section 2.4.2).

2.4.2 Mutagenesis

Mutagenesis was performed using the one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol described by Liu et al [Liu et al. 2008]. The primers used contained two components. The first is the mutation-containing overhang region, and the second is the plasmid-overlapping sequence that anneals to the plasmid undergoing mutagenesis. The mutation-containing overhang region for both the forward and reverse primers are complementary. Successful PCR amplification will generate forward and reverse products that contain the desired mutant sequence that complement each other perfectly. Primers that were used for PCR amplification are listed in Figure 2.2. PCR amplified

plasmids were purified using a PCR purification kit (Qiagen) according to manufacturer's instructions. Purified plasmid was mixed with water (6 µL), fast digest buffer (3 µL) (Fermentas), and fast digest Dpn1 (1 µL) (Fermentas). Reaction was incubated at 37 °C for 30 min. Reaction was then transformed into TOP10 cells and plated on LB agar plates containing ampicillin (0.1 mg/mL). Single colonies were selected for overnight cultures, which were grown in LB containing ampicillin (0.1 mg/mL) at 37 °C overnight. Overnight cultures were then harvested by centrifugation at 3005 x g for 10 min and the plasmid was extracted by using a miniprep kit (Geneaid). For DNA binding EMSAs, mutagenesis was performed on the truncated version of XRCC1$^{219-633}$. For cell-based functional studies, successive rounds of PCR, plasmid purification, TOP10 cell transformation, overnight culturing and minipreps were performed on the XRCC1-containing-ePYFP plasmid using P1, P3, and NLS forward and reverse primers. All mutations were confirmed by DNA sequencing covering the entire open reading frame.

2.4.3 Protein purification

All proteins were expressed in Rosetta pLysS cells. Cells were grown in 4 L of LB with added ampicillin (0.1 mg/mL) at 37˚C to an OD_{600} of 0.3-0.5 and induced with 1 mM IPTG at 20˚C overnight. The only exception was XRCC1$^{1-183}$, which was auto-induced at 16 ˚C over two days. Cells were then harvested by centrifugation at 3,315 x g for 15 min. Cells were

resuspended in NiA buffer (20 mM Tris pH 8, 500 mM KCl, 3 mM BME, 10% (v/v) glycerol, 10 mM imidazole) and lysed by sonication (3 x 1 min). The lysate was then clarified by centrifugation at 48, 384 x g for 45min to remove insoluble material. Proteins were initially purified by IMAC using a 5mL column (GE Healthcare). The bound protein was washed with NiA buffer containing 10mM imidazole (20 column volumes) followed by 30mM imidazole (10 column volumes) before elution with 300mM imidazole. Eluted protein (10 mL) was mixed with water (10 mL), 10 x buffer (2.5 mL of 500 mM Tris pH8, 10 mM EDTA) and TEV protease (2.5 mL, 1 mg/mL) to remove affinity tags. The sample was incubated at 4 °C overnight. Cleaved protein was further purified using ion exchange chromatography. An 8 mL MonoQ column was used for full length XRCC1, XRCC1$^{219-633}$ and mutants, while an 8 mL MonoS column for XRCC1$^{219-300,\ 219-415,\ 301-415}$, using 20 mM Tris (for MonoQ) or HEPES (for MonoS) pH 8, 3 mM BME, and 0 mM KCl (for binding buffer) or 500 mM KCl (for elution buffer). The relatively pure proteins were re-purified by a second IMAC purification step to remove any remaining His-tagged protein. An additional size exclusion chromatography step was required for full length XRCC1. Purified proteins were buffer exchanged into storage buffer (20 mM Tris pH8, 200 mM KCl, 3 mM BME, 10% (v/v) glycerol) and stored at -80°C.

2.4.4 DNA substrate generation

The design of DNA substrates followed those reported in Marintchev et al. 1999. 39, 24, 23, and 15 base DNA oligonucleotides were purchased from BioBasic/ IDT. Duplex DNA consisted of a 3' end fluorescently labelled 39mer, as well as non-labelled complementary oligos. Oligos were dissolved in water and annealed using a thermocycler (Thermo Scientific) starting at 100 °C and cooled 1 °C every minute to 25 °C final. Annealed oligos were purified using a 1 mL MonoQ column (GE Healthcare), using 20 mM Tris pH 8, and 0 mM KCl (for binding buffer) or 1 M KCl (for elution buffer). Purified DNA substrates were buffer exchanged into TE buffer (10 mM Tris pH 8, 1 mM EDTA).

2.4.5 DNA electrophoretic mobility shift assays (EMSAs)

DNA concentration was held constant at 5 nM for all reactions. Prior to the addition of protein, the reaction mixture (18 µL total volume) contained DNA substrate in binding buffer (10 mM Tris pH 8, 100 mM NaCl, 3 mM $MgCl_2$, 1 mM EDTA, 0.1% (w/v) CHAPS, 10 mM BME, 7.5% (v/v) glycerol, 0.02 mg/mL bovine serum albumin). Bovine serum albumin was initially stored in 20 mM potassium phosphate pH 7, 50 mM NaCl, 0.1 mM EDTA 5% (v/v) glycerol (from NEB). Protein (2 µL) was added to the reaction mixture to a final volume of 20 µL. All proteins were initially stored in 20 mM Tris pH 8, 200 mM KCl, 3 mM BME, 10% (v/v) glycerol. The final reaction mixture, including the no XRCC1 control, contained 12 mM Tris pH 8, 100

mM NaCl, 20 mM KCl, 3 mM $MgCl_2$, 1 mM EDTA, 0.1% (w/v) CHAPS, 10 mM BME, 8.5% (v/v) glycerol. Reactions were incubated at room temperature for 1h. Electrophoresis was performed on a 0.5 x TBE buffered 15% polyacrylamide non-denaturing gel with a stacking layer (4% polyacrylamide, 140 mM Tris pH 8). Electrophoresis was performed at room temperature, 100 V for 90-120 min. Gels were imaged using a typhoon scanner (GE) or Chemi-doc (Biorad) with settings optimized for fluorescein and the degree of DNA binding quantified by ImageJ. Each measurement was performed in triplicate and the average was calculated using Microsoft Excel 2007. Values were imported into Sigmaplot 12 and regression analysis was performed. Curve fitting seemed most reasonable using the 3 parameter Hill equation function: $f = \frac{ax^b}{c^b + x^b}$ where f is the fractional occupancy, x is the free ligand concentration, a is the maximum y-axis value, b is the Hill coefficient and c is the protein concentration where 50% of DNA is in complex (or the equilibrium dissociation constant).

2.4.6 Cell culture and repair foci monitoring

Cells were grown on coverslides. Lipofectamine 2000 (Life technologies) was used for transient transfections according to the manufacturer's instructions. 24 hours post transfection, cells were treated with 10 mM of H_2O_2 for 10 min, washed and incubated for 5 min in Dulbecco modified eagle medium (DMEM) (GIBCO-BRL, Invitrogen). Cells were then fixed with 2% (v/v) PFA for 20 min at room temperature. The DNA was

stained with 1 µg/ml DAPI (4',6'-diamidino-2-phenylindole) for 5 min at room temperature and mounted with Dako fluorescence mounting medium. Image acquisition was performed with a Nikon A1 inverted confocal microscope with the 63x objective.

2.4.7 Live cell imaging and micro-irradiation experiments

HeLa cells were seeded in 35 mm glass petri dishes (250, 000 cells per dish) and transfected 24 hours later with either XRCC1-YFP wild type or P1/3 mutant plasmid using Lipofectamine 2000 (Life technologies) according to manufacturer recommendations. Live cell images were captured using a Nikon A1 inverted confocal microscope equipped with an environmental chamber allowing the control of temperature, humidity and gas mixture. 24 hours post transfection, microirradiation was performed with a 405 nm diode laser set to 5% power. The laser power at the exit of the fiber was of 4 mW, and around 1 mW watts at the exit of the 10x Obj. A 488 nm Argon-laser was used to visualize the YFP fluorescence. Measurements were performed by immobilizing the laser at 100% in a point bleach with the Digital Handheld Optical Power PM100D from THORLABS. Stimulation and acquisition were performed with the 60x objective at a zoom of 4 using an image size of 512 x 512 pixels. A stimulation line of 5 µm was defined and microirradiation performed for 6 sec. Six images were taken before micro-irradiation to calculate the basal level of fluorescence. After micro-irradiation, an image was taken every 1 s over 2 min for short and every 2

min over 62 min for long experiments. Between 10 and 20 cells are micro-irradiated in each experiment. The fluorescence intensity at the micro-irradiated region is measured for each time point. Six images are taken before the micro-irradiation in order to quantify a mean of fluorescence that will be considered as the basal level of the protein in the region and used to normalize the measurements (this value is set to 1). In order to quantify the enrichment factor of XRCC1 in the micro-irradiation region the intensity observed for each time point is divided by the mean intensity measured before the micro-irradiation. The mean of at least 10 cells is displayed in the graph. Error bars represent the SEM.

2.5 Results

2.5.1 Residues 219-633 of XRCC1 confer DNA binding independent of the NTD

To systematically analyze and delineate the DNA binding function of XRCC1, we generated several deletions of XRCC1 based on known domain structure. Deletions analyzed included the N-terminal domain (NTD, residues 1-183), the first BRCT (BRCTI, residues 301-415), a proteolytic resistant fragment (residues 219-633) [Ali 2009], and various linker regions (Figure 2.3 A). As XRCC1 has been shown to interact with a variety of DNA substrates, we choose to use 39 base paired duplex substrates (duplex, nicked and gapped) similar to those previously reported [Marintchev 1999] (Figure 2.4). Full length XRCC1 was able to interact with

the DNA substrate as expected (Figure 2.5 B). Unexpectedly, the NTD failed to bind DNA despite using the same XRCC1 domain boundaries used in previous studies and reaction conditions (Figure 2.5 C). Rather truncation 219-633 was also able to interact with DNA with affinity comparable to full length XRCC1 (Figure 2.5 D). These results clearly demonstrate the DNA binding activity present in resides 219-633, and importantly that DNA binding can occur at wildtype levels in the absence of the N-terminal domain.

In order to further refine the boundaries of this DNA binding module a more extensive set of XRCC1 truncations were generated and tested for DNA binding activity. Since the BRCT1 domain of XRCC1 was previously demonstrated to interact with DNA through cross-linking studies [Nazarkina 2007], we tested DNA binding of this domain DNA in the absence of cross-linking agents. Consistent with previous reports, our BRCT1 construct (residues 301-415) was unable to stably interact with DNA (Figure 2.5 E). A construct spanning residues 219-415, to incorporate the N-terminal linker region and the following BRCT1 domain, was found to have DNA binding comparable to the 219-633 truncation (Figure 2.5 F). We then tested whether the 219-300 region alone was able to interact with DNA. The XRCC1 219-300 peptide was generated with an N-terminal NusA fusion tag to favour folding and solubility. A TEV protease cleavage site was also added to the N-terminus to allow removal of the NusA tag. Upon removal of

the NusA fusion, the 219-300 peptide was tested for DNA binding and was found to be unable to bind DNA (Figure 2.5 G). These results suggest that both the BRCT1 domain, as well as the N-terminal linker region 219-300, are required for stable DNA binding within XRCC1.

2.5.2 XRCC1$^{219-633}$ shows no DNA binding specificity

We next sought to determine whether DNA binding of XRCC1$^{219-633}$ displayed DNA binding specificity towards different forms of damaged DNA. Since the NTD of XRCC1 had previously been shown to possess DNA binding specificity, we hypothesized that XRCC1$^{219-633}$ may share a similar ability to distinguish between different repair intermediates. Such activity could be an asset for XRCC1 in order to recruit appropriate repair enzymes at different stages of repair. A series of 39 bp DNA substrates that resemble various intermediates observed during repair were generated by annealing appropriate DNA oligos (Figure 2.4). Specifically, these substrates included nicked and gapped DNA with 5' and 3'OH groups (resembling a direct single stranded break and a substrate requiring end processing respectively), as well as nicked and gapped DNA with a 5'PO_4 and 3'OH group (resembling substrates ready for ligation and gap filling respectively). DNA binding experiments performed with these substrates and XRCC1$^{219-633}$ revealed that binding occurred with similar affinity for all substrates tested, indicating that XRCC1$^{219-633}$ does not support DNA binding specificity (Figure 2.6 center).

For comparison, we also examined DNA binding of the N-terminal domain (XRCC1$^{1-183}$) with nicked/ gapped substrates. Surprisingly, the NTD not only lacked DNA binding specificity, but failed to bind any DNA with affinity comparable to XRCC1$^{219-633}$ (Figure 2.6 right). This result is in stark contrast to previously published findings [Marintchev 1999]. Taken together, results presented here suggest that the DNA binding observed for XRCC1 is attributed to the N-terminal linker region and the BRCT1 domain, and not the NTD.

2.5.3 Mutation of positively charged residues within XRCC1$^{219-300}$ abolish DNA binding

We next proceeded to further identify key residues responsible for mediating DNA binding within the XRCC1$^{219-633}$. An examination of the amino acid sequence within residues 219 to 300 revealed a large number of positively charged residues (Figure 2.7 A). Furthermore, these positively charged residues were found in clusters (2-3 residues in close proximity), which we designated as patch 1, 2, 3, 4 and 5 (P1, P2, P3, P4, P5). We hypothesized that these positively charged residues might mediate ionic interactions with the negatively charged sugar phosphate backbone of DNA, and that elimination of these charged residues would abolish DNA binding. To test this, mutagenesis was performed to generate clustered mutants within XRCC1$^{219-633}$ by replacing charged residues (Arg, Lys) with Ala. The purified mutants were then analyzed for DNA binding using EMSAs

(exemplar gels containing wild type, P1 and P3 mutants are shown in Figure 2.7 B).

The 39 bp DNA substrate was monitored as increasing concentrations of XRCC1 was added. Assuming that the disappearance of the substrate is due to a direct interaction with XRCC1 to form a DNA-protein complex, the amount of complex formed can be indirectly determined by measuring the disappearance of the substrate. The unbound substrate was quantified and compared to a DNA-only control to determine the fraction of unbound substrate, or 'unbound fraction'. The equation: 1 - 'unbound fraction' gives the quantity of the 'bound fraction', which was used to plot the DNA binding curves for each XRCC1 mutant cluster (Figure 2.7 C). Although these mutants were not entirely impeded for DNA binding, several were found to exhibit a lowered affinity. The concentration of protein that resulted in 50% complex formation (or the equilibrium dissociation constant) was 0.7 (+/- 0.01) μM for wild type and 1.62 (+/- 0.06), 1.47 (+/- 0.04), 1.87 (+/- 0.04), 1.22 (+/- 0.02), 1.57 (+/- 0.04) μM for mutants P1, 2, 3, 4, 5 respectively. These results indicate that altering the positive charge of side chains at residues within these clusters only diminished DNA binding to a minor extent. Nevertheless, the fact that individual cluster mutants showed reduced binding suggested that each of the tested residues might contribute additively to DNA binding, and that mutating multiple patches may be required to fully abolish binding. Comparison of the human XRCC1

sequence with that of hamster, frog and *Arabidopsis*, revealed that residues corresponding to P1 and P3 were more highly conserved (Figure 2.8 A). Since *Arabidopsis* XRCC1 is known to interact with DNA substrates [Martínez-Macías et al. 2013], we reasoned that corresponding residues in the human protein may be most important for DNA binding. Indeed, combined mutations in P1 and P3 (Ala substitutions at residues K243, K245, R246, K247, K271, R272, K274) resulted in significantly reduced levels of DNA binding (Figure 2.8 B). At a fixed protein concentration of 2 μM, the combo mutant P1/3 has no detectable DNA binding compared to wild type (Figure 2.8 C). This result strongly suggests that these two positively charged patches within XRCC1 are crucial for DNA binding function.

2.5.4 XRCC1 P1/3 mutation causes early release from sites of DNA damage and abolished foci formation

In order to evaluate the impact of these mutations within a cellular context, we compared XRCC1 localization of wild type and P1/3 mutant in CHO EM9 cells. Yellow fluorescent protein (YFP) was fused to the C-terminus of each protein to permit monitoring by microscopy. Since K243, K245, R246, K247, K271, R272 and K274 are residues previously implicated in forming part of the nuclear localization signal of XRCC1 [Kirby et al. 2015], we fused an additional DPKKKRKV nuclear localization signal after the YFP sequence to facilitate XRCC1 entry into the nucleus. This was

performed for both wild type and P1/3 variant to minimize discrepancies between plasmids.

We first determined if the addition of an NLS to P1/3 mutant protein would impact DNA binding *in vitro*. As shown in Figure 2.9, an EMSA was performed on purified XRCC1$^{219-633}$ P1/3 mutant with a C-terminally fused DPKKRKV sequence. The addition of an NLS sequence resulted in only a minor amount of DNA binding, approximately 20% of wild type. Since DNA binding remained low even with the positively charged NLS sequence present, further cell-based assays were performed using this construct.

The kinetics of XRCC1 recruitment (2 and 60 min) to DNA damage sites was monitored in cells that had been micro-irradiated to generate primarily DNA single stranded breaks. While both wild type and P1/3 variant proteins were initially recruited to damage sites, the P1/3 variant showed a faster dissociation from the site of damage compared to wild type (Figure 2.10). Furthermore, the number of cells with detectable levels of XRCC1 at damage sites was significantly reduced for the P1/3 variant compared to wild type XRCC1 at longer time points (Figure 2.11). Taken together these results suggests that while XRCC1 is able to initially respond to SSBR, the protein rapidly dissociates from the site of DNA damage when DNA binding is diminished.

We also evaluated the ability of the DNA binding-deficient P1/3 variant to respond to DNA single strand breaks generated through hydrogen

peroxide (H_2O_2) treatment. Peroxide treatment generates apurinic sites that result in single stranded breaks within the nucleus. Prior studies have shown the ability of XRCC1 to form repair foci at sites of peroxide-generated DNA damage [Kubota et al. 2009]. Cells expressing either wild type or P1/3 variant XRCC1 were treated with 10 mM H_2O_2 and monitored for XRCC1 foci formation. While wild type XRCC1 was able to form repair foci, the P1/3 variant showed no significant amount of foci formation (Figure 2.12). In agreement with the findings for microirradiation, results using peroxide as a source of damage suggest that DNA binding activity of XRCC1 is required for DNA repair foci formation.

```
XRCC1-Forward-1 (Gateway)
GGGGACAAGTTTGTACAAAAAAGCAGGCTTAGATTACGATATCCCAACGACCGAAAACCTG-
TATTTTCAGGGCATGCCGGAGATCCGCCTCC

XRCC1-Forward-219 (Gateway)
GGGGACAAGTTTGTACAAAAAAGCAGGCTTAGATTACGATATCCCAACGACCGAAAACCTG-
TATTTTCAGGGCTCTAGTGCTGCCTCCTCAGCC

XRCC1-Forward-301 (Gateway)
GGGGACAAGTTTGTACAAAAAAGCAGGCTTAGATTACGATATCCCAACGACCGAAAACCTG-
TATTTTCAGGGCGGAGAAGGCACCGAGCCCAG

XRCC1-Reverse-633 (Gateway)
GGGGACCACTTTGTACAAGAAAGCTGGGTCTCAATGGTGATGGTGATGGTGGGCCTGCGG-
CACCACCCC

XRCC1-Reverse-183 (Gateway)
GGGGACCACTTTGTACAAGAAAGCTGGGTCTCAGTTGGCGCTCTCATCCTC

XRCC1-Reverse-415 (Gateway)
GGGGACCACTTTGTACAAGAAAGCTGGGTCTCAGGCCTCATCCTCCTCACTG

XRCC1-Reverse-300 (Gateway)
GGGGACCACTTTGTACAAGAAAGCTGGGTCTCATCGGGGTTTGCCTGTCACTGC

XRCC1-Forward-219 (LIC)
ATTTCCAGGGAGCAGCCGCATCTAGTGCTGCCTCCTCAGCC

XRCC1-Reverse-633 (LIC)
CCTGCAAAGCACCGGCCTCAGGCCTGCGGCACCACCCC
```

Figure 2.1 Primers used to generate XRCC1 constructs. Each primer is oriented from 5' to 3'. The primers are labelled as forward where the number indicates that starting amino acid residue or reverse where the number indicates the ending residue of the construct. The parenthesis indicates which cloning method was used. Different combinations of forward and reverse primers were used to generate the desired protein construct.

P1
Forward- GCAGGGGCGGCGGCGTTGGATTTGAACCAAGAAGAAAAGAAGACCC
Reverse- CGCCGCCGCCCCTGCGGGAGACTCCTGGGGCTTGGAG

P2
Forward- GCGGCGACCCCCAGCGCACCACCAGCCCAGCTGTCGCCATCTG
Reverse- TGCGCTGGGGGTCGCCGCTTCTTCTTGGTTCAAATCCAACTTCCTCTTCCCTTT

P3
Forward- GCGGCACCTGCATTGCCAGCTCCAACTCGTACC
Reverse- CAATGCAGGTGCCGCGGGAACAGATGGCGACAG

P4
Forward- GCACCCGCAGGAGAAGGCACCGAGCC
Reverse- TCCTGCGGGTGCGCCTGTCACTGCCC

P5
Forward- GCAGCACCCGCAGCTGGCCCAGAGGAGCTGGG
Reverse- AGCTGCGGGTGCTGCGGGCTCGGTGCCTTCTCC

Figure 2.2 Primers used to generate XRCC1219-633 mutants. Each primer is oriented from 5' to 3'. The underlined region is the mutation-containing overhang and the remaining sequence is the plasmid-overlapping region.

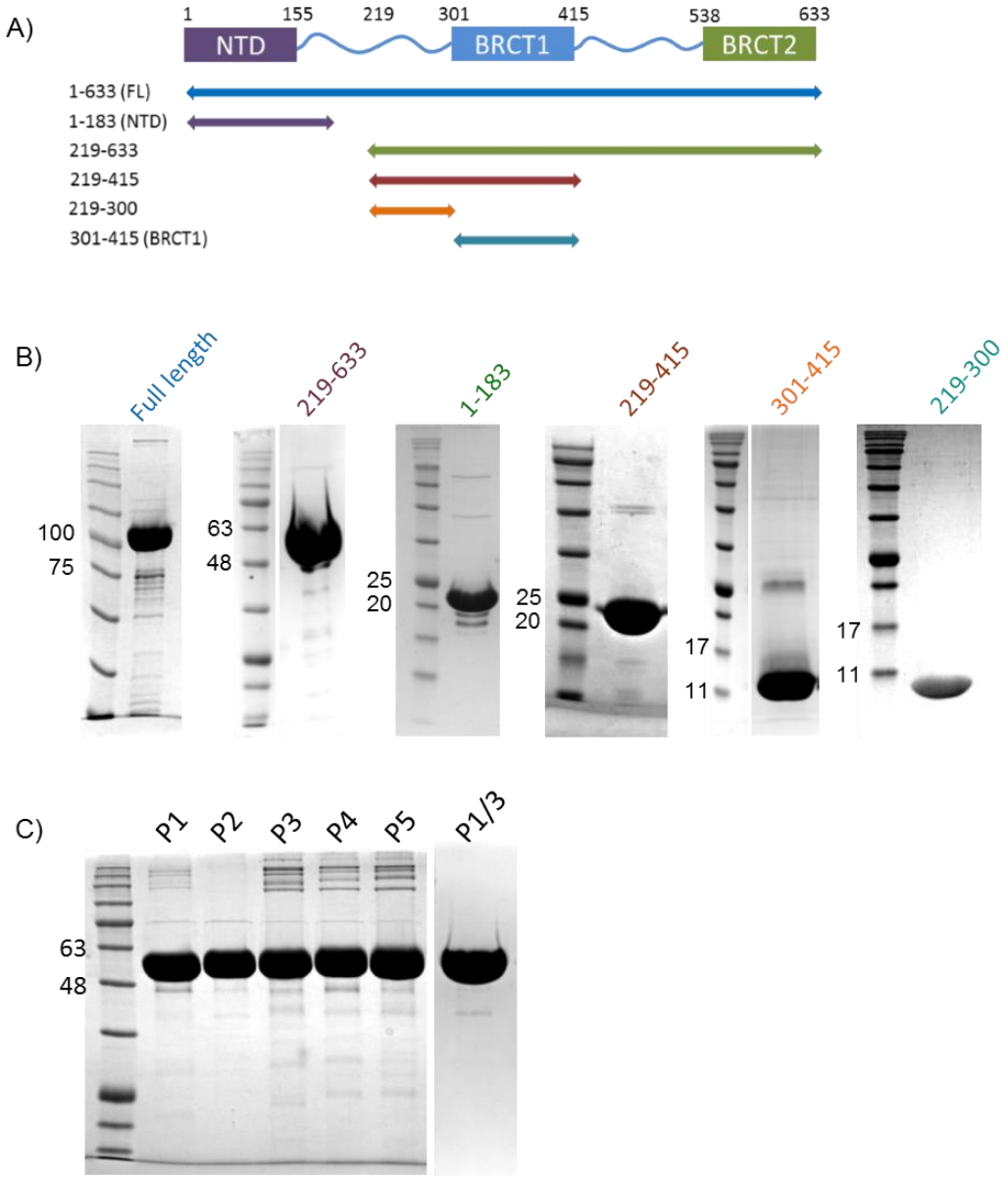


Figure 2.3 Purified XRCC1 truncations. A) Domain organization of XRCC1 with truncation boundaries indicated by arrows. B) SDS-PAGE gels illustrating purity of XRCC1 constructs used in DNA binding experiments: full length (70kDa), 219-633 (46kDa), 1-183 (20kDa), 219-415 (21kDa), 301-415 (13kDa), 219-300 (8kDa). C) SDS-PAGE gel showing purified XRCC1-219-633 mutants. All mutants are approximately 46kDa, but migrate with an apparent molecular weight of approximately 55kDa. A molecular weight marker is shown on left (kDa) for each coomassie stained gel.

