Insights into the recruitment of BRCA1 to double strand DNA breaks

by

Stephen John Campbell

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Biochemistry

Fall 2012

Edmonton, Alberta

For Roxie Jeanne ... and whomever is yet to come ...

Abstract

In response to genomic stress resulting from external and endogenous insults, the cell has acquired a set of complicated pathways that deal with the damage, which are collectively referred to as the DNA damage response (DDR). In order to repair double strand DNA breaks that may have occurred in G2 and M phase of cell growth, the cell employs a pathway called Homologous Recombination (HR) that utilizes the complementary chromatid as a template to ensure that the repair is error free. BReast CAncer 1 (BRCA1), an essential component of HR, is recruited through a large variety of different post translation modifications and protein-protein interactions. Upon recruitment to the site of a DNA double strand break, BRCA1 functions to initiate repair of the damaged strand. The goal of the thesis is to look in detail at the molecular mechanisms involved in certain aspects of BRCA1 recruitment. First, the crystal structures and *in vitro* activity of the RING Finger containing (RNF) E3-ubiquitin ligases RNF8 and RNF168 are discussed. Next, the peptide binding specificities of the phospho-peptide binding BRCA1 C-terminal (BRCT) domains of BRCA1 and MDC1 are compared. Finally, a preliminary screen to identify synthetic inhibitors of the BRCA1/phospho-peptide interaction is performed, as well as a discussion as to their therapeutic relevance.

Acknowledgements

First and foremost, I would like to thank my supervisor, Dr. Mark Glover. He has encouraged my independent thinking, supported good ideas and discouraged my poorer ones...gently. Mark has been a tremendous success personally and professionally; it is hard to imagine a better mentor and I will undoubtedly attempt to model my career after his.

My research on BRCA1 inhibition has been an ongoing collaboration with Amarnath Natarajan at the University of Nebraska. He has contributed valuable scientific and technical advice to this project for which I am thankful for.

My committee members, Dr. Charles Holmes, Dr. Bart Hazes and Dr. Michael Weinfeld, have been very supportive throughout my graduate studies. I would like to thank Dr. Georges Mer especially, for agreeing to be involved in my defence.

I would also like to thank all of the members of the Glover lab, past and present (Charles, Steve, Ross, Dave, Curtis, Nico, Danny, Nina, Megan, Gina, Jun, Joyce, Ruth, Yun, Zahra and Inbal), for their help and support.

Graduate school has encompassed some of the greatest times of my life, a period that I will always cherish. Words cannot express how grateful I am to have had a friend like Charles by my side, providing both emotional support and scientific contributions. He has become one of my best friends, and I will always remember the hours we spent discussing science and life at Tim Horton's. Angela, Roshani, Curtis and Ross have also become close friends, and I am grateful for them. Thanks to all of my friends in the department, particularly Danny, Steve, Ian, Tamara, Allison, OJ, Grant, and the EUPA Ultimate Frisbee team, for all of the good memories.

Most importantly, I would like to thank my wife, Avery, who is truly the love of my life and my best friend. She has supported and encouraged me every step of the way, even when it meant sacrifice on her part. Avery, I love you so much.

Lastly, my parents – Campbell's and Hagen's. Mom and Dad, you have been incredible parents; I still consider you infallible and trust your guidance implicitly in every aspect of my life. Mom and Dad H – I love you like my own parents, and am so grateful for your support as well as the guidance you have given me over the last 7 years. I am so thankful for my family, and hope that I will never let them down – particularly in this career path that I have chosen.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

Å	Angstrom (10-10) meters
ALS	Advanced Light Source
βME	β-mercaptoethanol
BRCT	BRCA1 C-terminal
CMCF	Canadian Macromolecular Crystallography Facility
DDR	DNA damage response
dsDNA	Double-stranded DNA
DSB	Double-stranded break
DTT	Dithiothreitol
EDTA	(Ethylenedinitrilo)-tetraacetic acid
EMSA	Electrophoretic mobility shift assay
FHA	Fork-Head Associated
FITC	Fluorescein isothiocyanate
FP	Fluorescence polarization
GST	Glutathione S-transferase
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HR	Homologous recombination
IR	Ionizing radiation
ITC	Isothermal titration calorimetry
IPTG	Isopropyl β-D-thiogalactopyranoside
LB	Luria Bertani

LC-MS	Liquid chromatography-mass spectrometry
MALDI-TOF	Matrix assisted laser desorption ionization-time of flight
MALLS	Multi angle laser light scattering
MIU	Motif that Interacts with Ubiquitin
NHEJ	Non-homologous end joining
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PDB	Protein Data Bank
PEG	Polyethylene glycol
PPI	Protein Protein Interaction
RING	Really Interesting New Gene
RNF	RING Finger
RMSD	Root Mean Square Deviation
ROS	Reactive Oxygen Species
SDS	Sodium dodecyl sulphate
SEC	Size exclusion chromotography
Se-Met	Seleno-methionine
SSB	Single-stranded break
SSBR	Single strand break repair
ssDNA	Single-stranded DNA
TRIS	Tris(hydroxymethyl)aminomethane
UIM	Ubiquitin Interaction Motif

Chapter 1

Introduction

1.1 DNA damage and Repair

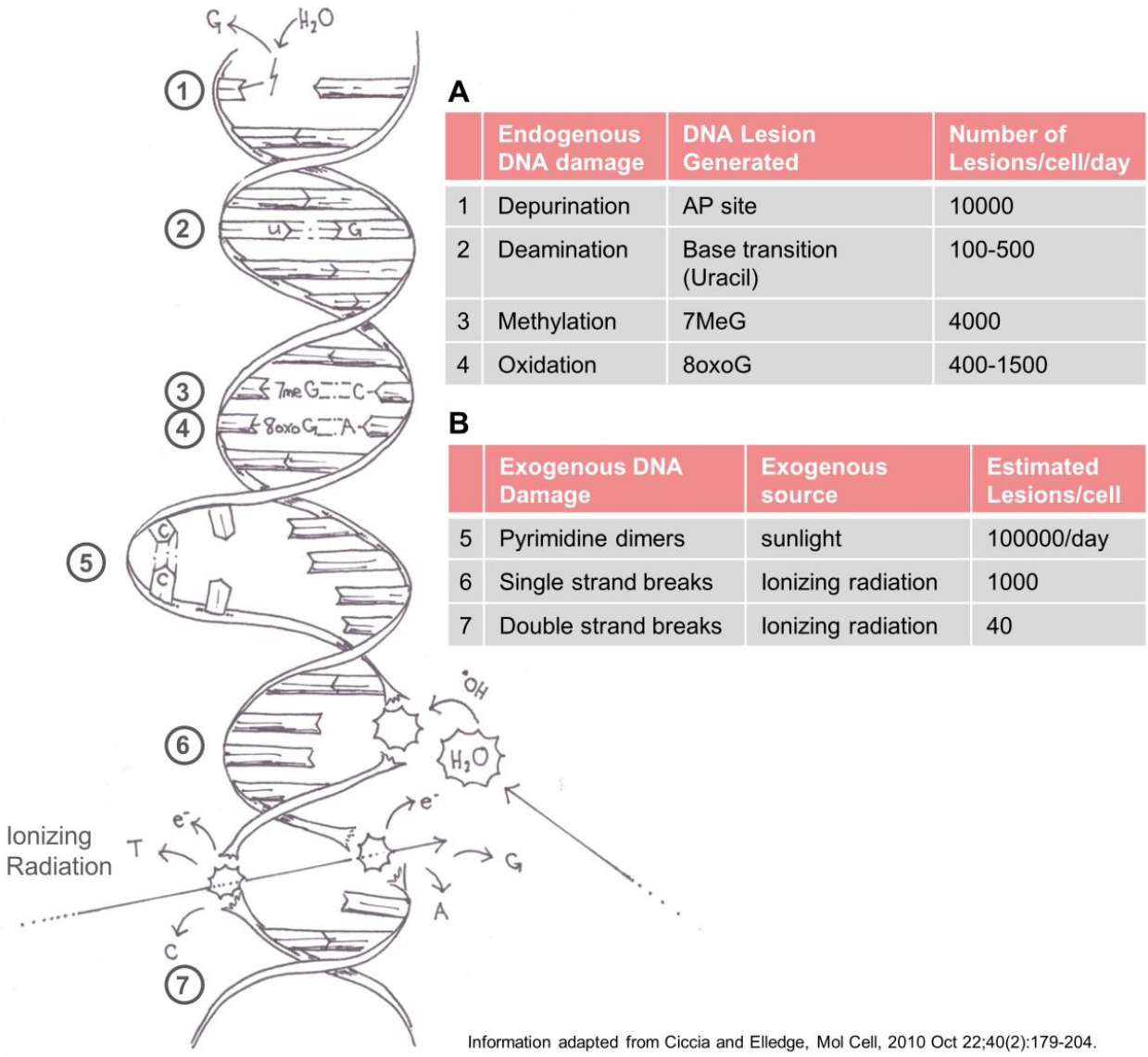
1.1a: Overview. As our genetic material, the main function of DNA is the long term storage of information that encodes many of the functional components of our cells. Given the importance that our genetic material remain unchanged, both external and endogenous insults that compromise its integrity can have lethal consequences. In addition, compromised DNA that is allowed to propagate into daughter cells can lead to transformation of those cells and tumourigenesis. To prevent this from occurring, our cells have evolved a complicated system of pathways that exist in order to repair damaged DNA - referred to as the DNA damage Response (DDR).

1.1b: An "extremely" brief history of DNA repair. During the 1920's and 30's, it became evident that certain types of radiation, particularly UV radiation, had damaging effects on the genetic material. These mutations were considered to be rare and random effects, whose biochemical properties were largely unknown [1]. The response of the cell to this damage was not considered a significant biological process, as protein was still widely accepted as the genetic material, and protein was considered to be very stable and certainly not prone to endogenous or environmental perturbations. In *Correcting the Blueprint of Life*, Errol Friedberg notes that "[if] investigators paid greater attention to the selectivity of short wavelength UV radiation for nucleic acids...our

understanding of the chemical nature of the gene might have been significantly accelerated". However, as observations of "rectification of genetic damage" arose, so did interest in the cell's biochemical response to this damage [2]. As it became widely accepted in the 1950's that DNA is in fact the genetic material, experimental evidence of the existence of DNA damage and the subsequent repair pathways began to emerge. Interest in the harmful biological effects of radiation in the wake of World War II, particularly the development of the first atomic bomb as a result of the Manhattan Project, further accelerated the beginning of the DNA repair field. Soon after, it began to gain notoriety as we observed the actual frequency of DNA damage and the cell's ability to maintain genome integrity in the face of these previously unknown dangers to our physical veracity.

1.1c: Types of DNA damage. Due to the aqueous environment within our cells, many types of damage that affect our genetic material arise spontaneously due to unavoidable interactions with water and oxygen (Figure 1A). Many different lesions result from normal DNA metabolism and happen relatively frequently, only a fraction of which will be mentioned here.

The absence of a 2' OH group in DNA results in its phospho-diester bond being less susceptible to cleavage by hydrolysis relative to RNA. However, this makes the N-glycosidic bond more labile, which is



A

	Endogenous DNA damage	DNA Lesion Generated	Number of Lesions/cell/day
1	Depurination	AP site	10000
2	Deamination	Base transition (Uracil)	100-500
3	Methylation	7MeG	4000
4	Oxidation	8oxoG	400-1500

B

	Exogenous DNA Damage	Exogenous source	Estimated Lesions/cell
5	Pyrimidine dimers	sunlight	100000/day
6	Single strand breaks	Ionizing radiation	1000
7	Double strand breaks	Ionizing radiation	40

Information adapted from Ciccia and Elledge, Mol Cell, 2010 Oct 22;40(2):179-204.

Figure 1. Cartoon representation of common types of DNA damage.

A Endogenous DNA damage is common, but much less detrimental to cellular function than damage incurred by external sources

B Pyrimidine dimers result as a direct result of everyday exposure to sunlight, and are typically repaired through the Nucleotide Excision Repair (NER) pathway. Strand breaks can occur either indirectly (6) or directly (7), largely depending on the density of the radiation.

thus prone to cleavage – usually resulting in the loss of purine bases. This depurination is perhaps the most common lesion, occurring up to 10,000 times a day [3]. Other types of endogenous damage that are common in our genome are base alkylation and interconversion of bases due to deamination [4]. Even single strand breaks (SSBs) can be caused by reactive oxygen species (produced during normal cellular metabolism) ionizing the covalent bonds of DNA. The amount of damage that actually occurs naturally as a result of cellular function has been estimated to be up to 10^5 lesions per day, which is not beyond the abilities of our inherent DNA repair pathways [5].

Natural external sources pose a significant threat to our genome stability (Figure 1B). The ultraviolet light we are exposed to throughout the day is alone responsible for an additional 10^5 DNA lesions, mostly consisting of pyrimidine dimers [5]. Our bodies are also exposed to a constant barrage of ionizing radiation (IR), such as cosmic rays and the decay of radionuclides, which can generate additional damage to bases and even strand breaks. The natural decay of trace amounts of uranium in our environment to radon is alone responsible for the majority of background ionizing radiation that we are exposed to [6].

1.1d: Strand breaks. It is particularly important that strand breaks resulting from both endogenous and exogenous sources are repaired accurately and efficiently. In eukaryotic cells double strand breaks (DSBs)

caused by ionizing radiation are the main cause of lethality due to DNA damage, and a single DSB may be sufficient to cause cell cycle arrest [7-9]. Ultimately, unrepaired strand breaks can cause chromosomal rearrangements and deleterious mutations in our DNA. This genomic instability can then result in tissue transformation and tumourigenesis.

Whether a SSB or DSB is incurred depends on the density of the radiation at the lesion site, and can come about either directly or indirectly [10, 11]. Direct breaks occur when the ionizing radiation breaks one or both strands of DNA by excitation of electrons involved in the covalent bonds within the sugar-phosphate backbone. Indirect breaks occur when free radicals generated by the reaction of radiation itself with water (in addition to metabolically produced radicals such as ROS), will ionize phospho-diester bonds causing both SSBs and DSBs. If IR breaks both strands of DNA cleanly and the functional groups required for ligation remain, the two ends can be religated with no loss of genome integrity. Unfortunately this in not often the case, as IR can cause a vast array of DSBs due to lost bases on both strands. In addition, DSBs can take on an assortment of different structural topologies where bases are oxidized, or SSBs and other lesions are present near the DSB. Collectively, these flawed DNA termini are sometimes referred to as ‘dirty ends’. As a result, repairing the genome accurately is not always as simple as ligating the ends back together, and more complicated processes must be initiated.

As we are exposed to a base amount of IR every day, damage to our genome is unavoidable. Consequently, the pathways involved in repairing strand breaks must also be accurate and efficient.

1.1e: The DNA Damage Response. In eukaryotic cells, this signal transduction pathway facilitates not only the repair of different types of DNA lesions, but also regulates cell growth and apoptosis depending on the severity of the damage that is incurred. In the event of DNA damage, the DDR activates cell cycle arrest at different checkpoints in the cell cycle, or, if the damage is extensive enough, apoptosis. To mark the damage for repair, IR induced foci are established and a signal cascade is initiated that results in different repair factors being recruited.

Different pathways control the repair of different types of lesions from pyrimidine dimers to double strand breaks. For example, small chemical alterations of DNA bases are repaired by Base Excision Repair (BER) [12] and SSBs are rectified by the Single Strand Break Repair (SSBR) pathway [13]. More complex lesions such as DSBs are repaired by one of two possible pathways – Non-Homologous End Joining (NHEJ) and Homologous Recombination (HR) [4, 13, 14]. These pathways exist together in a certain homeostasis, functioning largely in a cell cycle dependent manner. Both must be regulated very tightly in order to optimize repair, as well as to prevent unnecessary activity which may be detrimental to the structure of DNA during normal cellular functions.

While the DDR is extremely complex, the scope of this thesis is limited to the repair of DSBs, specifically through the Homologous Recombination (HR) pathway, which will be discussed in detail later.

1.1f: Ubiquitylation and the DDR. It is clear that many of the proteins involved in propagating the DDR are recruited through an interaction with a phosphorylated substrate. In particular, the BRCT domain is one of the most important phospho-peptide binding modules in the DDR, recognizing pSer/Thr target sequences. However, in recent years, data illustrating the importance of ubiquitin as a signalling molecule in the DDR are becoming much more prevalent.

The small, highly conserved signalling molecule ubiquitin plays an essential role in cell function in eukaryotic cells. The sequence of ubiquitin is conserved almost completely in animals from insects to cattle [15]. Initially thought to be involved primarily with the proteosome and protein degradation, it is becoming clear that the formation of ubiquitin chains acts as a signal for many other processes [16]. However, the idea of the modification of histones with ubiquitin to an end other than protein degradation was known soon after the identification of ubiquitin in the mid 1970's [17].

Ubiquitin typically acts by forming chains connected through the C-terminus of the donor ubiquitin and an internal Lys residue of the acceptor ubiquitin (Figure 2A). Depending on which Lys the donor ubiquitin is

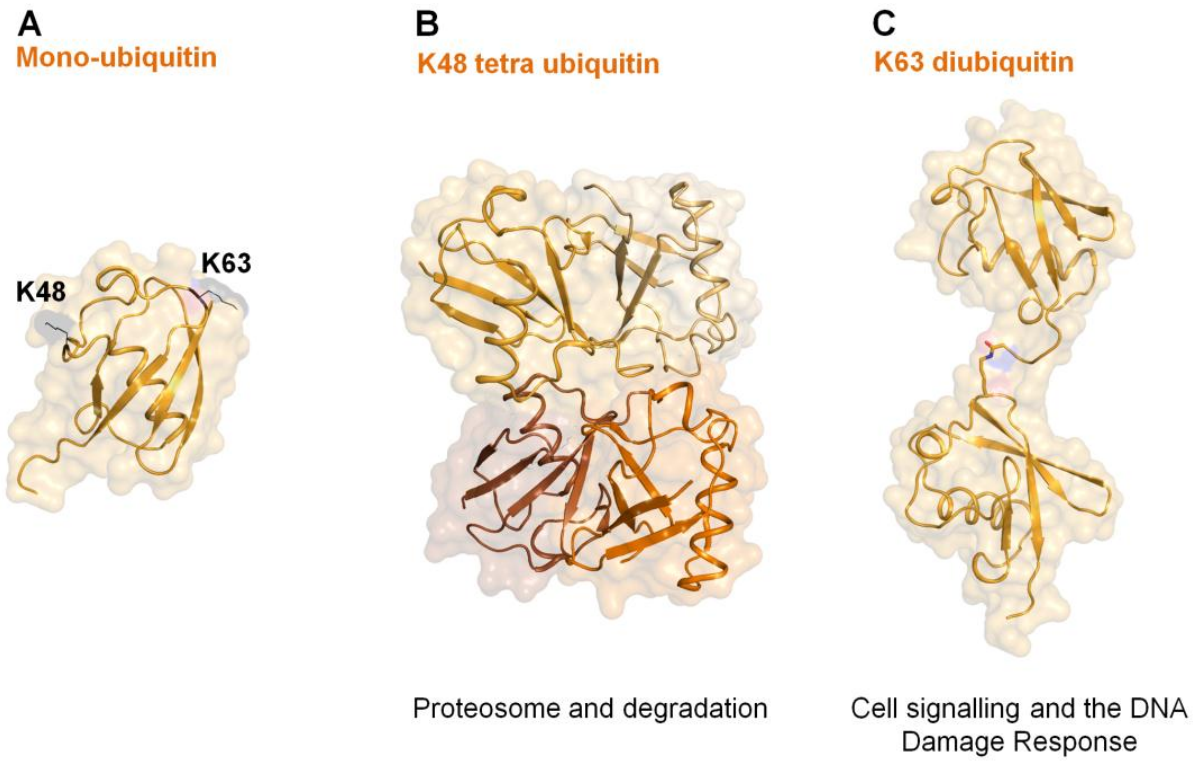


Figure 2. Topology of ubiquitin chains. Crystal structures are shown as cartoons with surface representation.

A Monoubiquitin with chain forming residues shown as black lines.

B K48 chains pack tightly, forming a compact chain

C K63 chains are elongated, exposing different regions for cofactor binding. Isopeptide bond is shown as sticks.

ligated to (often K48 or K63), the chains will have different topology, and as a result will have different functions (Figure 2) [18, 19]. When ubiquitin chains are linked through K48, the chains tend to be compact, packing together in a helical manner as each consecutive ubiquitin molecule is

added (Figure 2B). On the other hand, K63 linked chains tend to be extended, with the interactions between the consecutive ubiquitin molecules being quite limited (Figure 2C). As a result, different cofactors will recognize and bind to the different topologies, initiating diverse processes. K63 linked ubiquitin chains play a key role in the DDR, and are extremely important in the recruitment in repair factors to IR induced foci in the event of a DSB.

1.2 Homologous Recombination

1.2a: Double Strand break repair. NHEJ and HR are of particular importance, especially considering that the failure of SSBR to repair SSBs usually results in a DSB when the replication machinery reaches the lesion. The importance of these pathways is further supported by the fact that defects in these pathways often results in different types of degenerative diseases and cancer [4].

Once a DSB is detected, NHEJ works by first “cleaning” the broken DNA ends, that is, restoring the deoxy-ribose 5’ phosphate and 3’ hydroxyl groups on the terminal nucleotides at the break that are essential for ligation to occur. At this point, the broken DNA ends are stabilized and their flexibility is increased to encourage them to find each other and promote their ligation. Often domains of micro-homology are used to identify the complementary broken strand, improving accuracy [20, 21]. While efficient, this pathway has inherent drawbacks. By simply ligating

broken DNA, NHEJ does not account for the possibility of ‘dirty ends’. In addition, while ligation based on micro-homology can be accurate, it is quite possible for the wrong ends to be ligated together. As a result, NHEJ tends to be somewhat error prone, reliance on which can cause genomic instability [22-25].

HR, on the other hand, is a repair process that uses an identical chromatid as a template to repair damaged DNA. In the case of a double strand break, a homologous strand of DNA is brought within proximity of the DNA damage, and is copied to ensure that damage is repaired accurately. HR is limited to the G2 and M phases of the cell cycle, as it requires the presence of an undamaged complimentary chromatid for accurate repair of the DNA damage. NHEJ is thus more prevalent in G1 and S phase, as it is the only DSB repair pathway available when only a single chromatid is present. HR is not error prone, and is very important to our cells in order to maintain genomic stability.

Since it is dependent on the cell cycle, pathway choice between NHEJ and HR must be regulated very tightly. For example, it has been suggested that 53BP1 accumulation at sites of DSBs promotes NHEJ by increasing the flexibility of the broken DNA ends [20, 21]. Conversely, it is possible that BRCA1 accumulation helps to counteract the action of 53BP1 to promote HR. While the method of choice is unclear, it appears to depend largely on which repair factors are most prevalent at the lesion at the time [26, 27].

1.2b: Introduction to HR. HR uses a multitude of covalent modifications to aid in pathway choice and repair factor recruitment. Phosphorylation serves to initiate IR induced foci at sites of DSBs, as well as to sustain the DDR in general. Ubiquitylation also serves a very important role in the recruitment of repair factors, and is essential to propagation of the HR signal. In brief, when damage is sensed, broken DNA ends are stabilized and held in proximity to each other, while specific histones adjacent to sites of DNA damage are phosphorylated to signal that damage has occurred and to mark its location. Ubiquitylation then functions as a means to further amplify the DDR, and indirectly recruits very important repair factors, including the tumour suppressor protein, BReast CAncer 1 (BRCA1). A detailed description of the initiation of the HR pathway is discussed next.

1.2c: Initiation of the Homologous Recombination pathway. The exact mechanism with which HR repairs DSBs is still somewhat ambiguous, but the initiation of the pathway in the event of a DSB has been characterized extensively (Figure 3). Within seconds of incurred DNA damage (both DSBs and SSBs), Poly ADP-Ribose (PAR) structures are thought to mark the damage, catalyzed typically by PAR polymerase 1 (PARP1) [28]. The sensor protein complex Mre11-RAD50-NBS1 (MRN) recognizes the PAR chains and stabilizes the ends of double strand breaks through RAD50 while the other components of the complex (Mre11 and Nbs1) help to

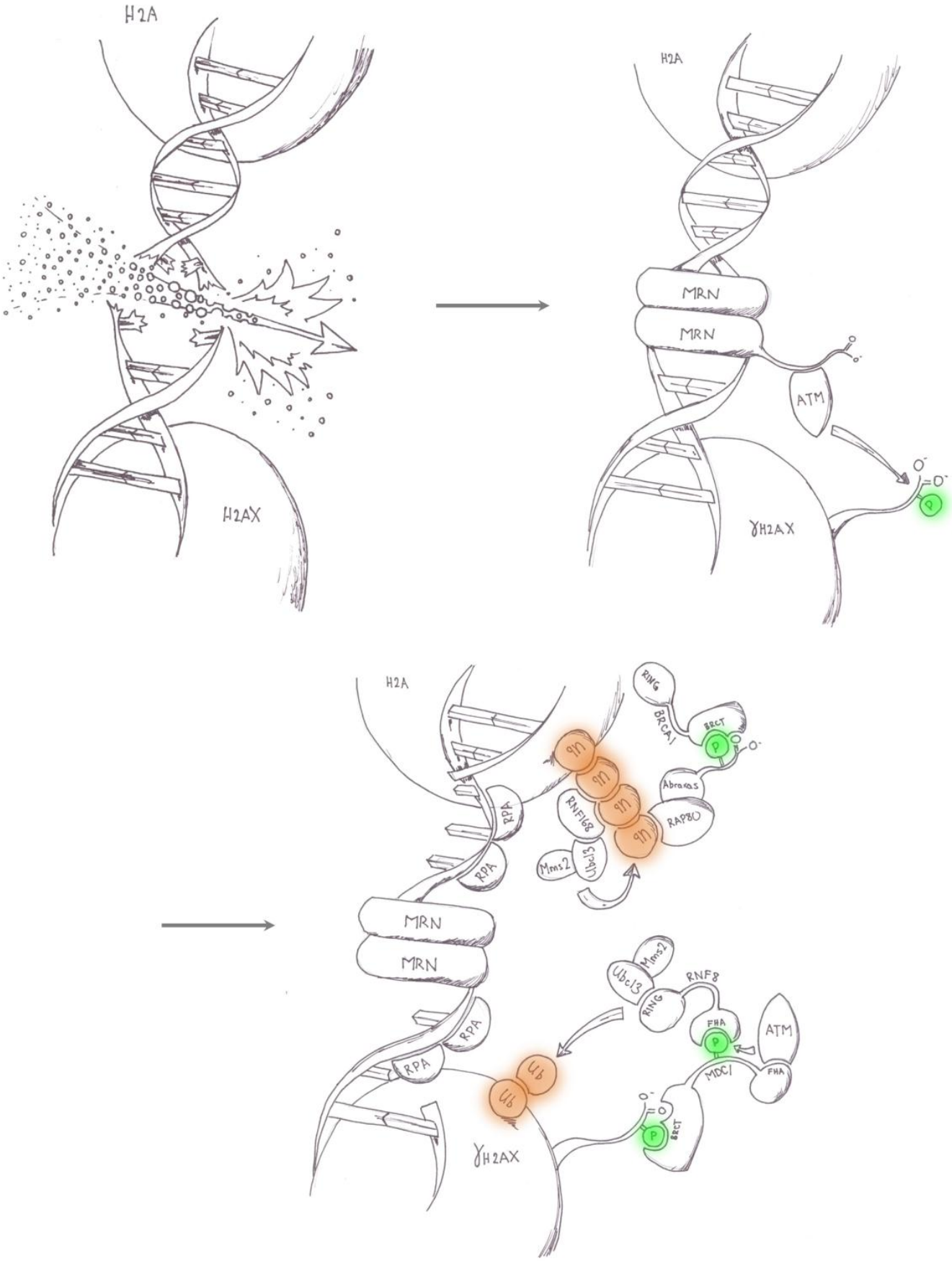

H2A
H2AX
MRN
MRN
ATM
γH2AX
RING
BRCA1
BRCT
Abraxas
RAP80
RNF168
Ubc13
Mms2
RPA
RPA
RPA
RNF8
FHA
MDC1
FHA
BRCT
Ub

Figure 3. Initiation of Homologous Recombination. If the IR dose is large enough, it has the energy to break the covalent bonds of both strands of DNA. MRN senses the break, and activates ATM. ATM then phosphorylates histone H2AX, recruiting MDC1. MDC1 acts as a scaffold to recruit the ubiquitin ligase complex RNF8/Ubc13/Mms2. In concert with RNF168, RNF8 forms K63 ubiquitin chains on adjacent histones, recruiting the BRCA1-A complex.

prepare them for HR [29, 30]. The PIKK Ser/Thr kinase Ataxia Telangiectasia Mutated (ATM) is recruited to the strand break through an interaction with the C-terminal region of NBS1, and is activated by the MRN complex [31, 32]. Pathway choice can occur at this point - if HR is activated and NHEJ is inhibited, the DNA ends must then be resected, exposing single stranded DNA, allowing access of the complementary strand. DNA end resection is regulated mainly through an interaction between ATM, CtIP and BRCA1 with the MRN complex [33-36].

In addition to playing a role in end resection, the kinase activity of ATM serves to covalently modify targets adjacent to the strand break, amplifying the signal of DNA damage. One of the most notable phosphorylation targets of ATM is Ser139 on the tail of Histone H2AX to form the covalently modified γH2AX [37]. It is at this point that a cascade is activated that recruits repair factors to the site of DSB. Mediator of DNA damage Checkpoint 1 (MDC1) binds to pSer139 through the pS_{139}QEY-COOH motif in γH2AX through its tandem BRCA1 C-terminal (BRCT)

domains [38, 39]. Interestingly, Tyr142 in H2AX is constitutively phosphorylated, and must be dephosphorylated in the event of a DSB to allow MDC1 binding, likely serving as another point of regulation of DNA repair through HR [40]. Once recruited, MDC1 will also interact with ATM, and serve to further activate ATM phosphorylation of H2AX to maintain the DDR [41].

Along with ATM-dependant phosphorylation, K63-ubiquitylation plays a key role in recruiting additional factors in HR. In order to recruit the ubiquitylation machinery, MDC1 is phosphorylated at any number of pTQ sites located near its N-terminus. The RING domain containing E3 ubiquitin ligase RNF8 is then recruited to the foci through an interaction of its N-terminal FHA domain with a pT on MDC1 [42-44]. It is thought that upon binding to MDC1, it recruits the E2 heterodimer, Ubc13/Mms2, to the breaks through an interaction with the RNF8 RING domain. Ubc13/Mms2, stimulated by RNF8, then initiates K63-ubiquitylation on the histone H2A [45, 46]. RNF8 likely exists as a dimer when bound to MDC1, with the FHA of the other protomer of RNF8 interacting with another E3 ubiquitin ligase, HERC2, which may facilitate the interaction of RNF8 with Ubc13/Mms2 [47]. Another E3, RNF168, is then recruited, likely by its Motifs that Interact with Ubiquitin (MIU), which bind to the growing ubiquitin chains and may further amplify ubiquitylation [48, 49]. Chain formation is regulated through an interaction of Ubc13 and OTUB1, a

deubiquitylating enzyme that suppresses ubiquitylation of the nucleosome by RNF168 [50].

The K63-ubiquitin chains are recognized by the RAP80/Abraxas(CCDC98) complex, which binds to the growing chains through the Ubiquitin Interacting Motifs (UIMs) of RAP80 [46, 51, 52]. BRCA1 is then recruited to the IR induced foci through an interaction of its BRCT repeat domains and the phosphorylated pSPTF motif at the C-terminal tail of Abraxas(CCDC98). BRCA1 has a RING domain, that when dimerized with BARD1, forms an E3 ubiquitin ligase [53, 54], however its role in the DDR is uncertain [55, 56]. Beyond this point, the exact mechanism by which the HR machinery is engaged to repair the DSB is also unclear. It has been proposed that the necessity of the multiple 'layers' of post-translational modifications (including phosphorylation and ubiquitylation) at the initiation of HR is in part due to the vast array of different types of DSBs, and the multiple points of tight regulation that are required to coordinate the different repair pathways to the appropriate damage [4].

While the role of BRCA1 in the downstream stimulation of HR remains unclear, we know that its recruitment is essential for the continuation of all aspects of the DDR. In the next section, the HR proteins that will be the focus of this thesis are discussed, including BRCA1.

1.2d: BRCA1. Human BRCA1 is a 1863 residue nuclear protein made up of an N-terminal RING domain and two C-terminal BRCT domains [57] (Figure 4A and Bi). BRCA1 has a multifaceted role in genome stability and the DDR, implicated in multiple cellular processes including transcriptional regulation, cell cycle control, and HR. It is currently understood that BRCA1 forms at least three complexes *in vivo*, each with separate functions in the DDR [58]. The BRCA1-A complex – the focus of this thesis – involves its interaction with Abraxas and its function in HR. The BRCA1-B complex involves an interaction with BACH1, involved in the Fanconia Anemia pathway and G2-M checkpoint regulation [59]. The BRCA1-C complex involves an interaction with CtIP, and as mentioned previously, may be involved in preparing the DNA breaks for HR by end resection.

The extent of BRCA1 involvement in many pathways related to genomic stability points to its significant role in tumorigenesis. For example, germline mutations in this gene confer a significant genetic predisposition for familial breast and ovarian cancers [60]. In addition, some sporadic cancers show dysfunction of BRCA1, suggesting that it may also play a role in the initiation of non-hereditary tumours [61-63]. It seems that the involvement of BRCA1 in HR in particular has been implicated in its connection to cancer. Cells that are deficient in BRCA1 have also been shown to be deficient in HR, and this tends to lead to a

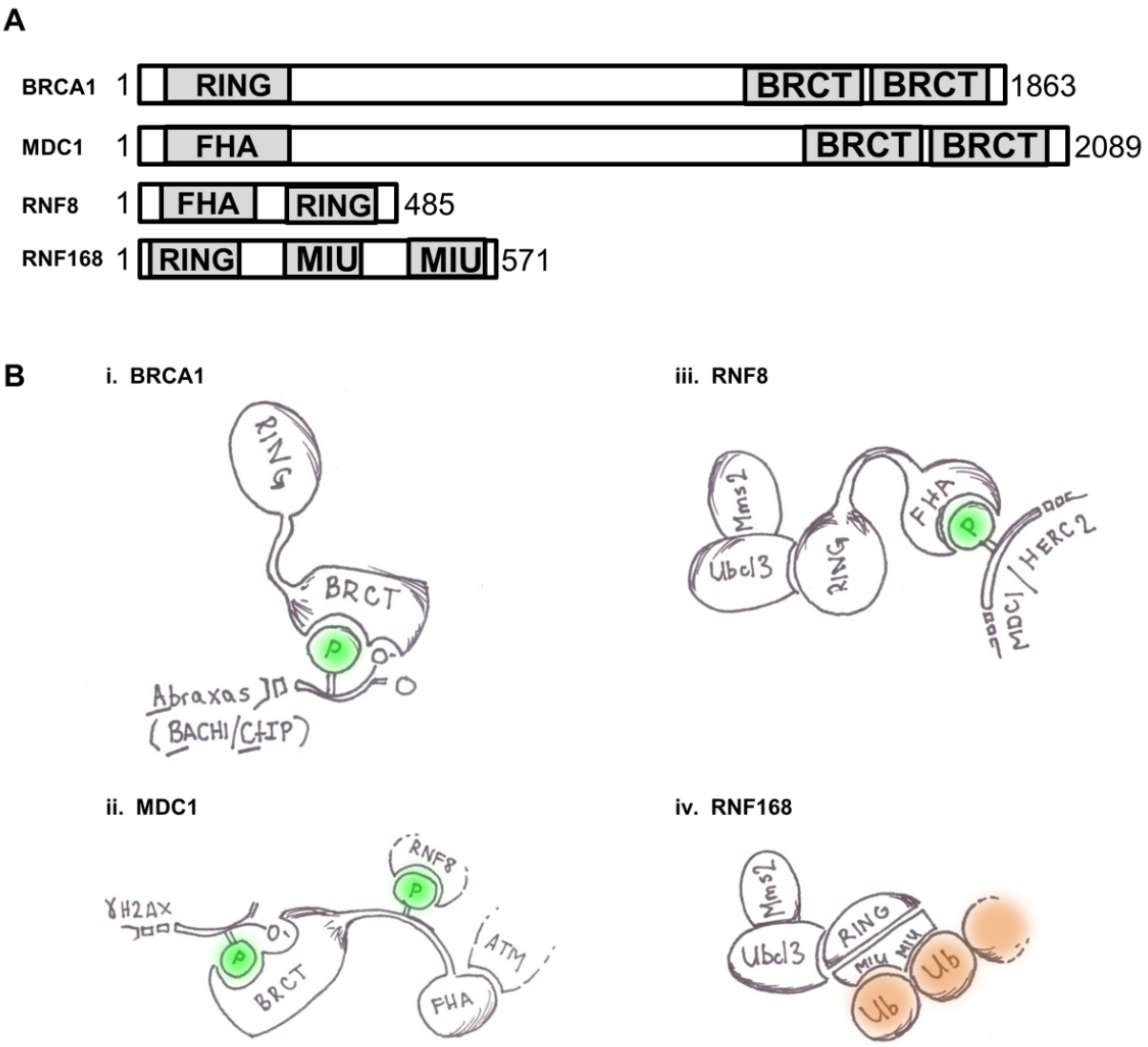


Figure 4. Domain structure of important HR proteins.

A The size and domain structure of BRCA1, MDC1, RNF8 and RNF168.

B Domain structure of important HR proteins, as well as the most notable interactions made by each domain during the DDR.

i. BRCA1 interacts with the phosphorylated tail (pSPTF) of Abraxas in the BRCA1-A complex, and is essential to HR. BRCA1 interacts with BACH1 and CtIP in the BRCA1-B and BRCA1-C complex, respectively. These interactions are also important in the DDR.

ii. MDC1 interacts with the C-terminal (pSQEY) tail of γH2AX, and recruits RNF8 to DSBs through a pThr in its N-terminal half. MDC1 also interacts with ATM through its FHA, though the mechanism is currently unknown.
iii. RNF8 likely initiates ubiquitylation at sites of DNA damage through an interaction with Ubc13/Mms2. RNF8 is recruited through an interaction of its FHA domain with a pThr on MDC1.
iv. RNF168 likely binds to the growing ubiquitin chains through its MIU domains. Through an interaction with Ubc13/Mms2, RNF168 further amplifies the K63 ubiquitylation signal.

reliance on other DNA repair pathways including NHEJ, which can lead to a loss of genome integrity.

Cancer causing mutations are concentrated in the RING and BRCT domains of BRCA1, suggesting the mutations in the BRCA1 gene cause either an inability to interact with phosphorylated targets (through the BRCT domain) or a defect in the ubiquitin ligase activity of its RING domain [60, 64-67]. In the presence of DNA damage, the BRCA1-BRCT domain interacts with phosphorylated partners such as the adaptor protein Abraxas(CCDC98) to form complexes that likely recruit required repair proteins, or are directly involved in HR [46, 51, 52]. Disruption of these phospho-protein interactions through mutations in the BRCT domains of BRCA1 [65, 68-71] may lead to a decrease in homologous recombination and an increased risk of genomic instability and cancer. The mechanism by which mutations in the RING domain cause cancer has yet to be determined.

1.2e: MDC1. Human Mediator of DNA damage checkpoint 1 (MDC1) is a 2089 amino acid protein and is involved in checkpoint-initiated cell cycle arrest, G2/M radiation-induced apoptosis and HR. While knock down of MDC1 causes hypersensitivity to IR and defects in the DDR, no disease related to hereditary MDC1 dysfunction has been identified [41, 72, 73]. MDC1 consists of an N-terminal FHA domain and a C-terminal BRCT repeat, which is structurally homologous to the BRCA1 BRCT (Figure 4A) [38, 39, 73]. Perhaps the most important function of MDC1 is the ability to bridge the nucleosomes near DNA damage to the massive protein complexes involved in HR through its FHA and BRCT domains (Figure 4Bii) [74]. In addition, the large S/TQ cluster domain that encompasses roughly the first 1200 residues are likely phosphorylated extensively upon activation of the DDR, a process that is essential for recruitment of important repair factors such as RNF8 [75].

1.2f: RNF8/RNF168. RiNg Finger 8 (RNF8) is an E3 ubiquitin ligase essential for the amplification of the HR signal at sites of DNA damage through the formation of K63-Ubiquitin chains [43, 44]. RNF8 is a ubiquitously expressed 485 residue protein, containing a N-terminal FHA domain and a C-terminal RING domain (Figure 4A and 4Biii) [76]. RNF8 has been reported to elongate K48-ubiquitin chains in complex with the E2 UbcH8 [45, 77], but the biological relevance has yet to be reported. It has

been suggested, though, that RNF8 does plays a significant role in NHEJ and telomere maintenance, in addition to its role in HR [78, 79].

RNF168 has a very different domain architecture compared to RNF8, containing an N-terminal RING domain, and two MIU domains (one near the middle of the protein, and a second near the C-terminus) (Figure 4A) [49]. It seems that RNF168 complements the role of RNF8, helping to 'concentrate' the K63 linked ubiquitin chains at sites of DNA damage through an interaction with Ubc13/Mms2 [48, 49]. The relative importance of RNF8 and RNF168 in ubiquitylation of histones H2A and H2AX is still controversial, but current opinion suggests that RNF8 is responsible for initiating K63 chain formation, and RNF168 is recruited to the growing chains in a manner that is dependent on both of its MIUs (Figure 4Biv). Defects in the RNF168 gene are responsible for RIDDLE (Radiosensitivity, Immunodeficiency, Dysmorphic features, and Learning Difficulties) syndrome, an immunodeficiency disorder associated with impaired DSB repair [80].

1.3 Important structural moieties in HR

1.3a: Phosphorylation dependent protein/protein interactions and the BRCT domain. The BRCT domain is relatively small (≈100 amino acids) and has been shown to be structurally conserved in an array of other proteins, many of them also involved in DNA repair [81, 82]. Proteins that contain a BRCT domain and have been implicated in the DDR include

BRCA1, MDC1, Topoisomerase Binding Protein 1 (TopBP1), XRCC1 and BARD1. BRCT domains tend to have low sequence homology, but the structure is well conserved, and is defined as a parallel β-sheet flanked by two α-helices on one side and one α-helix on the opposing side (Figure 5A) [83, 84]. Often, the BRCT domain occurs as a tandem repeat (Figure 4B), packing in a head to tail manner with an α-helical linker (as in BRCA1), but it may also exist as a single domain (XRCC1) [85].

The phospho-peptide binding action of BRCT domains seems to be a primary function of the tandem repeats [82, 86, 87]. Typically, they recognize a conserved pSer/Thr, with varying residues C-terminal to the phosphorylated residue. BRCA1 recognizes the pSPTF motif in all of its binding partners in the DDR, including Abraxas, BACH1 and CtIP [59, 68, 87-89]. The pSer/Thr interacts with a shallow hydrophilic pocket in the N-terminal BRCT repeat of BRCA1, while the phenylalanine (+3) to the pSer fits in to a hydrophobic pocket formed by both BRCT repeats (Figure 5C). The pSer binding pocket is lined with residues that are structurally conserved among many BRCT domains, while the phenylalanine binding groove is not as conserved, and is often involved in substrate specificity. Both of these interactions are essential for BRCA1 phospho-protein binding; mutations in the BRCT domains that disrupt these contacts are responsible for loss of function and are therefore the cause of increased cancer risk for women with corresponding mutations in the BRCA1 gene [65, 70, 71].

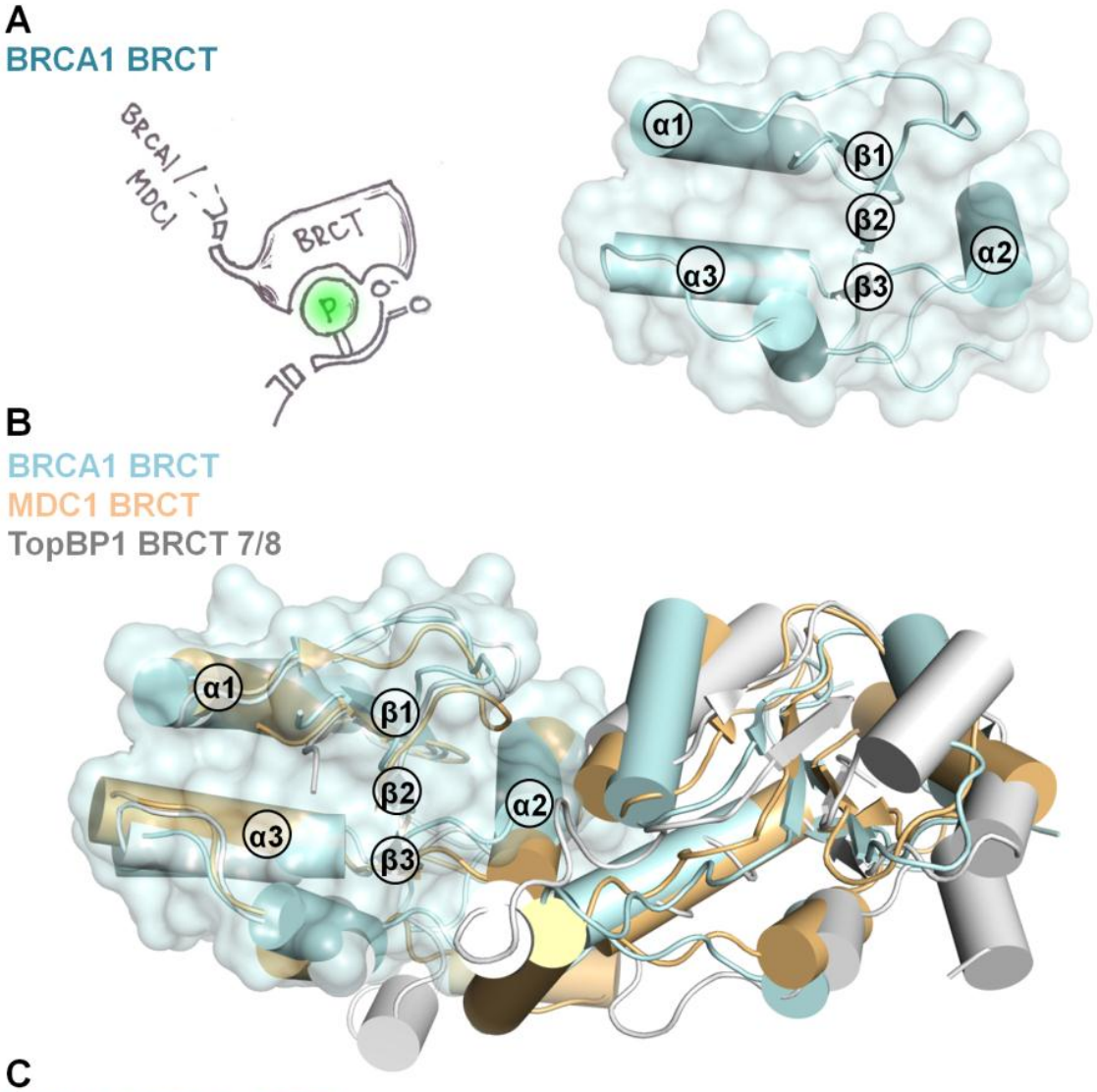

A
BRCA1 BRCT
BRCA1 / MDC1
BRCT
P
α1
β1
β2
α2
α3
β3
B
BRCA1 BRCT
MDC1 BRCT
TopBP1 BRCT 7/8
α1
β1
β2
α2
α3
β3
C
BRCA1 BRCT/pSPTF
MDC1 BRCT/pSQEY-COO-
Phe/Thr (+3)
G1656
G1899
S1655
T1898
M1775
P2009
pSer
K1702
K1936

Figure 5. Structural similarities of the BRCT domains.