Duplex undamaged substrate:

5'CGAGCCATGGCCGC**TA**GGCAGATTTTTTGCGGTGCCAGG

3'GCTCGGTACCGGCGATCCGTCTAAAAAACGCCACGGTCC

End processing substrate:

5'CGAGCCATGGCCGCT_GGCAGATTTTTTGCGGTGCCAGG

3'GCTCGGTACCGGCGATCCGTCTAAAAAACGCCACGGTCC

Gap filling substrate:

PO^4

5'CGAGCCATGGCCGCT_^GGCAGATTTTTTGCGGTGCCAGG

3'GCTCGGTACCGGCGATCCGTCTAAAAAACGCCACGGTCC

Direct single strand break substrate:

5'CGAGCCATGGCCGC**T^A**GGCAGATTTTTTGCGGTGCCAGG

3'GCTCGGTACCGGCGATCCGTCTAAAAAACGCCACGGTCC

Ligation substrate:

PO^4

5'CGAGCCATGGCCGC**T^A**GGCAGATTTTTTGCGGTGCCAGG

3'GCTCGGTACCGGCGATCCGTCTAAAAAACGCCACGGTCC

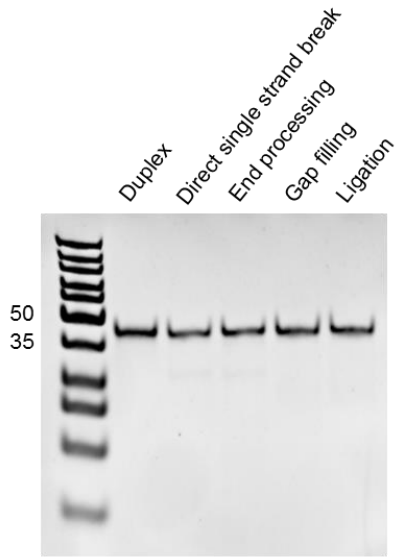


Figure 2.4 DNA substrates used in DNA binding studies (left). DNA substrates were chosen to resemble different repair intermediates. The duplex DNA substrate is 39bp long without any modifications. The End processing substrate is missing a base which is indicated by an underscore, and contains a 3' hydroxyl group on the thymine and 5' hydroxyl group on the guanine. The gap filling substrate contains a missing base which is indicated by an underscore as well as a 3' hydroxyl group on the thymine and a 5' phosphate group on the guanine. The direct single strand break substrate contains a nick which is indicated by an accent mark and also contains a 3' hydroxyl group on the thymine and a 5' hydroxyl group on the adenine. The ligation substrate contains a nick which is indicated by an accent mark and also contains a 3' hydroxyl group on the thymine and 5' phosphate group on the adenine. All substrates possess a 3' 6FAM fluorescent label on the bottom strand. Each sample was ran on a 15% 0.5 TBE gel and stained with ethidium bromide (right). A DNA ladder (in base pairs) is shown on the left most lane of gel for reference.

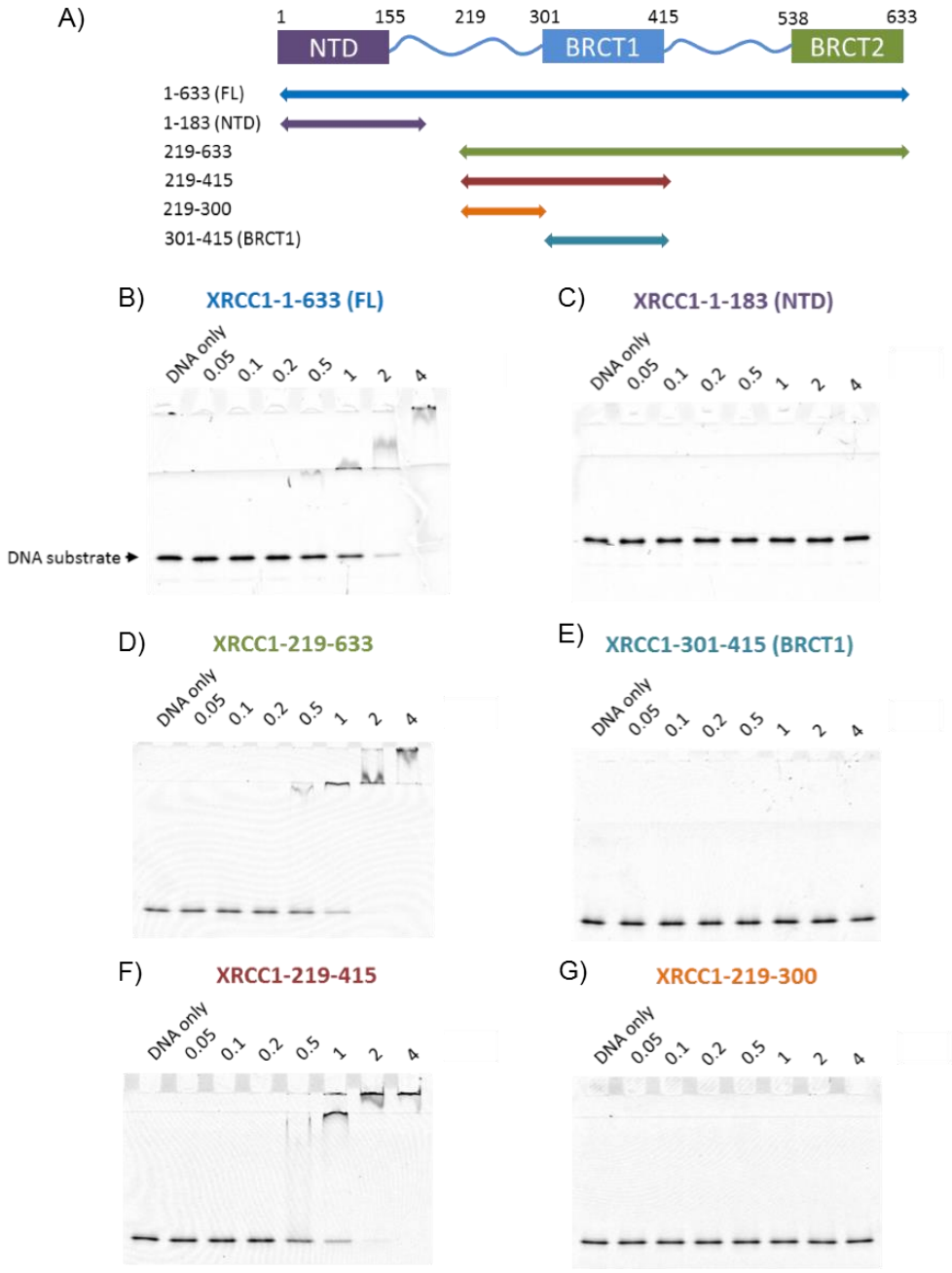


Figure 2.5 Comparison of DNA binding activities of XRCC1 truncations. A) Domain organization of XRCC1 with truncation boundaries indicated by arrows. Full length (blue), 1-183 (purple), 219-633 (green), 219-415 (red), 219-300 (orange) and 301-415 (cyan). B-G) DNA binding activity of XRCC1 truncations (μM concentrations) monitored by electrophoretic mobility shift using fluorescent 39bp duplex DNA substrate. The DNA substrate is indicated in B).

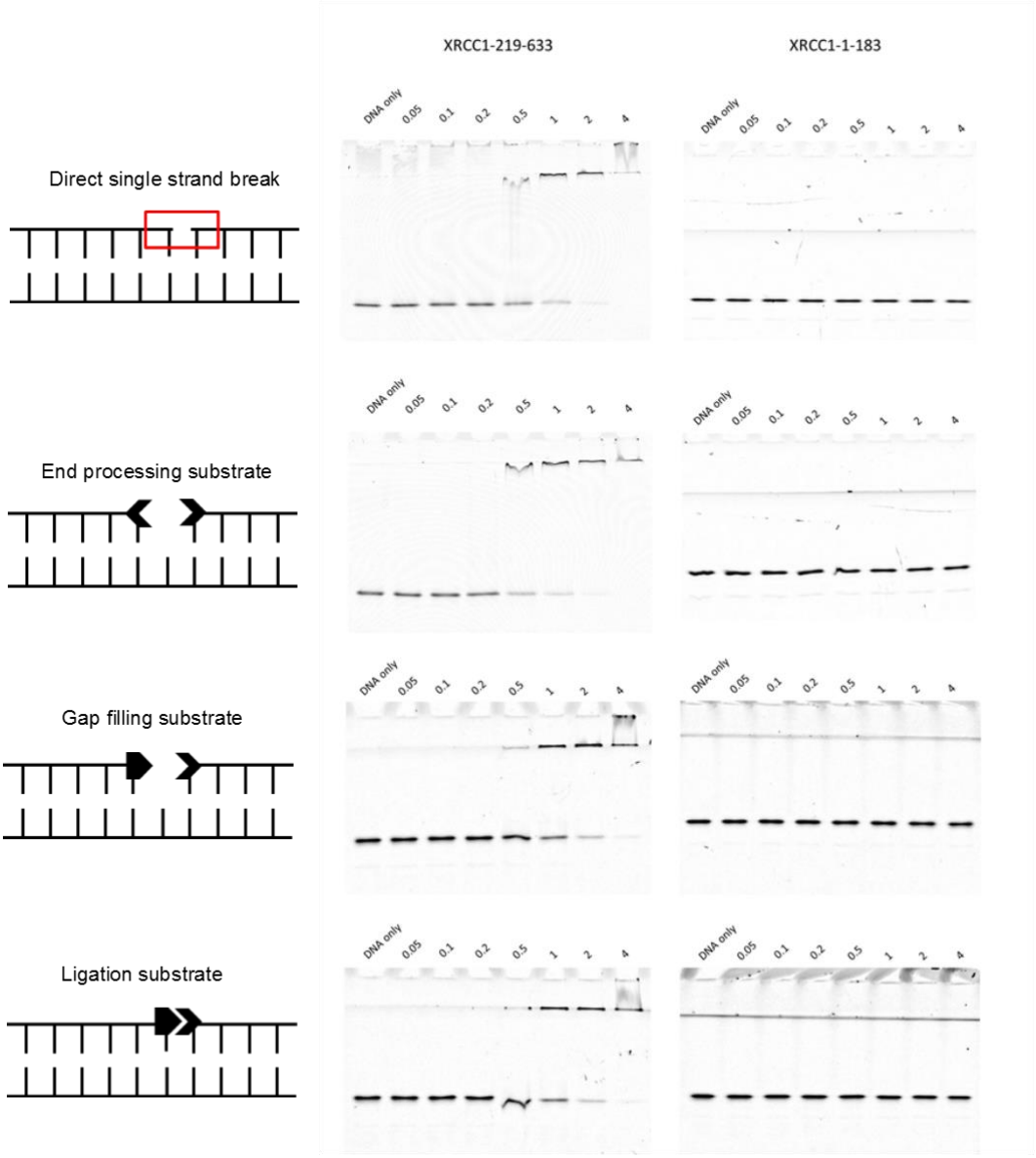


Figure 2.6 Comparison of truncated and full length XRCC1 DNA binding specificity. XRCC1$^{219-633}$ (center) or XRCC1$^{1-183}$ (right) at varying μM concentrations with different DNA repair intermediates (left). XRCC1$^{219-633}$ is able to shift DNA while the NTD, XRCC1$^{1-183}$ is not.

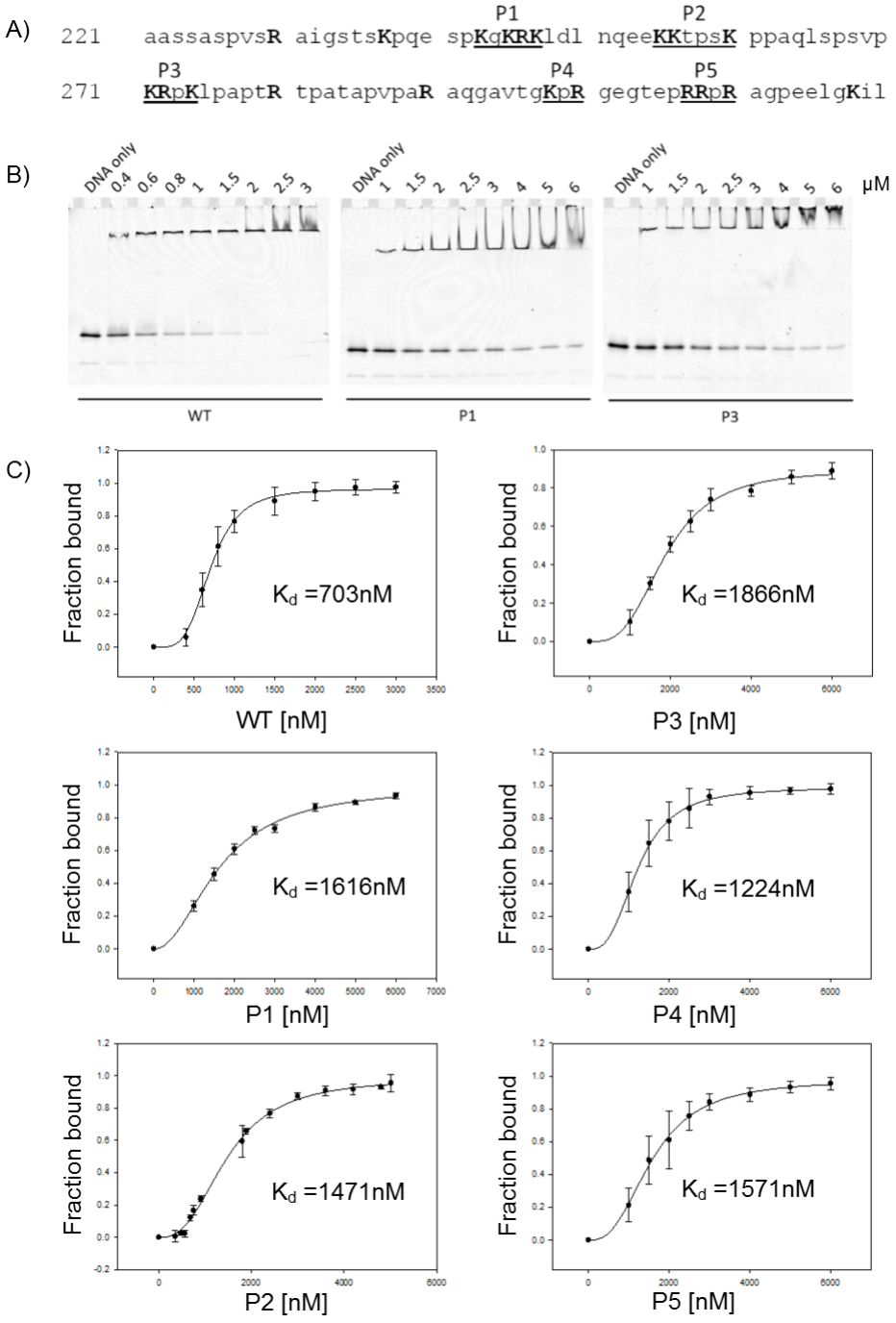


Figure 2.7 Binding curves of mutant P1-5. A) Sequence of the XRCC1 N-terminal linker. Positively charged residues that were targeted for alanine substitution are highlighted. B) Exemplar EMSA gels of XRCC1-219-633 WT, P1 and P3 mutant. C) Binding curves generated from EMSA for each alanine substituted mutant.

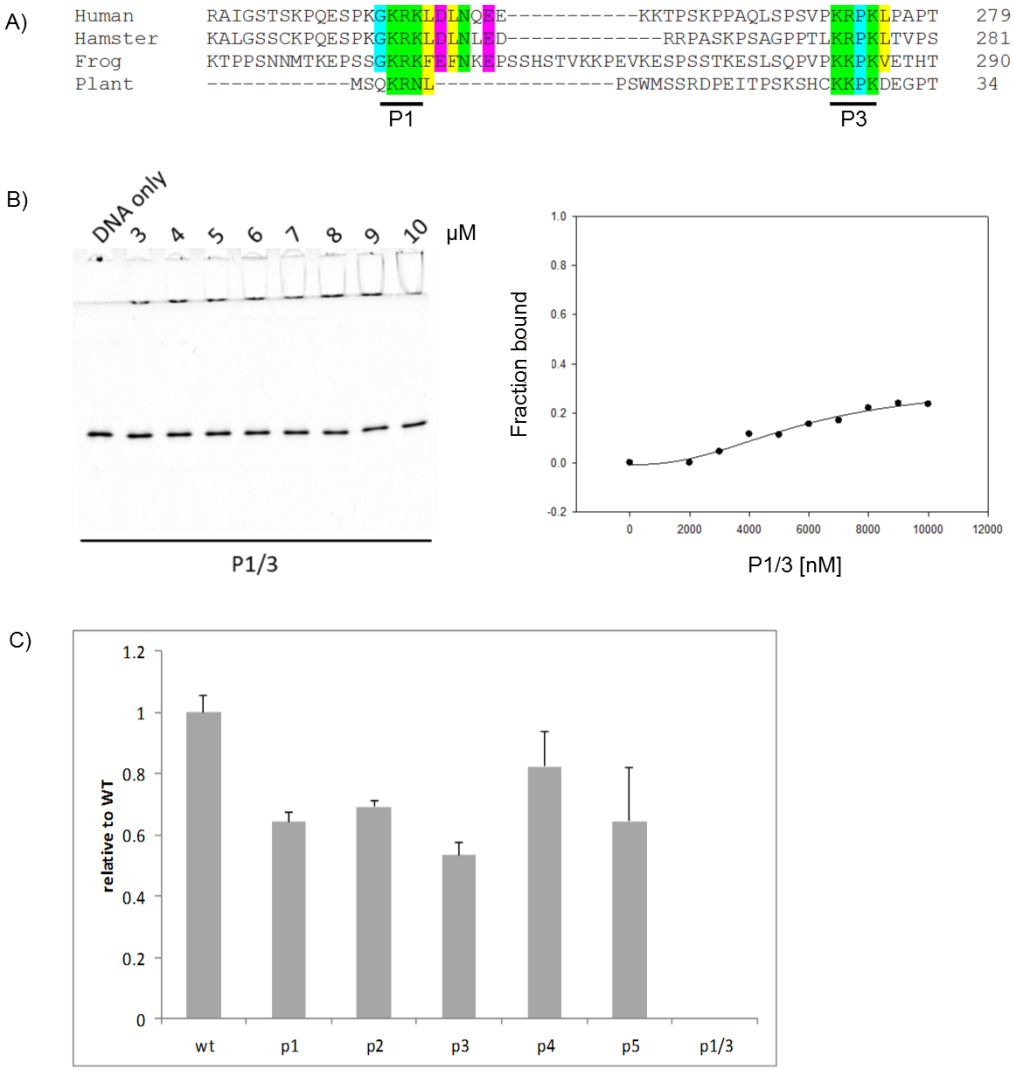
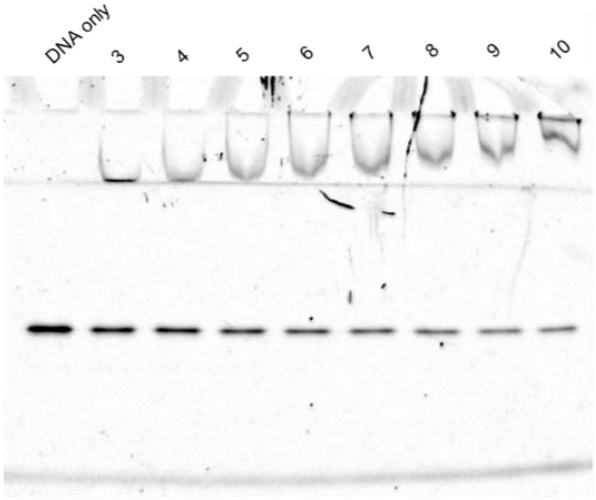


Figure 2.8 Comparison of DNA binding between XRCC1 wild type and mutant variants. A) Sequence alignment of XRCC1 from human, hamster, frog and Arabidopsis (plant). Conserved positively charged residues are highlighted green while other relatively conserved residues are coloured in cyan or purple (residues altering main chain) and yellow (hydrophobic residues). B) DNA binding of XRCC1$^{219-633}$ P1/3 mutant (left) and plotted binding curve (right). Protein concentration ranged from 3-10 μM. C) A comparison of DNA binding levels for mutant and wild type XRCC1 at 2 μM protein concentration. P1/3 possessed no measurable DNA binding at this concentration.

A)



B)

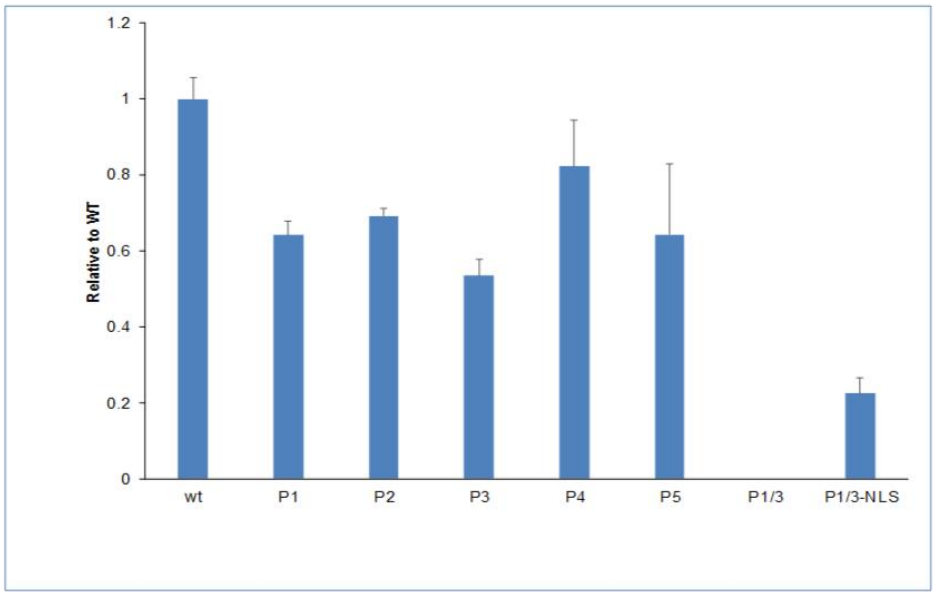


Figure 2.9 A) DNA binding activity of XRCC1$^{219-633}$ P1/3 containing added C-terminal NLS sequence. Protein concentration tested ranged from 3-10 µM. B) Comparison of DNA binding between wild type and various mutants of XRCC1 at 2 µM protein concentration. P1/3-NLS possessed approximately 20% of wild type DNA binding at the protein concentration tested.

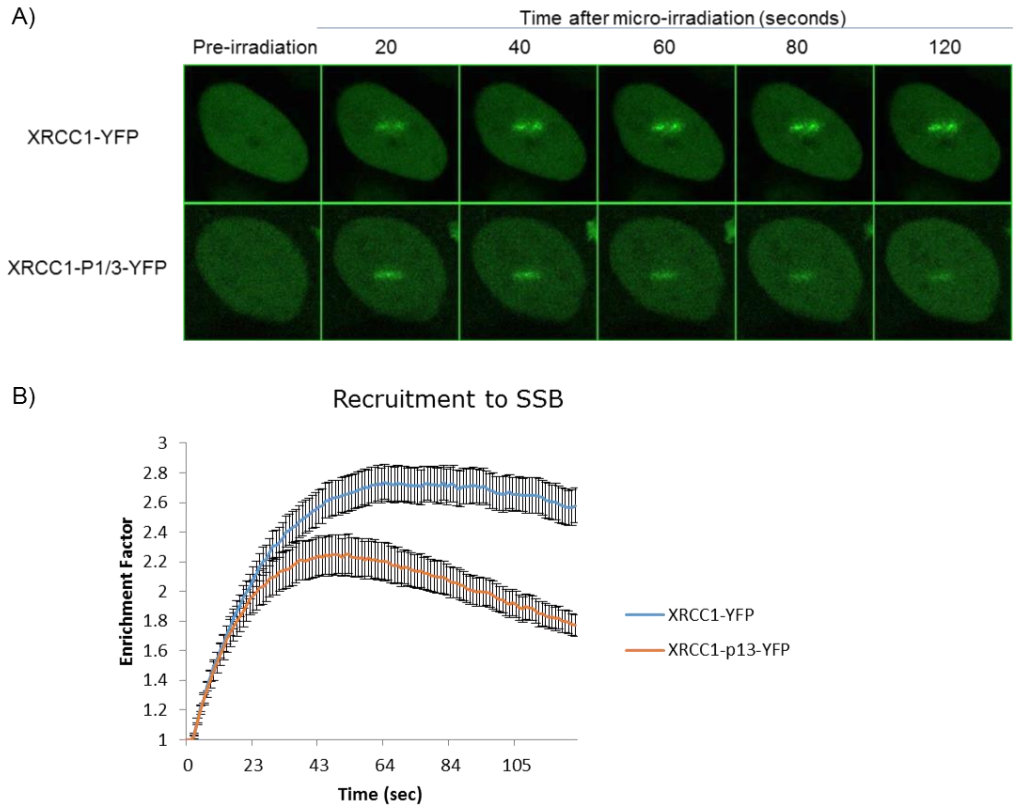


Figure 2.10 Recruitment of XRCC1-YFP and XRCC1-P1/3-YFP to sites of micro-irradiation damage (A). Time course of XRCC1 damage localization from 20-120 seconds post-irradiation. Quantified intensity of XRCC1 localization following micro-irradiation (B).