A The N-terminal BRCT domain of BRCA1. α-Helixes are shown as cylinders, and parallel β-sheets are shown with cartoon representation.

B An alignment of the BRCT domains of BRCA1 and MDC1 with BRCT domains 7/8 of TopBP1 (PDB ID: 1JNX, 2ADO and 3AL3, respectively). The N-terminal BRCT domain of BRCA1 is shown with surface representation. The α-helices are shown as cylinders and the β-sheets are shown as cartoon. The Cα rmsd of the BRCA1/MDC1 structure and the BRCA1/TopBP1 structure are 2.6 Å and 3.6 Å, respectively.

C An alignment of the BRCT domains of BRCA1 and MDC1 bound to their respective phospho-peptide target sequences (PDB ID: 1T2V and 2AZM, respectively). The N-terminal BRCT repeat of BRCA1 is shown with surface representation. Residues forming the phospho-serine binding pocket in the N-terminal repeat are labelled. Conserved residues forming the base of the hydrophobic pocket in the interface between the two BRCT domains are also shown. Important side chains are shown as sticks and hydrogen bonds are shown with dashed lines

The MDC1-BRCT recognition sequence, although similar to that of BRCA1, contains marked differences. MDC1 binds to the C-terminal tetrapeptide pSQEY-COO^- in the nucleosome variant γH2AX [38, 39, 73]. The pSer/Thr binding pockets in MDC1 contain structurally conserved residues with BRCA1; subsequently the interactions in these regions are similar (pSer and Phe for BRCA1, pSer and Tyr for MDC1) (Figure 5C). However, the free carboxy terminus of the H2AX peptide three residues C-terminal to the phosphorylated residue plays a much more substantial role

in MDC1 than it does in BRCA1, a characteristic that will be discussed in Chapter 3 [38, 39, 90].

1.3b: Ubiquitylation and the RING domain. The process of ubiquitylation can be summarized in three steps (1) activation by E1 (2) transfer to an E2 and (3) ligation by an E3 (Figure 6) [91]. The E1 activates ubiquitin by forming a thioester bond between an active site cysteine on the E1 and the C-terminus of ubiquitin in an ATP dependent manner. The E1~Ub complex facilitates the transfer of the ubiquitin molecule to the active site cysteine of an E2, through an interaction between the E1 and the E2. Finally, the E2 enzyme will interact with a specific E3 enzyme which will attach the ubiquitin to a specific substrate, initiating chain formation. While there are hundreds of different types of E3 enzymes, there are significantly fewer E2 enzymes. This is likely due to the fact that the E3 is the main determinate in selecting the wide variety of substrates. Conversely, E2 enzymes are essential in determining the type of chain to be built. The E2 heterodimer, Ubc13/Mms2 forms predominantly K63-linked chains, and is the primary poly-ubiquitin chain forming E2 in the DDR [42-45, 48, 49, 92]. The RING domains of RNF8 and RNF168 are the most prominent E3 ligases in HR, and are both likely involved in selecting the histones H2A and H2AX as substrates.

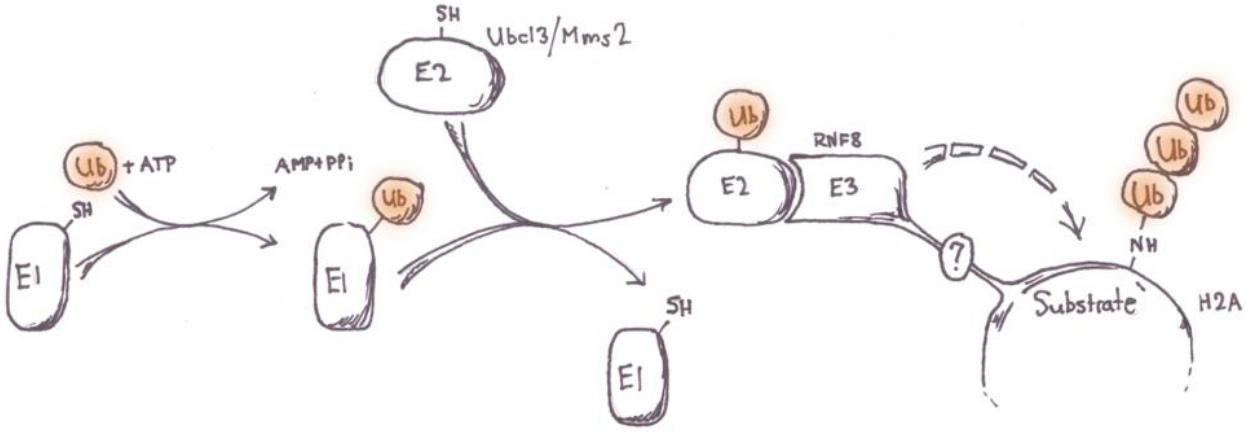


Figure 6. The enzymatic cascade leading to formation of ubiquitin chains. Ubiquitin as activated by a covalent interaction with an E1 enzyme. Upon transfer to an E2 enzyme, the E2 will interact with an E3, and form ubiquitin chains on a specific substrate.

Depending on the type of E3 ligase, ubiquitylation is catalyzed quite differently. There are three major families of E3s - HECT domain, U-box and RING domain type ligases [93]. Unlike the HECT type E3 ligases, the RING and U-box domain ligases do not form a covalent intermediate with ubiquitin [94, 95]. Exactly how the RING domain catalyzes chain formation is currently not clear. RING domains are structurally conserved; alignment of the core RING domains of Cbl, RAD18 and Traf6 with RNF4 gives an RMSD of ‹2.0 Å for all structures (Figure 7A, Table 1). They are typically defined by a conserved by a C_3HC_4 (N- to C-terminal) motif within the sequence, that form four C/C or C/H pairs [96]. Two zincs are coordinated, with two pairs each binding one zinc in a tetrahedral

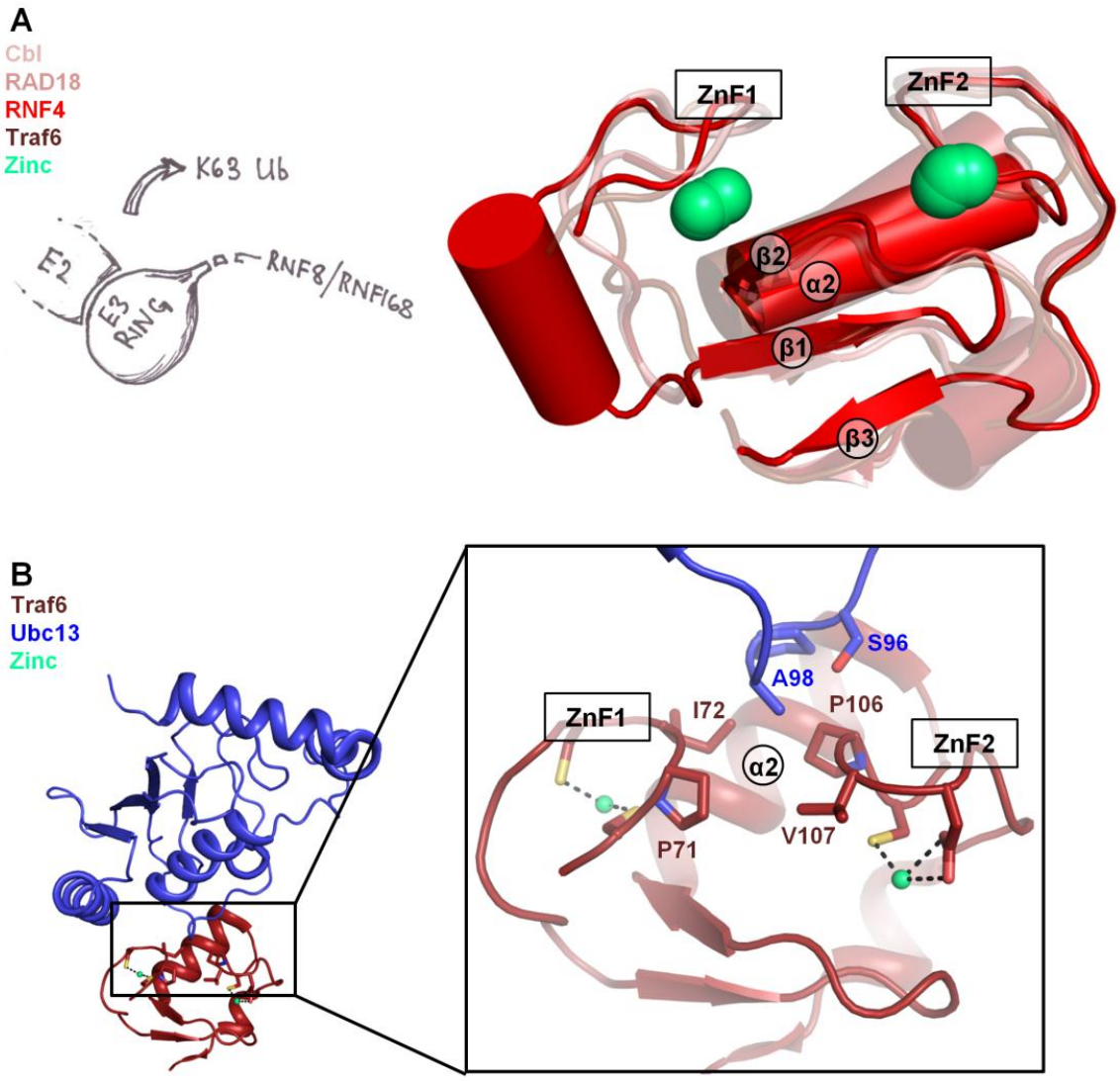

A
Cbl
RAD18
RNF4
Traf6
Zinc
K63 Ub
E2
E3
RING
RNF8/RNF168
ZnF1
ZnF2
β2
α2
β1
β3
B
Traf6
Ubc13
Zinc
S96
A98
P106
ZnF1
I72
ZnF2
α2
V107
P71
C
RAD18
BRCA1/BARD1
Zinc
α1
α3
α2
α1
α3

Figure 7. RING domain structural features.

A An alignment of the RING domains of Cbl, RAD18, RNF4 and Traf6 (PDB ID: 1FBV, 2Y43, 2XEU and 3HCT, respectively). The Cα RMSD of the alignments is ‹1.5 Å. Zincs are shown as green spheres.

B The crystal structure of the Traf6 RING/Ubc13 E2/E3 complex (PDB ID: 3HCT). Ubc13 is shown in blue and Traf6 in dark red. The binding interface is shown in detail on the right. Conserved residues are shown in sticks, and residues involved in the E2/E3 interaction are shown as sticks.

C The crystal structure of RAD18 (left) and the solution structure of BRCA1/BARD1 (right). N- and C-terminal helices are involved in dimerization of the protomers (PDB ID: 2Y43 and 1JM7, respectively) .

RING	RMSD of Cα alinged with RNF4 (Å)	E2 partner
RNF4	N/A	UbcH7
Cbl	1.7	RAD6B
RAD18	0.55	UbchH5a
Traf6	0.8	Ubc13

Table 1. RMSD of a selection of RING domains aligned to the core model RING domain of RNF4.

arrangement [97]. It is the coordination of the zinc that forms the structural integrity of the RING domain's two Zinc Fingers (ZnF), the most recognizable structural motifs in a RING domain (Figure 7A). The ZnFs are essential for the E2/E3 interaction, and if residues in this region are mutated, both binding and ubiquitylation activity is abrogated (Figure 7B) [98, 99]. In Ubc13, a conserved SPA motif interacts with conserved

hydrophobic residues at the tips of the ZnFs in the RING domain, as first observed in the crystal structure of the Traf6 RING domain bound to Ubc13 [98]. Other known E2/E3 interactions are identified in Table 1. Since RING domains are small (only about 50 amino acids) the secondary structure tends to be quite minimal. However, many RING domains do have an internal α-helix (α2), as well as an internal 3-strand anti-parallel β-sheet [97, 98, 100-103]. Many RING domains also have long N- and C-terminal α-helices involved in dimerization (α1 and α3) (Figure 7C).

1.4 Insights into BRCA1 recruitment: Thesis overview

The objective of this thesis is to look in detail at some of the factors that determine the recruitment of BRCA1 to double strand breaks. While there are many important repair factors in the DDR, BRCA1 stands out as being particularly important. The exact molecular mechanism by which BRCA1 executes its function is unclear, but its importance is highlighted by the devastating consequences incurred by dysfunctional BRCA1 recruitment to IR induced foci. This thesis will therefore address three different aspects of BRCA1 recruitment including the structural and biochemical explanations for the roles RNF8/RNF168, BRCT specificity of BRCA1 and MDC1, as well as a preliminary look into the synthetic inhibition of BRCA1 recruitment via its BRCT domains.

RNF8/RNF168 (Chapter 2). While the ubiquitylation involvement in HR and the recruitment of BRCA1 to DNA lesions is understood, the exact mechanism by which RNF8 and RNF168 function relative to each other is currently unclear. For example, it seems that RNF8 initiates ubiquitylation and RNF168 amplifies these chains, but the molecular basis of this is currently not known. In addition, the three dimensional structures of their RING domains are unknown, as well as the details of their interactions with Ubc13/Mms2. We describe in detail the first crystal structures of the RNF8 RING/Ubc13/Mms2 ternary complex and the RNF168 RING domain, in addition to providing further insights in to their interactions with Ubc13/Mms2. These data help to explain the order in which RNF8 and RNF168 function to initiate and amplify K63-ubiquitin chains, as well as providing insight into the mechanism of RING domain ubiquitylation in general.

BRCT specificity (Chapter 3). A point of interest in HR is the mode by which different BRCT domain containing proteins recognize and bind to their specific phosphorylated target sequences given the structural similarity that exists between BRCT domains. The BRCT domains of BRCA1 and MDC1 are very similar, and even recognize similar consensus sequences (pSPTF-COOH/NH_2 and pSQEY-COOH, respectively). We look into the structural mechanism that helps to dictate the specificity that

these two proteins have for their respective ligands, ensuring they properly carry out their specific functions.

BRCA1 BRCT inhibition (Chapter 4). Finally, we study the idea of BRCA1 as a therapeutic target by conducting a screen for small molecule inhibitors of the BRCA1/phospho-peptide interactions. Given its importance in genome stability, we hypothesize that inhibition of BRCA1 in tumor cells would be a viable therapeutic option. We present initial high through-put inhibitor screens for inhibitors of the BRCA1-BRCT, as well as preliminary peptidomimetics to design high affinity peptide inhibitors.

General Discussion (Chapter 5). The final chapter will be reserved primarily to discuss the future directions of many of the results described in this thesis, and to tie the findings together in a concise way.

1.5 References

1. Friedberg, E.C., *A brief history of the DNA repair field.* Cell Res, 2008. **18**(1): p. 3-7.
2. Friedberg, E.C., *Correcting the Blueprint of Life*. 1997: Cold Spring Harbor Press. 210.
3. Lindahl, T., *DNA repair enzymes.* Annu Rev Biochem, 1982. **51**: p. 61-87.
4. Ciccia, A. and S.J. Elledge, *The DNA damage response: making it safe to play with knives.* Mol Cell, 2010. **40**(2): p. 179-204.
5. Hoeijmakers, J.H., *DNA damage, aging, and cancer.* N Engl J Med, 2009. **361**(15): p. 1475-85.

6. NRC, *Committee on Health Effects of Exposure to Low Levels of Ionizing Radiations: Time for Reassessment?*, N.R. Council, Editor 1998, The National Academies Press: Washington, D.C.
7. Foray, N., C.F. Arlett, and E.P. Malaise, *Radiation-induced DNA double-strand breaks and the radiosensitivity of human cells: a closer look.* Biochimie, 1997. **79**(9-10): p. 567-75.
8. Frankenberg-Schwager, M., *Induction, repair and biological relevance of radiation-induced DNA lesions in eukaryotic cells.* Radiat Environ Biophys, 1990. **29**(4): p. 273-92.
9. Huang, L.C., K.C. Clarkin, and G.M. Wahl, *Sensitivity and selectivity of the DNA damage sensor responsible for activating p53-dependent G1 arrest.* Proc Natl Acad Sci U S A, 1996. **93**(10): p. 4827-32.
10. Goodhead, D.T., *Initial events in the cellular effects of ionizing radiations: clustered damage in DNA.* Int J Radiat Biol, 1994. **65**(1): p. 7-17.
11. Friedberg, E.C., Walker, G. C., Wolfram, S., Wood, R. D., Schultz, R. A., Ellenberger, T., *DNA Repair and Mutagenesis*. Second Edition ed. 2006: ASM Press.
12. Lindahl, T. and D.E. Barnes, *Repair of endogenous DNA damage.* Cold Spring Harb Symp Quant Biol, 2000. **65**: p. 127-33.
13. Caldecott, K.W., *Single-strand break repair and genetic disease.* Nat Rev Genet, 2008. **9**(8): p. 619-31.
14. West, S.C., *Molecular views of recombination proteins and their control.* Nat Rev Mol Cell Biol, 2003. **4**(6): p. 435-45.
15. Gavilanes, J.G., et al., *Isolation, characterization, and amino acid sequence of a ubiquitin-like protein from insect eggs.* J Biol Chem, 1982. **257**(17): p. 10267-70.
16. Haglund, K. and I. Dikic, *Ubiquitylation and cell signaling.* EMBO J, 2005. **24**(19): p. 3353-9.

17. Hershko, A., *Ubiquitin: roles in protein modification and breakdown.* Cell, 1983. **34**(1): p. 11-2.
18. Eddins, M.J., et al., *Crystal structure and solution NMR studies of Lys48-linked tetraubiquitin at neutral pH.* J Mol Biol, 2007. **367**(1): p. 204-11.
19. Komander, D., et al., *Molecular discrimination of structurally equivalent Lys 63-linked and linear polyubiquitin chains.* EMBO Rep, 2009. **10**(5): p. 466-73.
20. Difilippantonio, S., et al., *53BP1 facilitates long-range DNA end-joining during V(D)J recombination.* Nature, 2008. **456**(7221): p. 529-33.
21. Dimitrova, N., et al., *53BP1 promotes non-homologous end joining of telomeres by increasing chromatin mobility.* Nature, 2008. **456**(7221): p. 524-8.
22. Roth, D.B., T.N. Porter, and J.H. Wilson, *Mechanisms of nonhomologous recombination in mammalian cells.* Mol Cell Biol, 1985. **5**(10): p. 2599-607.
23. Roth, D.B. and J.H. Wilson, *Nonhomologous recombination in mammalian cells: role for short sequence homologies in the joining reaction.* Mol Cell Biol, 1986. **6**(12): p. 4295-304.
24. Adamo, A., et al., *Preventing nonhomologous end joining suppresses DNA repair defects of Fanconi anemia.* Mol Cell, 2010. **39**(1): p. 25-35.
25. Moynahan, M.E., et al., *Brca1 controls homology-directed DNA repair.* Mol Cell, 1999. **4**(4): p. 511-8.
26. Bouwman, P., et al., *53BP1 loss rescues BRCA1 deficiency and is associated with triple-negative and BRCA-mutated breast cancers.* Nat Struct Mol Biol, 2010. **17**(6): p. 688-95.
27. Bunting, S.F., et al., *53BP1 inhibits homologous recombination in Brca1-deficient cells by blocking resection of DNA breaks.* Cell, 2010. **141**(2): p. 243-54.

28. Haince, J.F., et al., *PARP1-dependent kinetics of recruitment of MRE11 and NBS1 proteins to multiple DNA damage sites.* J Biol Chem, 2008. **283**(2): p. 1197-208.
29. Williams, R.S., J.S. Williams, and J.A. Tainer, *Mre11-Rad50-Nbs1 is a keystone complex connecting DNA repair machinery, double-strand break signaling, and the chromatin template.* Biochem Cell Biol, 2007. **85**(4): p. 509-20.
30. Williams, R.S., et al., *Nbs1 flexibly tethers Ctp1 and Mre11-Rad50 to coordinate DNA double-strand break processing and repair.* Cell, 2009. **139**(1): p. 87-99.
31. Jazayeri, A., et al., *Mre11-Rad50-Nbs1-dependent processing of DNA breaks generates oligonucleotides that stimulate ATM activity.* EMBO J, 2008. **27**(14): p. 1953-62.
32. Lee, J.H. and T.T. Paull, *ATM activation by DNA double-strand breaks through the Mre11-Rad50-Nbs1 complex.* Science, 2005. **308**(5721): p. 551-4.
33. Sartori, A.A., et al., *Human CtIP promotes DNA end resection.* Nature, 2007. **450**(7169): p. 509-14.
34. Yu, X. and J. Chen, *DNA damage-induced cell cycle checkpoint control requires CtIP, a phosphorylation-dependent binding partner of BRCA1 C-terminal domains.* Mol Cell Biol, 2004. **24**(21): p. 9478-86.
35. Chen, L., et al., *Cell cycle-dependent complex formation of BRCA1.CtIP.MRN is important for DNA double-strand break repair.* J Biol Chem, 2008. **283**(12): p. 7713-20.
36. Schlegel, B.P., F.M. Jodelka, and R. Nunez, *BRCA1 promotes induction of ssDNA by ionizing radiation.* Cancer Res, 2006. **66**(10): p. 5181-9.
37. Rogakou, E.P., et al., *DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139.* J Biol Chem, 1998. **273**(10): p. 5858-68.

38. Lee, M.S., et al., *Structure of the BRCT repeat domain of MDC1 and its specificity for the free COOH-terminal end of the gamma-H2AX histone tail.* J Biol Chem, 2005. **280**(37): p. 32053-6.
39. Stucki, M., et al., *MDC1 directly binds phosphorylated histone H2AX to regulate cellular responses to DNA double-strand breaks.* Cell, 2005. **123**(7): p. 1213-26.
40. Cook, P.J., et al., *Tyrosine dephosphorylation of H2AX modulates apoptosis and survival decisions.* Nature, 2009. **458**(7238): p. 591-6.
41. Lou, Z., et al., *MDC1 maintains genomic stability by participating in the amplification of ATM-dependent DNA damage signals.* Mol Cell, 2006. **21**(2): p. 187-200.
42. Huen, M.S., et al., *RNF8 transduces the DNA-damage signal via histone ubiquitylation and checkpoint protein assembly.* Cell, 2007. **131**(5): p. 901-14.
43. Kolas, N.K., et al., *Orchestration of the DNA-damage response by the RNF8 ubiquitin ligase.* Science, 2007. **318**(5856): p. 1637-40.
44. Mailand, N., et al., *RNF8 ubiquitylates histones at DNA double-strand breaks and promotes assembly of repair proteins.* Cell, 2007. **131**(5): p. 887-900.
45. Plans, V., et al., *The RING finger protein RNF8 recruits UBC13 for lysine 63-based self polyubiquitylation.* J Cell Biochem, 2006. **97**(3): p. 572-82.
46. Wang, B., et al., *Abraxas and RAP80 form a BRCA1 protein complex required for the DNA damage response.* Science, 2007. **316**(5828): p. 1194-8.
47. Bekker-Jensen, S., et al., *HERC2 coordinates ubiquitin-dependent assembly of DNA repair factors on damaged chromosomes.* Nat Cell Biol, 2010. **12**(1): p. 80-6; sup pp 1-12.
48. Doil, C., et al., *RNF168 binds and amplifies ubiquitin conjugates on damaged chromosomes to allow accumulation of repair proteins.* Cell, 2009. **136**(3): p. 435-46.