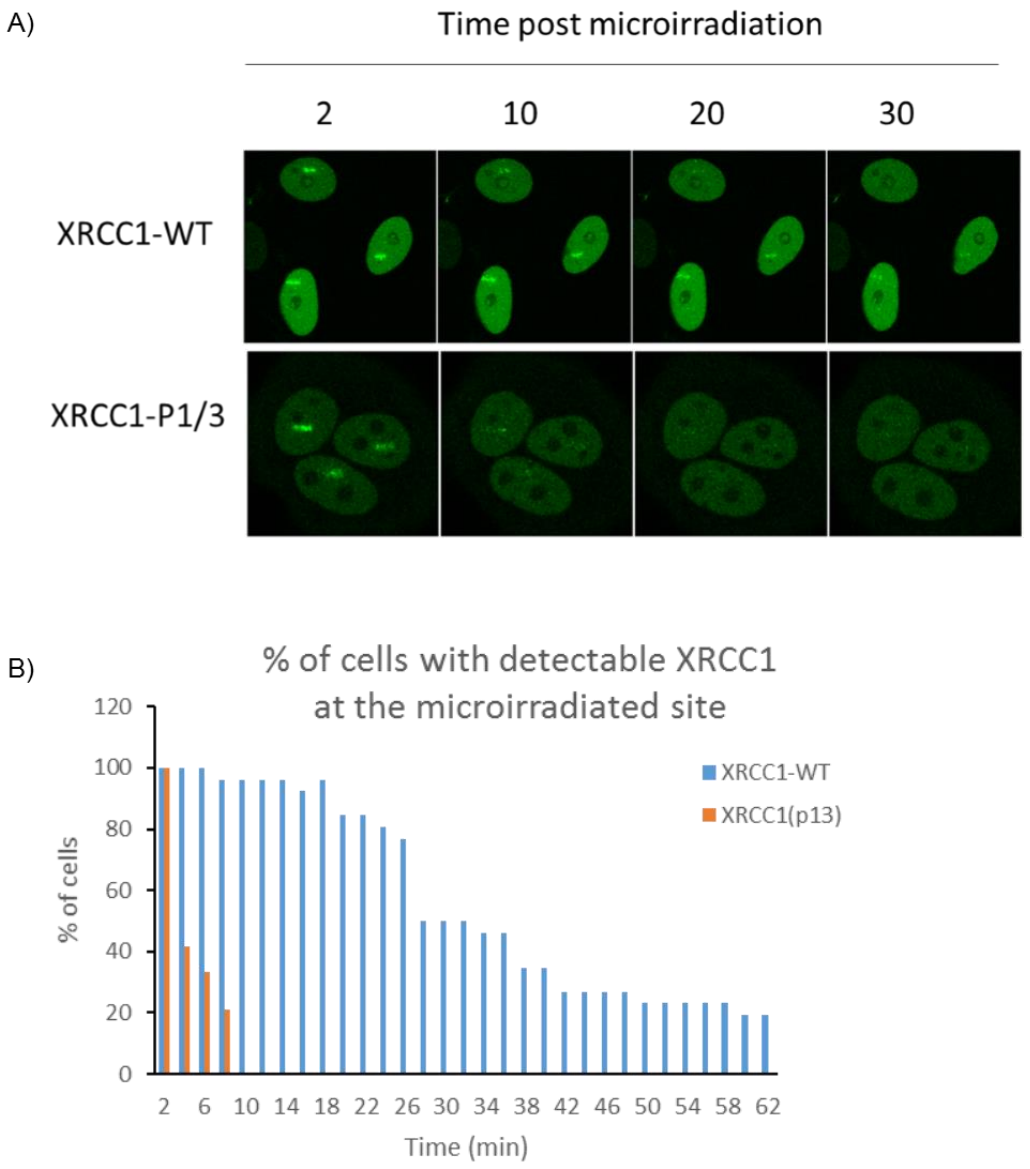


Figure 2.11 Recruitment of XRCC1-YFP and XRCC1-P1/3-YFP to sites of micro-irradiation damage (A). Time course of XRCC1 damage localization. (B) The number of cells showing detectable XRCC1 at damage sites was quantified and plotted as a relative percentage against the time monitored. Blue bars represent wild type protein while the orange bars represent the P1/3 variant.

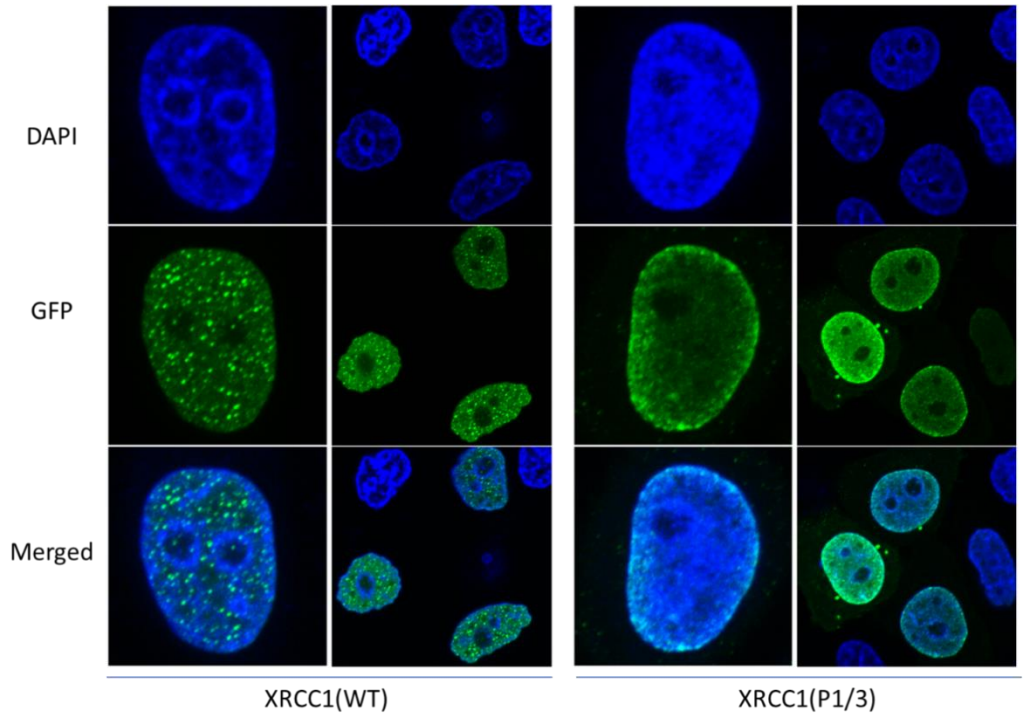


Figure 2.12 XRCC1 foci formation following 10mM H_2O_2 treatment. Cells expressing wild type XRCC1 (left) or P1/3 variant (right). DNA, stained with DAPI (blue colour); XRCC1, green.

2.6 Discussion

The repair of modified DNA bases and single stranded breaks in humans requires well coordinated events orchestrated by the scaffolding protein XRCC1. In order to effectively recruit and displace the many proteins performing repair, XRCC1 is thought to closely associate with DNA at all times during the repair process. A previously reported DNA binding domain in the N-terminal domain of XRCC1 was suggested to be capable of interacting directly with DNA [Marintchev 1999]. However, the significance of this interaction remains unclear as the NTD is also known to be responsible for binding PolB and the two interactions are suggested to be mutually exclusive [Cuneo 2010]. Other groups have also reported DNA binding activity for XRCC1; however, these interactions were dependent on the presence of a cross-linker. The use of chemical cross-linkers and the fact that all three domains of XRCC1 were found to be capable of forming a complex with DNA shed uncertainty on the meaning of these findings [Nazarkina 2007, Yamane 2000]. We therefore sought to further explore the DNA binding activity of XRCC1 in a more systematic way using native conditions. Since the NTD of XRCC1 was suggested to possess DNA binding specificity for damaged DNA using a surface also implicated in PolB binding, one of the key questions to be answered by work reported here is whether XRCC1 is able to engage DNA with regions outside the N-terminal

domain that may be responsible for its longer residency at damage sites throughout the entire repair process.

To answer this question, we first tested various truncations of XRCC1 for the ability to interact with DNA. Our results show that an alternate DNA binding module exists within XRCC1 between residues 219-415. This region, encompasses BRCT1 and the 80 residues N-terminal to it. We term this DNA binding region the **C**entral **D**NA **B**inding domain (CDB). This CDB region appears to function as a single entity as BRCT1 or the 80 residues proceeding it are not able to interact with DNA on their own. The CDB domain showed no DNA binding specificity and interacted with similar affinity for all forms of DNA tested (duplex, double stranded DNA with a nick or gap). This is consistent with an earlier report that demonstrated non-specific DNA binding in full length XRCC1 [Mani 2004]. Unexpectedly, we found that the NTD of XRCC1 did not significantly contribute to DNA binding and that it lacked DNA binding activity when isolated from the rest of the protein. These findings are in contrast to those reported previously [Marintchev et al. 1999]. The basis for the observed difference is unclear as the same domain boundaries and reaction conditions were used in both sets of experiments. It may be that the dominant DNA binding of XRCC1 occurs through the 219-415 region, and that the NTD domain functions in a more specialized way for interaction with damaged DNA.

Since the CDB domain does not appear to be involved in damage recognition it would appear that this DNA binding activity may be responsible for tethering XRCC1 to DNA during the entire repair process, presumably at a site, local but distinct from the actual DNA lesion. In this way XRCC1 could remain associated with DNA during repair while spatially opening up the damaged site. It is interesting to note that the CDB domain overlaps with several protein interacting regions in XRCC1 (PARP1, APE1, MPG, hNTH1, hNEIL2). Most of these protein partners are early factors in SSBR and BER that recognize and process damaged DNA substrates. Whether XRCC1 can interact simultaneously with both DNA and these proteins through the same region remains to be determined. It may be possible that XRCC1-CDB initially interacts with these protein partners when XRCC1 is being recruited, then later interacts with DNA after the partners are displaced. Since XRCC1 is known to increase the reactivity of certain enzymes, it is also possible that interaction with DNA physically alters DNA to favour enzyme processing [Vidal 2001, Whitehouse 2001]. Further experiments will be required to test the validity of these possibilities.

A close inspection of the amino acid sequence within residues 219-300 revealed a large number of positively charged residues. We reasoned that these residues may interact favourably with the DNA backbone to facilitate protein-DNA interactions. Indeed, mutations that removed these residues where found to abolish or weaken DNA binding. Our results indicate that

fully disrupting DNA binding is only possible after several charged residues are changed to alanine (specifically K243, K245, R246, K247, K271, R272 and K274). Altering the side chains of other residues within the 219-300 region had little to no impact on DNA binding. When we compared XRCC1 sequences between human, hamster, frog and *Arabidopsis*, it was found that positively charged residues in regions corresponding to human 243-247 and 271-274 are conserved. It will be particularly interesting to see if altering equivalent residues in *Arabidopsis* XRCC1 disrupts its recently reported DNA binding activity [Martínez-Macías 2013].

Although our findings demonstrate that residues 243-247 and 271-274 are important for DNA binding, another group showed that these residues form part of the NLS sequence needed for entry of XRCC1 into the nucleus [Kirby 2015]. Therefore, it would appear that the NLS sequence of XRCC1 serves two functions (transport into the nucleus and subsequent DNA binding) as has been observed in other proteins [Lacasse 1995].

The rapid dissociation of the XRCC1 P1/3 mutant from sites of micro-irradiated damage in cells suggests that the DNA binding of XRCC1 is important for its function in DNA repair. Since mutations within XRCC1 P1/3 are distal to the PARP1 binding site and are not expected to impact PARP1 binding, it is reasonable that the P1/3 variant may still be recruited to sites of DNA damage. However, the fact that XRCC1 P1/3 quickly dissociates from the damage site suggests that once PARP1 has been removed from

the damage site [Kim et al. 2015], XRCC1 relies on DNA binding from its CDB domain to ensure retention through the remainder of the repair process. It has been previously suggested that XRCC1 does not need to directly interact with DNA, but may indirectly maintain DNA contact through XRCC1 associated partners (e.g. APE1, PolB, Ligase3 etc.). Our results imply the opposite scenario. In particular, findings reported here suggest that XRCC1-partner interactions are insufficient to sustain an association with DNA, and that direct XRCC1-DNA interaction is necessary for it to remain at the site of damage and thereby recruit other repair factors.

Taken together, this work not only reports the discovery of the region within XRCC1 that is responsible for the majority of its DNA binding, but also localizes this interaction to a limited number of residues that mediate the ability of XRCC1 to remain stability bound at sites of DNA damage during repair.

3 Structural characterization of the XRCC1-DNA complex

3.1 Preface of chapter 3

Drs. Alba Guarne and Monica Pillon helped perform data collection, analysis and model generation for small angle x-ray scattering experiments. I performed all other experiments.

3.2 Abstract

XRCC1 is a major DNA repair factor in BER and SSBR. This scaffolding protein interacts with a multitude of repair enzymes as well as DNA. Here we have determined the small angle X-ray scattering (SAXS) model of the XRCC1 minimal DNA binding unit (CDB, which encompasses an 80 residue N-terminal linker and the central BRCT domain) alone, and in complex with DNA. The model reveals a surprisingly structured N-terminal linker that is enforced by a large number of proline residues. We also demonstrate that although these proline residues appear important for rigidity in the linker region, they are not required for DNA binding activity *in vitro*.

3.3 Introduction

XRCC1 regulates DNA base excision repair and single strand break repair by interacting with proteins as well as DNA. While many protein-protein interactions involve structural domains, the flexible linker regions have also been found to be critical for mediating these interactions. Each of the structural domains, namely the NTD and the two BRCT domains are connected by two flexible linker regions. The first is important for interaction with APE1, REV1 and Importin α, while the second is important for binding to PNKP and APTX [Vidal et al. 2001, Gabel et al. 2013, Kirby et al. 2015, Whitehouse et al. 2001]. These linkers have also been shown to undergo

post translational modifications that regulate interaction with repair factors such as APTX during repair.

Linkers serve a very important regulatory function in biological systems. As an example, a linker within the Src family of kinases is vital for repressing catalytic activity by mediating an interaction between the catalytic domain and Src homology domains (SH2 and SH3) [George et al. 2003]. Mutation to this linker eliminates the observed repression, highlighting the importance of linkers in enzymatic activity. Additionally, linkers in multi-domain proteins, such as those in polyketide synthases, are important for ensuring the efficiency of multi-step reactions. Furthermore, the linker length and composition can also influence the stability, folding and orientation of structural domains [George et al. 2003].

Large linkers predominantly adopt helical or coiled structures [George et al. 2003]. The rigidity imposed by helical linkers can be important for spatially isolating functional domains and ensuring minimal reactivity and/or proper folding. Residues such as Leu, Arg, Asp, Met, and Gln have a high propensity in helical linkers [George et al. 2003]. Non-helical linkers, however, show a different preference for amino acids, favoring Pro, Thr, Phe, His, Ser, Gly and Arg. Proline residues are particularly interesting as they provide a high level of rigidity within the polypeptide main chain by virtue of a sterically constrained ring structure. As well, due to the lack of an amide hydrogen, proline residues exhibit reduced residue-to-residue

interaction. Consequently, prolines can provide a rigid and 'isolated' region separated from other folded domains. Additionally, if several proline residues exist in close proximity they have the ability to from poly-proline type 2 (PPII) helices, which exists as extended left-handed helices with three residues per turn [Adzhubei et al. 2013]. In some instances, PPII helices have been shown to be important for mediating protein-protein, as well as protein-nucleic acid interactions [Adzhubei et al. 2013].

Within residues 219 to 415 of XRCC1 there are a large number of prolines. Since PPII helices have been implicated in protein-DNA interactions, we hypothesized that these proline residues may form a structurally stable unit that contributes to the observed XRCC1-DNA binding localized within this region. Here we show by SAXS analysis that a previously presumed flexible region of XRCC1 (residues 219-300) adopts a polyproline superstructure. Furthermore, we demonstrate that disruption of this superstructure has no significant impact on DNA binding activity within XRCC1.

3.4 Materials and methods

3.4.1 Expression vectors and mutagenesis

For SAXS experiments, the truncation XRCC1$^{219-415}$ was cloned into pDEST544 as described in section 2.4.1. To improve protein yield for crystallography and DNA binding experiments, XRCC1$^{219-415}$ was cloned into pDEST17 using the Gateway system. The BP entry clone containing

the XRCC1$^{219-415}$ gene (3.5 µL, described in section 2.4.1) was mixed with pDEST17 (0.5 µL, Invitrogen) and LR clonaseII (1 µL, Invitrogen). The reaction was incubated at room temperature overnight. The mixture was treated with proteinase K (1 µL, Thermo Scientific) at 37 °C for 30 min. The reaction was then transformed into TOP10 cells and plated on LB agar plates containing ampicillin (0.1 mg/mL). Single colonies were selected for overnight cultures, which were grown in LB containing ampicillin (0.1 mg/mL) at 37 °C overnight. Overnight cultures were then harvested by centrifugation at 3005 x g for 10 min. The plasmid was extracted by using a miniprep kit (Geneaid) according to manufacturer's instructions.

Mutagenesis was performed as described in section 2.4.2. Primers used for PCR are listed in Figure 3.1. One proline mutant PA (P267S, P270S, P273S, P276S) was generated in the XRCC1$^{219-633}$ truncation (described in section 2.4.1), while the remaining mutants, PB (P278S, P282S) and PC (P286S, P288S), were generated in the XRCC1$^{219-415}$ truncation (described in section 2.4.1). All mutations were confirmed via DNA sequencing.

3.4.2 Protein expression and purification

All proteins were expressed in Rosetta pLysS cells. Cells were grown in 4L of LB containing ampicillin (0.1 mg/mL) at 37˚C to an OD_{600} of 0.3-0.5 and induced with 1 mM IPTG at 20˚C overnight. Cells were then harvested by centrifugation at 3,315 x g for 15 min. Cells were resuspended in NiA

buffer (20 mM Tris pH 8, 500 mM KCl, 3 mM BME, 10% (v/v) glycerol, 10 mM imidazole) and lysed using a cell disruptor (Avestin EmulsiFlex-C5). The lysate was then centrifuged at 48, 384 x g for 45 min to remove debris. Proteins were initially purified by IMAC using a 5 mL column (GE Healthcare). The bound protein was washed with NiA buffer containing 10mM imidazole (20 column volumes) followed by 30 mM imidazole (10 column volumes) before elution with 300mM imidazole. Eluted protein (10 mL) was mixed with water (10 mL), 10 x buffer (2.5 mL of 500 mM Tris pH8, 10 mM EDTA) and TEV protease (2.5 mL, 1 mg/mL) to remove affinity tags. The sample was incubated at 4 °C overnight. Cleaved protein was further purified using ion exchange chromatography (Amersham). An 8 mL MonoQ column was used for XRCC1$^{219-633}$ PA mutant while an 8 mL MonoS column was used for XRCC1$^{219-415}$ wild type and mutants, using 20 mM Tris (for MonoQ) or HEPES (for MonoS) pH 8, 3 mM BME, and 0 mM KCl (for binding buffer) or 500 mM KCl (for elution buffer). The relatively pure proteins were re-purified by a second IMAC purification step to remove any remaining His-tagged protein. Purified proteins were buffer exchanged into storage buffer (20 mM Tris pH 8, 200 mM KCl, 3 mM BME, 10% (v/v) glycerol) and stored at -80°C.

3.4.3 DNA substrate generation

All DNA was purchased as single stranded oligonucleotides from BioBasic/IDT. Oligos were first dissolved in water and annealed using a

thermocycler (Thermo scientific) to generate the desired base pair lengths of 30, 25, 22, 20, 18, 15 and 10. Annealing was verified by polyacrylamide gel electrophoresis and concentration of each substrate was determined by Nanodrop (Thermo scientific).

3.4.4 Differential scanning fluorimetry

Commercial stock 5000 x sypro orange dye (Thermo Scientific) was diluted to 50 x with water. 5 µL of XRCC1$^{219-415}$ (56 µM originally in 20 mM Tris pH 8, 200 mM KCl, 3 mM BME, 10% (v/v) glycerol) was diluted with 45 µL of fourteen different buffers from the JBS solubility kit (Jenabioscience). These include 100 mM Na/KPO$_4$ pH 5, Sodium citrate pH 5.5, Na/KPO$_4$ pH 6, ADA pH 6.5, ammonium acetate pH 7, Na/KPO$_4$ pH 7, HEPES pH 7.5, Tris pH 7.5, EPPS pH 8, Bicine pH 8.5, Tris pH 8.5, CHES pH 9, CHES pH 9.5, CAPS pH 10. Diluted protein was then mixed with 5 µL of 50 x sypro orange and transferred to a 96 well PCR plate (Bio-Rad Laboratories) where fluorescence was monitored in a CFX96 Touch real time PCR machine (Bio-Rad Laboratories). The samples were heated starting at 25 °C and increased 0.5 °C every 10 seconds. Data was analyzed using CFX manager software (Bio-Rad Laboratories) to determine differences in thermal stability in the presence of different salts and buffers.

3.4.5 XRCC1-DNA complex crystallization

Purified XRCC1$^{219-415}$ was buffer exchanged into 20 mM CAPS pH10, 50 mM KCl. The protein was concentrated to 20-40 mg/mL (0.9-1.8

mM) as determined by Nanodrop (molecular weight 21.4k Da, extinction coefficient 15500 $M^{-1}cm^{-1}$). Equal volumes of protein (1.5 mM) and DNA substrate (1.8 mM) were mixed and incubated at room temperature for one hour. A slight cloudiness occurred during initial mixing but cleared after incubation. Hanging drop vapor diffusion was used for crystallization experiments. The first crystal form of XRCC1 was generated with a 20 bp substrate and mother liquor containing 50 mM MES pH 6.5, 1 mM spermine tetrahydrochloride, 25% (v/v) PEG400. A second crystal form was generated with 22F/18R DNA substrate and mother liquor containing 0.1 M sodium chloride, 0.2 M potassium chloride, 0.02 M magnesium chloride hexahydrate, 0.05 M Bis-Tris pH 7.0, 35% (w/v) PEG 2000.

3.4.6 Small angle X-ray scattering of DNA, XRCC1 and XRCC1-DNA complex

The $XRCC1^{219-415}$ was buffer exchanged on a size exclusion column (Superdex 200; GE Healthcare) equilibrated with 20 mM Tris pH 8.0, 100mM KCl, 3mM BME. Purified 39 base-pair duplex DNA substrate described in Figure 2.4 was buffer exchanged into 20mM Tris pH 8.0, 10 mM KCl, 3 mM BME using a 3k Da MWCO nanosep spin column (Pall). DNA was concentrated and the buffer exchanged four times to ensure complete buffer.

Scattering data for XRCC1 was measured at concentrations of 266, 201, 187 and 93 µM. Samples were prepared by diluting concentrated

protein to appropriate concentrations with 20 mM Tris pH 8.0, 100 mM KCl, 3 mM BME. Scattering data for the 39 bp duplex DNA substrate were measured at concentrations of 188, 142, and 96 µM by diluting the concentrated stock DNA in 20 mM Tris pH 8.0, 100 mM KCl, 3 mM BME. The XRCC1-DNA complex was prepared by mixing XRCC1 and DNA at a 1:1 molar ratio to give a final concentration for each component of 188, 141 and 94 µM. All samples were centrifuged at 10,000 x g for 10 min before loading into capillary tubes to remove any particulates from the solution. Scattering data was measured on a Rigaku BioSAXS-1000 instrument at 10 ^{o}C for 2 hours. SAXSLab 3.0.0r1 software (Rigaku) was used to generate scattering curves. The lack of radiation damage was confirmed by comparing the scattering data at the beginning and end of data collection. Comparison and analysis of the scattering curves was done using the ATSAS 2.6.0 suite [Petoukhov et al. 2012]. Samples were devoid of inter-particle interactions as judged from the Guinier plots, and folded as judged from the Kratky plots. Scattering curves were generated by merging the low q range from the most diluted samples with the higher q range from the most concentrated samples using the automerge tool in the ATSAS 2.6.0 suite. Radius of gyration and pair-distance distribution functions were determined using Primus and GNOM. The reported molecular weights were calculated based on the volume of correlation.

3.4.7 DNA electrophoretic mobility shift assays

DNA concentration was held constant at 5 nM for all reactions. Prior to the addition of protein, the reaction mixture (18 µL total volume) contained DNA substrate in binding buffer (10 mM Tris pH 8, 100 mM NaCl, 3 mM $MgCl_2$, 1 mM EDTA, 0.1% (w/v) CHAPS, 10 mM BME, 7.5% (v/v) glycerol, 0.02 mg/mL bovine serum albumin). Bovine serum albumin (New England Biolabs) was initially stored in 20 mM potassium phosphate pH 7, 50 mM NaCl, 0.1 mM EDTA 5% (v/v) glycerol. Protein (2 µL) was added to each reaction mixture to a final volume of 20 µL. All proteins were initially stored in 20 mM Tris pH 8, 200 mM KCl, 3 mM BME, 10% (v/v) glycerol. The final reaction mixture, including the no XRCC1 control, contained 12 mM Tris pH 8, 100 mM NaCl, 20 mM KCl, 3 mM $MgCl_2$, 1 mM EDTA, 0.1% (w/v) CHAPS, 10 mM BME, 8.5% (v/v) glycerol. The reactions were incubated at room temperature for 1h. Electrophoresis was performed on a 0.5 x TBE buffered 15% polyacrylamide non-denaturing gel with a stacking layer (4% polyacrylamide, 140 mM Tris pH 8). Electrophoresis was performed at room temperature, 100 V for 90-120 min. Gels were imaged using a Chemidoc (Biorad). For gels that were quantified, the relative intensity of the DNA substrate was quantified by ImageJ. Each measurement was performed in triplicate and the average was calculated using Microsoft Excel 2007. Values were imported into Sigmaplot 12 and regression analysis was performed. Curve fitting seemed most reasonable using the 3 parameter Hill

equation function: $f = \frac{ax^b}{c^b + x^b}$ where f is the fractional occupancy, x is the free ligand concentration, a is the maximum y-axis value, b is the Hill coefficient and c is the protein concentration where 50% of DNA is in complex (or the microscopic dissociation constant).

3.5 Results

3.5.1 Differential scanning fluorimetry reveals optimal buffers for XRCC1$^{219\text{-}415}$ stability

A crystal structure of XRCC1 bound to DNA would be highly informative in identifying key sites of interaction between protein and DNA. Prior to conducting crystallization experiments, differential scanning fluorimetry was performed on XRCC1$^{219\text{-}415}$ to identify an optimal buffer for protein stability. This technique uses a fluorescent dye (sypro orange) that exhibits fluorescence when bound to hydrophobic side chains typically abundant, but buried within the core of proteins [Niesen et al. 2007]. Thermal denaturation of a protein exposes the hydrophobic core and results in an increased fluorescence. Fluorescence can be easily monitored using a real time PCR machine and plotted to determine the melting temperature of the protein. Under protein stabilizing conditions, the temperature at which thermal denaturation occurs is increased. In contrast, melting temperature decreases under destabilizing conditions. The melting temperature can then be used to assess whether a given buffer is stabilizing or destabilizing toward a protein of interest. Fourteen different buffers were tested for their

ability to improve thermal stability of XRCC1. Analysis of the data revealed two general trends (Figure 3.2). First, XRCC1 is most stable around neutral pH, as shifts towards acidic or basic conditions reduced the melting temperature. Second, buffers with close to neutral pH containing sodium/potassium phosphate increase the stability of the protein, perhaps by mimicking phosphates present within the DNA backbone. This experiment indicated that sodium/potassium phosphate at neutral pH is able to stabilize XRCC1 and therefore these buffer conditions were initially chosen for XRCC1 crystallization trials.

3.5.2 Crystallization of DNA-bound XRCC1$^{219-415}$

Structural information of XRCC1$^{219-415}$ in complex with DNA would provide crucial insight for understanding how XRCC1 engages DNA during repair. Despite an extensive amount of effort to crystallize the XRCC1-DNA complex, no diffraction quality crystals were obtained. The HELIX screen (Molecular Dimensions) was chosen as the primary kit for crystallization trails because of its specific development for DNA/RNA samples. Since the ideal DNA substrate length was unknown, a series of DNA substrates with 25, 22, 20, 18, 15 and 10 base pair lengths were used along with varying lengths of complementing overhangs (Figure 3.3). Overhangs were included to potentially aid DNA substrate in assembly of larger polymers within the crystal lattice. From differential scanning fluorimetry experiments, XRCC1$^{219-415}$ was found to be most stable in the presence of phosphate and

neutral pH. Consequently, phosphate buffer at pH7 was initially used as the main buffer component for the protein. Unfortunately, this buffer lead to many false positive salt crystals (confirmed by X-ray diffraction) and therefore alternative buffers were used between pH 6-8, including Tris pH 6, Tris pH 8, HEPES pH 7.5. Many crystallization trials were observed to form significant amounts of phase separation (oil drops forming). Although this behavior does not preclude crystallization, attempts were made to disrupt phase separation by altering temperature and salt concentration. Unfortunately, changing these parameters had little impact on the formation of phase-separated drops.

XRCC1$^{219-415}$ has a high isoelectric point (PI) of 10.5. Prior studies with a variety of proteins have suggested that choosing a buffer with pH close to that of a protein's PI may increase the ability to form crystals [Kirkwood et al. 2015]. Hence, several buffers with a higher buffering range such as CAPS pH10 were tested with XRCC1 during crystallization trials. Increased buffer pH significantly reduced the amount of phase-separation observed. Using higher pH buffers allowed identification of several conditions that generated small crystals. These crystals only formed if the DNA substrate was greater than 20bp in length (Figure 3.4) which is consistent with our finding that smaller sized DNA substrates bind XRCC1 with reduced affinity. Although X-ray diffraction of these crystals indicated they were not salt, no visible diffraction could be seen even at lower

resolution. DNA-protein crystals frequently diffract poorly and often require significant optimization of DNA length and end structure to yield suitable crystals for data collection. Therefore, conditions identified here for crystallization of XRCC1-DNA may still prove useful once further optimization of DNA substrates is conducted.

3.5.3 XRCC1$^{219-415}$ SAXS analysis reveals a structured polyproline extension involved in DNA interaction

Although small angle X-ray scattering does not afford the high resolution potential of traditional X-ray crystallography, there are less restrictions on data acquisition and results are obtained from solution, which may provide less biased information compared to that from solid crystalline lattices. We therefore sought to characterize the structure of the CDB domain of XRCC1-DNA using Small Angle X-ray Scattering (SAXS). SAXS uses scattering of molecules in solution to construct low resolution models which in turn provide structural information. SAXS data were acquired for XRCC1$^{219-415}$ alone, 39bp DNA substrate alone, and the DNA-bound XRCC1 complex.

The 39bp DNA substrate used in the aforementioned EMSAs (in chapter 2) was analyzed at concentrations of 188 µM, 142 µM, and 96 µM (4.5, 3.4 and 2.3 mg/mL respectively). Guinier plot analysis of intensity versus low scattering angles showed a linear correlation, indicating the absence of inter-particle interactions. Using volume of correlation analysis, the molecular weight was estimated for each sample (Table 3.1). Samples

were observed to exhibit slight concentration effects, as revealed by the change in molecular weight with changing concentrations. Hence the scattering curves at 188 µM and 94 µM were merged and further analyzed. Since the theoretical molecular weight of DNA is 24k Da, and the experimental determined molecular weight ranged between 19.6 and 23.9 kDa, the sample was judged to be monomeric in solution. An *ab initio* model was generated using DAMMIN (Figure 3.5 A) [Petoukhov et al. 2012]. The shape of the P(r) function indicates the presence of an elongated molecule, as one would expect for a short segment of double stranded DNA. The calculated dimension of the 39bp DNA model (28 x 112 Å) was comparable to theoretical values (20 x 133 Å) [Pray 2008, Watson et al. 1953].

XRCC1$^{219\text{-}415}$ in the absence of DNA was analyzed at concentrations of 266 µM, 201 µM and 56 µM (5.7, 4.3, 1.2 mg/mL respectively). Guinier plot analysis of intensity versus low scattering angles showed a linear correlation, suggesting there are no inter-particle interactions. Using volume of correlation analysis [Rambo et al. 2013], the molecular weight was estimated for each sample (Table 3.1). Since the expected molecular weight of the XRCC1$^{219\text{-}415}$ is 21.4k Da, and the experimental molecular weight ranged between 25 to 26k Da it would appear that XRCC1 remained monomeric under the experimental conditions used. An *ab initio* model was generated using DAMMIF with Chi square value of 1.03 (Figure 3.5 B). Further rigid body modeling performed

using BUNCH [Petoukhov et al. 2012] with the known XRCC1-BRCT1 structure (PDB ID 2D8M). Chi square values for resulting model was 1.23. Analysis of the model suggests that the BRCT1 domain retains a globular form as expected, while the N-terminus exhibits an extended conformation. Interestingly, the N-terminal linker did not adopt a flexible random coil as was predicted by secondary structure prediction, rather it appears to have a rigid structure resembling a cylinder (Figure 3.6). Overall, the N-terminal region of XRCC1$^{219-415}$ forms a 'claw-like' extension that protrudes away from BRCT1 and may be important for positioning of positively charged residues within this region for XRCC1 DNA binding.

DNA-bound XRCC1$^{219-415}$ complex was analyzed by SAXS using a 1:1 molar ratio at concentrations of 188, 141 and 94 µM (4, 3, 2 mg/mL respectively). Guinier plot analysis of intensity versus low scattering angles indicated a linear correlation, suggesting there were no inter-particle interactions. Using volume of correlation analysis, the molecular weight was estimated for each sample (Table 3.1). Samples exhibited slight concentration effects, as revealed by the change in molecular weight with changing concentrations (Table 3.1). As a result, the scattering curves at 188 µM and 94 µM were merged and further analyzed. An *ab initio* model was generated using the multiphase modelling program MONSA (Figure 3.7 A) [Petoukhov et al. 2012]. Using the scattering intensities of the individual components as input (DNA and protein either separately or together),

MONSA was able to generate models that fit the observed scattering of the complex. Scattering intensity of XRCC1$^{219-415}$ alone, DNA alone, and both protein and DNA together were input into MONSA and the program allowed to run ten times. The resulting 10 models were remarkably similar regardless of the initial input scattering intensity, suggesting high confidence and quality of the final models. BRCT1 was readily identified in the model by virtue of its globular structure, thus permitting orientation of the extended XRCC1$^{219-415}$ molecule. All models predominantly showed the interaction between DNA and XRCC1 occurring at the extended N-terminus, with others suggesting the involvement of the BRCT1 domain (Figure 3.7 B). Taken together, these models imply that DNA interacts with XRCC1 using a combination of residues from both the N-terminal extension and BRCT1 domain.

3.5.4 Proline residues within XRCC1$^{219-415}$ N-terminal extension have minimal impact on DNA binding

Our SAXS model of XRCC1$^{219-415}$ suggested that proline residues within this region give rise to a rigid structural element in what was presumed to be a flexible region. Based on localization of DNA binding to this region of XRCC1 in Chapter 2, we hypothesized that the rigid nature afforded by proline residues may be important for mediating protein-DNA interactions, and that disruption of this structure might diminish DNA binding. We tested this possibility by substituting proline residues with

serine and then analyzing these mutants for DNA binding activity. Inspection of the primary amino acid sequence of XRCC1 indicated that several proline residues are clustered in close proximity, hence we substituted multiple proline residues to serine within the same mutant (Figure 3.8). In total, three proline cluster mutants were generated: PA (P267S, P270S, P273S, P276S), PB (P278S, P282S) and PC (P286S, P288S). Surprisingly, when evaluated for DNA binding, all three mutants (PA, PB and PC) showed little to no effect on the ability to interact with DNA. These results suggest that although proline residues within the N-terminal extension of XRCC1$^{219\text{-}415}$ may be responsible for formation of rigid structure in this region they do not appear to be required for DNA binding.

PA
Forward- TCATCTGTTTCCAAGAGATCTAAATTGTCAGCTCCAACTCGTACCCCAGCCACAG
Reverse- TGACAATTTAGATCTCTTGGAAACAGATGACGACAGCTGGGCTGGTGGTTTGCTG

PB
Forward- TCAACTCGTACCTCAGCCACAGCCCCAGTC
Reverse- TGAGGTACGAGTTGAAGCTGGCAATTTAGGTCT

PC
Forward- TCAGTCTCTGCCCGAGCACAGGGGGCAGTG
Reverse- CGGGCAGAGACTGAGGCTGTGGCTGGGGTA

Figure 3.1 Primers used to generate XRCC1$^{219-633}$ and XRCC1$^{219-415}$ mutants. Each primer is oriented from 5' to 3'. The underlined region is the mutation-containing overhang and the remaining sequence is the plasmid-overlapping region.

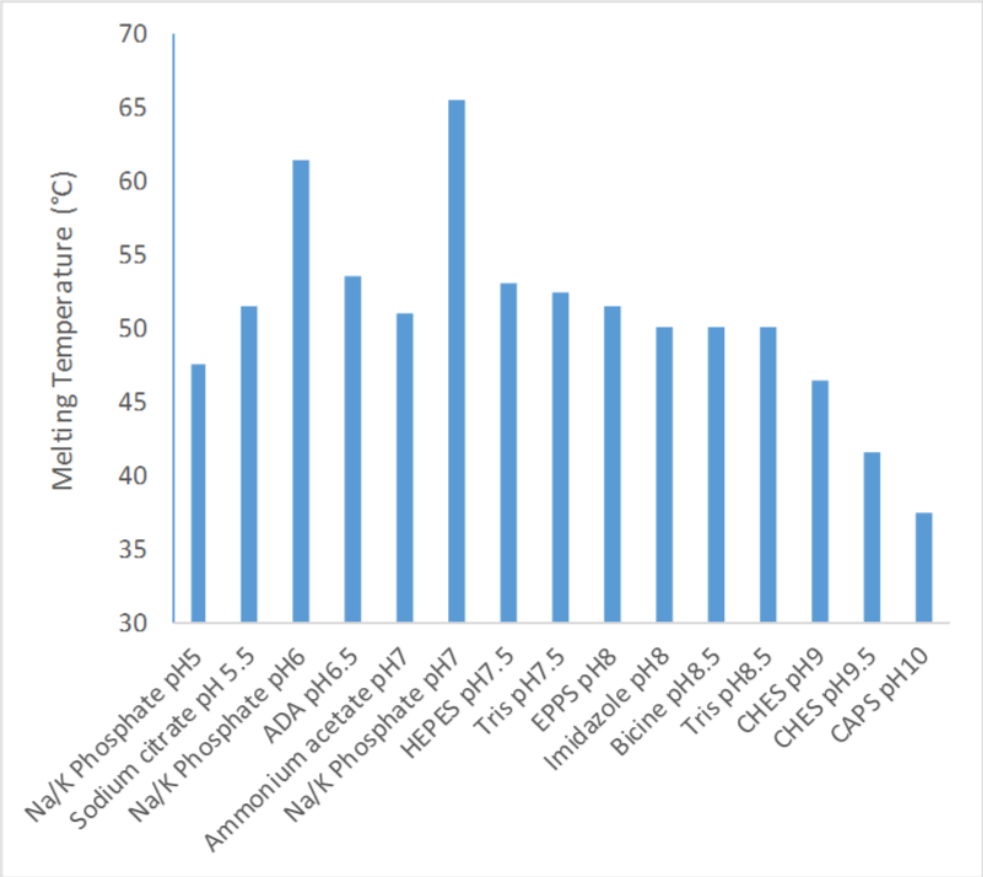


Figure 3.2 Melting temperature analysis of XRCC1$^{219-415}$ in response to different buffers. Buffers with sodium/potassium phosphate near neutral pH yield a higher melting temperature, suggesting an increased stability of XRCC1.

DNA substrates used for crystallization:

25bp substrates:
Forward 5'-TTTTTTTTTGGTGGGTGGTGGTGGG-3'
Reverse 3’-CCACCCACCACCACCCAAAAAAAAA-5’

22bp substrate:
Forward : 5’TTAGGCAGATTTTTTGCGGTGCC
Reverse : 5’AGGCACCGCAAAAAATCTGCCTA

22F/18R substrate:
Forward : 5’TTAGGCAGATTTTTTGCGGTGCC
Reverse : 5’ATGGCACCGCAAAAAATCT

20bp substrate:
Forward : 5’TAGATTTTTTGCGGTGCCAGG
Reverse : 5’ACCTGGCACCGCAAAAAATCT

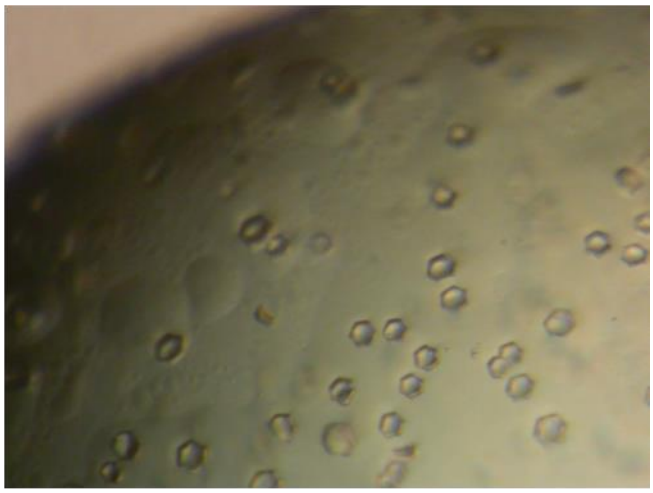
18bp substrate:
Forward : 5’TAGATTTTTTGCGGTGCCA
Reverse : 5’ATGGCACCGCAAAAAATCT

15bp substrate:
Forward : 5’TTTTTGCGGTGCCAGG
Reverse : 5’ACCTGGCACCGCAAAA

10bp substrate:
Forward : 5’TCGGTGCCAGG
Reverse : 5’ACCTGGCACCG

Figure 3.3 DNA substrates used for crystallization with XRCC1$^{219-415}$. DNA overhangs frequently allow intermolecular pairing to occur between two substrate molecules. This allows long polymers to form during crystallization. The remaining substrates have a single base over hang that can also facilitate polymer formation.

A)



B)

Figure 3.4 XRCC1$^{219-415}$ and DNA co-crystals. A) Crystals generated from 20 bp substrate and precipitated with 50 mM MES pH 6.5, 1 mM spermine tetrahydrochloride, 25% (v/v) PEG400. B) Crystals generated from 22F/18R substrate and precipitated with 0.1 M sodium chloride, 0.2 M potassium chloride, 0.02 M magnesium chloride hexahydrate, 0.05 M Bis-Tris 7.0, 35% (w/v) PEG 2000.

Table 3.1 SAXS data analysis table.

Sample	Io/C (mg/mL)	MW (kDa)		Rg PDF (Å)	Dmax (Å)
	Observed	Calculated	Expected		
XRCC1-219-415 (1.2mg/mL, 56 µM)	0.067	36.8	21.4	46.6	145
XRCC1-219-415 (4.3mg/mL, 201 µM)	0.073	26.7	21.4	37.9	130
XRCC1-219-415 (5.7mg/mL, 266 µM)	0.103	25.8	21.4	38	130
39bp DNA (2.3/4.5 mg/mL) merge	0.35	23.9	24	34.96	120
39bp DNA (2.3mg/mL, 94 µM)	0.35	23.9	24	28.81	120
39bp DNA (3.4mg/mL, 141 µM)	0.32	21.2	24	30.3	115
39bp DNA (4.5mg/mL, 188 µM)	0.3	19.6	24	33.42	110
XRCC1-219-415 + 39bp DNA (2/4mg/mL) merge	0.431	59.9	45.4	45.6	155
XRCC1-219-415 + 39bp DNA (2mg/mL, 93 µM)	0.43	50.8	45.4	45	147
XRCC1-219-415 + 39bp DNA (3mg/mL, 140 µM)	0.467	61.4	45.4	50.3	180
XRCC1-219-415 + 39bp DNA (4mg/mL, 187 µM)	0.495	66.1	45.4	53.8	190

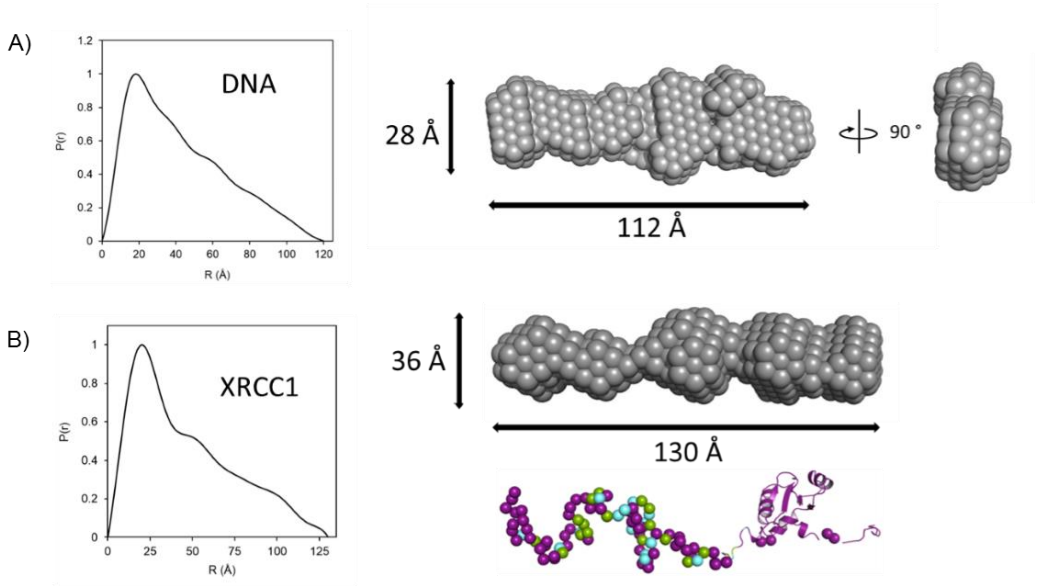


Figure 3.5 Small angle X-ray scattering for A)39bp DNA and B) XRCC1$^{219-415}$. The pair distribution curves are shown on the left, and the resulting molecular envelopes are shown on the right. From the molecular envelope of XRCC1, an *ab initio* model was generated with DAMMIF (grey model) and BUNCH (coloured model) to fill in atoms not structurally determined in the BRCT1 structure.

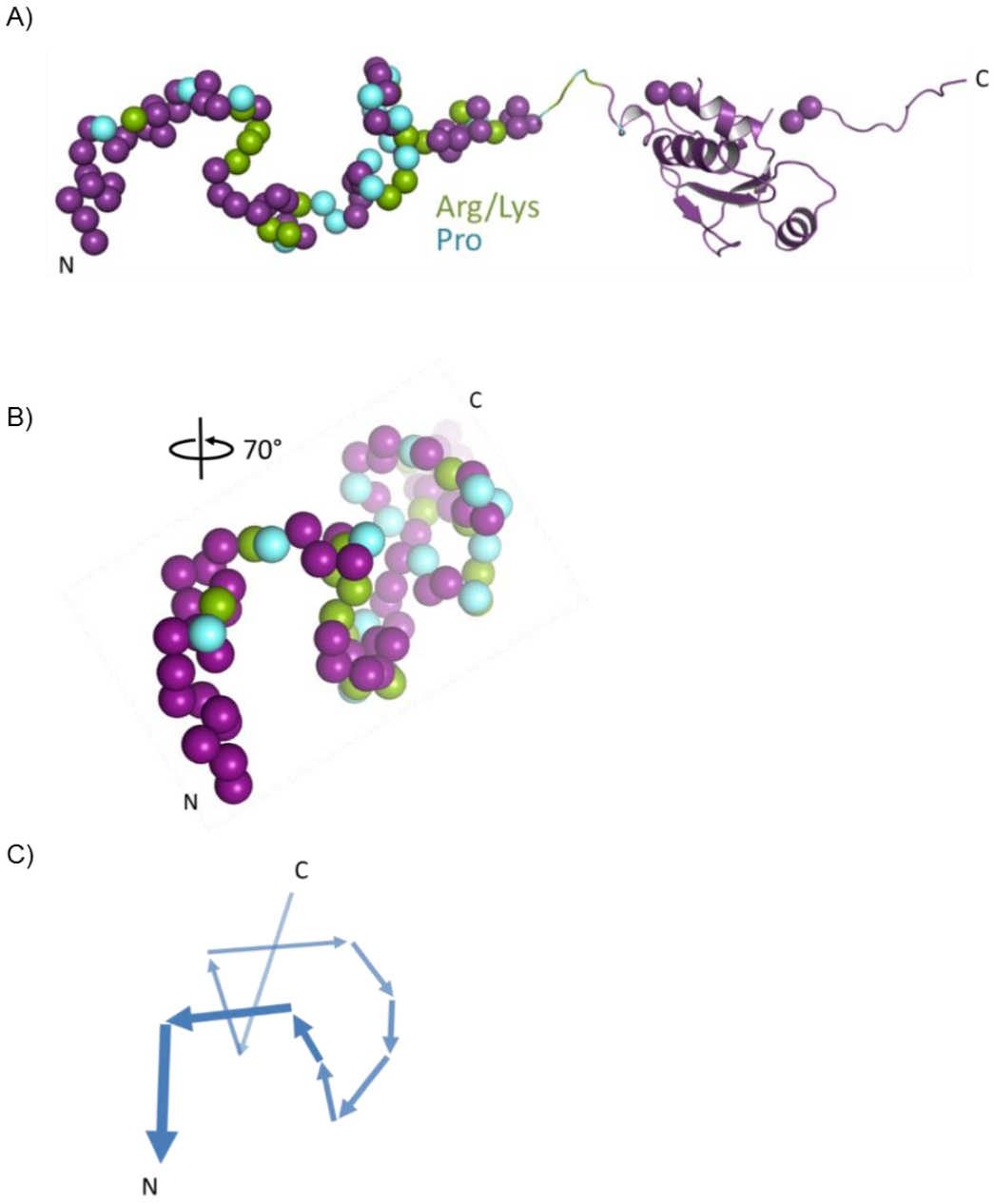


Figure 3.6 A) Model of XRCC1$^{219-415}$. Arg and Lys residues are coloured green while Pro are coloured cyan. B) Different orientation of model shows that the N-terminal linker forms a cylindrical claw superstructure, with many positively charged residues positioned inside the claw region. C) A simplified schematic of the N-terminal linker. Each arrow depicts a segment of peptide.

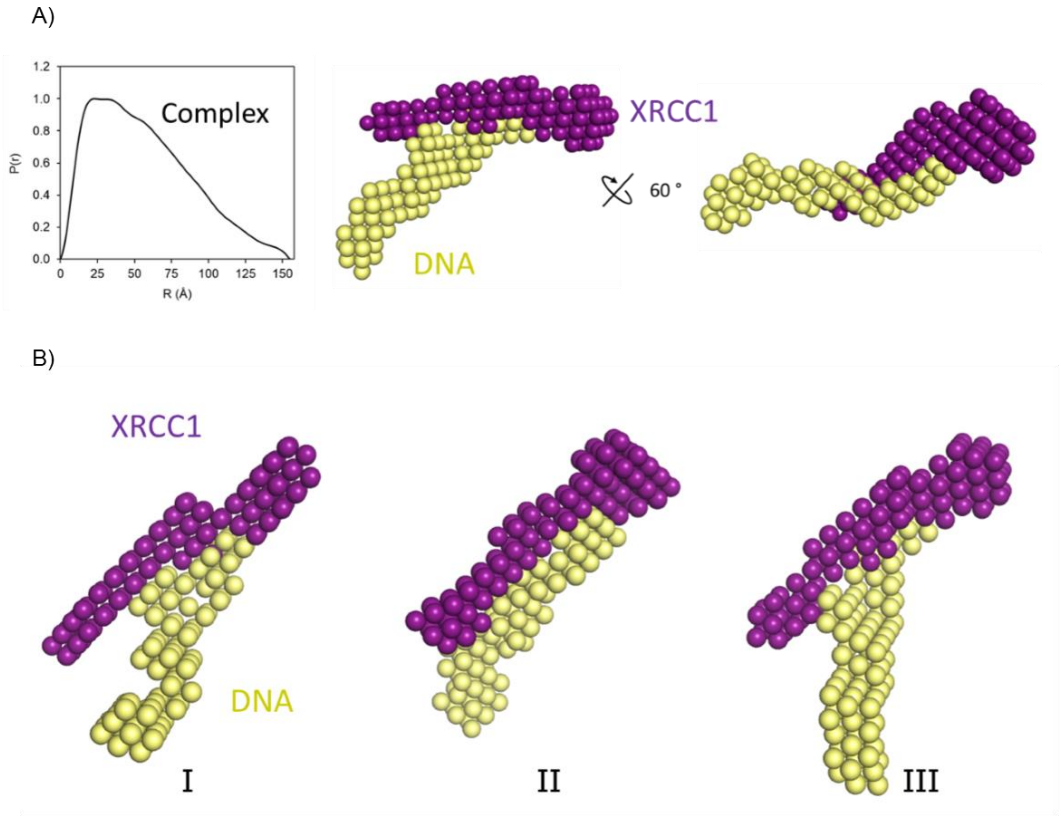


Figure 3.7 A) SAXS analysis of XRCC1/DNA complex. The pair distribution function (left) and the MONSA generated model in two different orientations (right). Purple spheres correspond to XRCC1$^{219-415}$ while yellow spheres correspond to 39 bp DNA. B) Three different XRCC1/DNA complex models generated by MONSA. All models consistently show XRCC1 engages DNA by the N-terminal linker.