49. Stewart, G.S., et al., *The RIDDLE syndrome protein mediates a ubiquitin-dependent signaling cascade at sites of DNA damage.* Cell, 2009. **136**(3): p. 420-34.
50. Nakada, S., et al., *Non-canonical inhibition of DNA damage-dependent ubiquitination by OTUB1.* Nature, 2010. **466**(7309): p. 941-6.
51. Kim, H., J. Huang, and J. Chen, *CCDC98 is a BRCA1-BRCT domain-binding protein involved in the DNA damage response.* Nat Struct Mol Biol, 2007. **14**(8): p. 710-5.
52. Liu, Z., J. Wu, and X. Yu, *CCDC98 targets BRCA1 to DNA damage sites.* Nat Struct Mol Biol, 2007. **14**(8): p. 716-20.
53. Hashizume, R., et al., *The RING heterodimer BRCA1-BARD1 is a ubiquitin ligase inactivated by a breast cancer-derived mutation.* J Biol Chem, 2001. **276**(18): p. 14537-40.
54. Meza, J.E., et al., *Mapping the functional domains of BRCA1. Interaction of the ring finger domains of BRCA1 and BARD1.* J Biol Chem, 1999. **274**(9): p. 5659-65.
55. Shakya, R., et al., *BRCA1 tumor suppression depends on BRCT phosphoprotein binding, but not its E3 ligase activity.* Science, 2011. **334**(6055): p. 525-8.
56. Reid, L.J., et al., *E3 ligase activity of BRCA1 is not essential for mammalian cell viability or homology-directed repair of double-strand DNA breaks.* Proc Natl Acad Sci U S A, 2008. **105**(52): p. 20876-81.
57. Miki, Y., et al., *A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1.* Science, 1994. **266**(5182): p. 66-71.
58. Huen, M.S., S.M. Sy, and J. Chen, *BRCA1 and its toolbox for the maintenance of genome integrity.* Nat Rev Mol Cell Biol, 2010. **11**(2): p. 138-48.

59. Clapperton, J.A., et al., *Structure and mechanism of BRCA1 BRCT domain recognition of phosphorylated BACH1 with implications for cancer.* Nat Struct Mol Biol, 2004. **11**(6): p. 512-8.
60. Futreal, P.A., et al., *BRCA1 mutations in primary breast and ovarian carcinomas.* Science, 1994. **266**(5182): p. 120-2.
61. Turner, N.C., et al., *BRCA1 dysfunction in sporadic basal-like breast cancer.* Oncogene, 2007. **26**(14): p. 2126-32.
62. Rakha, E.A., et al., *Expression of BRCA1 protein in breast cancer and its prognostic significance.* Hum Pathol, 2008. **39**(6): p. 857-65.
63. Turner, J.M., et al., *BRCA1, histone H2AX phosphorylation, and male meiotic sex chromosome inactivation.* Curr Biol, 2004. **14**(23): p. 2135-42.
64. Brzovic, P.S., et al., *BRCA1 RING domain cancer-predisposing mutations. Structural consequences and effects on protein-protein interactions.* J Biol Chem, 2001. **276**(44): p. 41399-406.
65. Lee, M.S., et al., *Comprehensive analysis of missense variations in the BRCT domain of BRCA1 by structural and functional assays.* Cancer Res, 2010. **70**(12): p. 4880-90.
66. Gayther, S.A., et al., *Germline mutations of the BRCA1 gene in breast and ovarian cancer families provide evidence for a genotype-phenotype correlation.* Nat Genet, 1995. **11**(4): p. 428-33.
67. Friedman, L.S., et al., *Confirmation of BRCA1 by analysis of germline mutations linked to breast and ovarian cancer in ten families.* Nat Genet, 1994. **8**(4): p. 399-404.
68. Williams, R.S., et al., *Structural basis of phosphopeptide recognition by the BRCT domain of BRCA1.* Nat Struct Mol Biol, 2004. **11**(6): p. 519-25.
69. Williams, R.S., et al., *Detection of protein folding defects caused by BRCA1-BRCT truncation and missense mutations.* J Biol Chem, 2003. **278**(52): p. 53007-16.

70. Williams, R.S. and J.N. Glover, *Structural consequences of a cancer-causing BRCA1-BRCT missense mutation.* J Biol Chem, 2003. **278**(4): p. 2630-5.
71. Coquelle, N., R. Green, and J.N. Glover, *Impact of BRCA1 BRCT domain missense substitutions on phosphopeptide recognition.* Biochemistry, 2011. **50**(21): p. 4579-89.
72. Lou, Z., et al., *MDC1 is coupled to activated CHK2 in mammalian DNA damage response pathways.* Nature, 2003. **421**(6926): p. 957-61.
73. Stewart, G.S., et al., *MDC1 is a mediator of the mammalian DNA damage checkpoint.* Nature, 2003. **421**(6926): p. 961-6.
74. Coster, G. and M. Goldberg, *The cellular response to DNA damage: a focus on MDC1 and its interacting proteins.* Nucleus, 2010. **1**(2): p. 166-78.
75. Matsuoka, S., et al., *ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage.* Science, 2007. **316**(5828): p. 1160-6.
76. Seki, N., et al., *Isolation, tissue expression, and chromosomal assignment of a novel human gene which encodes a protein with RING finger motif.* J Hum Genet, 1998. **43**(4): p. 272-4.
77. Lok, G.T., et al., *Differential regulation of RNF8-mediated Lys48- and Lys63-based poly-ubiquitylation.* Nucleic Acids Res, 2012. **40**(1): p. 196-205.
78. Rai, R., et al., *The E3 ubiquitin ligase Rnf8 stabilizes Tpp1 to promote telomere end protection.* Nat Struct Mol Biol, 2011. **18**(12): p. 1400-7.
79. Feng, L. and J. Chen, *The E3 ligase RNF8 regulates KU80 removal and NHEJ repair.* Nat Struct Mol Biol, 2012. **19**(2): p. 201-6.
80. Stewart, G.S., et al., *RIDDLE immunodeficiency syndrome is linked to defects in 53BP1-mediated DNA damage signaling.* Proc Natl Acad Sci U S A, 2007. **104**(43): p. 16910-5.

81. Glover, J.N., R.S. Williams, and M.S. Lee, *Interactions between BRCT repeats and phosphoproteins: tangled up in two.* Trends Biochem Sci, 2004. **29**(11): p. 579-85.
82. Leung, C.C. and J.N. Glover, *BRCT domains: easy as one, two, three.* Cell Cycle, 2011. **10**(15): p. 2461-70.
83. Williams, R.S., R. Green, and J.N. Glover, *Crystal structure of the BRCT repeat region from the breast cancer-associated protein BRCA1.* Nat Struct Biol, 2001. **8**(10): p. 838-42.
84. Zhang, X., et al., *Structure of an XRCC1 BRCT domain: a new protein-protein interaction module.* EMBO J, 1998. **17**(21): p. 6404-11.
85. Dulic, A., et al., *BRCT domain interactions in the heterodimeric DNA repair protein XRCC1-DNA ligase III.* Biochemistry, 2001. **40**(20): p. 5906-13.
86. Yu, X., et al., *The BRCT domain is a phospho-protein binding domain.* Science, 2003. **302**(5645): p. 639-42.
87. Manke, I.A., et al., *BRCT repeats as phosphopeptide-binding modules involved in protein targeting.* Science, 2003. **302**(5645): p. 636-9.
88. Shiozaki, E.N., et al., *Structure of the BRCT repeats of BRCA1 bound to a BACH1 phosphopeptide: implications for signaling.* Mol Cell, 2004. **14**(3): p. 405-12.
89. Varma, A.K., et al., *Structural basis for cell cycle checkpoint control by the BRCA1-CtIP complex.* Biochemistry, 2005. **44**(33): p. 10941-6.
90. Campbell, S.J., R.A. Edwards, and J.N. Glover, *Comparison of the structures and peptide binding specificities of the BRCT domains of MDC1 and BRCA1.* Structure, 2010. **18**(2): p. 167-76.
91. Hershko, A., et al., *Components of ubiquitin-protein ligase system. Resolution, affinity purification, and role in protein breakdown.* J Biol Chem, 1983. **258**(13): p. 8206-14.

92. Wang, B. and S.J. Elledge, *Ubc13/Rnf8 ubiquitin ligases control foci formation of the Rap80/Abraxas/Brca1/Brcc36 complex in response to DNA damage.* Proc Natl Acad Sci U S A, 2007. **104**(52): p. 20759-63.
93. Hatakeyama, S. and K.I. Nakayama, *U-box proteins as a new family of ubiquitin ligases.* Biochem Biophys Res Commun, 2003. **302**(4): p. 635-45.
94. Scheffner, M., et al., *The HPV-16 E6 and E6-AP complex functions as a ubiquitin-protein ligase in the ubiquitination of p53.* Cell, 1993. **75**(3): p. 495-505.
95. Scheffner, M., U. Nuber, and J.M. Huibregtse, *Protein ubiquitination involving an E1-E2-E3 enzyme ubiquitin thioester cascade.* Nature, 1995. **373**(6509): p. 81-3.
96. Lovering, R., et al., *Identification and preliminary characterization of a protein motif related to the zinc finger.* Proc Natl Acad Sci U S A, 1993. **90**(6): p. 2112-6.
97. Barlow, P.N., et al., *Structure of the C3HC4 domain by 1H-nuclear magnetic resonance spectroscopy. A new structural class of zinc-finger.* J Mol Biol, 1994. **237**(2): p. 201-11.
98. Yin, Q., et al., *E2 interaction and dimerization in the crystal structure of TRAF6.* Nat Struct Mol Biol, 2009. **16**(6): p. 658-66.
99. Zheng, N., et al., *Structure of a c-Cbl-UbcH7 complex: RING domain function in ubiquitin-protein ligases.* Cell, 2000. **102**(4): p. 533-9.
100. Brzovic, P.S., et al., *Structure of a BRCA1-BARD1 heterodimeric RING-RING complex.* Nat Struct Biol, 2001. **8**(10): p. 833-7.
101. Bellon, S.F., et al., *Crystal structure of the RAG1 dimerization domain reveals multiple zinc-binding motifs including a novel zinc binuclear cluster.* Nat Struct Biol, 1997. **4**(7): p. 586-91.
102. Huang, A., et al., *Symmetry and asymmetry of the RING-RING dimer of Rad18.* J Mol Biol, 2011. **410**(3): p. 424-35.

103. Plechanovova, A., et al., *Mechanism of ubiquitylation by dimeric RING ligase RNF4.* Nat Struct Mol Biol, 2011. **18**(9): p. 1052-9.

Chapter 2

[1]Structural insights into the function of RNF8 and RNF168 in Ubc13/Mms2-dependent ubiquitylation

[1]A version of this chapter is in submission as Campbell, S.J., [¥]Edwards, R.A., [¥]Leung, C.C.Y., [¥]Neculai, D., Hodge, C.D., Dhe-Paganon, S., and Glover, J.N.M., **Structural insights into the function of RNF8 and RNF168 in Ubc13/Mms2-dependent ubiquitylation.** *J. Biol. Chem.*, 2012. [¥]Equal Contribution

2.1 Introduction

Amongst the various forms of DNA damage that must be identified and repaired to maintain genomic stability, double strand breaks (DSBs) are particularly deleterious, as even a single DSB can cause cell cycle arrest [1, 2]. Homologous recombination (HR) is an essential pathway for the repair of these inevitable DNA lesions. HR relies on a diverse set of post-translational modifications to recruit downstream repair factors, including the breast cancer-associated protein BRCA1, to chromatin surrounding the DSB in order to promote its repair.

HR is initiated by the recognition of the DSB by factors, such as the MRN complex, that lead to the activation of Ataxia Telangiectasia Mutate (ATM) kinase, which phosphorylates multiple targets within the region of the damage [3]. Phosphorylation of the C-terminal tail of the histone variant H2AX by ATM facilitates the association of Mediator of DNA Damage checkpoint 1 (MDC1) at these sites [4]. MDC1 is then phosphorylated on multiple residues, which results in the binding of the E3 ubiquitin ligase, RNF8, via a selective interaction between the RNF8 Fork Head Associated (FHA) domain and one of three **pThr**-Gln-X-Phe motifs in MDC1 [5, 6]. RNF8 also contains a RING domain that functions as an E3 ubiquitin ligase, and interacts with the E2 ubiquitin conjugating enzyme heterodimer, Ubc13/Mms2, to ubiquitylate histones near the lesion [6-8]. It is thought that RNF168 is then recruited to the growing ubiquitin chains via its Motifs that Interact with Ubiquitin (MIU), and, through an interaction

between its RING domain and Ubc13/Mms2, amplifies the ubiquitin chains initiated by RNF8 [9, 10]. Receptor Associated Protein 80 (RAP80) then specifically binds to the ubiquitin chains through its tandem UIM (ubiquitin interaction motif), and through an interaction with the adapter protein Abraxas, BRCA1 is recruited [11-13]. Finally, the ubiquitin signal is attenuated by the Ubc13 inhibitory factor, OTU domain Ubiquitin Binding 1 (OTUB1) [14].

Ubiquitylation catalyzed by Ubc13/Mms2 generates a unique form of polyubiquitin characterized by linkages between Lys63 of one ubiquitin and the C-terminal carboxylate of the next ubiquitin in the chain, and unlike K48-linked chains, K63-linked polyubiquitylation does not target the substrate for proteolysis. Ubc13/Mms2 consists of a catalytic E2, Ubc13, which contains an active site cysteine that forms a thioester linkage with the C-terminus of ubiquitin, thereby activating it for poly-ubiquitin chain formation. Mms2 is an E2-like protein that binds Ubc13 and facilitates the binding of the acceptor ubiquitin in an orientation to promote isopeptide bond formation between Lys63 of the acceptor ubiquitin and the C-terminal thioester of the donor ubiquitin [15-19]. The participation of Ubc13/Mms2 in distinct pathways is dependent on its interactions with different E3 partners. For example, interactions with the dimeric RING E3, TRAF6, recruits Ubc13 to the NF-KB signalling pathway, while interactions with the U-box E3, CHIP, enables Ubc13 to ubiquitylate chaperone-bound proteins [20, 21]. Ubc13, as well as related E2s such as UbcH5, are

selectively inhibited by OTUB1, a deubiquitylating enzyme (DUB), which binds the Ubc13~Ub conjugate, likely directly inhibiting attack of the acceptor ubiquitin and blocking binding of RING E3 proteins [22, 23].

While both RNF8 and RNF168 are important Ubc13 partners in the HR DNA repair pathway, the basis for their distinct functions is unclear. To address the roles of the RING domains of these E3 ligases, we probed their structures and abilities to catalyze K63-linked polyubiquitylation. RNF8 adopts a dimeric structure, stabilized by an extended coiled-coil that is novel in the RING protein family, and interacts with Ubc13 through a surface that is conserved in the TRAF6 RING domain as well as the CHIP U-box. The interaction between RNF8 and Ubc13/Mms2 markedly catalyzes polyubiquitylation. RNF168, in contrast, is a monomeric RING domain, and, while its RING domain adopts a similar structure to RNF8, it does not show binding Ubc13 or catalyze polyubiquitylation to the same extent as RNF8. Our data show that the RING domain of RNF168 alone is not sufficient for its activity and that it is likely dependent on its additional structural moieties. The ability of RNF8 to build *de novo* K63 ubiquitin chains points to its role in the initiation of ubiquitylation in HR.

2.2 Results

2.2a: RNF8$_{345-485}$ interacts with and activates Ubc13/Mms2 in vitro. We set out to uncover a minimal, RING-containing construct of RNF8 for detailed structural and functional studies. Secondary structure predictions

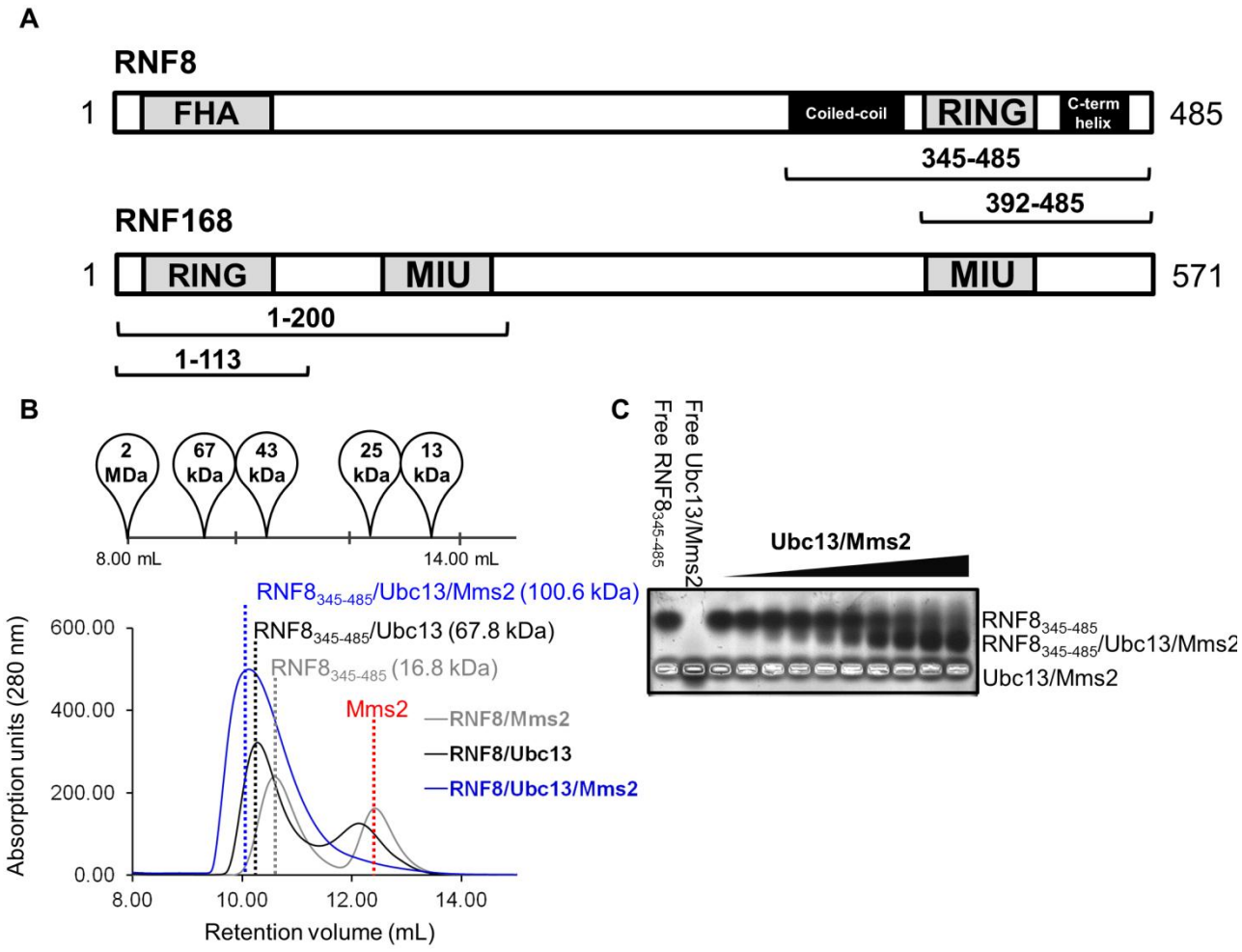


Figure 1. Domain structure and E2/E3 complex formation.

A Domain map of RNF8 and RNF168.

B Size Exclusion Chromatography (SEC) showing the different retention volumes of the RNF8 complexes. Dotted lines indicate peak position, and the peak composition is labelled above.

C Native agarose Electromobility Shift Assay (EMSA) of Ubc13/Mms2 titrated into constant $RNF8_{345-485}$ from 2.3 µM to a maximum of 64 µM. Lanes one and two show free $RNF8_{345-485}$ and Ubc13/Mms2, respectively.

suggested that large N and C-terminal helices flank the RING domain (Figure 1A). A construct encompassing both these predicted helices and the RING domain ($RNF8_{345-485}$) was expressed, purified and used in an *in vitro* ubiquitylation assay to test its ability to build K63-linked ubiquitin chains in combination with the E2 heterodimer, Ubc13/Mms2. $RNF8_{345-485}$ catalyzes the formation of ubiquitin chains efficiently *in vitro*, while Ubc13/Mms2 alone catalyzes the formation of only diubiquitin after 20 hours at 37°C (Figure 2A).

Next, we used size exclusion chromatography (SEC) to test for stable interactions between $RNF8_{345-485}$ and Ubc13/Mms2. $RNF8_{345-485}$/Ubc13 elutes with a lower retention volume as compared to free $RNF8_{345-485}$, indicating a direct interaction between Ubc13 and RNF8 (Figure 2B). The ternary complex of $RNF8_{345-485}$/Ubc13/Mms2 forms a complex with a slightly lower retention volume than the RNF8/Ubc13 complex (Figure 1B). The presence of all three components in the complex was verified by mass spectrometry. As expected, $RNF8_{345-485}$ shows no interaction with Mms2 alone confirming that the ternary complex between $RNF8_{345-485}$ and Ubc13/Mms2 occurs between RNF8 and Ubc13 (Figure 1B). An electrophoretic mobility shift assay (EMSA) titrating increasing Ubc13/Mms2 heterodimer into a constant concentration of RNF8 confirmed complex formation of $RNF8_{345-485}$ and Ubc13/Mms2 (Figure 1C).

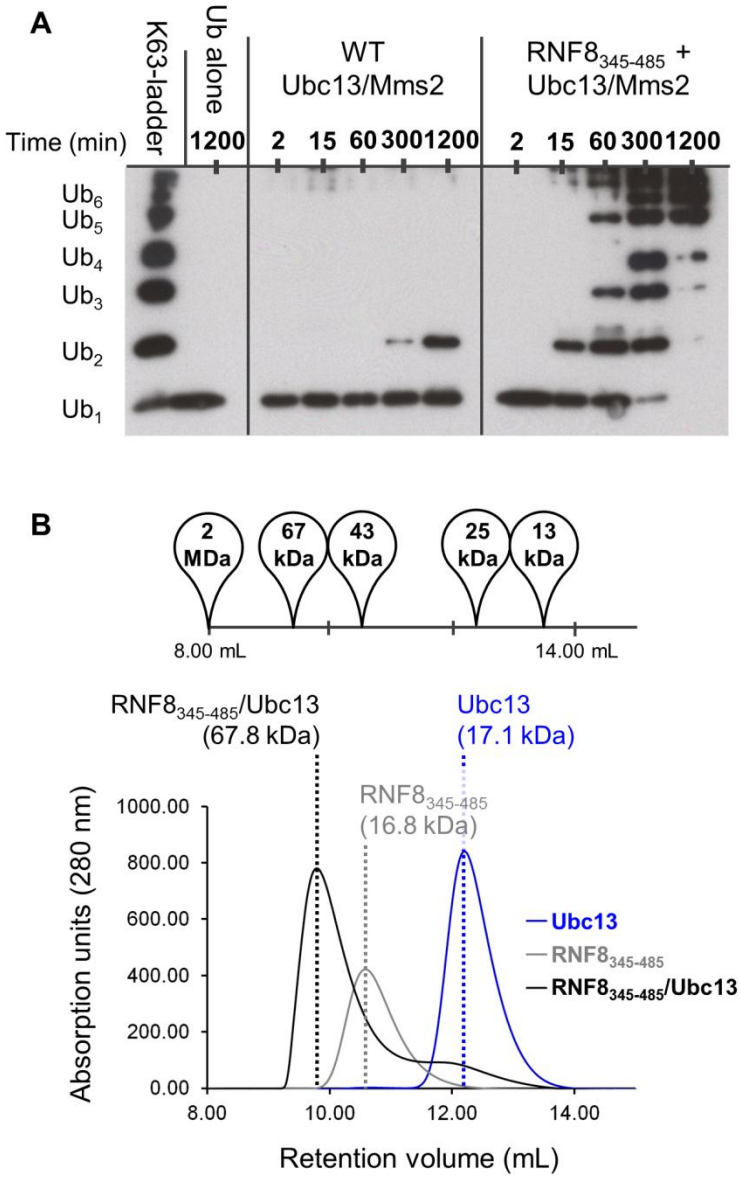


Figure 2. $RNF8_{345-485}$ binds Ubc13/Mms2 to catalyze K63-linked polyubiquitylation.