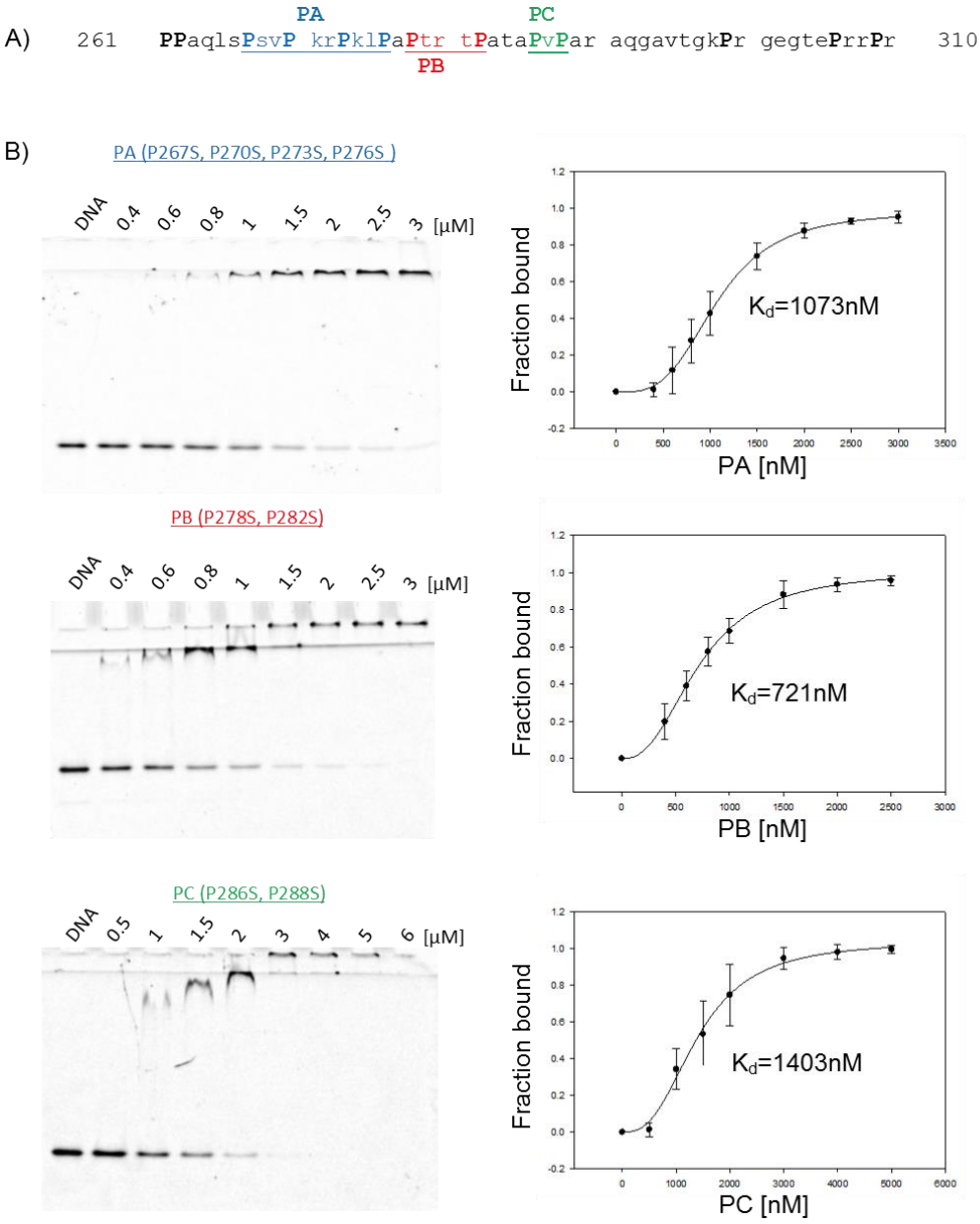


Figure 3.8 EMSA of proline mutants with DNA. A) Sequence within XRCC1$^{219-415}$ highlighting Pro residues that were substituted with Ser. B) EMSA of Ser substituted mutants with 39 bp duplex DNA substrate (left). Binding curves were generated for each protein (right).

3.6 Discussion

The SAXS model of XRCC1$^{219-415}$ generated in the absence of DNA revealed the presence on an elongated molecule with structure extending past the BRCT1 domain. Although the N-terminal extension appears to lack any traditional secondary structure elements (i.e. alpha helices or beta strands) or compact folding, there seems to be some rigid cylindrical super-structure present. The amino acid sequence within the 219-300 region contains a large number of proline residues (15 out of 81 residues). Proline imposes conformational rigidity to a peptide due to its ring structure and has been observed to form a third structural element known as a poly-proline type 2 helix [Adzhubei et al. 2013]. As supported by SAXS analysis, the close proximity of proline residues appear to induce structurally stable sharp bends within the region, giving rise to rigid features. Notably, this region forms a ‘claw-like’ structure that contains many of the positively charged residues that are important for nuclear localization and DNA binding [Kirby et al. 2015]. This non-traditional structure may play an undefined role in DNA repair. Although results presented here indicate that substitution of select prolines into serines within residues 219-300 has minimal impact on DNA binding *in vitro*, this rigid structure could still be biologically important, since this region is also implicated in mediating interactions with other DNA repair proteins such as APE1 and REV1. It will be interesting to determine if proline residues and/or positively charged residues within the extended

region of XRCC1$^{219-300}$ are required for other interactions necessary for BER and SSBR.

SAXS models of DNA-bound XRCC1$^{219-415}$ suggest that the N-terminal linker, which extends away from the BRCT1 domain, functions as the dominant contributor for DNA interaction. Indeed, all MONSA generated models showed the DNA interaction predominantly occurring between the N-terminal linker and DNA. DNA within these models appears to adopt an elongated helical structure as expected, suggesting that interaction with XRCC1 does not grossly distort DNA geometry. Interestingly, the model of DNA-bound XRCC1 indicates that the N- terminal linker remains relatively rigid and does not show signs of wrapping around DNA as one might have expected. This finding is in line with the protein alone model where the proline rich region is found to exist as a rigid structure in the absence of DNA. Since this region of XRCC1 has been shown to mediate both DNA and protein interactions it seems possible that the polyproline region may form a pseudo-domain, which serves as an interaction platform for DNA and other enzymes, thereby facilitating formation of higher-order nucleoprotein repair complexes.

XRCC1 truncation studies outlined in chapter 2 revealed that BRCT1 is required for DNA binding. Although SAXS models suggest the predominant DNA binding activity in XRCC1 resides within the polyproline region (residues 219-300), closer examination of these models suggests

there may be some direct interaction between BRCT1 and DNA. Interestingly, in SAXS models where BRCT1 interacts with DNA binding only occurs through contacts with the face of BRCT1 closest to the polyproline extension. Inspection of the NMR structure of BRCT1 shows a continuous positively charged landscape on one face of the domain (PDBID: 2D8M, Figure 3.9 A), consisting of residues R307, R308, R310, R391, R393, R394, R395 and R400. These arginine residues are readily poised to interact with a negatively charged surface, which the DNA backbone can provide (Figure 3.9 B, C). Since this region is upstream of the PARP1 binding site, and active PARP1 produces long chains of negatively charged ADP ribose polymers, these positive residues may be important for interacting with active PARP1 and its PAR product. Further exploration into the function of these residues within BRCT1 is needed.

The structure of the BRCT domain from RFC in the presence of DNA has been determined. Comparison of the structure of BRCT1 with that of the DNA-bound RFC-BRCT structure (Figure 3.10) reveals that the two BRCT structures not only maintain similarity at the fold level, but also in the presence of an elongated N-terminus. In the case of XRCC1-BRCT1, the extension was not present in the NMR structure because the domain boundaries used for determination excluded this region. Nevertheless, our SAXS model clearly indicates the presence on an ordered N-terminal extension of BRCT1 similar to what was observed for RFC-BRCT in the

presence of DNA. The N-terminus of RFC-BRCT consists of a coil-helix-coil, where the second coil makes extensive hydrophobic contact with helix 1 and 3 of a canonical BRCT domain. This interaction directs the N-terminus towards one face of the BRCT domain, where it provides additional contacts to the BRCT-DNA interaction. Considering that the XRCC1-BRCT also possess an extended N-terminus it is possible that the N-terminus exhibits similar behaviour to RFC-BRCT. However, unlike the hydrophobic residues present in RFC, the XRCC1 surface is highly charged and would therefore not likely adopt a similar folded N-terminal extension observed with RFC-BRCT. Unfortunately, due to the low resolution of SAXS models the trajectory of the XRCC1 N-terminal region cannot be accurately determined and therefore a more accurate comparison of the functionally similar BRCT domains of XRCC1 and RFC will have to await X-ray crystallographic determination.

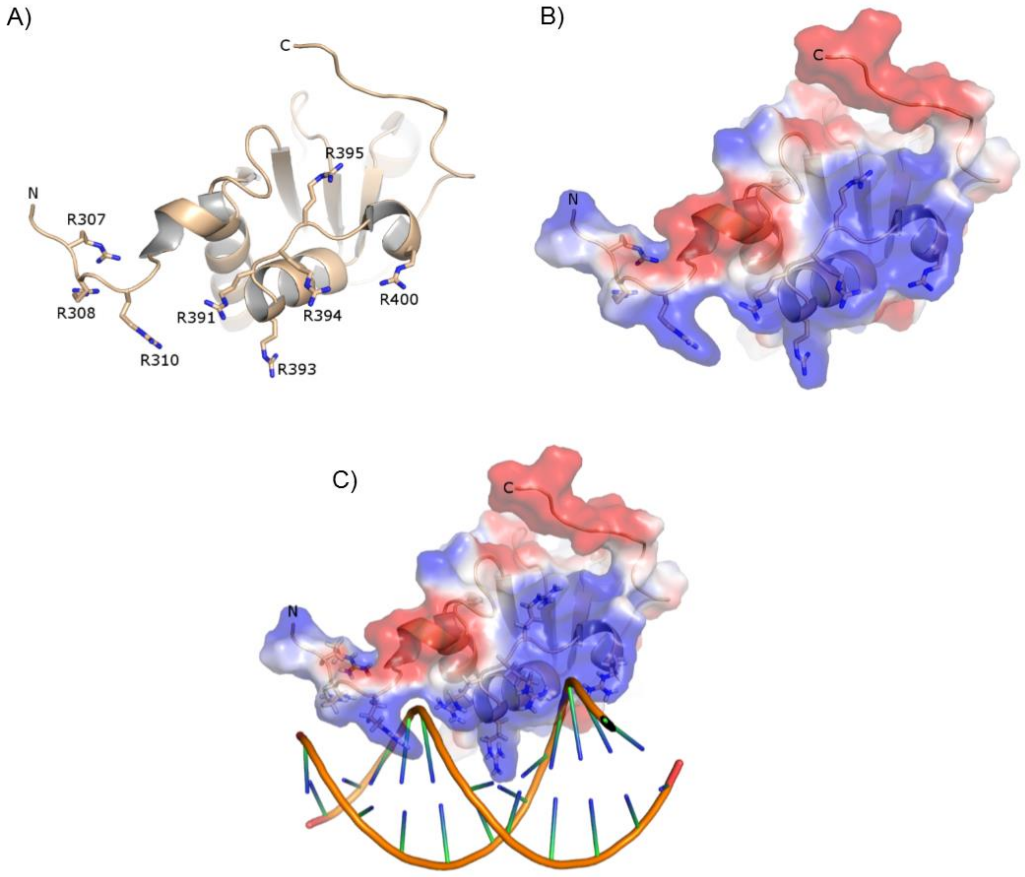


Figure 3.9 XRCC1 BRCT1 structure with positively charged residues shown. B) Electrostatic surface of BRCT1, where blue indicates positive regions and red indicates negative. C) Hypothetical model of BRCT1 interacting with DNA.

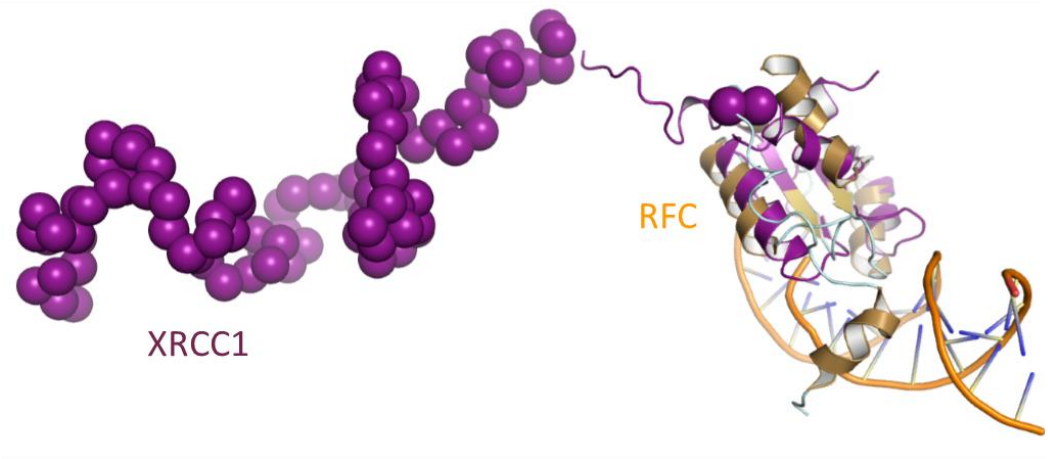


Figure 3.10 XRCC1-219-415 SAXS model (purple) superimposed on RFC-BRCT/DNA complex (yellow). The BRCT domain was used for structural alignment.

4 Disruption of the XRCC1/Ligase3 interaction with small molecule modulators

4.1 Preface of chapter 4

Dr. Meghan Mcfadden helped perform the ‘magnetic fishing’ assay, compound screening and IC50 determination. I performed all other experiments.

4.2 Abstract

Increased capacity for DNA repair in tumor cells has been shown to result in resistance of tumors to conventional cancer therapeutics. Thus, DNA repair proteins serve as targets for inhibition to improve clinical outcomes of current treatments. Using a ‘magnetic fishing’ assay designed to report on the status of protein-protein interactions, we have identified 6 compounds: phenylmercuric acetate, chloranil, merbromin, chelerythrine, 3,4-dimethoxydalbergione and celastrol able to disrupt the XRCC1/Ligase3 complex, a crucial component of the base excision and single strand break repair pathways.

4.3 Introduction

Although repair of DNA is necessary for genomic stability in normal cells, it can prove counterproductive during cancer therapy where DNA damage is an integral component of eliminating tumor cells. Currently, the majority of first line chemotherapeutic agents, including alkylating agents, platinum compounds and radiotherapy, operate by inducing DNA damage. Elevated repair of DNA damage in tumor cells negates these therapeutic effects and contributes to tumor resistance [2001, Fojo]. Hence, recent efforts have increasingly focused on disrupting DNA repair to improve conventional treatments, and proteins of DNA repair pathways have become attractive targets for inhibition and modulation [Madhusudan 2005, Nero 2014, Abbotts 2014].

Two major DNA repair pathways, base excision repair (BER) and single strand break repair (SSBR), require XRCC1 to coordinate repair via protein-protein interactions (PPIs). XRCC1 is particularly important because loss of expression is associated with absence of hormone receptors (ER-/PgR-/AR-) and triple negative phenotypes in invasive breast cancers [Sultana 2013]. Interestingly, cancer cells that lack XRCC1 expression are found to be vulnerable to inhibitors of DNA double strand break repair, inducing a synthetic lethal condition. This suggests that in the absence of base excision repair and single strand break repair, double strand break repair becomes the compensatory pathway. More importantly, it shows that targeting these compensatory pathways is a viable form of therapeutic treatment.

Devoid of catalytic activity, XRCC1 functions in DNA repair as a scaffold, orchestrating and coordinating protein-protein interactions between other repair factors such as APE1, PNKP, polB and Ligase3 [Caldecott 1996, Marintchev 2000, Masson 1998, Vidal 2001, Campalans 2005, Nash 1997]. Once recruited to the site of DNA damage, XRCC1 remains at the site until the DNA backbone is sealed by its binding partner, Ligase3. XRCC1 and Ligase3 form a constitutive heterodimeric complex. This interaction serves to stabilize Ligase3, as a reduction in XRCC1 expression leads to reduced levels of Ligase3 in cells [Caldecott 1995]. Importantly, disruption of XRCC1/Ligase3 interaction via point mutations

results in a reduction of single strand break repair in G_1 phase [Caldecott 2000]. It is therefore reasonable to assume that disrupting the XRCC1/Ligase3 interaction by small molecule modulators would similarly inhibit single strand break repair in a cell cycle dependent manner. Such modulators would not only serve as useful probes for studying DNA repair, but may also have potential as cancer therapeutic adjuvants.

In order to identify small molecule modulators of the XRCC1/Ligase3 interaction, we established a magnetic 'fishing' assay to identify small molecules capable of disrupting the XRCC1/Ligase3 BRCT-BRCT complex. Screening a bioactive library of small molecule compounds with this assay, resulted in identification of 6 compounds that showed dose-dependent response. Mode of action studies using structural and other biophysical methods are in progress.

4.4 Materials and methods

4.4.1 Expression vectors and protein purification

XRCC1$^{219-633}$ and Ligase3$^{931-1009}$ were cloned into pLic-His as described in section 2.4.1. Ligase3 gene was purchased from Open Biosystems (Accession # BC068005). XRCC1$^{538-633}$ was first cloned into pDONR201 (Invitrogen) and subsequently cloned into pDEST17 (Invitrogen) using the Gateway system as described in section 2.4.1. The primers used for PCR amplification of genes are listed in Figure 4.1.

All proteins were expressed in Rosetta pLysS cells. Cells were grown in 4 L of LB with added ampicillin (0.1 mg/mL) at 37˚C to an OD_{600} of 0.3-0.5 and induced with 1 mM IPTG at 20˚C overnight. The only exception was $XRCC1^{1-183}$, which was auto-induced at 16 ˚C over two days. Cells were then harvested by centrifugation at 3,315 x g for 15 min. Cells were resuspended in NiA buffer (20 mM Tris pH 8, 500 mM KCl, 3 mM BME, 10% (v/v) glycerol, 10 mM imidazole) and lysed by sonication (3 x 1 min). The lysate was then clarified by centrifugation at 48, 384 x g for 45min to remove insoluble material. Proteins were initially purified by IMAC using a 5mL column (GE Healthcare). The bound protein was washed with NiA buffer containing 10mM imidazole (20 column volumes) followed by 30mM imidazole (10 column volumes) before elution with 300mM imidazole. Eluted protein (10 mL) was mixed with water (10 mL), 10 x buffer (2.5 mL of 500 mM Tris pH8, 10 mM EDTA) and TEV protease (2.5 mL, 1 mg/mL) to remove affinity tags. The sample was incubated at 4 °C for overnight for $XRCC1^{219-633}$ and 2 days for $XRCC1^{538-633}$ and $Ligase3^{931-1009}$ due to slow proteolytic cleavage. Cleaved protein was further purified using ion exchange chromatography (Amersham). An 8 mL MonoQ column was used for $XRCC1^{219-633}$ and $XRCC1^{538-633}$, while an 8 mL MonoS column was used for $Ligase3^{931-1009}$, using 20 mM Tris (for MonoQ) or HEPES (for MonoS) pH 8, 3 mM BME, and 0 mM KCl (for binding buffer) or 500 mM KCl (for elution buffer). The relatively pure proteins were re-purified by a second

IMAC purification step to remove any remaining His-tagged protein. An additional size exclusion chromatography step was required for Ligase3$^{931-1009}$. XRCC1$^{219-633}$ and Ligase3$^{931-1009}$ was buffer exchanged into storage buffer (20 mM Tris pH8, 200 mM KCl, 3 mM BME, 10% (v/v) glycerol). XRCC1$^{538-633}$ was buffer exchanged into 20 mM Tris pH8, 500 mM KCl, 3 mM BME, 10% (v/v) glycerol. All proteins were stored at -80°C.

4.4.2 Size exclusion chromatography

Size exclusion chromatography was performed using a SEC650 column (BioRad Laboratories) equilibrated with 20 mM Tris pH 8, 100 mM KCl, 3 mM BME. 100 µL of XRCC1$^{219-633}$ or Ligase3$^{931-1009}$ (50 µM each) was injected onto the column and separated at a flow rate of 0.5 mL/min. For analysis of the XRCC1-Ligase3 complex, individual proteins were mixed in equal volumes (75 µL of 100 µM stock) to give a final solution of 150 µL 50 µM complex. The mixture was incubated for 1h at room temperature prior to SEC analysis. Elution fractions corresponding to peaks in absorbance (280 nm) were collected and analyzed by SDS PAGE.

4.4.3 Far western blotting

The protocol used for far western blotting was adapted from Wu et al. 2007. Briefly, XRCC1$^{219-633}$ without His-tag was run on an 11% SDS PAGE gel at 140 V for 60 min. XRCC1$^{219-633}$ was transferred to an activated PVDF membrane (PVDF membrane soaked in methanol for 5 min prior) by running for 75 min at 225 mA. Renaturing of XRCC1$^{219-633}$ transferred to

PVDF was performed by sequentially washing the membrane with 25 mL of renaturing buffer (20 mM Tris pH 7.5, 100 mM NaCl, 1 mM EDTA, 0.1% (v/v) Tween 20, 10% (v/v) glycerol, 2% (w/v) skim milk, 1 mM DTT) supplemented with 5 M, 3 M, 1 M, or 0.1 M guanidine hydrochloride, respectively. Starting with the highest concentration of guanidine hydrochloride (5 M), each wash solution was used to treat the PVDF membrane for 30 min at room temperature. A final treatment with renaturing buffer without guanidine hydrochloride was performed for 1 h at 4 °C. His-tagged Ligase3$^{931\text{-}1009}$ (10 µL, 3 mg/mL) was then added to renaturing buffer lacking guanidine (25 mL) and used to treat the PVDF membrane overnight at 4 °C. The membrane was washed three times with 15 mL of TBST buffer (50 mM Tris pH 7.5, 150 mM NaCl, 0.1% (v/v) Tween 20). Membrane was then treated with a mouse-anti-His antibody (Abcam) (4 µL of 1 mg/mL) in 15 mL of TBSTM buffer (50 mM Tris pH 7.5, 150 mM NaCl, 0.1% (v/v) Tween 20, 2% (w/v) skim milk) for 3 h at room temperature. The membrane was washed with 15 mL of TBST buffer (50 mM Tris pH 7.5, 150 mM NaCl, 0.1% (v/v) Tween 20) three times. The membrane was then treated with an alkaline phosphatase conjugated goat anti-mouse antibody (Abcam) (4 µL, 1 mg/mL) in 15 mL of TBSTM buffer. The membrane was washed with 15 mL of TBST buffer (50 mM Tris pH 7.5, 150 mM NaCl, 0.1% (v/v) Tween 20) three times and developed with an alkaline phosphatase substrate conjugate kit (BioRad).

4.4.4 Magnetic 'fishing' assay

XRCC1$^{219-633}$ and Ligase3$^{931-1009}$ were buffer exchanged into 20 mM Tris pH 8, 150mM KCl. To test specificity of the assay, Ligase3$^{931-1009}$ (100 µL of 1 µM) was mixed with His-tagged XRCC1$^{219-633}$ or His-tagged TEV protease (100 µL of 0, 1, 2, 5, 10 or 15 µM) and incubated for 15 min at room temperature. PureProteome Ni-NTA conjugated magnetic beads (Millipore) (15 µL) were added to the mixture and incubated for 5 min at low speed on a platform shaker (VWR). The beads were then sequentially washed three times with 100 µL of binding buffer. Proteins were eluted by incubating magnetic beads with 100 µL of 50% (v/v) methanol, 1% (v/v) acetic acid for 30 min. Eluates were analyzed using a Thermo Scientific LCQ Fleet mass spectrometer fitted with a 10 µL sample loop of a 250 µm i.d. (3-aminopropyl)-triethoxysilane (APTES)-coated fused silica capillary. Using the online syringe pump, 0.5% (v/v) acetic acid in 1:1 (v/v) methanol/water was delivered at a flow rate of 15 µL/min as 10 µL injections were made for each eluate sample. Total MS/MS signal for the fragmentation of the *m*/*z* 850 peak (corresponding to Ligase3$^{931-1009}$) was monitored overtime. The total integrated peak area of each sample injection was compared to a calibration curve to determine the amount of Ligase3$^{931-1009}$ recovered.

The reproducibility of the magnetic fishing assay was validated by repeating the assay ten times and applying the Z' test with the following

equation: $Z' = 1 - \frac{(3\sigma_L + 3\sigma_H)}{|\mu_H - \mu_L|}$, where σ_L is the standard deviation of the low control, σ_H is the standard deviation of the high control, μ_L is the average signal of the low control and μ_H is the average signal of the high control. The high control represents the amount of Ligase3$^{931\text{-}1009}$ recovered by native XRCC1$^{219\text{-}633}$, while the low control corresponds to the amount of Ligase3$^{931\text{-}1009}$ recovered using thermally denatured XRCC1$^{219\text{-}633}$. A final concentration of 2 µM was used for all proteins. Washing steps and MS/MS analysis was performed as described above.

4.4.5 Compound screening

A mass-encoded compound library, previously described by Mcfadden et al. (2010), was used to screen for modulators of XRCC1/Ligase3 interaction. Prior to screening, a 1000 compound collection was subdivided into mixtures containing 10 compounds each. XRCC1$^{219\text{-}633}$ (100 µL, 2 µM) was incubated with each mixture of compounds (10 µM of each compound in each mixture) in 20 mM Tris pH 8, 150 mM KCl, 2% (v/v) DMSO at room temperature for 30 min with shaking. Ligase3$^{931\text{-}1009}$ (10 µL, 20 µM) was added to the mixture to a final concentration of 2 µM and allowed to incubate with gentle shaking for 5min. Ni-NTA conjugated magnetic beads (15 µL) were subsequently added to the mixture for 5 min to capture His-tagged XRCC1. Beads were sequentially washed three times with 100 µL of 20 mM Tris pH 8, 150 mM KCl, prior to protein elution of bound proteins with 100 µL of 50/50 (v/v)

methanol water solution with 1% (v/v) acetic acid for 30 min. The amount of Ligase3$^{931\text{-}1009}$ recovered was determined by MS/MS analysis. Library screening was performed in duplicate, and any mixture reducing Ligase3 recovery by >75% relative to control was deemed at hit.

Mixtures that were identified as hits were deconvoluted by injecting 10 µL of the original eluate and measuring the mass spectrum over the *m*/*z* range of 100-1000. This spectrum was compared to a control sample without any compound present. Additional ion peaks present in the sample spectrum, but absent in the control spectrum, were identified and compared to the list of mass-encoded compounds in the mixture to identify the ligand by molecular mass. Discrete screening of individual compounds against the XRCC1/Ligase3 complex was performed to validate each hit. A mixture containing XRCC1$^{219\text{-}633}$ (2 µM) and each compound (10 µM) in 100 µL of 20 mM Tris pH 8, 150 mM KCl, 2% (v/v) DMSO was incubated at room temperature for 30 min with shaking before adding Ligase3$^{931\text{-}1009}$ (10 µL, 20 µM). Washing steps and MS/MS were performed as described above.