A $RNF8_{345-485}$ enhances Ubc13/Mms2 polyubiquitylation.

B $RNF8_{345-485}$ forms a complex with Ubc13 as determined by SEC. Peak elution volume is marked by a dotted line, and composition and molecular weight is labelled above. The elution volumes of size standards are shown above.

2.2b: The crystal structure of the RNF8$_{345-485}$/Ubc13/Mms2 complex. We crystallized and determined the structure of the RNF8$_{345-485}$/Ubc13/Mms2 complex to 4.8 Å resolution, with three complexes in the asymmetric unit. The Ubc13/Mms2 heterodimer was used as a search model, and maps phased with two copies of Ubc13/Mms2 revealed clear F_o-F_c difference electron density of a two-fold symmetric RNF8$_{345-485}$ dimer (Figure 3A). Analysis of an anomalous difference map revealed peaks for the positions of the pairs of zinc atoms coordinated within each of the RING domains, and an anomalous difference map calculated from a crystal containing seleno-methionine-substituted RNF8$_{345-485}$ was used to establish the amino acid sequence register of RNF8 (Figure 3A and B). Immediately evident is a coiled-coil mediated dimer formed between the helices N-terminal to the RING domain of the two protomers of RNF8 (Figure 3A, 4A). This is consistent with the observations from SEC, which show that RNF8$_{345-485}$ elutes at a much smaller retention volume than that expected for a 16.8 kDa protein, likely because of dimer formation and a larger R_g due to an elongated structure (Figure 2B). The N-termini of the coiled-coil helices form crystal contacts with the Mms2 of a complex related by crystal symmetry. Two of the complexes observed in the crystallographic asymmetric unit contain one RNF8 dimer and one Ubc13/Mms2 heterodimer. In the third complex a single RNF8 protomer lies on a crystallographic two fold axis, forming a complex that consists of one RNF8 dimer and two Ubc13/Mms2 heterodimers. In the two

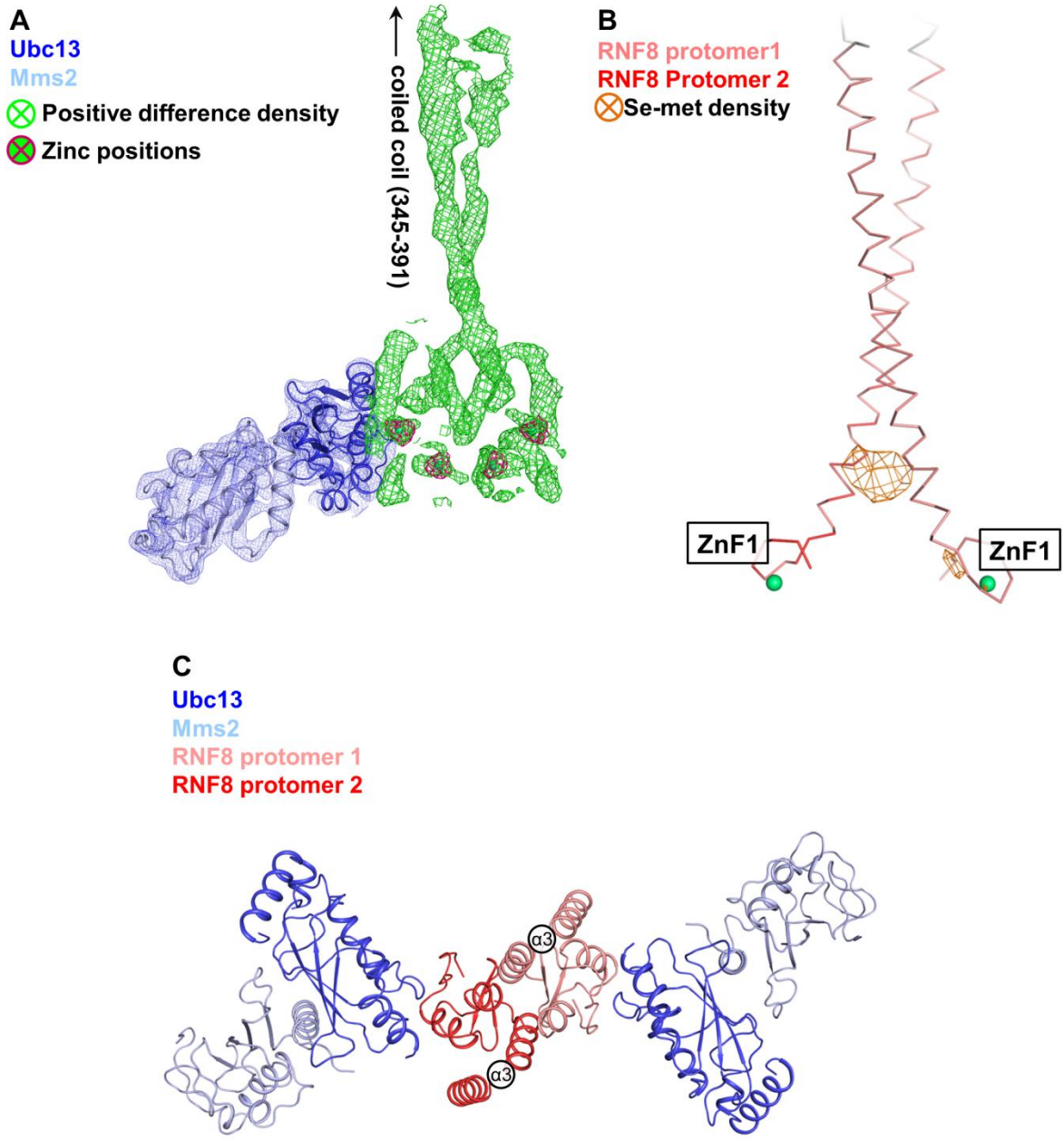


Figure 3. The crystal structure of RNF8$_{345-485}$/Ubc13/Mms2.

A Electron density of RNF8 homodimer visible after phasing with two Ubc13/Mms2 heterodimers. The 2Fo-Fc map for Ubc13/Mms2 is shown in blue, and the positive difference density corresponding to the RNF8 dimer contoured to 2σ is shown in green. The zinc density resulting from an anomalous difference map is shown in pink.

B Anomalous difference map contoured to 2σ showing the position of Met393 in the coiled-coil. Difference density is shown in orange, and the RNF8 protomers are shown as lines, colored salmon and red.
C The 1:1 complex of $RNF8_{345-485}$/Ubc13/Mms that lies on the two fold crystallographic axis shown as a cartoon. $RNF8_{345-485}$ is shown in red, Ubc13 in dark blue and Mms2 in light blue. The C-terminal α3 in $RNF8_{354-485}$ is labelled.

complexes that contain only one bound Ubc13 per RNF8 dimer, the Ubc13 binding surface of the unbound RNF8 protomer is not occluded, and based on the relative zinc positions in the two RNF8 RING domains does not seem to be distorted. Nevertheless, no interaction with Ubc13 is visible in the electron density. Conversely, the third complex that sits on the two fold axis is necessarily symmetrical, with both protomers of the RNF8 dimer interacting with a Ubc13/Mms2 heterodimer (Figure 3C). In this structure, the coiled-coil is not visible in the electron density. This is likely because the crystal packing is such that the tip of the coiled-coil cannot make contacts with symmetry mates – as is observed in the other two complexes – resulting in increased flexibility.

The structural similarity of RNF8 with other RING domains is shown with an alignment of the RING domain of RNF4 (Figure 4C). The RING domain of RNF8 contains two prominent Zinc fingers (ZnF1 and ZnF2) and an α-helix (α2) immediately following an internal two stranded anti-parallel β-sheet (β1 and β2). RNF8 contains an extended N-terminal helix

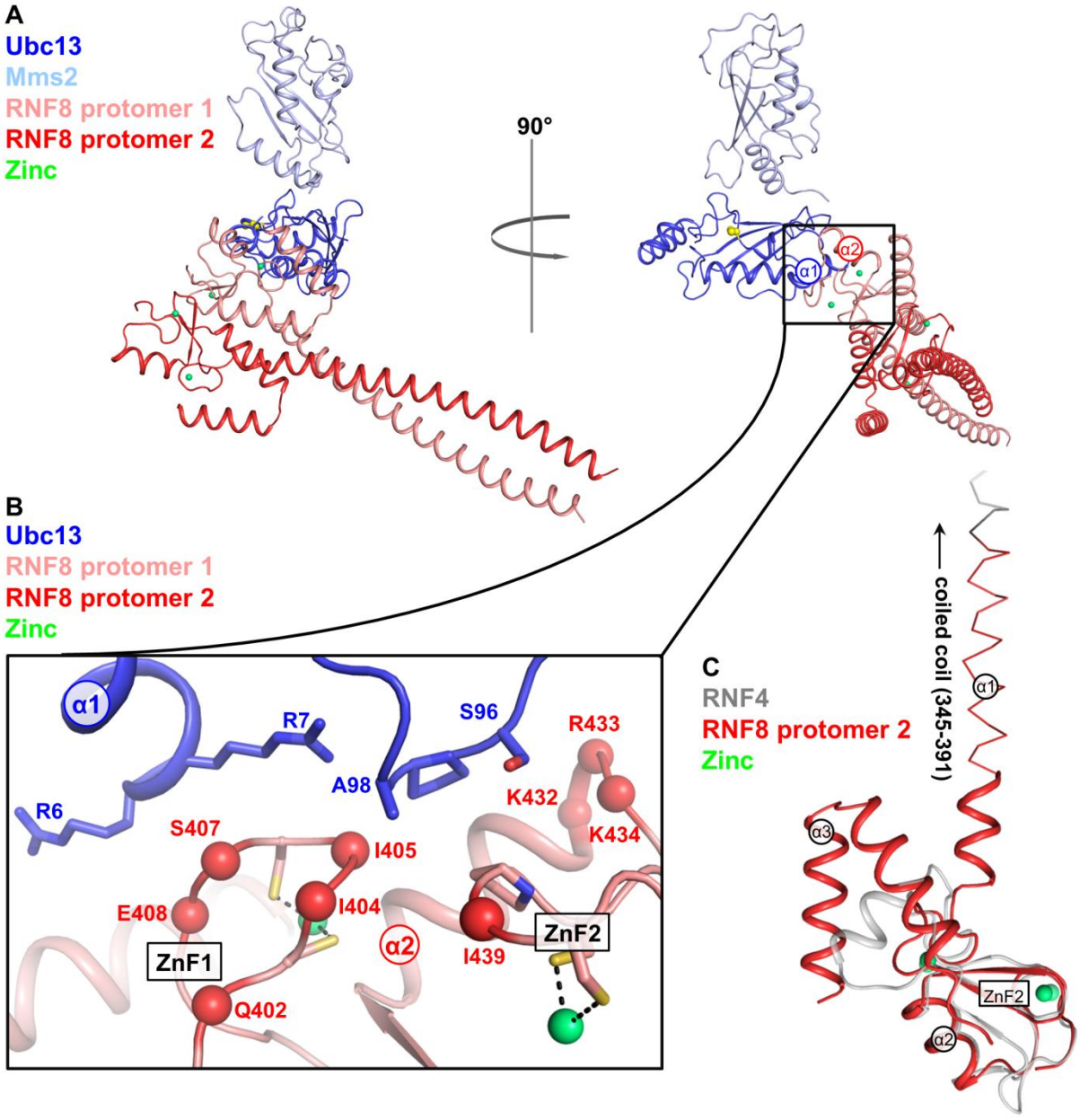


Figure 4. The crystal structure of RNF8$_{345-485}$/Ubc13/Mms2.

A Two orientations of the crystal structure of the RNF8/Ubc12/Mms2 ternary complex shown in cartoon representation. The RNF8 protomer interacting with Ubc13 is shown in salmon, and the unbound RNF8 protomer is shown in red. Ubc13 is shown in dark blue, and Mms2 is shown in light blue. The catalytic Cys87 in Ubc13 is shown as yellow

spheres, and the Zn atoms coordinated by the ZnFs of RNF8 are shown as green spheres.

B The RNF8/Ubc13 binding interface. Important residues in complex formation are shown as spheres on RNF8 and sticks on Ubc13. RNF8 is shown in salmon, and Ubc13 is shown in dark blue.

C Alignment of a single RNF8 protomer with the crystal structure of RNF4 (PDB ID: 2XEU). RNF8 is shown in red, and RNF4 is shown in gray. The N-terminal coiled-coil is shown as a line, and important structural features are labelled. The truncated construct of RNF8 lacks the coiled-coil as indicated by the transition between lines and cartoon.

(α1) preceding ZnF1 that forms the coiled-coil, as well as a C-terminal helix (α3), both of which form the RNF8 dimerization interface. N- and C-terminal helices are also found in other RING domain containing protein dimers, including BRCA1-BARD1, TRAF6, and CHIP, although none of these contain an extended dimeric coiled-coil [20, 21, 24].

Although the crystal structure is low resolution, certain potential contacts can be identified between RNF8 and Ubc13. The ZnFs of the bound protomer of RNF8 interact with Ubc13, with Pro438, Ile439, Ile404 and Ile405 positioned to contact the Ser-Pro-Ala motif on Ubc13, similar to interactions observed between the RING domains of both TRAF6 and CHIP with Ubc13 (Figure 4B). In the crystal structure of TRAF6 bound to Ubc13, Asp57 just N-terminal to the first ZnF also makes electrostatic contacts with basic residues in α1 of Ubc13 [20]. The conformation around Asp57 in TRAF6 is not conserved in RNF8, but polar and acidic

residues near ZnF1 (including Gln402, Ser407 and Glu408) still come into proximity with the basic surface of Ubc13 α1.

2.2c: The stoichiometry of the RNF8/Ubc13 complex is 1:1 in vitro. Interestingly, complexes with both 2:1 and 1:1 RNF8:Ubc13 ratios are observed in the crystal structure. Previously, RNF4 and RAD18 were shown to form an asymmetric complex with their corresponding E2 enzymes, with one RING domain dimer binding to one E2 [25, 26]. To identify whether or not RNF8 forms an asymmetric complex, we performed a titration of Ubc13 into a constant concentration of $RNF8_{345-485}$ on a gel filtration column (Figure 5A). As the amount of Ubc13 is increased, the complex peak reaches its maximum height at a RNF8:Ubc13 ratio of 1:1. Additional complex is not formed in the presence of excess Ubc13, suggesting a 1:1 stoichiometry. To verify this result, we ran Multi Angle Laser Light Scattering (MALLS) on purified samples of the RNF8/Ubc13 complex at three different molar ratios (Figure 5B). At a molar ratio of $RNF8_{345-485}$:Ubc13 of 2:1, the MW prediction across the complex peak is sloped, indicating a heterogenous species consisting of both RNF8/Ubc13 complexes and free RNF8. At a molar ratio of 1:1 $RNF8_{345-485}$:Ubc13, the MW prediction is a constant 65.3 ± 2 kDa across the peak, corresponding to an RNF8 dimer bound to two Ubc13 monomers, and this stoichiometry does not change under conditions of excess Ubc13 (Figure 5C). Taken

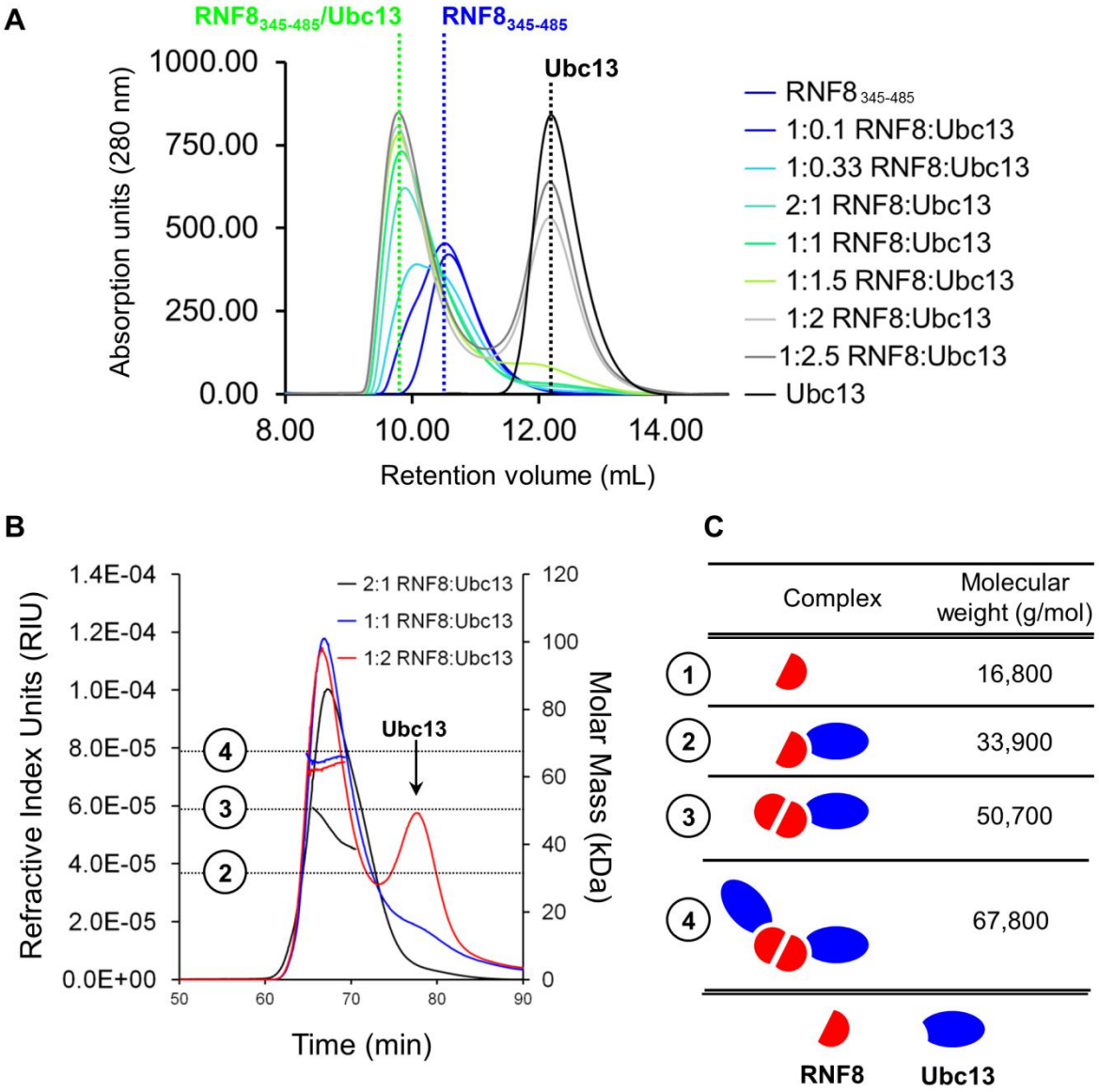


Figure 5. The stoichiometry of the RNF8/Ubc13 interaction.

A Titration of increasing Ubc13 (20 µM to 500 µM) into a constant concentration of RNF8 (200 µM). Peak elution volumes are indicated by dotted lines, and the composition of each peak is labelled above.

B Multi Angle Light Scattering (MALLS) of purified samples of RNF8/Ubc13 complexes. The Gaussian curves indicate protein elution based on refractive index, and the horizontal curves indicated predicted molecular weight.

C Cartoon showing the possible complexes forming in solution. Only a dimer of RNF8 interacting with a two Ubc13 molecules fits the molecular weight predictions determined by MALLS.

together, these data suggest that each protomer of RNF8 within a dimer can independently bind a Ubc13/Mms2 heterodimer in solution.

2.2d: RNF8$_{345-485}$ interacts with Ubc13 through its ZnF motifs. Our structure reveals that the RNF8 RING finger interacts with Ubc13 in a manner that is similar to that of the RING type E3 ubiquitin ligase TRAF6 and the U-box type E3 ubiquitin ligase, CHIP. This conserved binding interface involves four hydrophobic residues at the tips of the ZnFs in the RING domain (Ile404, Ile405, Pro438 and Ile439 in RNF8) that contact a conserved Ser-Pro-Ala motif in Ubc13 (Figure 6A) [20, 21, 27]. To demonstrate the function of this interaction in Ubc13-catalyzed ubiquitylation, we generated several point mutations at positions involved in the E2/E3 interaction and tested their activities in ubiquitylation assays. In Ubc13, we mutated the conserved serine and alanine to aspartic acid, generating S96D and A98D. In RNF8, three residues at the tips of the fingers, I404, I405 and I439, were mutated to aspartic acid. I405D was not stable and did not over express, so an alternate mutant, I405A, was purified. This mutation was previously shown to retain RNF8 activity as well as the interaction between RNF8 and Ubc13 [28].

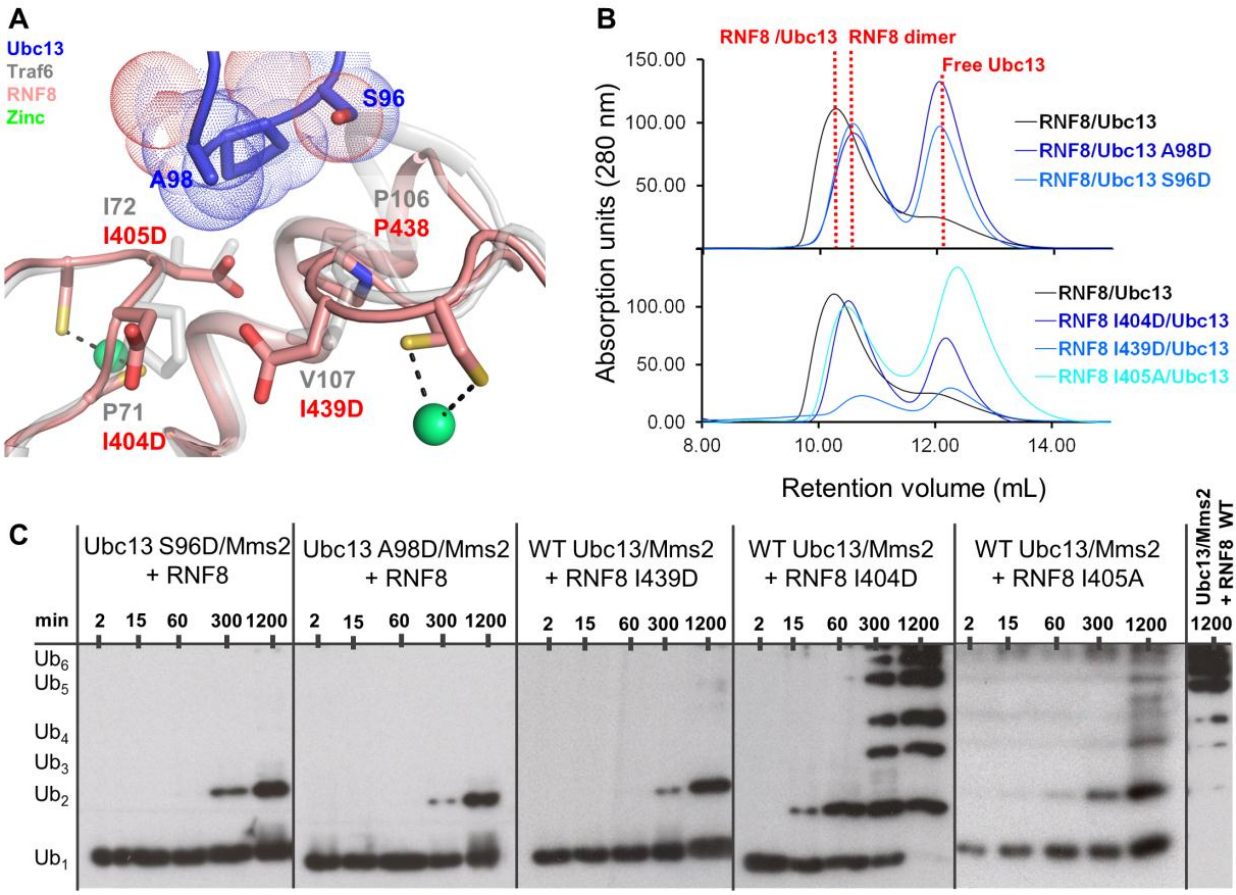


Figure 6. RNF8 interacts with Ubc13 through its ZnFs.