4.4.6 Determination of IC_{50} values

Dose-dependent response curves were determined for each hit. A mixture containing XRCC1$^{219\text{-}633}$ (2 µM) and varying concentrations (0.1 µM – 2000 µM) of each compound in 100 µL of 20 mM Tris pH8, 150 mM KCl, 2% (v/v) DMSO was incubated at room temperature for 30 min with shaking before adding Ligase3$^{931\text{-}1009}$ (10 µL, 20 µM). Washing steps and MS/MS

were performed as described above. The amount of Ligase3 recovered was plotted against the log concentration of each compound and fitted using Prism7 with the following equation: Y=a+(b-a)/(1+10^((LogIC50-X)*HillSlope)), where 'a' is the minimum and 'b' is the maximum Y plateau value.

4.4.7 Structure determination of protein-ligand complexes

Purified Ligase3$^{931\text{-}1009}$ was dialyzed into 20 mM HEPES pH 7.5 and concentrated to 13 mg/mL. Crystallization trials were performed using the hanging drop vapor diffusion method by mixing 3 µL of protein with 1 µL of precipitant (0.14 M Ammonium Formate pH 6.6, 14% (w/v) PEG 3350, 30% (v/v) glycerol) and incubating over a solution of 1.5 M ammonium sulfate at 4 ˚C. Crystals appeared within 2-3 days and were allowed to grow over two weeks. Compounds of interest (phenylmercuric acetate, chloranil, merbromin, chelerythrine and celastrol) were first dissolved to a final concentration of 100 mM in a solution containing 25% (v/v) diethylene glycol, 25% (v/v) ethylene glycol, 25% (v/v) glycerol, and 25% (v/v) 1, 4 dioxane (Molecular Dimensions). Compound soaking solution was prepared by mixing 1 µL of dissolved compound with 99 µL of cryo-protectant buffer (20 mM HEPES pH 7.5, 0.14 M ammonium formate pH 6.6, 14% (w/v) PEG 3350, 30% (v/v) glycerol), bringing the final compound concentration to 1 mM. Suitable Ligase3 crystals were transferred to 1 µL of the soaking solution and incubated for varying lengths of time (1 – 16 h). Crystals were

frozen with liquid nitrogen and data collected at the Canadian Light Source, Canadian Macromolecular Crystallography Facility beamline 08ID-1. Diffraction data were processed using *iMosfilm* (CCP4). Molecular replacement was performed using the crystal structure of Ligase3 (931-1009) (PDB ID: 3PC8 chain A) as a search model in *Phenix-AutoMR.* The model resulting from molecular replacement was further improved by iterative rounds of model building and refinement carried out using *Coot* and *Phenix-Refine*.

Crystallization of XRCC1$^{538-633}$ was performed by the hanging drop vapor diffusion method using protein dialyzed into 20 mM HEPES pH 7.5, 500 mM KCl and concentrated to 1 mg/mL. Four volumes of protein were mixed with 1 volume of precipitant (0.1 M Citric Acid: NaOH pH 3.5, 2 M Ammonium Sulfate) and incubated over 2 M ammonium sulfate at 20 °C. Long rectangular crystals formed after 1 week, and were allowed to reach a maximum growth after ~ 2 months. Five compounds: phenylmercuric acetate, chloranil, merbromin, chelerythrine and celastrol were directly added to the crystallization drop and incubated for times ranging from (1 - 12 h). Compound-soaked crystals were flash frozen in liquid nitrogen prior to data collection using a MicroMax-007 HF X-ray generator mounted with VariMax optics and Raxis 4^{++} detector (Rigaku). Diffraction data were processed using *iMosfilm* (CCP4). Molecular replacement was performed using the crystal structure of XRCC1-BRCT2 (PDB ID: 1CDZ chain A) as

an initial search model in *Phenix-AutoMR*. Resulting models from molecular replacement were further improved by iterative rounds of model building and refinement carried out using *Coot* and *Phenix-Refine*.

4.5 Results

4.5.1 Verification of XRCC1/Ligase3 interaction

In order to ensure truncations of XRCC1 and Ligase3 would be suitable for further compound screening it was necessary to first determine that these constructs could form complex efficiently. We first generated a series of XRCC1 truncations and tested them for their ability to interact with the Ligase3 BRCT domain. Prior reports have indicated that the interaction between BRCT domains of XRCC1 and Ligase3 occur between the most C-terminal BRCT domains of both proteins, hence our constructs were designed to contain those regions. The Ligase3 BRCT domain (residues 931-1009) expressed well in *E. coli* and was stable. The XRCC1 BRCT (residues 538-633) was less well behaved, and therefore we used a more soluble and stable construct (residues 219-633). This construct was stable and easily purified (Figure 4.2 A).

Interaction between XRCC1$^{219-633}$ and Ligase3$^{931-1009}$ was first tested using size exclusion chromatography. XRCC1, Ligase3 and a mixture of both proteins were resolved on a size exclusion column as described in methods (SEC650, BioRad Laboratories). To assess interaction, protein elution volumes were compared and protein from elution fractions analyzed

by SDS PAGE. Ligase3$^{931-1009}$ has a significantly lower molecular weight compared to XRCC1$^{219-633}$ and therefore these proteins do not overlap during elution. XRCC1$^{219-633}$ and Ligase3$^{931-1009}$ eluted at volumes of 11.5 mL and 15.7 mL, respectively (Figure 4.2 B). The mixture contained one major peak, with elution volume 11.25 mL. If a complex formed, the elution volume of the complex would be expected to be less than 11.25 mL. Nevertheless, SDS PAGE analysis indicated that Ligase3$^{931-1009}$ co-eluted with XRCC1$^{219-633}$, suggesting that the domain boundaries chosen for XRCC1 and Ligase3 were sufficient for stable complex formation.

To further verify interaction between XRCC1$^{219-633}$ and Ligase3$^{931-1009}$ we performed far western blotting [Wu et al. 2007]. For this assay, we purposefully left the His-tag on Ligase3$^{931-1009}$ while removing the His tag from XRCC1$^{219-633}$. XRCC1$^{219-633}$ (bait) was run on an SDS PAGE gel and transferred to a PVDF membrane. XRCC1 was renatured using a series of guanidine-HCl washes and then incubated with Ligase3$^{931-1009}$ (prey). Since Ligase3 contains a His-tag, detection of an interaction with XRCC1 is possible using an anti-His antibody. Only if the two proteins interact, would one be able to detect the complex on the PVDF membrane. Consistent with results from SEC, XRCC1$^{219-633}$ and Ligase3$^{931-1009}$ were found to interact by far western blotting (Figure 4.3) indicating these domain boundaries are suitable for further studies involving protein-protein interactions.

4.5.2 Identification of XRCC1/Ligase3 complex modulators

In order to screen for compounds able to disrupt the BRCT-mediated interaction of XRCC1/Ligase3, a 'magnetic fishing' assay was designed similar to that reported for monitoring of other PPIs [Mcfadden et al. 2010] (Figure 4.4). In this assay a His-tagged protein is first bound to Ni-NTA conjugated magnetic beads. Protein-protein interaction with a non-His tagged protein is then monitored by recovery from magnetic beads following washing and elution steps. The amount of target protein recovered is determined using mass spectrometry (MS). Since Ligase3$^{931\text{-}1009}$ can be easily detected by MS this domain was chosen as the non-His tagged target protein. As shown in Figure 4.5 A, recovery of Ligase3 appeared to be specific to His-XRCC1, since another His-tagged protein (His-TEV) failed to recover Ligase3. Further analysis of the assay indicated a Z' value of 0.66 suggesting that this assay is sufficiently robust for use in screening of PPI modulators (Figure 4.5 B). Because this assay was designed for use with compounds dissolved in dimethyl sulfoxide (DMSO), and DMSO at even low concentrations can alter protein stability, magnetic fishing was performed in the presence of increasing concentrations of DMSO (up to 2% (v/v)) to determine the tolerable amount of DMSO. From this analysis it was apparent that DMSO does not interfere with Ligase3 recovery, suggesting that 2% (v/v) DMSO is tolerable for screening compounds (Figure 4.5 C).

Using the magnetic fishing assay, 1000 compounds were screened for their ability to disrupt the XRCC1/Ligase3 complex. His-tagged XRCC1 was incubated with compounds prior to the incubation with Ligase3. If any compound was able to disrupt complex formation, the amount of Ligase3 recovered after washes would be less than control experiments. Any compound found to reduce Ligase3 recovery by >75% was considered a compound of interest or 'hit'. 1000 compounds were chosen from a subset of the Canadian Compound Selection belonging to the High Throughput Screening facility at McMaster University (Hamilton, ON, Canada). In order to improve screening efficiency compounds were initially screened in mixtures containing 10 different compounds. When a given mixture of 10 compounds was found to reduce Ligase3 recovery, discrete screening was performed to identify the compound(s) of interest. The initial screen was carried out in duplicate and resulted in 4 complex mixtures being identified as hits (Figure 4.6 A). Following deconvolution by discrete screening, 6 compounds of interest were identified: phenylmercuric acetate, chloranil, merbromin, celastrol, 3,4-dimethoxydalbergione and chelerythrine (Figure 4.6 B, C). Each of the 6 compounds were found to reduce Ligase3 recovery in a dose-dependent manner. Half-maximal inhibitory concentration values (IC50) obtained for these compounds were as follows: phenylmercuric acetate (2.5 µM), chloranil (42 µM), merbromin (10 µM), celastrol (10 µM), 3,4-dimethoxydalbergione (13 µM), and chelerythrine (1.8 mM) (Figure 4.7).

4.5.3 Crystallization of protein-modulator complexes

In order to begin characterizing the mode of action of XRCC1/Ligase3 interaction modulators identified above, the crystal structure of BRCT domains from XRCC1 and Ligase3 were determined in the presence of modulating compounds. Since screening did not reveal if modulators were bound to XRCC1 or Ligase3, crystals of both proteins were soaked with modulators of interest. Unfortunately, 3,4-Dimethoxydalbergione could not be purchased commercially, hence the crystallization trials were only performed with the five remaining compounds. Ligase3$^{931\text{-}1009}$ crystals formed readily and diffracted to 2.0 Å. These crystals were transferred to a solution containing 1 mM of each compound and allowed to soak for varying amounts of time. Crystals were then flash frozen with liquid nitrogen and sent to the CLS-CMCF facility for data collection. Crystals that were soaked overnight typically resulted in poor diffraction (less than 3.5 Å) and were considered too low quality for structure determination. Only crystals that were soaked for 1 hour generated good quality diffraction (2.0-2.6 Å). Such data sets were then used for structure determination. Statistics for data processing and structure refinement are tabulated in table 4.1. Of the structures solved, no electron density resembling the 5 compounds tested could be found. Although it is possible that the binding site for compound interaction was occluded

through crystal packing, overall these results indicate that the compounds do not interact with Ligase3$^{931-1009}$ in crystals (Figure 4.8).

Attempts to generate XRCC1$^{219-633}$ crystals were unsuccessful. This protein contains BRCT1 and BRCT2 domains joined by a ~100 residue linker. As such, it may be very dynamic in solution and therefore difficult to crystalize. Based on prior reports that indicate XRCC1 and Ligase3 interact through their C-terminal BRCT domains, it seemed reasonable to remove BRCT1 and the following linker region of XRCC1. Crystals of the XRCC1 C-terminal BRCT domain (residues 538-633) were generated, and soaked with compounds. In this case, solid compounds were added directly to protein crystals and allowed to soak for up to 12 hours since all solutions that were capable of dissolving compounds were found to destroy crystals. Of the five compounds tested, merbromin and phenylmercuric acetate appeared to damage the crystal over time, with appearance of crystal cracking. Furthermore, these crystals diffracted poorly (7-8 Å) compared to the apo crystals (2.5 Å). Chelerythrine-soaked crystals did not show damage, however the crystals only diffracted to 4Å resolution. Chloranil and celastrol did not show signs of crystal damage and data sets were collected for each. Both data sets were processed to 2.8 Å resolution. Unfortunately, the solved structures did not contain electron density for the compounds. Taken together, these results suggest that the compounds did not interact with crystallized XRCC1 BRCT2 (Figure 4.9, Table 4.2). It remains possible

that different soaking conditions and/or use of a different crystal form may yield more favorable results.

```
XRCC1-Forward-538 (Gateway)
GGGGACAAGTTTGTACAAAAAAGCAGGCTTAGATTACGATATCCCAACGACCGAAAAC-
CTGTATTTTCAGGGCCTCCCAGATTTCTTCCAGGGCAAGC

XRCC1-Reverse 633 (Gateway)
GGGGACCACTTTGTACAAGAAAGCTGGGTCTCAGGCCTGCGGCACCACCCC

Ligase3-Forward-931 (LIC)
ATTTCCAGGGAGCAGCCGCAACAAAGGTATTGCTGGACATC

Ligase3-Reverse-1009 (LIC)
CCTGCAAAGCACCGGCCCTAGCAGGGAGCTACCAGTCT
```

Figure 4.1 Primers used to generate XRCC1 constructs. Each primer is oriented from 5' to 3'. The primers are labelled as forward where the number indicates that starting amino acid residue or reverse where the number indicates the ending residue of the construct. The parenthesis indicates which cloning method was used.

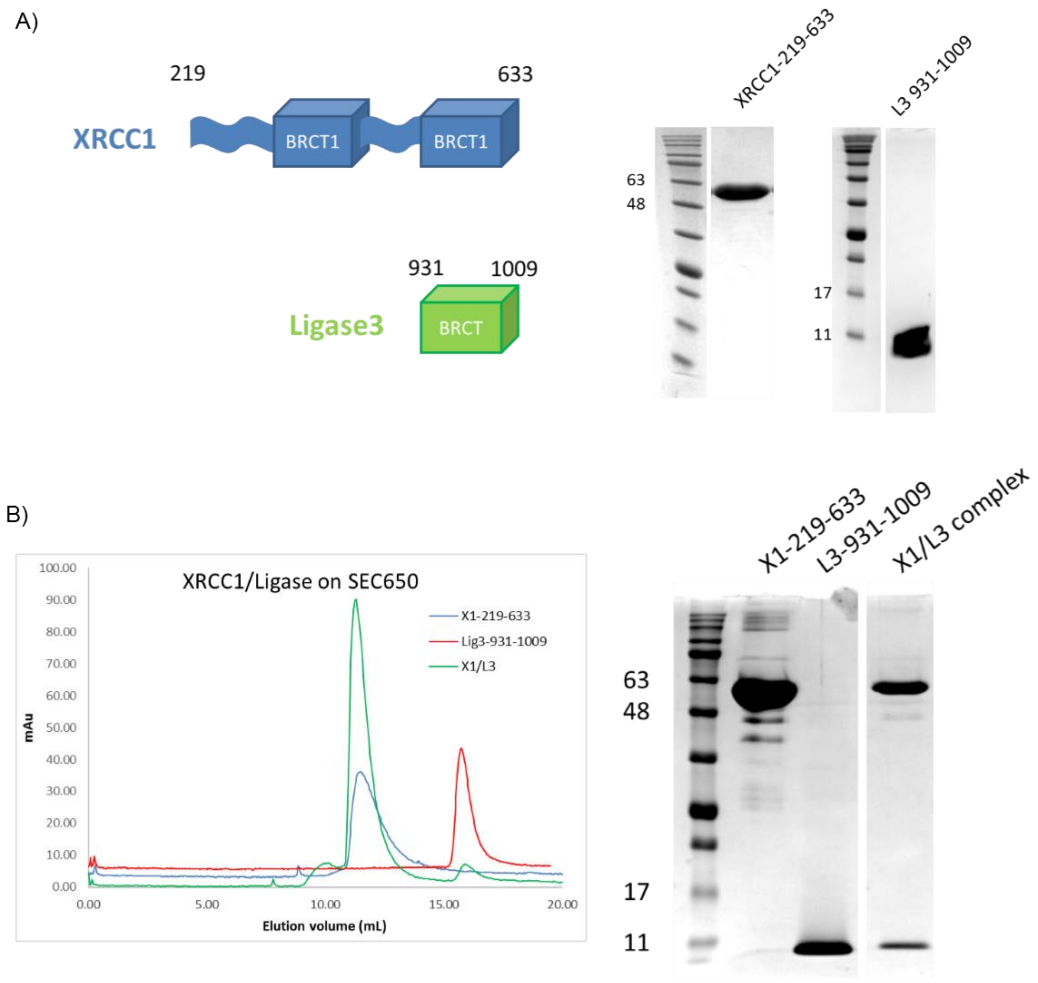


Figure 4.2 Analysis of XRCC1/Ligase3 interaction by Size exclusion chromatography. A) Domain organization within truncations of XRCC1$^{219-633}$ and Ligase3$^{931-1009}$ (left). XRCC1 contains both BRCT1 and 2 domains while Ligase3 contains only a single C-terminal BRCT domain. Purity of proteins used for complex formation analysis (right). B) Size exclusion chromatogram of XRCC1$^{219-633}$ alone (blue), Ligase3$^{931-1009}$ alone (red) and XRCC1/Ligase3 mixture (green). Fractions corresponding to the XRCC1/Ligase3 complex (green peak) were analyzed by SDS-PAGE (right). Molecular weight markers (in kDa) are shown on left of these coomassie stained gels.

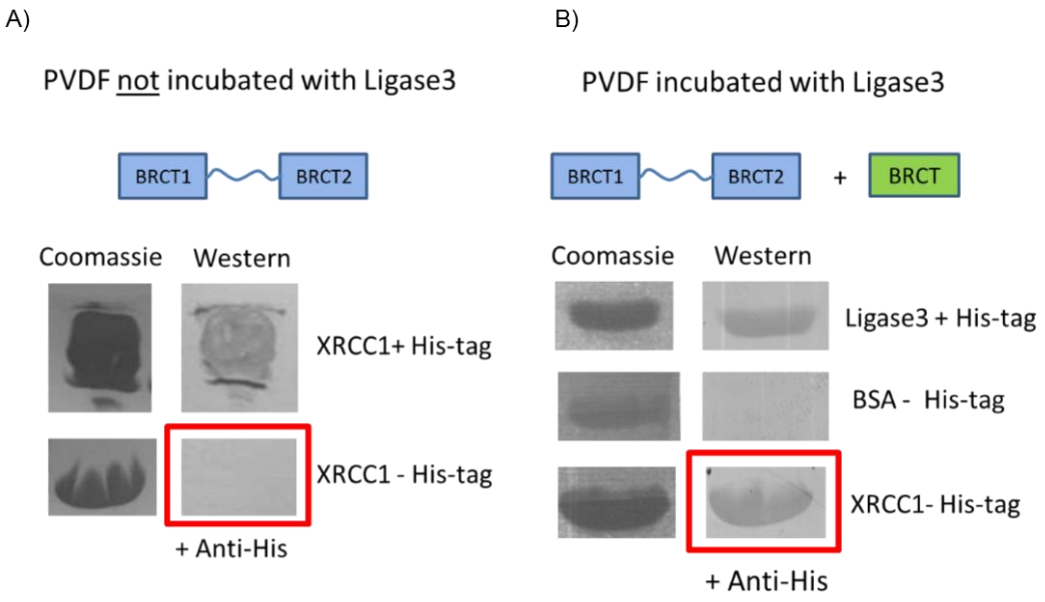


Figure 4.3 Far western blot analysis of XRCC1/Ligase3 complex formation. A) Western blot of His-XRCC1$^{219-633}$ (top) and XRCC1$^{219-633}$ without His tag (bottom) probed with anti-His antibody. B) Far western blot of XRCC1$^{219-633}$ without a His tag (bait), BSA without a His tag (negative control), Ligase3$^{931-1009}$ with a His tag (prey and positive control). XRCC1$^{219-633}$ without a His tag can only be detected after incubation with His-Ligase3$^{931-1009}$. The red box from A) and B) are bands of interest being compared.

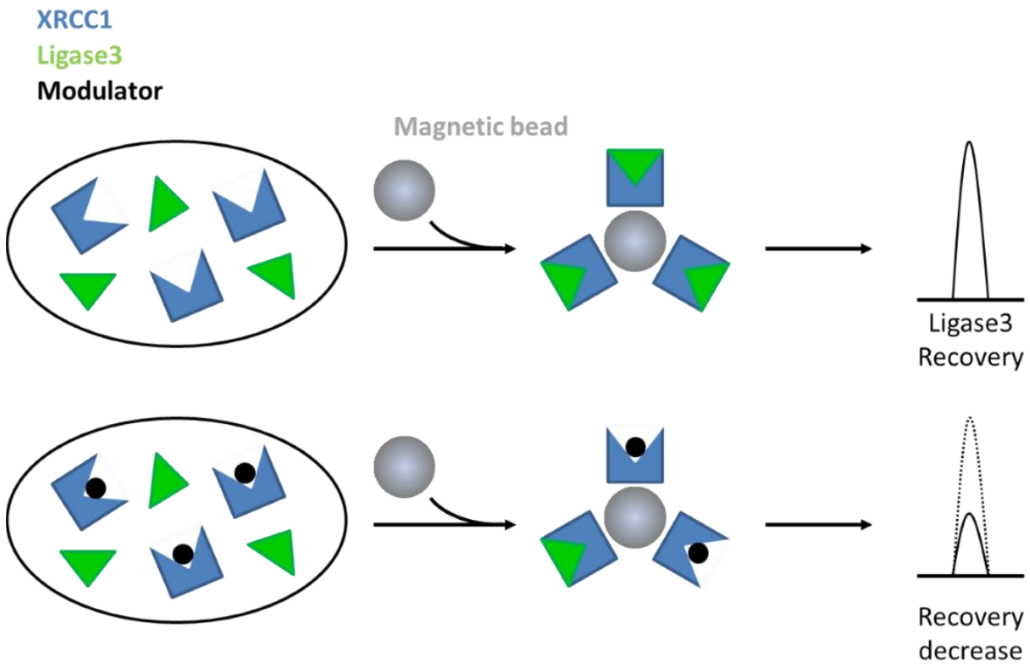


Figure 4.4 Schematic of 'magnetic fishing' assay. In the absence of a modulator (black circle), XRCC1 efficiently recovers Ligase3 (top). In the presence of a modulator, the XRCC1/Ligase3 is not formed, leading to reduction in Ligase3 recovery (bottom).

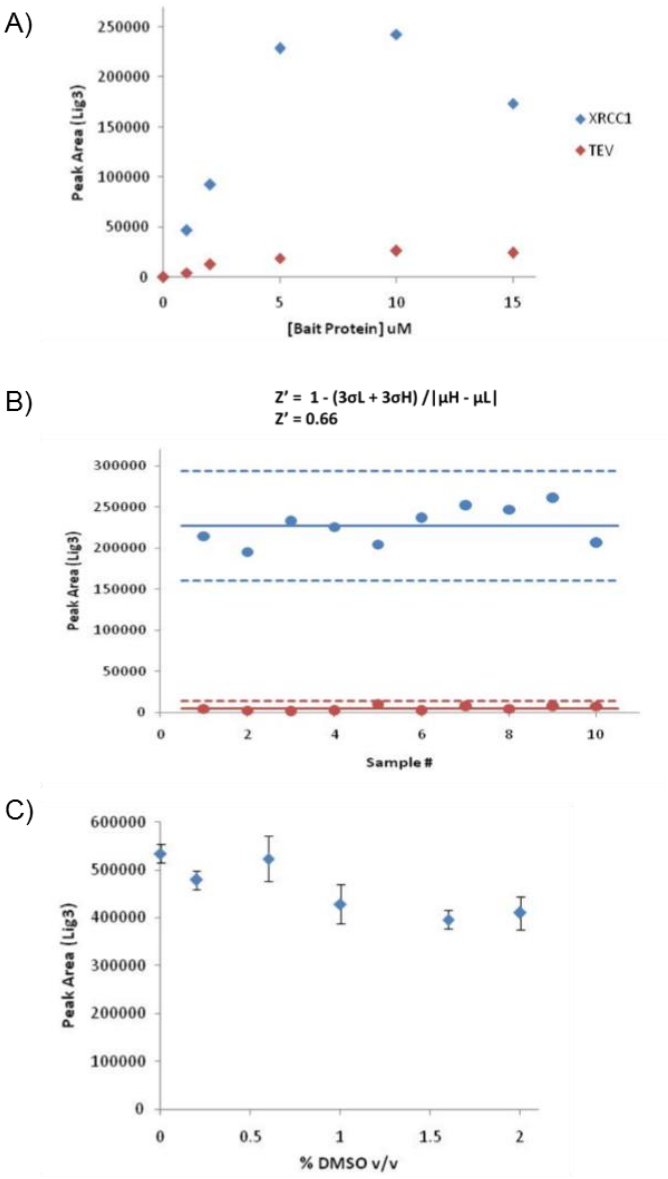


Figure 4.5 Assessment of 'magnetic fishing' assay quality. A) Magnetic fishing assay using either His-XRCC1 or His-TEV protease as bait. TEV protease recovers significantly lower amounts of Ligase3$^{931\text{-}1009}$ indicating recovery of Ligase3 was specific to XRCC1. B) 10 trials of 'magnetic fishing assay' to demonstrate reproducibility. The Z' score determined was 0.66, indicating this assay is suitable for high throughput screening. C) Magnetic fishing assay performed with his-tagged XRCC1 in the presence of increasing concentrations of DMSO. The recovery of Ligase3$^{931\text{-}1009}$ was not significantly altered by DMSO up to 2% (v/v).