A An alignment of the RNF8 RING domain with the crystal structure of the TRAF6 RING domain bound to Ubc13 (PDB ID: 3HCU) showing important residues involved in E2/E3 complex formation. RNF8 is shown in red, TRAF6 RING in gray, and Ubc13 in blue. Aspartic acid side chain rotomers predicted based on the relative side chain positions in the high resolution TRAF6 RING structure.

B Gel filtration of mutant Ubc13 (above) and RNF8 (below). Composition is shown above each peak, and peak elution volume is indicated by a dotted line.

C Ubiquitylation assays of mutant Ubc13 and RNF8 carried out for the indicated times.

The Ubc13 mutants, S96D and A98D, abolished complex formation as shown by SEC, as did the RNF8 mutants I404D, I405A and I439D (Figure 6B). The Ubc13 mutants, as well as RNF8 I439D, do not show any substantial ubiquitylation activity over the levels observed in the absence of RNF8, indicating that these residues are essential for the E2/E3 interaction. Conversely, RNF8 I404D shows ubiquitylation activity only slightly weaker than WT, and RNF8 I405A shows weak ubiquitylation activity, intermediate between that observed for RNF8 I439D and RNF8 I404D. This suggests that, while the affinity of the I404D and I405A RNF8 mutants for Ubc13 is reduced, sufficient residual binding remains to catalyze ubiquitin chain formation (Figure 6C). These results confirm that the ZnFs of RNF8 are likely involved in the interactions with the conserved Ser-Pro-Ala motif in Ubc13, but that not all of the residues are involved equally in the interaction. This is supported by the fact that the residue in the position +1 to Cys^1 in ZnF1 (Ile404 in RNF8) is not as conserved as the other residues in the RING domain [29].

2.2e: The coiled-coil is required for dimerization and complex formation. In order to investigate the function of the coiled-coil, we made $RNF8_{392-485}$, which contains the RING domain and the C-terminal helix, but a truncated N-terminal helix (Figure 4C). This construct, lacking the first 47 residues of the active RNF8 (11.2 kDa) has a greatly increased retention time in SEC, consistent with a monomer, and does not show any

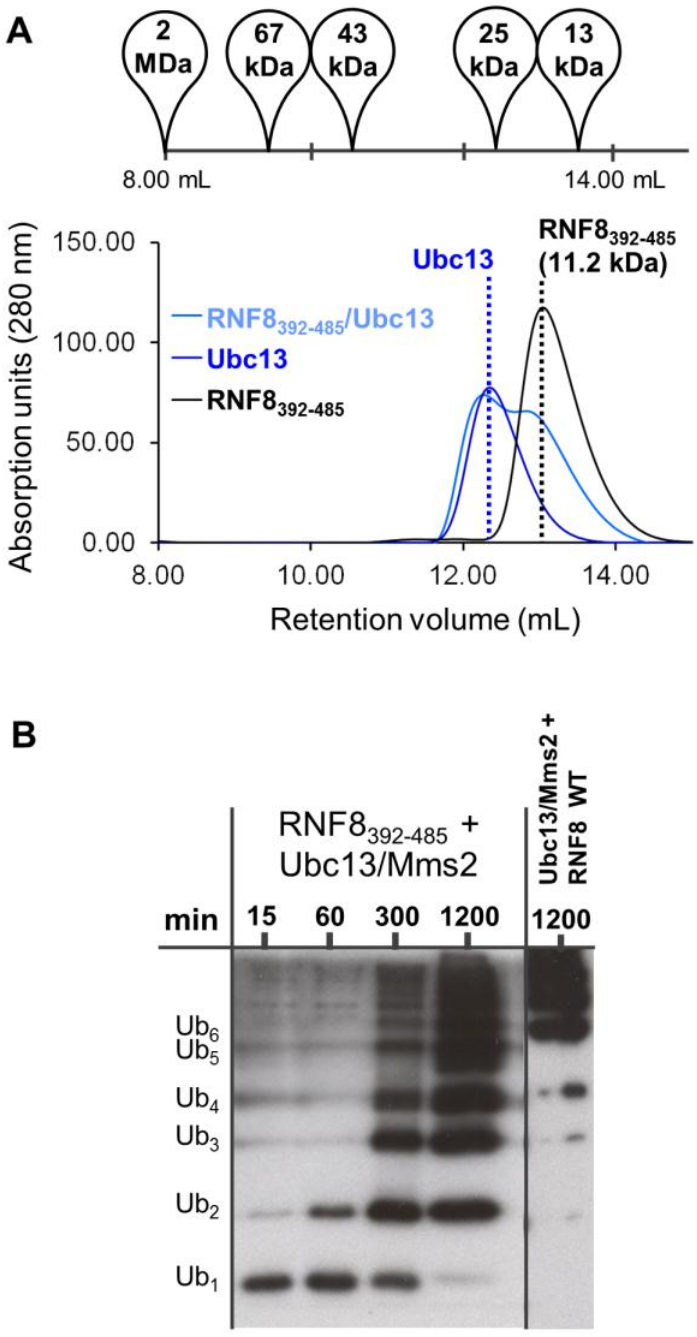


Figure 7. $RNF8_{392-485}$ does not interact tightly with Ubc13 and is deficient in K63-Ub chain formation.

A Gel filtration of $RNF8_{392-485}$ free and with Ubc13. Peak elution volume is indicated with a dotted line, and composition and molecular weight is shown above the peaks. Elution volumes of standards are shown above the graph.

B Ubiquitylation time course indicates $RNF8_{392-485}$ catalyzes polyubiquitylation less efficiently than $RNF8_{345-485}$.

interaction with Ubc13 (Figure 7A). However, this construct does have significant catalytic activity, albeit reduced significantly compared to $RNF8_{345-485}$ (Figure 7B).

Based on solvent exposed surface area calculations, the coiled-coil accounts for roughly one third of the total dimerization interface of $RNF8_{345-485}$, which explains the reduction in stability of the RNF8 dimer upon its truncation. In addition, the N-terminal helix that forms the coiled-coil extends directly into ZnF1, which contains the residues Ile404 and Ile405, both of which are important for the interaction with Ubc13 (Figure 3B). It is possible that deletion of the coiled-coil results in a collateral destabilization of ZnF1, reducing Ubc13 binding affinity and loss of ubiquitylation activity.

We also examined the possibility that the coiled-coil could harbour cryptic tandem Ubiquitin Interacting Motifs (UIMs), similar to the tandem UIMs of RAP80, which facilitate recruitment of RAP80 to K63-ubiquitylated chromatin [8, 13, 30] (Figure 8). We mutated Ala384 on RNF8 to an aspartic acid, which corresponds to a critical ubiquitin binding residue in RAP80, however this mutation did not interfere with complex formation or ubiquitylation [30]. In addition, unlike GST-RAP80, GST-RNF8 immobilized on glutathione-sepharose beads did not pull down K63-

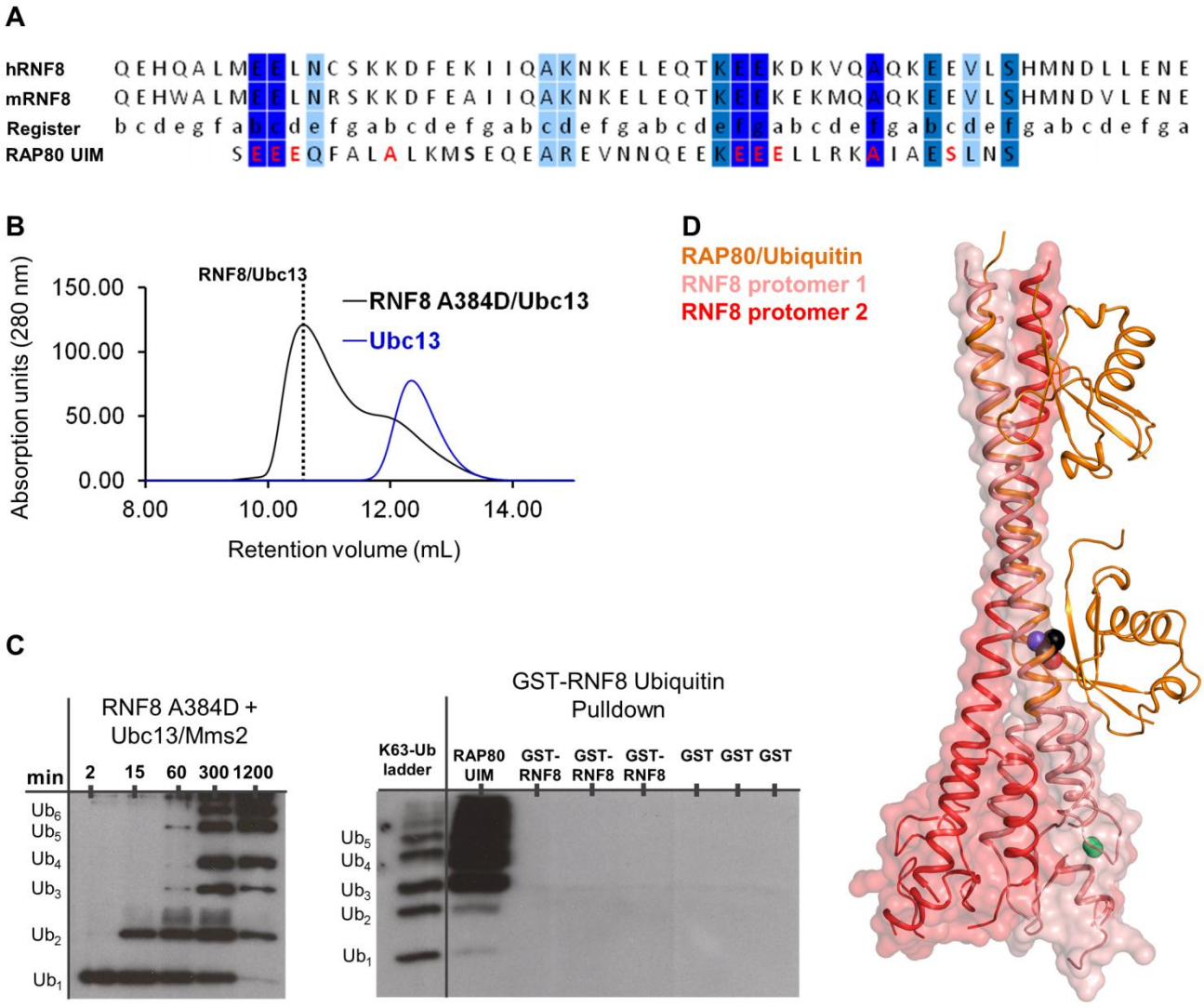


Figure 8. The RNF8$_{345-485}$ coiled-coil as a UIM.

A Alignment of the coiled-coil of hRNF8 and mRNF8 coiled-coil with the UIM of RAP80. Conserved residues are shown in blue. Residues required for ubiquitin binding are shown in red. The relative positions of the residues in the coiled-coil is predicted by the register shown.

B Complex formation between RNF8 A384D and Ubc13 as shown by SEC.

C Samples were incubated at 37°C and quenched with SDS loading buffer, then resolved with SDS-PAGE. GST-RNF8$_{345-485}$ was bound to GST-beads and incubated with K63-Ub ladders, then resolved with SDS-PAGE.

Ubiquitin is detected by incubation with horse radish peroxidase-conjugated mouse anti-Ub.
D Alignment of the crystal structure of RAP80/UIM with the coiled-coil of RNF8. Ub1 makes potential contact with the absolutely conserved UIM A384. RNF8 is shown in red, and ubiquitin is shown in orange.

polyubiquitin ladders, indicating that the RNF8 coiled-coil is unlikely to function as a ubiquitin-binding module.

2.2f: RNF168 has deficient ubiquitylation activity in vitro. Two RNF168 fragments were successfully purified. The construct 1-113 (12.8 kDa) contains only the RING domain and a small C-terminal extension, and 1-200 (23.1 kDa) contains the RING domain as well as one of two Motifs that Interact with Ubiquitin (MIU) thought to be important for RNF168 function [10] (Figure 1A). We were unable to purify full length RNF168 containing the second MIU. Both constructs of RNF168 were deficient in forming K63 ubiquitin chains *in vitro*; both $RNF168_{1-113}$ and $RNF168_{1-200}$ catalyze the formation of only very low levels of K63-polyubiquitin ladders after 24 hours at 37° (Figure 9B). This data agrees with previous reports that suggest that the second MIU in RNF168 is essential to its function [9]. $RNF168_{1-200}$ elutes at a volume consistent with a dimer on SEC, which is in agreement with a predicted coiled-coil in the region surrounding the first MIU domain. $RNF168_{1-113}$, which lacks the predicted coiled-coil, runs as a monomer on SEC (Figure 9A).

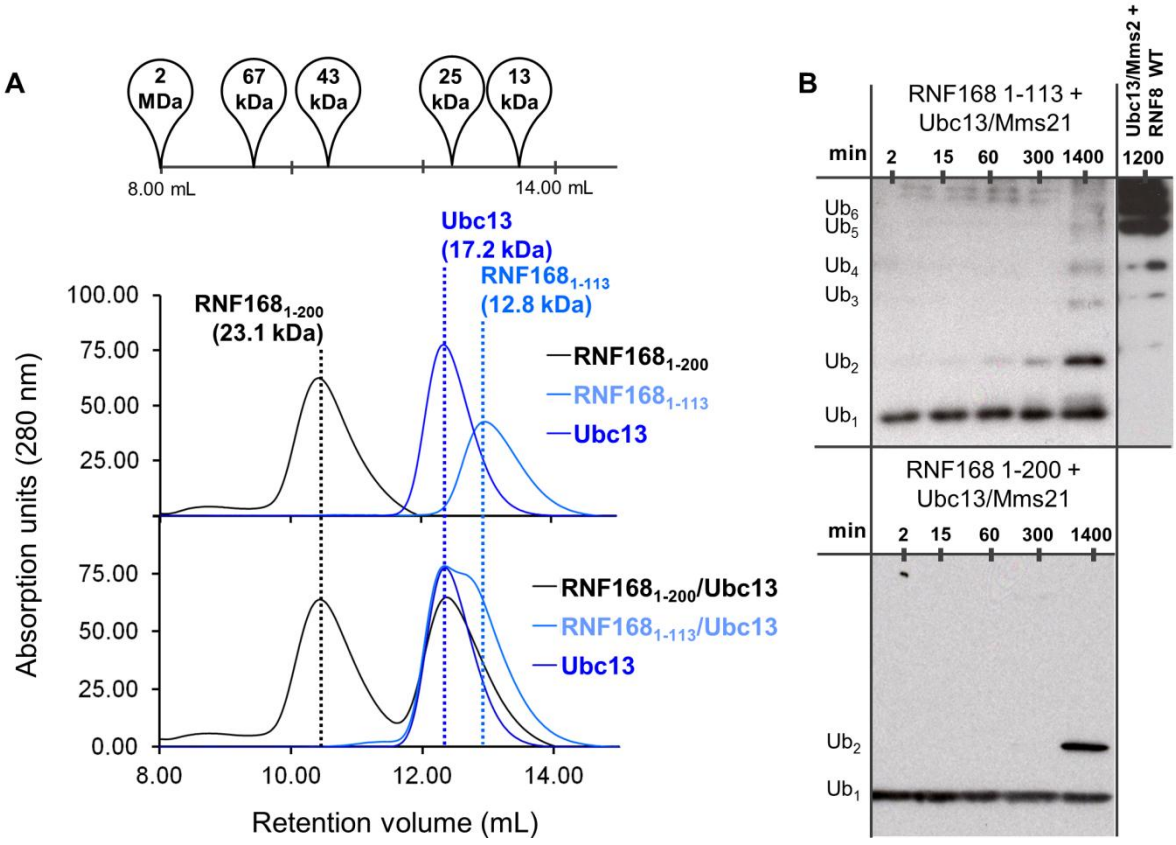


Figure 9. Complex formation and ubiquitylation activity of RNF168$_{1-113}$ and RNF168$_{1-200}$.

A Gel filtration of purified RNF168 constructs. Peak elution volume is indicated by a dotted line. Size and composition is labelled above each peak. Oligomerization is determined by elution volume relative to the standard elution volumes, shown above.

B Ubiquitylation assays of of RNF168$_{1-113}$ and RNF168$_{1-200}$.

In addition, neither RNF168$_{1-113}$ or RNF168$_{1-200}$ show any detectable binding to the Ubc13/Mms2 heterodimer *in vitro* by SEC,

suggesting that the affinity for Ubc13 of these constructs is likely much weaker than that of $RNF8_{345-485}$ (Figure 9A).

2.2g: RNF168 crystal structure reveals possible self-regulation. The crystal structure of $RNF168_{1-113}$ was solved to 2.12 Å resolution with a single molecule of RNF168 in the asymmetric unit. The sequence identity between RNF8 and RNF168 is approximately 40%, and their core RING domains are structurally similar with an RMSD of 1.13 Å (Figure 10A). The ZnFs, helix α2 and β-sheets 1 and 2 align very closely with RNF8. The N-terminal and C-terminal helices found in RNF8 are also conserved in RNF168. However, the N-terminal helix is very short – only about 5 residues – and the C-terminal helix, while kinked as in RNF8, is also shorter (Figure 10A). This truncation of the dimerization interface in RNF168 compared to RNF8 may explain why this RING domain does not form a dimer in solution. A coiled-coil is predicted in RNF168 C-terminal to the RING domain around resides 170-200 [10], and is likely the main interface required for RNF168 dimerization.

Although the RNF168 RING domain shows only very weak ubiquitylation activity *in vitro*, its catalysis of Ubc13/Mms2 mediated K63-ubiquitin chains in HR has been shown previously [9, 10]. ZnF1 in RNF168 retains the absolutely conserved hydrophobic residue (Ile18) that is important for the RNF8/Ubc13 interaction, as well as Pro52 in ZnF2 (Figure 10B). The region near ZnF1 in RNF8 that interacts with Ubc13 is

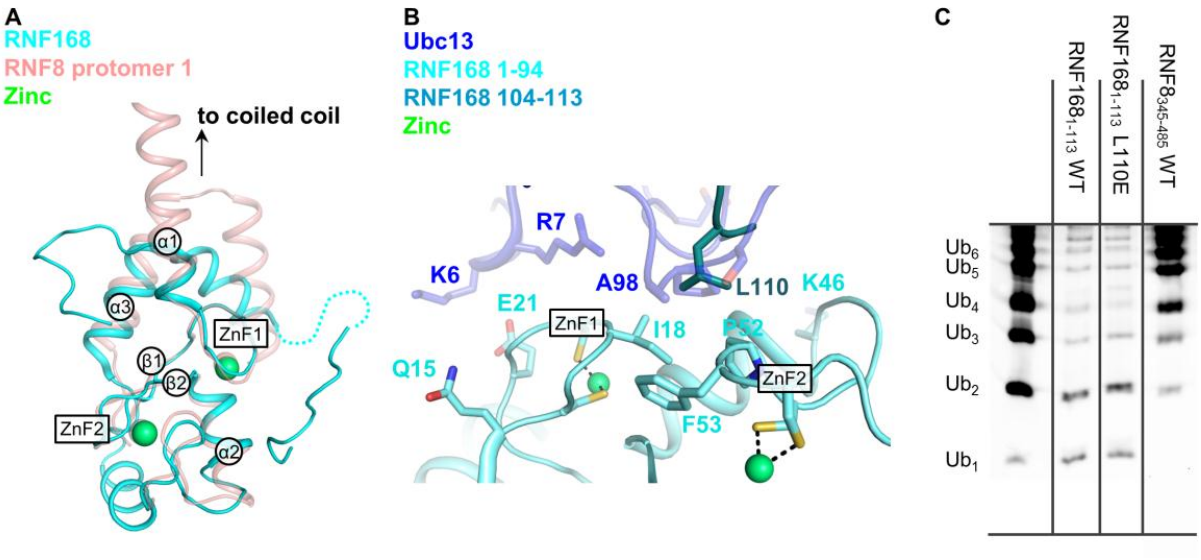


Figure 10. Crystal structure of RNF168$_{1-113}$ shows partial occlusion of the Ubc13 binding site.

A Crystal structure of RNF168$_{1-113}$ (cyan) aligned with a single protomer of RNF8345-485 (red) in cartoon representation. Important structural features are labelled. Zn atoms are shown as green spheres.

B Potential binding interface of RNF168 and Ubc13 created by a structural alignment of RNF168$_{1-113}$ with RNF8$_{345-485}$. Conserved residues in the ZnFs of RNF168 are shown as sticks. The C-terminal tail of RNF168$_{1-113}$ occluding the Ubc13 binding site is shown in dark green. RNF168 is shown in cyan, and Ubc13 is shown in semi-transparent dark blue.

C Ubiquitylation assay using RNF168$_{1-113}$ L110E.

very similar in RNF168, with conserved residues Gln15 and Glu21 in a position to interact with α1 of Ubc13. In RNF168, it is possible that these residues may also contribute to binding.

Intriguingly, an alignment of the RNF168 structure onto one protomer of RNF8 positions Ubc13 into the likely binding site of RNF168, with the conserved Ser-Pro-Ala motif contacting the ZnF motifs in RNF168. However, residues C-terminal to the RNF168 RING domain (104-113) loop back around and Leu110 partially occludes the Ubc13 binding site (Figure 10B). To test whether this is biologically relevant, we mutated Leu110 to a glutamic acid in an attempt to disrupt this interaction and thereby enhance Ubc13 binding. However, this mutant still fails to bind to Ubc13 as assayed by SEC (data not shown), and does not show increased ubiquitylation activity when compared to WT (Figure 10C). The relevance of the interaction of Leu110 with the RNF168 RING ZnFs observed in the crystal structure is unclear.

2.3 Discussion

The specific roles of RNF8 and RNF168 in the generation of K63-linked polyubiquitin chains near the sites of DNA damage are still unclear. Here we have probed the function of the structured regions of these proteins which encompass their RING domains. We show that *in vitro*, the dimeric RNF8 RING domain binds tightly to Ubc13 and catalyzes polyubiquitylation, in a manner that is partially dependent upon its ability to form a dimer through its coiled-coil. Conversely, the monomeric RNF168 RING domain, $RNF168_{1\text{-}113}$, does not interact tightly with Ubc13, resulting in deficient ubiquitylation activity compared to RNF8. The larger $RNF168_{1\text{-}}$

$_{200}$, which contains a region predicted to form a coiled-coil and likely dimerizes, does not interact with Ubc13 and does not catalyze significant ubiquitylation activity.

Previous reports have suggested that while RNF8 initiates ubiquitylation at sites of DNA damage, RNF168 is required to amplify the signal [9, 10]. The fact that the RNF8 RING has a much higher intrinsic affinity for Ubc13 than does the RNF168 RING is consistent with the proposed roles of RNF8 and RNF168 in the generation of K63-polyubiquitin chains in the DNA damage response. We propose that the highly active RNF8 RING is able to build *de novo* chains, whereas RNF168 appears to be much more dependent on the presence of its MIU domains, which likely serve to target RNF168 to pre-initiated chains. The affinity of full-length RNF168 for the poly-ubiquitin chains may make up for the inherently poor Ubc13 binding activity of its RING domain.