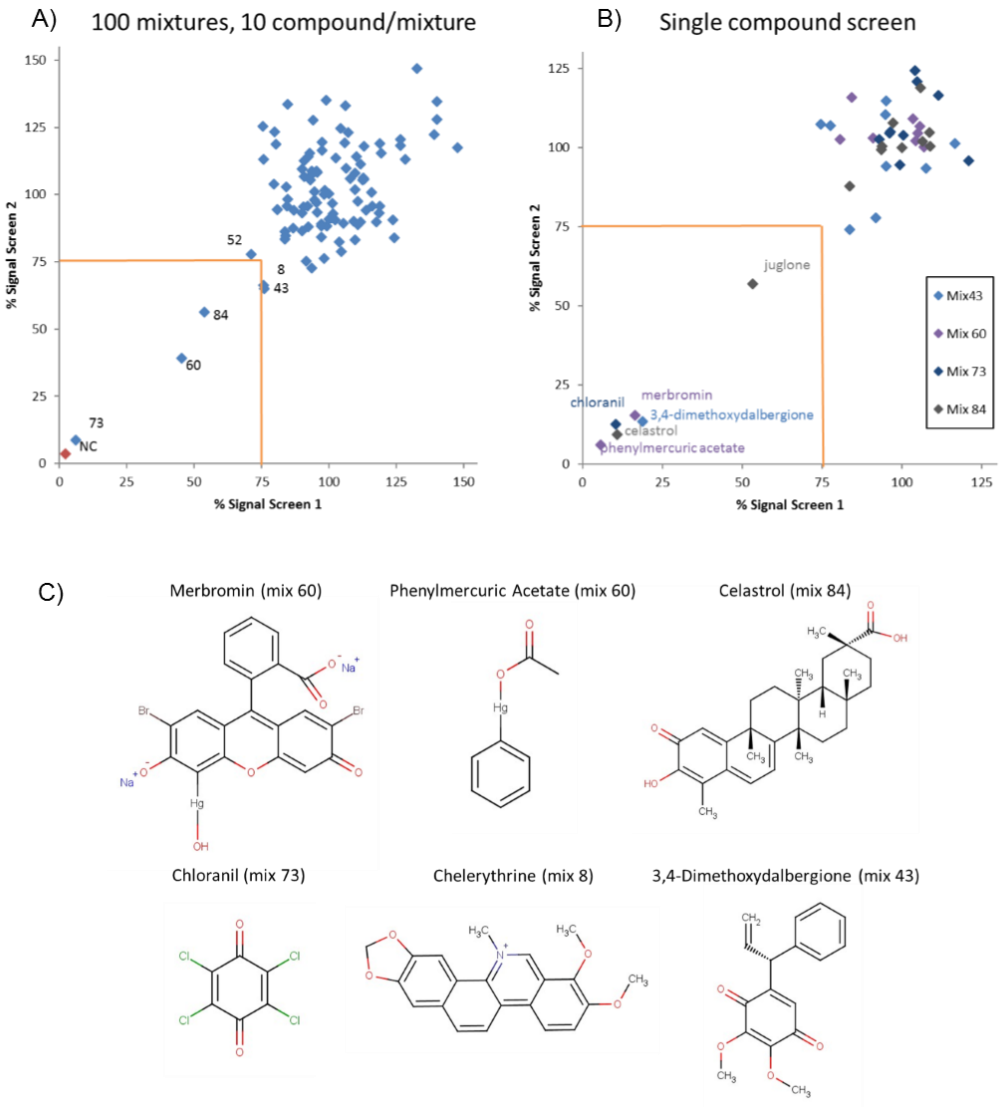


Figure 4.6 Plots of Ligase3 recovery resulting from duplicate screens of compound mixtures (A). Each mixture contained 10 compounds. Mixtures that reduced ligase3 recovery by >75% were deemed to contain a modulator. B) Duplicate plot from discrete screening to identify individual compounds that are capable of disrupting XRCC1/Ligase3 complex. C) Structures of compounds that were found to disrupt XRCC1/Ligase3 interaction.

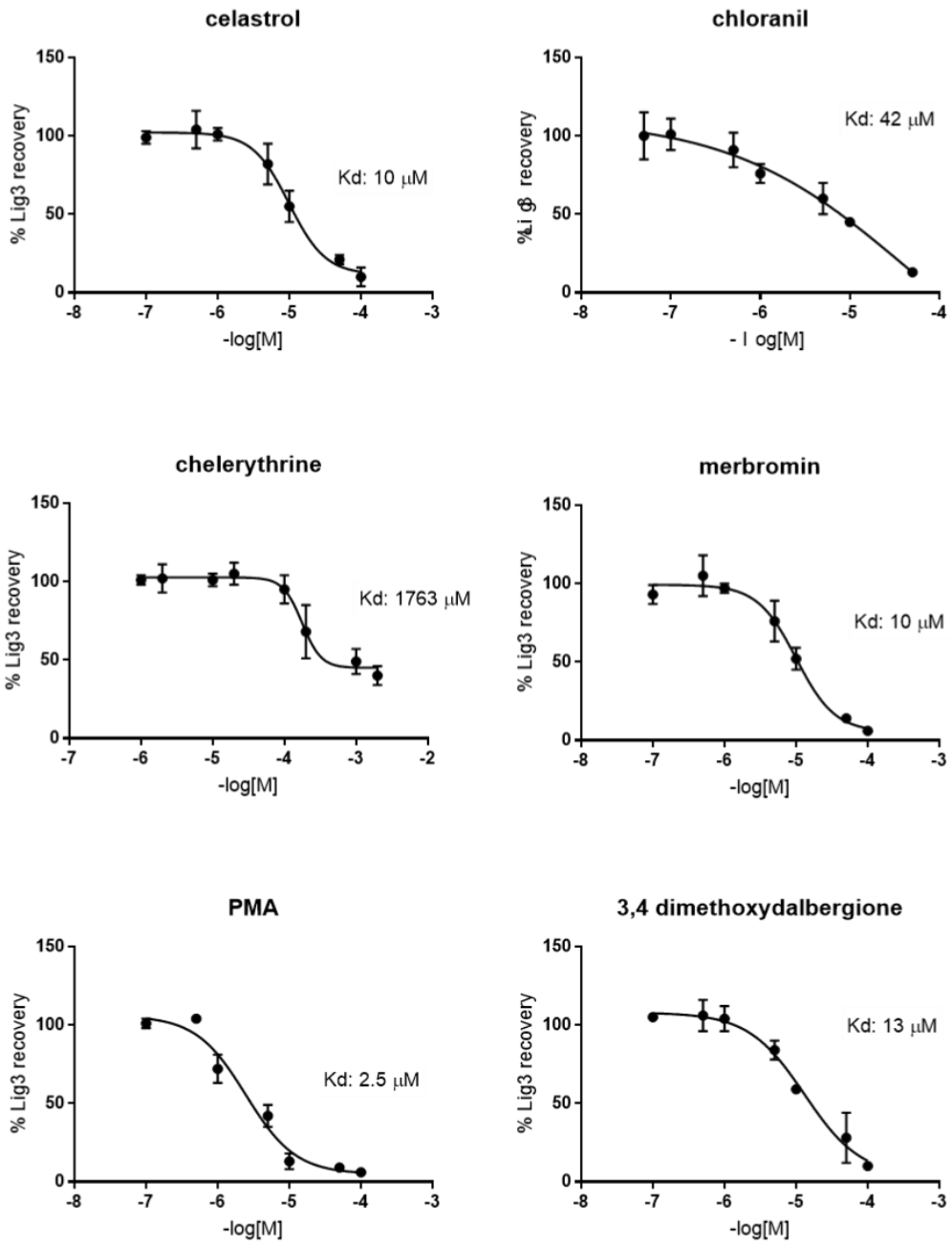


Figure 4.7 Dose-response analysis for XRCC1/Ligase3 modulators to determine IC_{50} values. Curves were fitted with this equation: Y=a+(b-a)/(1+10^((LogIC50-X)*HillSlope))

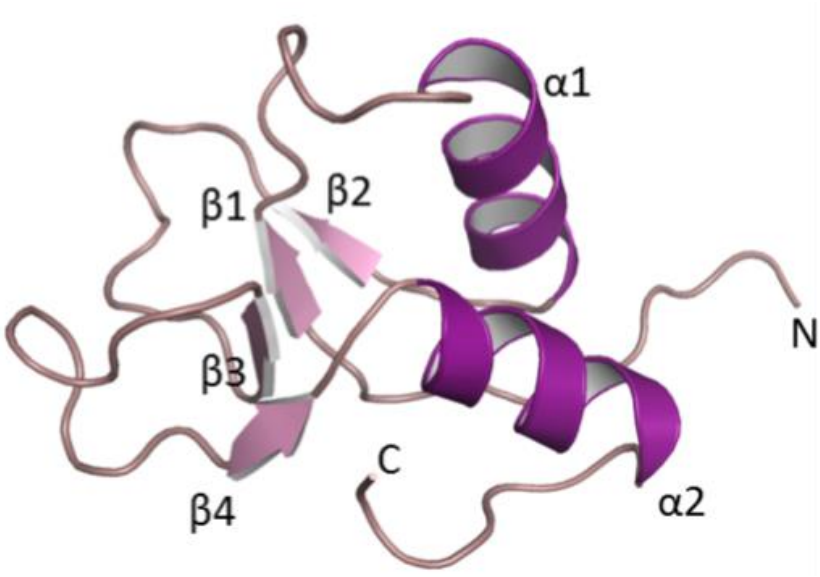


Figure 4.8 Structure of apo Ligase3$^{931\text{-}1009}$. None of the compounds tested were present in the refined structure.

Table 4.1 Data processing and model refinement statistics of Ligase3$^{931-1009}$ inhibitor-soaked crystals.

	Phenyl mercuric acetate	Merbromin	Celastrol	Chloranil	Chelerythrine
Resolution range	71.03-2.2	71.09-2.1	70.89-2.1	70.61-2.59	70.76-2.5 (2.6-2.5)
Unique reflections	8778	10025	9986	5351	6022
Redundancy	7.3 (7.5)	8.6 (8.8)	9.4 (9.6)	8.9 (8.8)	13.8 (14.2)
Mean I/sigI	16.4 (3.7)	14.7 (4.8)	19 (5.3)	17.7 (4.8)	16.6 (4.4)
Rmerged (%)	6.2 (52.9)	7 (33)	5.4 (33.1)	7 (36.9)	7.4 (51.9)
Completeness (%)	100 (100)	99.5 (100)	99.6 (100)	99.0 (96.5)	100 (100)
Unit cell	71.03/71.03/63.90	71.09/71.09/64.05	70.96/70.96/63.96	70.61/70.61/64.5	70.76/70.76/63.9
Angles	90/90/90	90/90/90	90/90/90	90/90/90	90/90/90
Space group	$P4_32_12$	$P4_32_12$	$P4_32_12$	$P4_32_12$	$P4_32_12$
Rwork/Rfree (%)	20.5/23.2	19.4/23.9	19.7/22.8	20/23.4	19.7/23.5
R.M.S.D bonds (Å)	0.007	0.009	0.007	0.009	0.008
R.M.S.D angles (°)	0.849	0.881	0.792	1.055	0.860

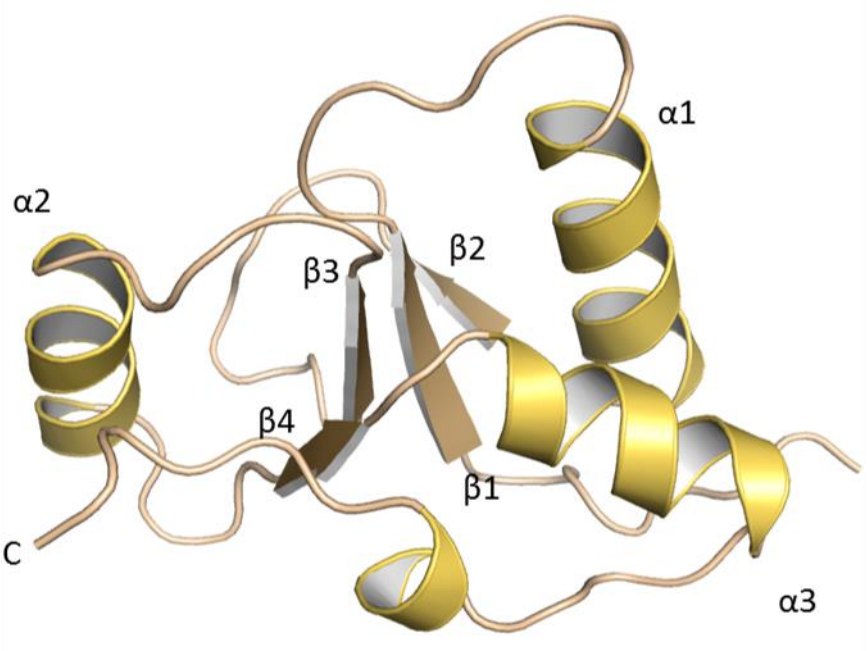


Figure 4.9 Structure of apo XRCC1$^{538-633}$. None of the compounds tested were present in the refined structure.

Table 4.2 Data processing and refinement statistics of XRCC1$^{538-633}$ and soaked inhibitors.

	Celastrol	Chloranil
Resolution range	54.05-2.80	54.16-2.8
Unique reflections	5379	5494
Redundancy	5.5 (5.0)	6.5 (6.7)
Mean I/sigI	10.9 (5.3)	11 (3.5)
Rmerged (%)	8.6 (21.6)	14.8 (52.5)
Completeness (%)	98.2 (94.7)	99.9 (100)
Unit cell	37.77/54.05/100.20	37.98/54.16/100.18
Angles	90/90/90	90/90/90
Space group	$P2_12_12_1$	$P2_12_12$
Rwork/Rfree (%)	18.1/26.2	18.5/24.7
R.M.S.D bonds (Å)	0.009	0.008
R.M.S.D angles (°)	1.041	0.958

4.6 Discussion

The interaction between XRCC1 and Ligase3 was first characterized by Nash et al. and was suggested to be localized to the C-terminal BRCT domains of both proteins [Nash et al. 1997]. Interaction of XRCC1 and Ligase3 is important for stabilizing Ligase3 in cells and disruption of this interaction leads to reduction of DNA repair capacity [Moore et al. 2000]. We hypothesized that disrupting this final step of single strand break repair could be exploited to induce synthetic lethality in a double strand break repair compromised cancer cell. We first generated constructs of XRCC1 and Ligase3 ($XRCC1^{219-633}$ and $Ligase3^{931-1009}$) and verified their interaction by size exclusion chromatography and far western blotting. We then developed a 'magnetic fishing assay' that permitted detection of small molecule modulators capable of disrupting the XRCC1/Ligase3 complex. This assay demonstrated good reproducibility (Z' score of 0.66) and was subsequently used to screen for modulators.

We chose to use a subset of the Canadian Compound Selection for our screen since most of the compounds are bioactive molecules that have demonstrated biological activity *in vivo*. Of the 6 compounds capable of disrupting XRCC1/Ligase3 interaction (phenyl mercuric acetate, chloranil, merbromin, celastrol, 3,4-dimethoxydalbergione and chelerythrine), some have documented biological activities. In particular, phenyl mercuric acetate has been widely used as an anti-fungal agent; merbromin as an antiseptic;

while celastrol has been shown to possess anti-inflammatory, anti-oxidant and anti-cancer activities; and finally, chelerythrine has been used as an anti-bacterial compound. While the effects of these compounds have been documented, their modes of action have remained elusive. Results presented here show that at least one way these compounds could impact biology is by disrupting the XRCC1/Ligase3 complex, and possibly single strand break repair. Nevertheless, at this point it is not yet clear if any of the compounds identified here will have an impact on DNA repair in cells. Cell-based studies will be required to further assess the ability of identified modulators to disrupt BER and SSBR. If favorable results were to be obtained from such studies, these compounds could be tested for the ability to sensitize cancer cells harboring various genetic deficiencies.

Since crystals of Ligase$^{931\text{-}1009}$ and XRCC1$^{538\text{-}633}$ were readily generated, we soaked each of the available compounds into both proteins. Unfortunately, none of the compounds interacted with crystallized Ligase3, and at least two compounds did not interact with crystallized XRCC1. It is possible that a surface buried by crystal lattice contacts may also be where compounds bind. Interestingly, XRCC1 crystals soaked with merbromin or phenyl mercuric acetate were rapidly degraded suggesting these compounds might interact at a region important for crystal lattice contacts.

The structure of the BRCT-mediated interacting region of XRCC1/Ligase3 has been determined [Cuneo et al. 2011]. This structure

demonstrated that the interaction of BRCT domains from XRCC1 and Ligase3 buries approximately 1750 $Å^2$ and involves both ionic and hydrophobic interactions. The primary interface in this PPI is mediated through the α1 helix of XRCC1 and α1 helix of Ligase3, where five inter-protein hydrogen bonds form. Interestingly, within the crystal, XRCC1 forms different oligomers, a homotetramer with itself and also a dimer of heterodimers (two XRCC1 BRCT2 domains interacting with two Ligase3 BRCT domains). In the dimer of heterodimers, one XRCC1 BRCT2 domain can simultaneously interact with a Ligase3 BRCT domain as well as another XRCC1 BRCT domain. Additionally, the structure indicates that an N-terminal region preceding the XRCC1 BRCT domain provides additional contacts to the hetero-complex. Therefore, this region may provide selectivity in favoring interaction of XRCC1/Ligase3 over XRCC1/XRCC1. Since our $XRCC1^{538-633}$ construct lacked this important N-terminal region, it is possible that our identified modulators interact at this site. Attempts at crystallization including an extended N-terminus of $XRCC1^{-}$BRCT2 may prove useful in future endeavors to obtain structural information of modulator compounds in complex with XRCC1.

In this work we identify the first compounds to be reported with ability to disrupt XRCC1/Ligase3 interaction. Although these compounds have not been tested for activity in cells, the fact that most identified compounds have already been reported to be bioactive suggests they should be accessible

to and tolerated by cells. It will be particularly exciting to determine if these compounds are able to selectively sensitize cancer cells toward DNA damaging chemotherapeutic agents.

5 Summary, future directions and conclusions

5.1 Significance of XRCC1-DNA binding

XRCC1 has been described as a scaffolding protein in the literature, where its main role is to localize and orchestrate enzymatic partners that perform repair of damaged DNA. When work on this thesis began, interacting partners of XRCC1 had already been documented and shown to encompass all SSBR and BER proteins identified to date. However, little was known about how these protein-protein interactions govern the repair process. Several studies found that aside from a stable interaction with Ligase3, other protein complexes formed by XRCC1 are highly dynamic and appear to be formed only as needed during repair. The molecular mechanisms underlying the series of decisive recruitment and displacement events necessary to control repair events remains unclear.

In addition to mediating protein-protein interactions, several reports have demonstrated that XRCC1 is able to bind DNA *in vitro*. Consequently, it has been suggested that this DNA binding activity may be important for localizing repair factors to the site of DNA damage; however, this was never directly demonstrated. Key findings in the literature relevant to XRCC1 DNA binding can be summarized as follows: 1) the NTD of XRCC1 exhibits preferential binding to damaged DNA (i.e. nicked and gapped containing); 2) full length XRCC1 binds DNA non-specifically, but shows a slight preference for damaged DNA; and 3) both of the BRCT domains of XRCC1 can be crosslinked to DNA and DNA-containing protein complexes (i.e.

BRCT1 with APE1/DNA complex and BRCT2 with DNA alone). Work presented in this thesis demonstrates the presence of a DNA binding module encompassing BRCT1 and the preceding 80 residues. Our findings indicate that a key mediator of this interaction is the preceding N-terminal linker, where Ala substitution of selected residues (K243, K245, R246, K247, K271, R272, K274) abolishes DNA binding *in vitro*. The robustness of this interaction was underscored by the difficulty in disrupting the interaction. Substitution of less than seven residues did not significantly reduce DNA binding. This unique region contains many positively charged residues, which appear to facilitate binding to the negatively charged DNA back bone. Contrary to earlier reports, DNA binding of XRCC1 was found to be entirely non-specific (Figure 2.4). Since XRCC1 is involved in two major DNA repair pathways, which each generate several unique DNA structural intermediates, non-specific DNA binding of XRCC1 may permit retention to the damage region at all stages of repair independent of the type of DNA intermediate present. Additionally, the lack of DNA binding specificity may be important to prevent XRCC1 from occluding the damage site from other repair enzymes that require access. Regardless, findings reported in this thesis indicate that DNA binding activity of XRCC1 is required in cells in order to remain anchored to damage sites as repair progresses. Such a mechanism may help prevent further damage to the DNA repair intermediate. Prolonged residence of XRCC1 at the damage

site has been suggested as a potential means to ensure a given DNA repair intermediate remains protected and is passed directly to the next repair enzyme when appropriate [Prasad et al. 2010].

Since the dominant DNA binding activity of XRCC1 is non-specific, the question remains how would XRCC1 remain localized to the damage site? It is possible that subtle conformational changes occur within XRCC1 and/or DNA as XRCC1 encounters different DNA substrates during repair. These effects might be difficult to capture in standard EMSA assays used in this thesis. Nevertheless, staged conformational changes in the XRCC1-DNA complex could enhance specificity of enzyme binding; however, no evidence currently exists to support this possibility. It seems more likely that the controls for enzyme recruitment lie in post translational modifications that are imposed upon XRCC1 and/or its binding partners (e.g. phosphorylation, PAR formation). Structural studies of XRCC1/polB have shown that the NTD of XRCC1 can be oxidized to generate intramolecular disulfide bonds and that this process enhances polB binding [Cuneo et al. 2010]. This observation provides a clear example of how an induced conformational change controls binding specificity during repair and further supports the plausibility that additional ‘induced’ conformational changes may be important for binding specificity during other stages of repair. Structural studies of XRCC1 in a complex with other repair enzymes and DNA will be helpful in evaluating possible control mechanisms.

Our SAXS analysis of XRCC1's DNA binding region (residues 219-415), indicates that the presumed flexible linker (residues 219-300) forms a stable structured state imposed by a series of resident proline residues. The spacing of these proline residues is highly favorable for formation of poly-proline type 2 helices, which are largely ignored in secondary structure prediction algorithms such as psi-pred. While poly-proline helices do not adopt geometrically constrained structures like alpha helices and beta strands, they do provide structural integrity to a region and facilitate formation of stable structures. Consistent with results from our SAXS analysis, the 219-300 linker has been shown to be surprisingly resistant to *in vitro* proteolysis [Ali et al. 2009].

At this time, it is not clear what purpose proline-induced rigidity serves within the 219-300 linker. Our mutational studies indicate that ordered structure is not required for DNA binding activity of XRCC1. The possibility remains, however, that rigidity within the linker could be important for spatially constraining domains of XRCC1 from each other or from interaction with partner repair proteins. Separation of domains, and by extension their protein partners, may be important to prevent unauthorized access of repair enzymes to inappropriate DNA repair intermediate. Since miss-timed repair events have severe consequences (e.g. abortive ligation by DNA Ligase3 and 4), it is vital to accurately choreograph repair events. Although substitution of several proline residues did not impact DNA

binding, it is possible that these substitutions may dis-regulate other repair events and it will be very interesting to explore these possibilities further.

5.2 Significance of XRCC1/Ligase3 interaction

XRCC1 and Ligase3 are believed to remain constitutively bound *in vivo*. Knocking out XRCC1 has been shown to result in a reduction of Ligase3 in cells, implying that this interaction is highly important for Ligase3 stability. XRCC1/Ligase3 interaction is further suggested to be responsible for ensuring success of the final ligation steps in both BER and short patch SSBR. Indeed, mutations that disrupt XRCC1/Ligase3 interaction reduce SSBR in a cell cycle dependent manner. When work first began on this project, the structure of the interacting complex was not yet determined. Therefore, I initially sought to co-crystallize the XRCC1/Ligase3 BRCT complex. Although we were able to crystallize each protein separately, the BRCT-BRCT complex proved more elusive. During my structural characterization of these domains it became clear that isolated BRCT domains of both XRCC1 and Ligase3 showed a propensity to form homo-dimers (in solution) and homo-tetramers (within the crystal). It was not clear why self-association was being favored over heterocomplex formation until the efforts of another group [Cuneo et al. 2011] resulted in successful structural determination of the XRCC1/Ligase3 complex. Their structure revealed that a significant contributor to complex stability and specificity involves a short N-terminal extension preceding the XRCC1 BRCT domain,

which was lacking from my construct. Nevertheless, from this analysis it became clear that even small extensions of unstructured regions adjoining BRCT domains could play important biological functions. At this time, it is not clear if XRCC1 and Ligase3 adopt quaternary structures in excess of a simple homodimer, but further studies should be conducted to investigate this possibility.

Work presented here demonstrates that the interaction of XRCC1 and Ligase3 can be disrupted by small molecule modulators. Our 'magnetic fishing' assay was able to identify six bioactive compounds that are capable of disrupting this complex: phenyl mercuric acetate, chloranil, merbromin, celastrol, 3,4-dimethoxydalbergione and chelerythrine. At this point, we have had little success in crystallizing these compounds in complex with Ligase3 and XRCC1. All conditions we have tested result in severe crystal damage when compounds are present, suggesting that compounds interact with the protein in a fashion that prohibits crystal lattice contacts. The next priority will be to confirm which protein each of the compounds interact with, and to determine the mode of action of these compounds in disrupting the XRCC1/Ligase3 complex.

Even if none of these compounds proves useful as a chemotherapeutic adjuvant, they may still be useful as chemical probes to further study DNA repair. Our *in vitro* analysis indicates that these compounds can disrupt XRCC1/Ligase3 interaction. Since several

modulators have already been demonstrated to be bioactive, it seems likely that they may be accessible to and tolerated by cells. If so, there is a reasonable chance that they may be effective inhibitors of BER and short patch SSBR in cells. Use of these modulators as chemical probes of BER and SSBR may be particularly enlightening when studying mature neurons, where the majority of the cell population are post-mitotic [Rulten et al. 2013, Kole et al. 2013]. Since these cells are no longer replicating their chromosome for mitosis, they lack any replication coupled DNA repair such as homologous recombination and hence heavily rely on SSBR and BER for repair. The biological impact of disrupting SSBR and BER in these mature neurons would be insightful to both DNA repair and neurobiology fields. Furthermore, this cell type represents a class of cells that have reduced/impaired replication coupled DNA repair, which could extend to certain populations of cancer cells such as BRCA1 deficient cells. Comparing and contrasting these two different types of cells, and their behavior in the presence of our identified modulators could answer and raise interesting new questions.

5.3 Future directions

5.3.1 DNA binding and repair kinetics *in vivo*

We have identified mutations that disrupt the ability of XRCC1 to interact with DNA *in vitro* and *in vivo*. The next important step will be to validate the biological significance of this DNA binding in regards to DNA

repair. Such an experiment would involve complementation studies of P1/3 in a human XRCC1 knockout cell line. These cells could be challenged with a variety of DNA damaging agents (alkylating or oxidative) and compared for repair efficiency. If the DNA binding ability of XRCC1 is indeed important for its biological function, we would expect to see a difference in repair between those cells complemented with wild type and the P1/3 mutant. Repair efficiency could be monitored by comet assay [Collins 2004]. In this experiment, we would expect to see an increased amount of unrepaired DNA damage in P1/3 complemented cells.

Since mutations that abolish DNA binding are located at or near other protein-protein interactions sites (e.g. BRCT1 interacts with PARP1 and other DNA glycosylase), we need to investigate whether these mutations disrupt other interactions. This is particularly important if *in vivo* assays show a difference between wild type and P1/3 mutant, because the disruption of protein-protein interactions would also be expected to contribute to decreased repair efficiency. Since PARP1 is usually the first protein to respond to DNA single strand breaks, we would need to verify that the interaction between PARP1 and XRCC1 is unaffected by the P1/3 mutation. With purified PARP1 and XRCC1 (both wild type and mutants), we could analyze their interaction using simple biophysical analysis such as size exclusion chromatography or far western blotting (as described in chapter 4). Ultimately, the interaction between the P1/3 mutant and other

members of SSBR and BER, namely Ligase3, polB, PNKP and APE1, will need to be verified to ensure the observed consequences of the P1/3 mutant is attributed to the abolishment of DNA binding alone.

Given the lack of an effect of proline mutants (residues 219-300) on DNA binding, the similar experiments to those outlined above could also be performed to further investigate potential roles for this interesting region of XRCC1. The fact that proline mutants selectively disrupt local structure without altering DNA binding could be highly useful for teasing out specific roles for this segment of XRCC1 (i.e. through co-localization studies of other repair factors with XRCC1 at sites of damage).

5.3.2 Structural characterization of XRCC1's DNA binding domain

It is currently unclear whether the interaction between XRCC1 and DNA alters the overall conformation of either component. Since XRCC1 is known to stimulate enzymatic activity of its partners, one possibility is that XRCC1 alters the structure of DNA to favor enzyme binding and/or activity. A high resolution X-ray structure of XRCC1 bound to DNA would identify any changes to DNA once XRCC1 is bound. Additionally, this structure would be the first to show the conformation and any secondary structure present in the N-terminal linker region.

Current efforts in crystallization of the XRCC1/DNA complex have yielded some preliminary crystals; however, they do not show any diffraction. It seems most important that the ideal DNA substrate be

identified, as the substrate seems to contribute most to crystal quality. A systematic approach of altering the DNA substrate, centering around 20bp in length with varying number of base over hangs would be an appropriate start. Since XRCC1$^{219-415}$ is highly soluble in low salt conditions (30 mg/mL) NMR studies should also be attempted.

5.3.3 Determining the mode of action of XRCC1/Ligase3 modulators

We have identified six compounds that are capable of disrupting the XRCC1/Ligase3 interaction, however the mode of action remains to be determined. First, the protein with which each compound interacts must be identified. To this end, we could use circular dichroism temperature melt analysis of individual proteins in the presence of compounds to monitor change in thermal stability dependent on compound binding. Isothermal scanning calorimetry and/or thermofluor assays could also be used for similar purposes.

Since our XRCC1 construct is rather large, locating the binding site may require more in-depth experiments. This would involve generating smaller truncations within the 219-633 region, and testing for the ability to interact with compounds. An ideal experiment to test this would be saturation transfer difference (STD) NMR, where a signal of the compound will only be detected if a direct interaction between compound and protein exists [Viegas et al. 2011]. Additionally, this experiment could also provide dissociation constants, as well as highlight which component of the

molecule is important for the interaction. This could lead to structure-activity relationship studies and further development of our identified compounds (i.e. chemical group modification) for higher binding affinities and optimal pharmacokinetics.

The impact of our identified compounds on DNA repair *in vivo* will have to be verified. We would need to compare SSBR and BER repair in the presence or absence of compounds in a variety of cell lines using the comet assay. Since XRCC1/Ligase3-dependent deficiency of SSBR and BER can be overcome when the cell enters S phase, care would need to be taken to monitor repair before this phase. Alternatively, we could artificially arrest the cell cycle and observe the impact of these compounds by serum starvation or amino acid starvation [Rosner et al. 2013].

5.4 Conclusions

Work presented in this thesis furthers our functional understanding of the DNA repair protein XRCC1. Our work has challenged published results that indicated the DNA binding of XRCC1 is dependent on the NTD, and purposes an alternative DNA binding site. As well, cell-based studies with DNA binding mutants provide the first experimental evidence that DNA binding of XRCC1 is important for function *in vivo*. Our biochemical and structural work have also provided the groundwork for future experiments that seek to understand the importance of this new-found DNA binding in the context of DNA metabolism.

Importantly, we have identified the first reported small molecule inhibitors of XRCC1/Ligase3 interaction. These compounds require further characterization, but show promise for use as probes of XRCC1/Ligase3 function in cells, and may have further potential as adjuvant for chemotherapeutic treatment.

References

Adams, P. D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L., Kapral, G.J., Grosse-Kunstleve, R.W., McCoy, A.J., Moriarty, N.W., Oeffner, R., Read, R.J., Richardson, D.C., Richardson, J.S., Terwilliger, T.C., Zwart, P.H. PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66,** 213–221 (2010).

Adzhubei, A. a, Sternberg, M. J. E. & Makarov, A. a. Polyproline-II helix in proteins: structure and function. *J. Mol. Biol.* **425,** 2100–32 (2013).

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

Ali, A. a E., Jukes, R. M., Pearl, L. H. & Oliver, A. W. Specific recognition of a multiply phosphorylated motif in the DNA repair scaffold XRCC1 by the FHA domain of human PNK. *Nucleic Acids Res.* **37,** 1701–12 (2009).

Allen, B. L., Uhlmann, F., Gaur, L.K., Mulder, B.A., Posey, K.L., Jones, L.B., Hardin, S.H. DNA recognition properties of the N-terminal DNA binding domain within the large subunit of replication factor C. *Nucleic Acids Res.* **26,** 3877–3882 (1998).

Almeida, K. H. & Sobol, R. W. A unified view of base excision repair: lesion-dependent protein complexes regulated by post-translational modification. *DNA Repair (Amst).* **6,** 695–711 (2007).

Asherie, N. Protein crystallization and phase diagrams. *Methods* **34,** 266–272 (2004).

Battye, T.G.G., Kontogiannis, L., Johnson, O., Powell, H.R., Leslie, A.G.W. IMosflm: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Cryst*. **D67**, 271-281 (2011).

Beernink, P. T., Hwang, M., Ramirez, M., Murphy, M.B., Doyle, S.A., Thelen, M.P. Specificity of protein interactions mediated by BRCT domains of the XRCC1 DNA repair protein. *J. Biol. Chem.* **280,** 30206–13 (2005).

Berquist, B. R., Singh, D.K., Fan, J., Kim, D., Gillenwater, E., Kulkarni, A., Bohr, V.A., Ackerman, E.J., Tomkinson, A.E., Wilson III, D.M. Functional capacity of XRCC1 protein variants identified in DNA repair-deficient

Chinese hamster ovary cell lines and the human population. *Nucleic Acids Res.* **38,** 5023–5035 (2010).

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

Bruner, S. D., Norman, D. P. & Verdine, G. L. Structural basis for recognition and repair of the endogenous mutagen 8-oxoguanine in DNA. *Nature* **403,** 859–66 (2000).

Bryant, H. E., Schultz, N., Thomas, H.D., Parker, K.M., Flower, D., Lopez, E., Kyle, S., Meuth, M., Curtin, N.J., Helleday, T. Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. *Nature* **434,** 913–7 (2005).

Cabrita, L. D., Dai, W. & Bottomley, S. P. A family of E. coli expression vectors for laboratory scale and high throughput soluble protein production. *BMC Biotechnol.* **6,** 12 (2006).

Caldecott, K. W., Tucker, J. D., Stanker, L. H. & Thompson, L. H. Characterization of the XRCC1-DNA ligase III complex in vitro and its absence from mutant hamster cells. *Nucleic Acids Res.* **23,** 4836–43 (1995).

Caldecott, K. W., Aoufouchi, S., Johnson, P. & Shall, S. XRCC1 polypeptide interacts with DNA polymerase β and possibly poly (ADP-ribose) polymerase, and DNA ligase III is a novel molecular 'nick-sensor' in vitro. **24,** 4387–4394 (1996).

Caldecott, K. W. XRCC1 and DNA strand break repair. *DNA Repair (Amst).* **2,** 955–969 (2003).

Campalans, A., Marsin, S., Nakabeppu, Y., O'Connor, T.R., Boiteux, S., Radicella, J.P. XRCC1 interactions with multiple DNA glycosylases: a model for its recruitment to base excision repair. *DNA Repair (Amst).* **4,** 826–35 (2005).

Campalans, A., Kortulewski, T., Amouroux, R., Menoni, H., Vermeulen, W., Radicella, J.P. Distinct spatiotemporal patterns and PARP dependence of XRCC1 recruitment to single-strand break and base excision repair. *Nucleic Acids Res.* **41,** 3115–29 (2013).

Campalans, A., Moritz, E., Kortulewski, T., Biard, D., Epe, B., Radicella, J.P. Interaction with OGG1 is required for efficient recruitment of XRCC1 to base excision repair and maintenance of genetic stability after exposure to oxidative stress. *Mol. Cell. Biol.* **35,** 1648–58 (2015).

Cherry, A. L., Nott, T.J., Kelly, G., Rulten, S.L., Caldecott, K.W., Smerdon, S.J. Versatility in phospho-dependent molecular recognition of the XRCC1 and XRCC4 DNA-damage scaffolds by aprataxin-family FHA domains. *DNA Repair (Amst).* **35,** 116–125 (2015).

Cuneo, M. J. & London, R. E. Oxidation state of the XRCC1 N-terminal domain regulates DNA polymerase beta binding affinity. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 6805–10 (2010).

Cuneo, M. J., Gabel, S. A., Krahn, J. M., Ricker, M. A. & London, R. E. The structural basis for partitioning of the XRCC1/DNA ligase III-α BRCT-mediated dimer complexes. *Nucleic Acids Res.* **39,** 7816–7827 (2011).

Collins, A. R. The comet assay for DNA damage and repair: principles, applications, and limitations. *Mol. Biotechnol.* **26,** 249–261 (2004).

Date, H., Igarashi, S., Sano, Y., Takahashi, T., Takahashi, T., Takano, H., Tsuji, S., Nishizawa, M., Onodera, O. The FHA domain of aprataxin interacts with the C-terminal region of XRCC1. *Biochem. Biophys. Res. Commun.* **325,** 1279–85 (2004).

Della-Maria, J., Hedge, M.L., McNeill, D.R., Matsumoto, Y., Tsai, M.S., Ellenberger, T., Wilson III, D.M., Mitra, S., Tomkinson, A.E. The interaction between polynucleotide kinase phosphatase and the DNA repair protein XRCC1 is critical for repair of DNA alkylation damage and stable association at DNA damage sites. *J. Biol. Chem.* **287,** 39233–39244 (2012).

Dianova, I. I., Sleeth, K.M., Allinson, S.L., Parsons, J.L., Breslin, C., Caldecott, K.W., Dianov, G.L. XRCC1-DNA polymerase β interaction is required for efficient base excision repair. *Nucleic Acids Res.* **32,** 2550–2555 (2004).

Fan, J., Otterlei, M., Wong, H.-K., Tomkinson, A. E. & Wilson, D. M. XRCC1 co-localizes and physically interacts with PCNA. *Nucleic Acids Res.* **32,** 2193–201 (2004).

Farmer, H., McCabe, N., Lord, C.J., Tutt, A.N.J., Johnson, D.A., Richardson, T.B., Santarosa, M., Dillon, K.J., Hickson, I., Knights, C., Martin, N.M.B., Jackson, S.P., Smith, G.C.M., Ashworth, A. Targeting the

DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature* **434,** 917–21 (2005).

Feng, H., Parker, J. M., Lu, J. & Cao, W. Effects of deletion and site-directed mutations on ligation steps of NAD+-dependent DNA ligase: A biochemical analysis of BRCA1 C-terminal domain. *Biochemistry* **43,** 12648–12659 (2004).

Fojo, T. Journal of the National Cancer Institute. **93,** 1434–1436 (2001).
The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.
Abbotts, R., Thompson, N. & Madhusudan, S. DNA repair in cancer: Emerging targets for personalized therapy. *Cancer Manag. Res.* **6,** 77–92 (2014).

Fortini, P. & Dogliotti, E. Base damage and single-strand break repair: Mechanisms and functional significance of short- and long-patch repair subpathways. *DNA Repair (Amst).* **6,** 398–409 (2007).

Gabel, S. a., DeRose, E. F. & London, R. E. XRCC1 interaction with the REV1 C-terminal domain suggests a role in post replication repair. *DNA Repair (Amst).* **12,** 1105–1113 (2013).

George, R. a & Heringa, J. An analysis of protein domain linkers: their classification and role in protein folding. *Protein Eng.* **15,** 871–879 (2003).

Ghosh, S., Canugovi, C., Yoon, J.S., Wilson III, D.M., Croteau, D.L., Mattson, M.P., Bohr, V.A. Partial loss of the DNA repair scaffolding protein, Xrcc1, results in increased brain damage and reduced recovery from ischemic stroke in mice. *Neurobiol. Aging* **36,** 2319–2330 (2015).

Glover, J. N. M., Williams, R. S. & Lee, M. S. Interactions between BRCT repeats and phosphoproteins: Tangled up in two. *Trends Biochem. Sci.* **29,** 579–585 (2004).

Good, M. C., Zalatan, J. G. & Lim, W. A. Scaffold Proteins: Hubs for Controlling the Flow of Cellular Information. *Science (80-.).* **332,** 680–686 (2011).

Hecht, S. S. DNA adduct formation from tobacco-specific N-nitrosamines. *Mutat. Res. - Fundam. Mol. Mech. Mutagen.* **424,** 127–142 (1999).

Hosfield, D. J., Mol, C. D., Shen, B. & Tainer, J. A. Structure of the DNA repair and replication endonuclease and exonuclease FEN-1: Coupling DNA and PCNA binding to FEN-1 activity. *Cell* **95,** 135–146 (1998).

Iles, N., Rulten, S., El-Khamisy, S. F. & Caldecott, K. W. APLF (C2orf13) is a novel human protein involved in the cellular response to chromosomal DNA strand breaks. *Mol. Cell. Biol.* **27,** 3793–803 (2007).
Jacobs, A. L. & Schär, P. DNA glycosylases: In DNA repair and beyond. *Chromosoma* **121,** 1–20 (2012).

Jekimovs, C., Bolderson, E., Suraweera, A., Adams, M., O'Byrne, K.J., Richard, D.J. Chemotherapeutic compounds targeting the DNA double-strand break repair pathways: the good, the bad, and the promising. *Front. Oncol.* **4,** 86 (2014).

Jones, S. & Thornton, J. M. Principles of protein-protein interactions. *Proc. Natl. Acad. Sci. U. S. A.* **93,** 13–20 (1996).

Kim, I. K., Stegeman, R. A., Brosey, C. A. & Ellenberger, T. A quantitative assay reveals ligand specificity of the DNA scaffold repair protein XRCC1 and efficient disassembly of complexes of XRCC1 and the poly(ADP-ribose) polymerase 1 by poly(ADP-ribose) glycohydrolase. *J. Biol. Chem.* **290,** 3775–3783 (2015).

Kirby, T. W., Gassman. N.R., Smith, C.E., Pedersen, L.C., Gabel, S.A., Sobhany, M., Wilson, S.H., London, R.E. Nuclear Localization of the DNA Repair Scaffold XRCC1: Uncovering the Functional Role of a Bipartite NLS. *Sci Rep* **5,** 13405 (2015).

Kirkwood, J., Hargreaves, D., O'Keefe, S. & Wilson, J. Using isoelectric point to determine the pH for initial protein crystallization trials. *Bioinformatics* **31,** 1444–1451 (2015).

Kobayashi, M., Ab, E., Bonvin, A. M. J. J. & Siegal, G. Structure of the DNA-bound BRCA1 C-terminal region from human replication factor C p140 and model of the protein-DNA complex. *J. Biol. Chem.* **285,** 10087–10097 (2010).

Kubota, Y., Nash, R.A., Klungland, A., Schär, P., Barnes, D.E., Lindahl, T. Reconstitution of DNA base excision-repair with purified human proteins: interaction between DNA polymerase beta and the XRCC1 protein. *EMBO J.* **15,** 6662–70 (1996).

Kubota, Y., Takanami, T., Higashitani, A. & Horiuchi, S. Localization of X-ray cross complementing gene 1 protein in the nuclear matrix is controlled by casein kinase II-dependent phosphorylation in response to oxidative damage. *DNA Repair (Amst).* **8**, 953–960 (2009).

Kole, A. J., Annis, R. P. & Deshmukh, M. Mature neurons: equipped for survival. *Cell Death Dis.* **4,** e689 (2013).

LaCasse, E. C. & Lefebvre, Y. A. Nuclear localization signals overlap DNA- or RNS-binding domains in nucleic acid-binding proteins. *Nucleic Acids Res.* **23,** 1647–1656 (1995).

Lee, Y. Katyal, S., Li, Y., El-Khamisy, S.F., Russell, H.R., Caldecott, K.W., McKinnon, P.J. The genesis of cerebellar interneurons and the prevention of neural DNA damage require XRCC1. *Nat. Neurosci.* **12,** 973–80 (2009).

Levin, D. S., McKenna, A. E., Motycka, T. A., Matsumoto, Y. & Tomkinson, A. E. Interaction between PCNA and DNA ligase I is critical for joining of Okazaki fragments and long-patch base-excision repair. *Curr. Biol.* **10,** 919–922 (2000).

Leung, C. C. Y. & Glover, J. N. M. BRCT domains: Easy as one, two, three. *Cell Cycle* **10,** 2461–2470 (2011).

Li, M., Lu, L.-Y., Yang, C.-Y., Wang, S. & Yu, X. The FHA and BRCT domains recognize ADP-ribosylation during DNA damage response. *Genes Dev.* **27,** 1752–68 (2013).

Liu, H. & Naismith, J. H. An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnol.* **8,** 91 (2008).

Lu, M., Mani, R.S., Karimi-BUsheri, F., Fanta, M., Wang, H., Litchfeld, D.W., Weinfeld, M. Independent mechanisms of stimulation of polynucleotide kinase/phosphatase by phosphorylated and non-phosphorylated XRCC1. *Nucleic Acids Res.* **38,** 510–521 (2010).

Luo, H., Chan, D.W., Yang, T., Rodriguez, M., Chen, B.P.C., Leng, M., Mu, J.J., Chen, D., Songyang, Z., Wang, Y., Qin, J. A new XRCC1-containing complex and its role in cellular survival of methyl methanesulfonate treatement. *Mol. Cell. Biol.*, 24, 8356-8365 (2004).

Luqmani, Y. A. Mechanisms of drug resistance in cancer chemotherapy. *Med. Princ. Pract.* **14,** 35–48 (2005).

Madabhushi, R., Pan, L. & Tsai, L. H. DNA damage and its links to neurodegeneration. *Neuron* **83,** 266–282 (2014).

Madhusudan, S. & Hickson, I. D. DNA repair inhibition: a selective tumour targeting strategy. *Trends Mol. Med.* **11,** 503–11 (2005).

Mani, R. S., Karimi-Busheri, F., Fanta, M., Caldecott, K.W., Cass, C.E., Weinfeld, M. Biophysical characterization of human XRCC1 and its binding to damaged and undamaged DNA. *Biochemistry* **43,** 16505–14 (2004).

Marintchev, A., Mullen, M.A., Maciejewski, M.W., Pan, B., Gryk, M.R., Mullen, G.P. Solution structure of the single-strand break repair protein XRCC1 N-terminal domain. *Nat. Struct. Biol.* **6,** 884–93 (1999).

Marintchev, A., Robertson, A., Dimitriadis, E.K., Prasad, R., Wilson, S.H., Mullen, G.P. Domain specific interaction in the XRCC1 – DNA polymerase β complex. *Nucleic Acids Res.* **28,** 2049–2059 (2000).

Marsin, S., Vidal, A.E., Sossou, M., Murcia, J.M., Page, F.L., Boiteux, S., Murcia, G.D., Radicella, J.P. Role of XRCC1 in the Coordination and Stimulation of Oxidative DNA Damage Repair Initiated by the DNA Glycosylase hOGG1. *J. Biol. Chem.* **278,** 44068–44074 (2003).

Martínez-Macías, M. I., Córdoba-Cañero, D., Ariza, R. R. & Roldán-Arjona, T. The DNA repair protein XRCC1 functions in the plant DNA demethylation pathway by stimulating cytosine methylation (5-meC) excision, gap tailoring, and DNA ligation. *J. Biol. Chem.* **288,** 5496–505 (2013).

Masson, M., Niedergang, C., Schreiber, V., Muller, S., Murcia, J.M., Murcia, G.D. XRCC1 is specifically associated with poly(ADP-ribose) polymerase and negatively regulates its activity following DNA damage. *Mol. Cell. Biol.* **18,** 3563–71 (1998).

Maynard, S., Fang, E. F., Scheibye-Knudsen, M., Croteau, D. L. & Bohr, V. A. DNA damage, DNA repair, aging, and neurodegeneration. *Cold Spring Harb. Perspect. Med.* **5,** a025130 (2015).

McFadden, M. J., Junop, M. S. & Brennan, J. D. Magnetic 'fishing' assay to screen small-molecule mixtures for modulators of protein - protein interactions. *Anal. Chem.* **82,** 9850–9857 (2010).

Milroy, L., Grossmann, T. N., Hennig, S., Brunsveld, L. & Ottmann, C. Modulators of Protein − Protein Interactions. *Chem. Rev.* **114**, 4695-4748 (2014).

Mol, C. D., Izumi, T., Mitra, S. & Tainer, J. A. DNA-bound structures and mutants reveal abasic DNA binding by APE1 and DNA repair coordination. *Nature* **403,** 451–6 (2000).

Nash, R. a, Caldecott, K. W., Barnes, D. E. & Lindahl, T. XRCC1 protein interacts with one of two distinct forms of DNA ligase III. *Biochemistry* **36,** 5207–11 (1997).

Nazarkina, Z. K., Khodyreva, S. N., Marsin, S., Lavrik, O. I. & Radicella, J. P. XRCC1 interactions with base excision repair DNA intermediates. *DNA Repair (Amst).* **6,** 254–64 (2007).

Nero, T. L., Morton, C. J., Holien, J. K., Wielens, J. & Parker, M. W. Oncogenic protein interfaces: small molecules, big challenges. *Nat. Rev. Cancer* **14,** 248–62 (2014).

Niesen, F. H., Berglund, H. & Vedadi, M. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat. Protoc.* **2,** 2212–2221 (2007).

Parsons, J. L., Dianova, I.I., Finch, D., Tait, P.S., Ström, C.E., Helleday, T., Dianov, G.L. XRCC1 phosphorylation by CK2 is required for its stability and efficient DNA repair. *DNA Repair (Amst).* **9,** 835–841 (2010).

Petoukhov, M. V., Franke, D., Shkumatov, A.V., Tria, G., Kikhney, A.G., Gajda, M., Gorba, C., Mertens, H.D.T. Konarev, P.V, Svergun, D.I. New developments in the ATSAS program package for small-angle scattering data analysis. *J. Appl. Crystallogr.* **45,** 342–350 (2012).

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

Prasad, R., Shock, D. D., Beard, W. A. & Wilson, S. H. Substrate channeling in mammalian base excision repair pathways: Passing the baton. *J. Biol. Chem.* **285,** 40479–40488 (2010).

Pray, L. Discovery of DNA structure and function: Watson and Crick. *Nature Education* **1**(1):100 (2008).

Rambo, R. P. & Tainer, J. A. Accurate assessment of mass, models and resolution by small-angle scattering. *Nature* **496,** 477–81 (2013).

Rosner, M., Schipany, K., Hengstschläger, M. Merging high-quality biochemical fractionation with a refined flow cytometry approach to monitor

nucleocytoplasmic protein expression throughout the unperturbed mammalian cell cycle. *Nature Protocols*, **8**, 602–626 (2013).

Rulten, S. L. & Caldecott, K. W. DNA strand break repair and neurodegeneration. *DNA Repair (Amst).* **12,** 558–567 (2013).

Sultana, R., Abdel-Fatah, T., Abbotts, R., Hawkes, C., Albarakati, N., Seedhouse, C., Ball, G., Chan, S., Rakha, E.A., Ellis, I.O., Madhusudan, S. Targeting XRCC1 deficiency in breast cancer for personalized therapy. *Cancer Res.* **73,** 1621–34 (2013).

Taylor, R. M., Wickstead, B., Cronin, S. & Caldecott, K. W. Role of a BRCT domain in the interaction of DNA ligase III-alpha with the DNA repair protein XRCC1. *Curr. Biol.* **8,** 877–880 (1998).

Thompson, L. H. & West, M. G. XRCC1 keeps DNA from getting stranded. *Mutat. Res.* **459,** 1–18 (2000).

Vidal, A. E., Boiteux, S., Hickson, I. D. & Radicella, J. P. XRCC1 coordinates the initial and late stages of DNA abasic site repair through protein-protein interactions. *EMBO J.* **20,** 6530–6539 (2001).

Viegas, A., Manso, J., Nobrega, F. L. & Cabrita, E. J. Saturation-transfer difference (STD) NMR: A simple and fast method for ligand screening and characterization of protein binding. *J. Chem. Educ.* **88,** 990–994 (2011).

Wardlaw, C. P., Carr, A. M. & Oliver, A. W. TopBP1: A BRCT-scaffold protein functioning in multiple cellular pathways. *DNA Repair (Amst).* **22,** 165–174 (2014).

Watson, J. D. & Crick, F. H. C. Molecular structure of nucleic acids. *Nature* **171,** 737–738 (1953).

Wei, L., Nakajima, S., Hsieh, C.L., Kanno, S., Masutani, M., Levine, A.S., Yasui, A., Lan, Li. Damage response of XRCC1 at sites of DNA single strand breaks is regulated by phosphorylation and ubiquitylation after degradation of poly(ADP-ribose). *J. Cell Sci.* 126, 4414–23 (2013).

Whitehouse, C. J. *et al.* XRCC1 stimulates human polynucleotide kinase activity at damaged DNA termini and accelerates DNA single-strand break repair. *Cell* **104,** 107–17 (2001).

Wong, H. K. & Wilson, D. M. XRCC1 and DNA polymerase β interaction contributes to cellular alkylating-agent resistance and single-strand break repair. *J. Cell. Biochem.* **95,** 794–804 (2005).

Wu, Y., Li, Q., Chen, X.Z. Detecting protein-protein interaction by far western blotting. *Nature Protocols* **2**, 3278 - 3284 (2007).

Yamane, K., Katayama, E. & Tsuruo, T. The BRCT regions of tumor suppressor BRCA1 and of XRCC1 show DNA end binding activity with a multimerizing feature. *Biochem. Biophys. Res. Commun.* **279,** 678–84 (2000).

Zhang, X., Moréra, S., Bates, P.A., Whitehead, P.C., Coffer, A.I., Hainbucher, K., Nash, R.A., Sternberg, M.J.E., Lindahl, T., Freemont, P.S. Structure of an XRCC1 BRCT domain: a new protein-protein interaction module. *EMBO J.* **17,** 6404–11 (1998).