The differences in the Ubc13 binding and ubiquitylation activities of RNF8 and RNF168 may also explain why the Ubc13 inhibitor OTUB1 appears to act downstream of RNF168, without impeding RNF8 function at DNA damage foci [14]. It is possible that the higher Ubc13 binding affinity of RNF8 may render it less susceptible to OTUB1 inhibition than RNF168. Also, the high intrinsic ubiquitin ligase activity of RNF8 could lead to depletion of the local mono-ubiquitin pool at nascent DNA damage foci. As OTUB1 is allosterically regulated by free mono-ubiquitin [22, 23], this could further reduce OTUB1-mediated inhibition. However,

after hand-off of ubiquitylation from RNF8 to RNF168, the weaker inherent Ubc13 binding activity of RNF168 could render this enzyme more susceptible to OTUB1 inhibition. The resultant reduction in ubiquitylation could lead to an increase in the mono-ubiquitin pool, initiating a feedback mechanism that would lead to shut down of ubiquitylation.

Mutation of RNF8 Ile405 to an aspartic acid results in an unstable protein that does not express, but mutating this same residue to an alanine results in a stable protein that does not form a tight complex with Ubc13, but is weakly active in a ubiquitylation assay. Because of this residue's proximity to the hydrophobic face of α2 and Pro438, it is likely that an aspartic acid cannot be accommodated at this position, and therefore disrupts RNF8 folding. Hydrophobic residues are strongly conserved at this position, which could explain how a substituted alanine may still make, albeit weaker, hydrophobic interactions with the Ser-Pro-Ala motif in Ubc13. Alternatively, an aspartic acid at position Ile404 disrupts the complex, but does not abrogate ubiquitylation. This position is slightly more solvent exposed, and therefore could more reasonably accommodate a carboxylate group. In addition, a charged residue is sometimes found at this position. For example, the U-box domain of CHIP has a lysine in this position and it is the aliphatic carbons of this side chain that interact with Ubc13. Finally, I439D abolishes complex formation as well as ubiquitylation. In this position, the negatively charged carboxylate group must be solvent exposed, and the only solvent access is towards

the Ubc13 binding site. This would effectively abrogate binding with Ubc13.

Our structural information also provides implications about the mechanism of RNF8-driven ubiquitylation. The crystal structures of other E2/E3 complexes with ubiquitin have revealed that many E3 enzymes may contact both the E2 and the ubiquitin. For example, both the E3 enzymes Rabex-5 and the HECT domain of NEDD4L contact ubiquitin in different ways, possibly to position ubiquitin into a catalytically favourable position [31-33]. In addition, it has been shown that Ubc13/Mms2 alone can form K63-linked polyubiquitin chains without the participation of an E3 [34]. It is possible that an E3 ligase is required simply to increase the rate of the reaction. In their crystal structure of ubiquitin covalently bound to Ubc13 C87S/Mms2, Eddins et al. showed that the Ubc13/Mms2 heterodimer is poised to form the K63-isopeptide bond between ubiquitin units. They show that the K63 of a non-covalently bound acceptor ubiquitin is aligned to attack the C-terminus of the donor ubiquitin [17]. The authors suggest that the covalently bound donor ubiquitin is flexibly tethered to Ubc13, and indeed, alternative donor ubiquitin orientations are observed in other covalent E2~Ub crystal structures [32, 35-37]. Alignment of the crystal structure of the covalent Ub~Ubc13/Mms2 structure with our ternary structure of RNF8 indicates potential contact between RNF8 and the flexible ubiquitin molecule. Lys432, Arg433 and Lys434 (KRK) on α2 of RNF8 (Figure 11) form a positive patch that would clash with Lys6 and

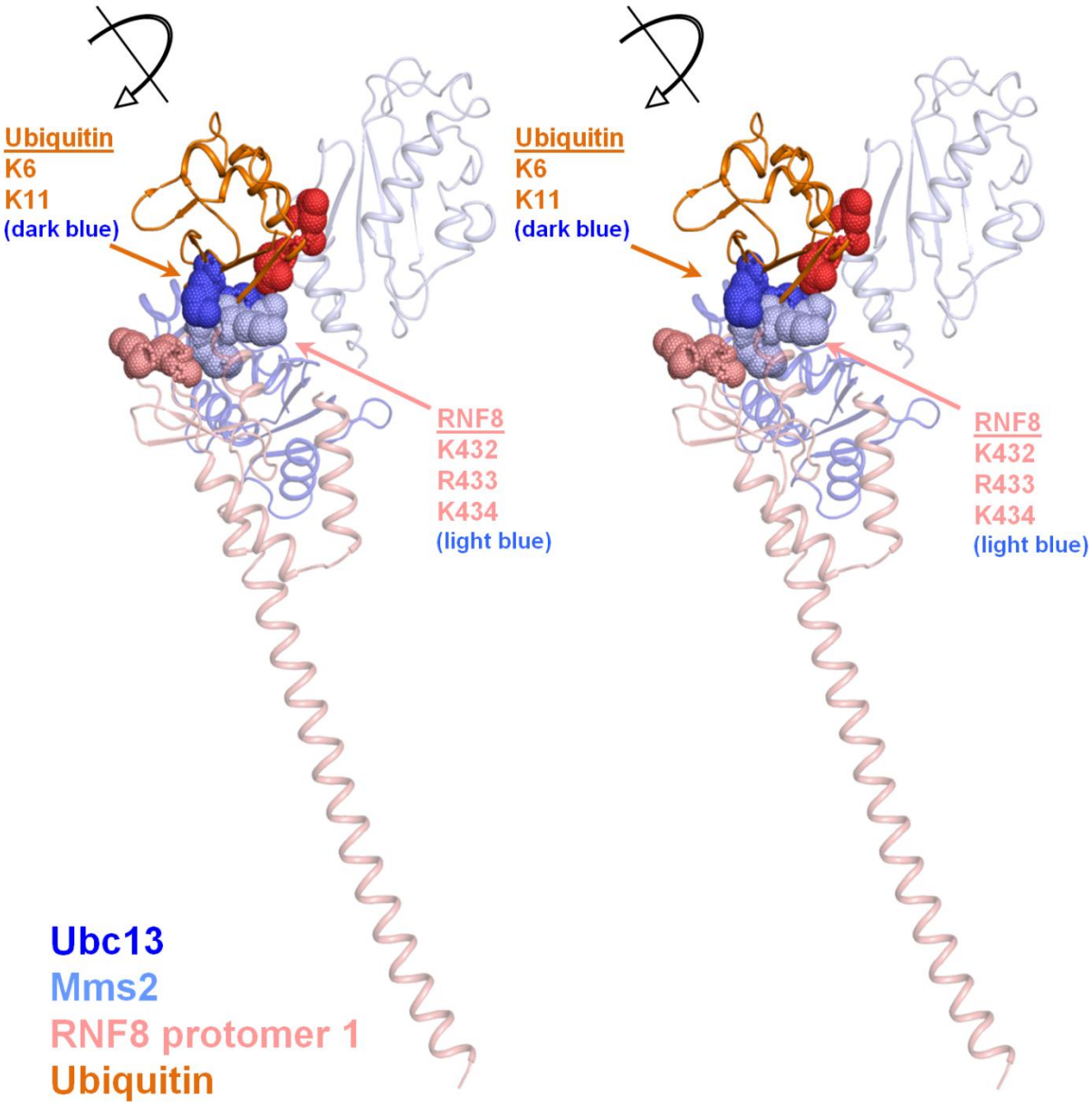


Figure 11. Predicted clashing of the RNF8 RING domain and the Ubc13/Mms2 covalently bound ubiquitin. RNF8 is shown as a red cartoon, and Ubc13 and Mms2 are shown as a dark blue and a light blue cartoon, respectively. Ubiquitin is shown as an orange cartoon, and the positive residues are shown as light blue spheres. The clashing residues in RNF8 are shown as light blue spheres.

Lys11 in β1 and β2 of ubiquitin as it is positioned in the Eddins et al. structure, thus requiring a change in the ubiquitin orientation. We speculate that upon binding of the RNF8 to Ubc13/Mms2, the RING domain makes contact with the covalently bound ubiquitin, orienting the donor ubiquitin in a manner that enhances its reactivity for nucleophilic attack by the Lys63 of the acceptor ubiquitin. Since Ubc13 is primarily charged with mono-ubiquitin, this mode of catalysis is consistent with the idea that RING domains catalyze the formation of ubiquitin chains by sequentially adding a single ubiquitin monomer at a time [38].

Although coiled-coils have been predicted in RING domain-containing proteins previously [8, 9], we have presented the first crystal structure of a homodimeric RING domain stabilized by a coiled-coil. A question that remains is the specific role of the coiled-coil and dimerization of RNF8. One role appears to be to stabilize the RNF8 RING domain for interactions with Ubc13, however other roles are also possible. For example, an overall dimeric oligomeric state of RNF8 could allow the two FHA domains of the dimer to interact with different phosphorylated protein partners. Indeed, the RNF8 FHA not only interacts with MDC1, but also binds phosphorylated HERC2, a HECT E3 ligase, and RNF8 is required for the participation of both MDC1 and HERC2 in HR [39]. It is also possible that both protomers within the RNF8 dimer may participate more directly in the catalysis of ubiquitylation. For example, another dimeric

RING E3, RNF4, uses one RING to contact its E2, UbcH5a, while the other RING domain appears to directly contact the covalently attached donor ubiquitin [26]. While such a mechanism is formally possible in the RNF8-Ubc13 system, it would require a significant shift in the orientation of the donor ubiquitin from that observed in the Ub~Ubc13/Mms2 crystal structure [17]. Thus, it may be that RNF8 dimerization is essential for the optimal function of both its FHA and RING domains.

3.4 Materials and Methods

Expression, purification and mutagenesis. OpenBiosystems cDNA of RNF8 (MHS4771-99611437) was used as a template to clone $RNF8_{345-485}$ and $RNF8_{392-485}$ into the vector pGEX-6P-1 (GE). $RNF168_{1-113}$ used for structure solution was cloned from an OpenBiosystems cDNA template (MHS1010-7508235) into the pET28a-LIC (Genbank) vector using the In-Fusion CF Dry-Down PCR Cloning Kit (Clontech). $RNF168_{1-200}$ was cloned using the OpenBiosystems cDNA template (MHS1010-7508235) and ligated into pGEX-6P-1. $RNF168_{1-113}$ used for biochemical assays was also ligated into the vector pGEX-6P-1.

Recombinant GST fusion RNF8 and RNF168 constructs and full length Ubc13 and Mms2 were expressed in the vector pGEX-6P1 after transformation into *E. coli* BL21 Gold (Invitrogen). Cells were pelleted, and resuspended in lysis buffer (20 mM HEPES pH 6.8, 400 mM NaCl, 10 μM $ZnSO_4$ and 1 μL/mL βME) and 1x HALT Protease Inhibitor solution

(Peirce). Cells were lysed using sonication in the presence of lysis buffer and the lysate was cleared by centrifugation at 40000 rcf for 30 minutes. GST-RNF8 and GST-RNF168 were separated from the cleared lysate via a glutathione-S-sepharose column, and washed with lysis buffer. Recombinant proteins were cleaved from the GST on column with Precision Protease, and further purified by gel filtration chromatography. RNF8, Ubc13 and Mms2 were concentrated in gel filtration buffer (20 mM HEPES pH 6.8, 200 mM NaCl, 10 μM $ZnSO_4$ and mM DTT) quantified by a BCA assay, and mixed in a 1:1:1 molar ratio, respectively. The complex was then purified by gel filtration, concentrated in gel filtration buffer and quantified. Ubc13 and Mms2 were cloned as previously described [19], expressed in pGEX-6P-1 and purified in an identical manner to RNF8 and RNF168. Mutants of RNF8, Ubc13 and RNF168 were made using PCR mutagenesis and the recombinant pGEX-6P-1 plasmids as a template. Inserts containing point mutants were then religated into pGEX-6P-1 and sequenced.

$RNF168_{1-113}$ used in structure solution was purified alternatively. After resuspension in 30 mL per liter bacterial culture of lysis buffer (20 mM Hepes pH 7.5, 500 mM NaCl, 2 mM imidazole pH 8.0, 1mM PMSF, and 1X protease inhibitor (Sigma, P2714-1BTL)), cells were lysed using sonication. A volume of 2.0 mL settled Talon resin per 40 mL lysate (Clontech, 635504) was rocked with unclarified lysate for 60 min at 4 °C and transferred to a column. Protein was eluted with 30 mL of elution

buffer (20 mM Hepes(pH 7.5), 500 mM NaCl, 5% glycerol, 250 mM imidazole(pH 8.0)) and dialyzed overnight at 4 ^{o}C against 50 volumes of dialysis buffer (20 mM Hepes(pH 7.5), 150 mM NaCl, 5 mM DTT). The His6 tag was cut with thrombin, and the resultant protein was further purified by gel-filtration and concentrated to 0.75mM (~10 mg/mL).

Crystallization and structure solution

The RNF8$_{345-485}$/Ubc13/Mms2 complex was concentrated to 10 mg/mL in storage buffer (20 mM HEPES pH 6.8, 200 mM NaCl, 10 µM $ZnSO_4$ and 1mM DTT). Crystals were grown using vapour diffusion at 25°C in 0.075 M sodium acetate pH 4.5 and 1.0 M ammonium phosphate and grew to approximately 200 µm after one month. Data were collected at beamline CMCF 08ID-1 at the Canadian Light Source and beamline 12.3.1 at the Advanced Light Source. The complex structure was solved by Phaser using Ubc13/Mms2 crystal structure (PDB ID: 1J7D) as a search model [40]. Two molecules of Ubc13/Mms2 were found initially, and RNF8 was built manually into the positive Fo-Fc density. A homology model made up of multiple RING domains built manually in Pymol and the zinc positions from an anomalous difference map were used as a guide for model building. The model was refined by rigid body refinement in PHENIX [41]. The coiled-coil was built using and ideal model generated by the program CCCP [42]. Statistics of data collection, processing, and refinement are provided in Table SI. Se-Met derivative data were collected to 9.0 Å. The

$RNF8_{345-485}$/Ubc13/Mms2 complex structure was used as a search model in Phaser against the Se-Met data, and the resulting .pdb file was used to calculate an anomalous difference map and determine the methionine positions.

$RNF168_{1-113}$ crystals were grown at 18°C using the sitting drop method by mixing equal volumes of protein (10 mg/ml) and Crystallization Buffer (1.5 Na malonate, pH 7.3). Suitable crystals were cryoprotected by immersion in 2.4 M Na malonate (pH 7.3) supplemented with 10% TMAO (v/v) prior to freezing in liquid nitrogen. Diffraction data were collected using a home source (Rigaku FRE SuperBright). All data sets were integrated and scaled using the XDS package [43]. The structure was solved using both molecular replacement and the weak phases derived from a Sulfur-SAD. MOLREP was used for molecular replacement with the model (PDB ID: 3FL2) [44]. Automated model building using ARP/wARP, combined with iterative model building using the graphics program Coot were used to build the final model [45]. Maximum-likelihood and TLS refinement in REFMAC5 was used to refine the model [46, 47]. Statistics of data collection, processing, and refinement are provided in Table I. The Protein Data Bank accession codes for the RNF8345-485/Ubc13/Mms2 complex and RNF1681-113 are 4EPO and 3L11, respectively.

Size Exclusion Chromatography. All analytical size exclusion chromatography was done using a Superdex 75 10/300 gel filtration

Table 1. Data Collection and refinement statistics

	$RNF8_{345-483}$/Ubc13/MMs2	$RNF8_{345-485}$/Ubc13/Mms2 Se-Met	$RNF168_{1-113}$
Data Collection			
Space Group	$p4_22_12$	$p4_22_12$	$p4_32_12$
Cell Dimensions			
a,b,c (Å)	205.26, 205.26, 235.37	203.96, 203.96, 234.179	49.7, 49.7, 110.15
α,β,γ (°)	90, 90, 90	90, 90, 90	90, 90, 90
Wavelength (Å)	0.97934	0.97934	1.54
Resolution (Å)	123-4.80	200-9.00	45.3-2.12
$^2R_{sym}$	6.3 (51.3)*	13.5 (35.8)*	12.4 (64.3)*
I/σI	24.3 (2.68)*	7.2 (2.66)*	40.1 (5.17)*
Completeness (%)	97.8 (99.4)*	74.1 (	99.4 (99.7)*
Redundancy	6.8 (5.6)*	4.3 (4.4)*	53.4(54.3)*
Refinement			
Resolution (Å)	123-4.80	N/A	45.3-2.12
No. Reflections	45,858	N/A	8,384
R_{work}/R_{free}	30.1/31.7	N/A	18.9/21.5
No. Atoms			
Protein	10,547	N/A	831
Water	0	N/A	34
Zinc	10	N/A	2
Overall B-factor ($Å^2$)	175	N/A	23.5
R.M.S. Deviations			
Bond Lengths	0.009	N/A	0.024
Bond Angles	1.44	N/A	2.42
Ramachandran			
Preferred (%)	93.2	N/A	96.0
Allowed (%)	6.8	N/A	4.0
Disallowed (%)	0	N/A	0

*Values in parentheses' are for highest resolution shell
^{2}Rsym= $\sum|(I_{hkl}) - \langle I \rangle| / \sum(I_{hkl})$, where I_{hkl} is the integrated intensity of a given reflection.

column (Amersham) at 4°C. Various concentrations of free protein and complex were loaded onto the column with a volume of 100 µL, and run at 0.5 mL/min using an Akta Purifier (GE). An Amersham UV-900 was used to detect protein eluting off the column. Column was standardized by running mixed standards (Dextrin, BSA, Ovalbumin, Chymotrypsinogen and RNase A), each at a concentration of 1 mg/mL and brought to an injection volume of 250 µL.

Multi Angle Laser Light Scattering (MALLS). Three samples of quantified $RNF8_{345-485}$ and Ubc13 were brought up to a concentration of 200 µM of $RNF8_{345-485}$ and increasing concentration of Ubc13 (100, 200 and 400 µM) in gel filtration buffer. The injection volume of the mixed samples was 100 µL. Samples were run over a Superose 6 column, in line with a Wyatt Systems REX to quantify eluted protein concentration via refractive index, and a DAWN MALLS to detect light scattering. Data was analysed with ASTRA, and molecular weight calculations were normalized to a BSA standard.

Ubiquitylation assays. 100 nM human E1 and 200 nM Ubc13/Mms2 was incubated with 0.25 µM of recombinant E3 (RNF8 or RNF168) and brought up to 50 µL in ubiquitylation buffer (20 mM HEPES pH 6.8, 200 mM NaCl, 2.5 mM MgSO4, 10 µM ZnSO4, 0.1 mM DTT, 2mM ATP, 5 mM creatine phosphate, 0.6 units/mL creatine kinase and 0.6 units/mL

inorganic phosphatase) and incubated at 37°C. Samples quenched with SDS-PAGE loading buffer at 2, 15, 60, 300 and 1200 minutes and resolved on a 15% (v/v) polyacrylamide gel. The protein was transferred to a Immobilin-P PVDF membrane and probed with mouse anti-ubiquitin IgG (Santa Cruz) over night at 4°C. Membranes were washed, and incubated with a 1:10000 dilution of anti-mouse linked horse radish peroxidise, and the chemiluminescence reaction was activated using the Super Signal West Hisprobe kit (Pierce). Membranes were imaged by exposure to Kodak Biomax film and scanned.

Acknowledgements

We would like to thank P. Grochulski, S. Labiuk and J. Gorin at the Canadian Light Source beamline 08ID-1 (Saskatoon, SK), and S. Classen at the SYBILS beamline (Lawrence Berkeley National Laboratory, Berkeley, CA) for assisting in the collection of our diffraction data. The authors would also like to thank Laila Yermekbayeva for protein expression and Lissete Crombet for plasmid cloning services of RNF168. This work was supported by a research grant from the Canadian Cancer Society and Canadian Breast Cancer Research Alliance.

3.5 References

1. Huang, L.C., K.C. Clarkin, and G.M. Wahl, *Sensitivity and selectivity of the DNA damage sensor responsible for activating*

p53-dependent G1 arrest. Proc Natl Acad Sci U S A, 1996. **93**(10): p. 4827-32.

2. Ciccia, A. and S.J. Elledge, *The DNA damage response: making it safe to play with knives.* Mol Cell, 2010. **40**(2): p. 179-204.
3. Bennett, E.J. and J.W. Harper, *DNA damage: ubiquitin marks the spot.* Nat Struct Mol Biol, 2008. **15**(1): p. 20-2.
4. Stucki, M., et al., *MDC1 directly binds phosphorylated histone H2AX to regulate cellular responses to DNA double-strand breaks.* Cell, 2005. **123**(7): p. 1213-26.
5. Huen, M.S., et al., *RNF8 transduces the DNA-damage signal via histone ubiquitylation and checkpoint protein assembly.* Cell, 2007. **131**(5): p. 901-14.
6. Mailand, N., et al., *RNF8 ubiquitylates histones at DNA double-strand breaks and promotes assembly of repair proteins.* Cell, 2007. **131**(5): p. 887-900.
7. Kolas, N.K., et al., *Orchestration of the DNA-damage response by the RNF8 ubiquitin ligase.* Science, 2007. **318**(5856): p. 1637-40.
8. Wang, B. and S.J. Elledge, *Ubc13/Rnf8 ubiquitin ligases control foci formation of the Rap80/Abraxas/Brca1/Brcc36 complex in response to DNA damage.* Proc Natl Acad Sci U S A, 2007. **104**(52): p. 20759-63.
9. Doil, C., et al., *RNF168 binds and amplifies ubiquitin conjugates on damaged chromosomes to allow accumulation of repair proteins.* Cell, 2009. **136**(3): p. 435-46.
10. Stewart, G.S., et al., *The RIDDLE syndrome protein mediates a ubiquitin-dependent signaling cascade at sites of DNA damage.* Cell, 2009. **136**(3): p. 420-34.
11. Kim, H., J. Huang, and J. Chen, *CCDC98 is a BRCA1-BRCT domain-binding protein involved in the DNA damage response.* Nat Struct Mol Biol, 2007. **14**(8): p. 710-5.

12. Liu, Z., J. Wu, and X. Yu, *CCDC98 targets BRCA1 to DNA damage sites.* Nat Struct Mol Biol, 2007. **14**(8): p. 716-20.
13. Wang, B., et al., *Abraxas and RAP80 form a BRCA1 protein complex required for the DNA damage response.* Science, 2007. **316**(5828): p. 1194-8.
14. VanDemark, A.P., et al., *Molecular insights into polyubiquitin chain assembly: crystal structure of the Mms2/Ubc13 heterodimer.* Cell, 2001. **105**(6): p. 711-20.
15. Moraes, T.F., et al., *Crystal structure of the human ubiquitin conjugating enzyme complex, hMms2-hUbc13.* Nat Struct Biol, 2001. **8**(8): p. 669-73.
16. Eddins, M.J., et al., *Mms2-Ubc13 covalently bound to ubiquitin reveals the structural basis of linkage-specific polyubiquitin chain formation.* Nat Struct Mol Biol, 2006. **13**(10): p. 915-20.
17. McKenna, S., et al., *An NMR-based model of the ubiquitin-bound human ubiquitin conjugation complex Mms2.Ubc13. The structural basis for lysine 63 chain catalysis.* J Biol Chem, 2003. **278**(15): p. 13151-8.
18. McKenna, S., et al., *Noncovalent interaction between ubiquitin and the human DNA repair protein Mms2 is required for Ubc13-mediated polyubiquitination.* J Biol Chem, 2001. **276**(43): p. 40120-6.
19. Yin, Q., et al., *E2 interaction and dimerization in the crystal structure of TRAF6.* Nat Struct Mol Biol, 2009. **16**(6): p. 658-66.
20. Zhang, M., et al., *Chaperoned ubiquitylation--crystal structures of the CHIP U box E3 ubiquitin ligase and a CHIP-Ubc13-Uev1a complex.* Mol Cell, 2005. **20**(4): p. 525-38.
21. Brzovic, P.S., et al., *Structure of a BRCA1-BARD1 heterodimeric RING-RING complex.* Nat Struct Biol, 2001. **8**(10): p. 833-7.

22. Hibbert, R.G., et al., *E3 ligase Rad18 promotes monoubiquitination rather than ubiquitin chain formation by E2 enzyme Rad6.* Proc Natl Acad Sci U S A, 2011. **108**(14): p. 5590-5.
23. Plechanovova, A., et al., *Mechanism of ubiquitylation by dimeric RING ligase RNF4.* Nat Struct Mol Biol, 2011. **18**(9): p. 1052-9.
24. Zheng, N., et al., *Structure of a c-Cbl-UbcH7 complex: RING domain function in ubiquitin-protein ligases.* Cell, 2000. **102**(4): p. 533-9.
25. Lok, G.T., et al., *Differential regulation of RNF8-mediated Lys48- and Lys63-based poly-ubiquitylation.* Nucleic Acids Res, 2011.
26. Lovering, R., et al., *Identification and preliminary characterization of a protein motif related to the zinc finger.* Proc Natl Acad Sci U S A, 1993. **90**(6): p. 2112-6.
27. Sato, Y., et al., *Structural basis for specific recognition of Lys 63-linked polyubiquitin chains by tandem UIMs of RAP80.* EMBO J, 2009. **28**(16): p. 2461-8.
28. Penengo, L., et al., *Crystal structure of the ubiquitin binding domains of rabex-5 reveals two modes of interaction with ubiquitin.* Cell, 2006. **124**(6): p. 1183-95.
29. Kamadurai, H.B., et al., *Insights into ubiquitin transfer cascades from a structure of a UbcH5B approximately ubiquitin-HECT(NEDD4L) complex.* Mol Cell, 2009. **36**(6): p. 1095-102.
30. Lee, S., et al., *Structural basis for ubiquitin recognition and autoubiquitination by Rabex-5.* Nat Struct Mol Biol, 2006. **13**(3): p. 264-71.
31. Hofmann, R.M. and C.M. Pickart, *Noncanonical MMS2-encoded ubiquitin-conjugating enzyme functions in assembly of novel polyubiquitin chains for DNA repair.* Cell, 1999. **96**(5): p. 645-53.
32. Serniwka, S.A. and G.S. Shaw, *The structure of the UbcH8-ubiquitin complex shows a unique ubiquitin interaction site.* Biochemistry, 2009. **48**(51): p. 12169-79.

33. Hamilton, K.S., et al., *Structure of a conjugating enzyme-ubiquitin thiolester intermediate reveals a novel role for the ubiquitin tail.* Structure, 2001. **9**(10): p. 897-904.
34. Sakata, E., et al., *Crystal structure of UbcH5b~ubiquitin intermediate: insight into the formation of the self-assembled E2~Ub conjugates.* Structure, 2010. **18**(1): p. 138-47.
35. Pierce, N.W., et al., *Detection of sequential polyubiquitylation on a millisecond timescale.* Nature, 2009. **462**(7273): p. 615-9.
36. Bekker-Jensen, S., et al., *HERC2 coordinates ubiquitin-dependent assembly of DNA repair factors on damaged chromosomes.* Nat Cell Biol, 2010. **12**(1): p. 80-6; sup pp 1-12.
37. McCoy, A.J., *Solving structures of protein complexes by molecular replacement with Phaser.* Acta Crystallogr D Biol Crystallogr, 2007. **63**(Pt 1): p. 32-41.
38. Adams, P.D., et al., *PHENIX: a comprehensive Python-based system for macromolecular structure solution.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 2): p. 213-21.
39. Grigoryan, G. and W.F. Degrado, *Probing designability via a generalized model of helical bundle geometry.* J Mol Biol, 2011. **405**(4): p. 1079-100.
40. Kabsch, W., *Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants.* Journal of Applied Crystallography, 1993(26): p. 795-800.
41. Vagin, A. and A. Teplyakov, *Molecular replacement with MOLREP.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 1): p. 22-5.
42. Cohen, S.X., et al., *ARP/wARP and molecular replacement: the next generation.* Acta Crystallogr D Biol Crystallogr, 2008. **64**(Pt 1): p. 49-60.
43. Murshudov, G.N., A.A. Vagin, and E.J. Dodson, *Refinement of macromolecular structures by the maximum-likelihood method.* Acta Crystallogr D Biol Crystallogr, 1997. **53**(Pt 3): p. 240-55.

44. Winn, M.D., M.N. Isupov, and G.N. Murshudov, *Use of TLS parameters to model anisotropic displacements in macromolecular refinement.* Acta Crystallogr D Biol Crystallogr, 2001. **57**(Pt 1): p. 122-33.

Chapter 3

[1]Comparison of the structures and peptide binding specificities of the BRCT domains of MDC1 and BRCA1

[1]A version of this chapter has been published in Campbell, S.J., Edwards, R.A., and Glover, J.N.M., **Comparison of the structures and peptide binding specificities of the BRCT domains of MDC1 and BRCA1**. *Structure*, 2010. 18(2): p. 167-76.

3.1 Introduction

In the face of genotoxic events, the DDR relies on nuclear signalling systems that coordinate the repair of DNA damage with regulation of the cell cycle and apoptosis. Protein phosphorylation plays a central part in these signalling pathways, involving conserved Ser/Thr protein kinases and classes of proteins which have evolved to specifically recognize phosphorylated protein targets [1].

As mentioned in chapter 1, one of the most important classes of protein modules that recognize phospho-proteins in the DNA damage response are BRCT domains, named after the tandem repeats found at the C-terminus of the breast cancer associated protein, BRCA1 [2-4]. This region is critical to the tumour suppressor function of BRCA1, as cancer-associated mutations tend to cluster in this region of the protein [2, 5, 6]. The tandem BRCA1 BRCT repeats pack in a head-to-tail manner [7], forming a single protein domain that specifically binds peptide motifs containing a phospho-serine followed by a phenylalanine 3 residues toward the C-terminus [8] (Figure 1A). Recognition of pSer-x-x-Phe motifs govern interactions between BRCA1 and target proteins such as the BRIP1/BACH1 DNA helicase [9], the DNA end processing nuclease CtIP [10], and the double strand break-associated protein, Abraxas [11-13]. The phospho-serine of the peptide target is recognized by a cluster of residues in the N-terminal BRCT repeat that provide ligands for the phosphate,

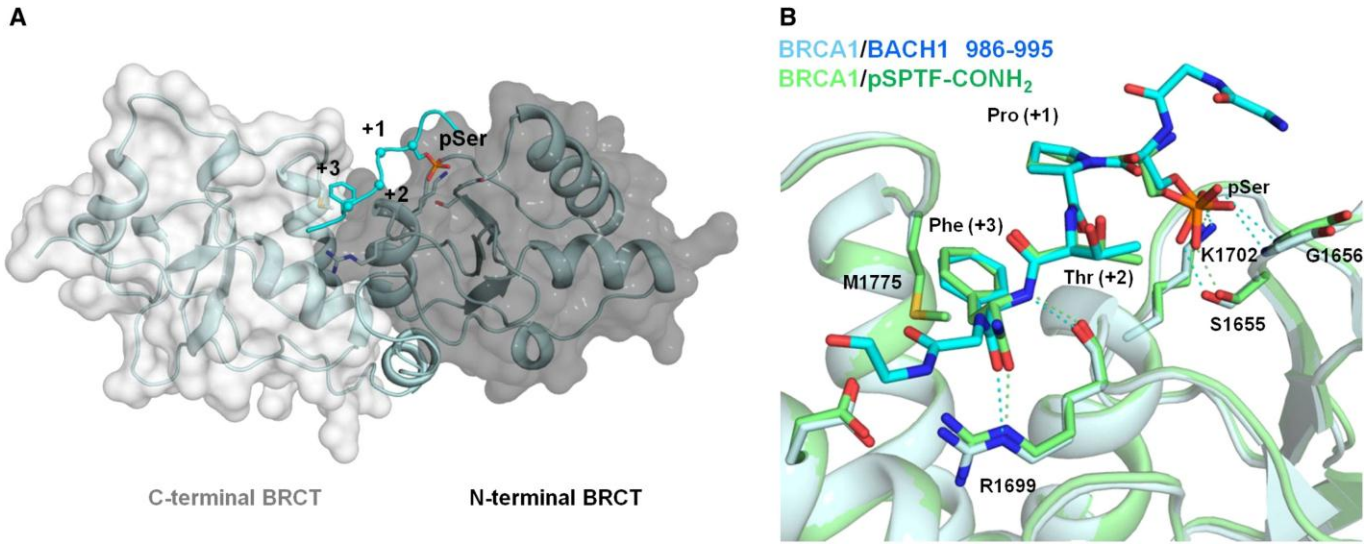


Figure 1. Structure of the BRCA1 BRCT domain bound to phospho-peptide targets.

A Overview of the structure of the BRCA1 BRCT domain bound to phospho-BACH1 (residues 986-995). BACH1 peptide is shown as cyan cartoon with Cα atoms shown as spheres and the phospho-serine and phenylalanine (+3) residues shown as sticks.

B An alignment of the BRCA1/BACH1 structure (teal) with that of BRCA1 BRCT bound to its minimal tetrapeptide recognition sequence (pSPTF-$CONH_2$) (green). The BRCT domain is shown as a cartoon and the peptides are shown as sticks. Important residues involved in peptide binding are labelled and hydrogen bonds are shown with dashed lines.

while the phenylalanine is recognized by a pronounced pocket at the interface of the two repeats [14-17] (Figure 1B). Critical for recognition of the phenylalanine is Arg1699, which interacts with the main chain of the peptide phenylalanine, positioning the side chain in the hydrophobic recognition pocket. The phenylalanine recognition pocket is essential for BRCA1 tumour suppression as is illustrated by the finding that the

missense mutant M1775R specifically perturbs the structure of this pocket and abrogates phospho-peptide recognition [18].

Another BRCT protein, MDC1, plays an essential early role in signalling from DNA double strand breaks (DSBs) through its recognition of the ATM-phosphophorylated histone variant, H2AX [19-21]. Phospho-H2AX (also known as γH2AX) ultimately leads to the recruitment of DNA repair, replication and cell cycle regulatory factors to the site of the damage [22, 23]. The C-terminal phosphorylated tail of γH2AX serves as the specific binding site for MDC1 [24, 25], which initiates a cascade of phosphorylation and ubiquitination events that recruit BRCA1 to the DSB foci [26].

Intriguingly, the BRCA1 and MDC1 tandem BRCT repeats recognize very similar phospho-peptide substrates: pSer–x-x-Phe/Tyr for BRCA1, pSer-Gln-Glu-Tyr for MDC1. In spite of this similarity, MDC1 and BRCA1 bind different phosphorylated proteins to fulfill distinct roles in the DNA damage response. The most striking difference in specificity is likely the high degree of selectivity that MDC1 shows for the C-terminal carboxylate of the γH2AX tail, which is precisely conserved at the +3 position with respect to the pSer in all H2AX orthologs from humans to yeast. Here we probe the structural basis for the differences in selectivity for phospho-peptide C-terminus structure between BRCA1 and MDC1. We demonstrate, through structural and binding analysis of wild type and mutant BRCT domains, characteristics that give MDC1 a large degree of

specificity for the free C-terminus of H2AX and in contrast, the seemingly more promiscuous nature of BRCA1.

3.2 Results

3.2a: Specificity of BRCA1 and MDC1 BRCT domains for the C-termini of their phospho-peptide targets. The MDC1 BRCT domain specifically binds the γH2AX peptide tail, pSer-Gln-Glu-Tyr-COO^-, in a manner that is highly dependent on the free carboxyl-terminus at the +3 position. In contrast, BRCA1 has been shown to bind internal pSer –x-x-Phe peptide targets (as in BACH1 or CtIP) as well as the pSer-x-x-Phe motif at the C-terminus of the protein (as in Abraxas). In order to compare the specificities of the BRCA1 and MDC1 BRCT domains for the chain termini at the +3 position, we used a competitive fluorescence polarization (FP) assay [27]. Each domain was bound to a fluorescein-labelled phospho-peptide target specific for that BRCT domain. The BACH1 target sequence was used for BRCA1, while the γH2AX tail peptide was used for MDC1. The labelled peptide was then competed off the BRCT domain using tetrapeptides containing the minimal recognition sequences: pSPTF for BRCA1, and pSQEY for MDC1. To test the role of the C-terminus in peptide binding, competition experiments were performed with tetrapeptides containing either free carboxylate C-termini, or amidated (and therefore uncharged) C-termini, and the inhibition constants (K_i) of each of the tetrapeptides was determined. The results demonstrate that

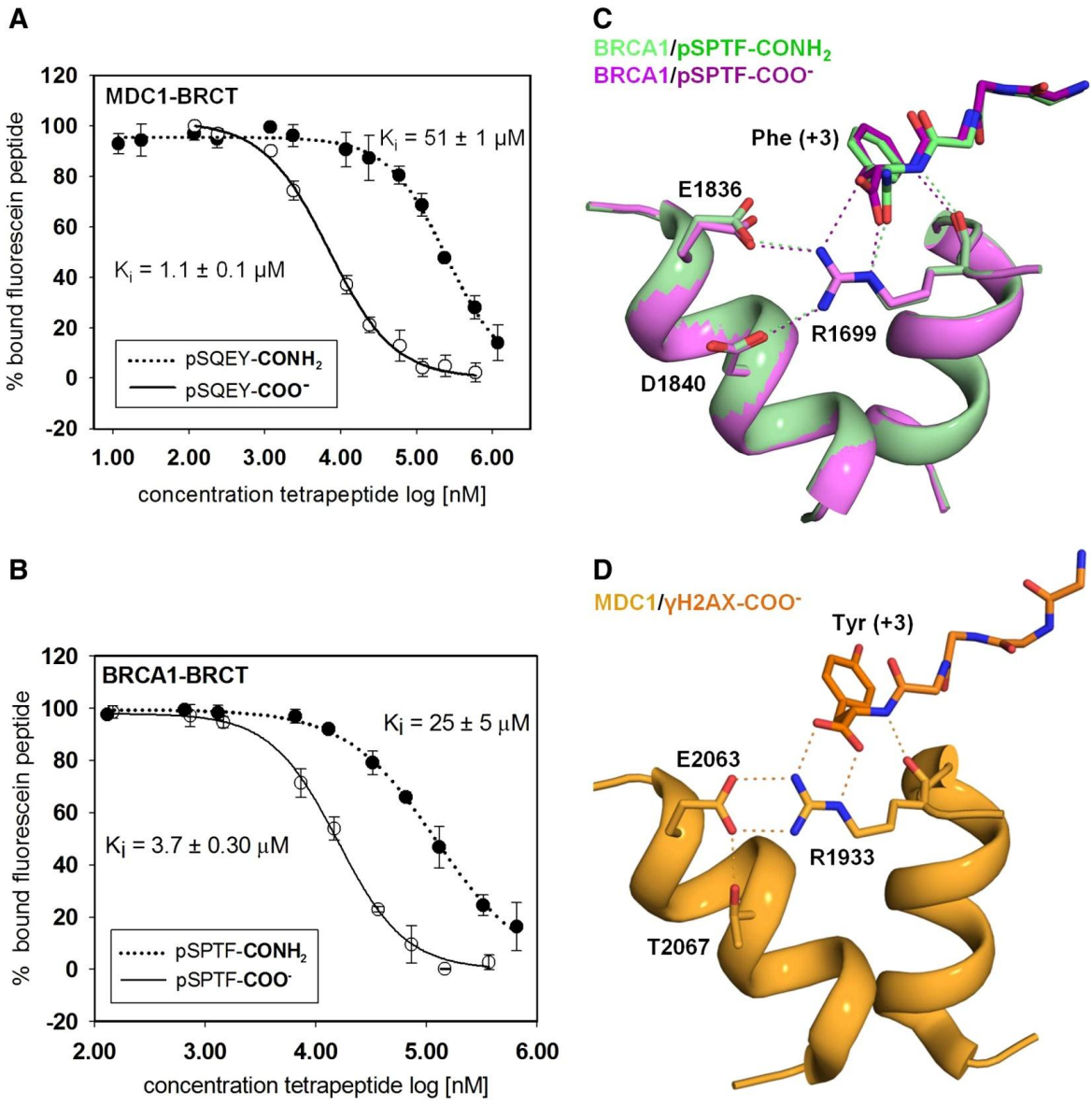


Figure 2. Effect of the structure of the phopsho-peptide C-terminus structure on interactions with BRCA1 and MDC1.

A FP experiments comparing the ability of γH2AX-derived tetrapeptides with amidated or free C-terminal ends to compete for MDC1 binding with a native γH2AX phospho-peptide. K_i values are calculated as described in Experimental Procedures. The dashed line indicates a tetrapeptide with an amidated C-terminus, and the solid line indicates a peptide with a free carboxylate C-terminus.

B FP experiments comparing the ability of BACH1-derived tetrapeptides with amidated or free C-terminal ends to compete for BRCA1 binding with a native BACH1 phospho-peptide. The dashed line indicates a tetrapeptide with an amidated C-terminus, and the solid line indicates a tetrapeptide with a free carboxylate C-terminus.
C A structural alignment of BRCA1 BRCT in complex with the BACH1 tetrapeptide amidated at the C-terminus (green) with the structure of a complex containing tetrapeptide with a C-terminal carboxylate (magenta). Important hydrogen bonds are indicated with dashed lines.
D Structure of the MDC1 BRCT domain bound to γH2AX. Hydrogen bonds and salt bridges involved in recognition of the γH2AX C-terminus are shown with dashed lines.

the tetrapeptide pSQEY-COO^- competes very effectively for binding to MDC1, with a K_i of 1.1 ± 0.1 μM (Figure 2A). However, amidation of the C-terminal tail (pSQEY-$CONH_2$) reduced the K_i approximately 50-fold, to 51 ± 1 μM. These findings are consistent with previously reported values from direct binding measurements [24]. BRCA1 also showed a significant, although less dramatic, preference for a negatively charged C-terminal carboxylate at the +3 position. The tetrapeptide pSPTF-COO^- inhibited binding of the fluorescein-labelled phospho-peptide very effectively (K_i = 3.7 ± 0.3 μM) while the C-terminally amidated version (pSPTF-$CONH_2$) inhibited binding about 7-fold less efficiently (K_i = 25 ± 5 μM) (Figure 2B). The micro molar K_i value for the C-terminally amidated tetrapeptide is

similar to previously reported values using direct binding measurements [27].

3.2b: Structural mechanism of recognition of carboxy chain termini by BRCA1. To investigate the differences in affinity of BRCA1 for a free carboxylate *versus* an amidated tail at the +3 position, we solved the crystal structures of both tetrapeptides bound to the BRCA1 BRCT domain. Complexes of BRCA1 BRCT bound to either tetrapeptide crystallized in the same space group as the unliganded BRCT domain [7] and the structures were determined to 2.70 Å (Table 1). The amidated tetrapeptide (pSPTF-$CONH_2$) binds in a manner that is almost identical to that observed in the previously reported structures of BRCA1 BRCT bound to longer BACH1 peptides [16] (Figure 1B). In these structures, the backbone oxygen of the C-terminal amide makes a single hydrogen bond with the Nε of the conserved Arg1699. The peptide amide nitrogen rotates vertically above the oxygen, in a position that minimizes clashes with the terminal NH_2 of the guanidinium group of Arg1699. In the structure of free carboxylate peptide (pSPTF-COO^-) the C-terminal tail rotates slightly, so that the free carboxylate group can form a double salt bridge with the guanidinium group of Arg1699 (Figure 2C). The enhanced interaction of the free carboxylate group with Arg1699 compared with the amidated C-terminus likely explains the increased K_i of the free carboxylate tetrapeptide for the BRCA1 BRCT.

3.2c: A network of salt bridges controls the specificity of recognition for target peptide chain termini. A comparison of the structure of the MDC1 BRCT domain bound to the γH2AX tail [23] with the BRCA1 BRCT bound to the pSPTF-$CONH_2$ and pSPTF-COO^- tetrapeptides suggests a mechanism that might explain the enhanced specificity of MDC1 for a free C-terminal carboxylate group (Figure 2C and D). In all the structures, a conserved arginine residue (Arg1699 in BRCA1, Arg1933 in MDC1) contacts the backbone of the +3 residue. In the structures of BRCA1 and MDC1 bound to peptides with free carboxylate termini, the analogous arginine residues make similar, double salt bridge interactions with the peptide C-terminal carboxylate and are therefore likely key in recognition of the peptide C-terminus. Differences in the environment of these arginine residues therefore may regulate the differential recognition of the peptide backbone. In both BRCT domains, the arginine is largely buried and is held in place through internal salt bridging interactions. In BRCA1, Arg1699 interacts with both Asp1840 and Glu1836. In MDC1, Glu2063 is equivalent to BRCA1 Glu1836, however Asp1840 in BRCA1 has been replaced with a threonine in MDC1 (Thr2067). As a result, Arg1933 forms a dual salt-bridge interaction with the conserved Glu2063. We suggest that the difference in the environment of the analogous arginine residues impacts recognition of the C-terminus of the phospho-peptide chain. The arrangement observed in MDC1 allows the Arg1933 guanidinium group to make a dual salt bridging interaction with a highly favourable geometry. In

contrast, in BRCA1, the shift in Glu1836 positions one of its carboxylate oxygens closer to the peptide +3 carboxylate (~4.5 Å in BRCA1 *vs* ~5.5 Å in MDC1). This difference may help to explain the reduced selectivity of BRCA1 for a target with a free carboxylate group at the +3 position.

The structures of BRCA1 BRCT indicate that the interaction of Arg1699 with Asp1840 influences its interactions with Glu1836, and therefore may affect peptide binding selectivity. To test the importance of Asp1840 in BRCA1, and the analogous Thr2067 in MDC1, we created BRCA1 and MDC1 mutants in which these residues were exchanged for the corresponding residue in the other BRCT protein, generating the mutants BRCA1 D1840T, and MDC1 T2067D. The effects of these mutations on the structures and peptide binding specificities of the mutant proteins were determined.

3.2d: Structure and peptide binding specificity of BRCA1 D1840T. The BRCA1 D1840T mutant in complex with both the pSPTF-$CONH_2$ and pSPTF-COO^- peptides was crystallized and the structures were determined to 2.8 and 3.0 Å resolution, respectively (Table 1). The structures reveal that the D1840T substitution facilitates an MDC1-like interaction between Arg1699 and Glu1836 in which the two side chains contact one another through double salt bridging interactions (Figure 3A and B). In the BRCA1-BRCT/pSPTF-COO^- structure, the C-terminal

peptide terminus rotates almost completely horizontal relative to Arg1699, resulting in a more optimal salt bridge compared to the wild type structure.

Fluorescence polarization was used to assess the effect of the BRCA1 D1840T mutation on recognition of the pSPTF-COO^- and pSPTF-$CONH_2$ peptides (Figure 3C). The K_i of D1840T BRCA1 for pSPTF-$CONH_2$ increases to 12 ± 2 μM, while the K_i for pSPTF-COO^- increases to 1.6 ± 0.3 μM. The results indicate a subtle 2-fold increase in the binding affinity of the mutant for either tetrapeptide compared to wild type BRCA1. In spite of the fact that the mutation has induced a more MDC1-like conformation in the mutant, there is no corresponding increase in the selectivity of the mutant for peptides with a free C-terminal carboxylate.

3.2e: Structure and peptide binding specificity of MDC1 T2067D. Fluorescence polarization spectroscopy was used to characterize the changes in affinity and specificity induced by the T2067D mutation in MDC1 (Figure 4A). Direct binding experiments suggested a small, 2-fold decrease in the affinity of the T2067D mutant for that of the fluorescein-labelled γH2AX peptide compared to wild type. However, no significant differences between the mutant and wild type were observed in the context of competition studies using the γH2AX tetrapeptide (K_i = 1.4 ± 0.2 μM vs 1.1 ± 0.1 μM), suggesting the mutation induces little if any change in recognition of the γH2AX C-terminal carboxylate. In contrast, the amidated γH2AX peptide (pSQEY-$CONH_2$) was a much more effective

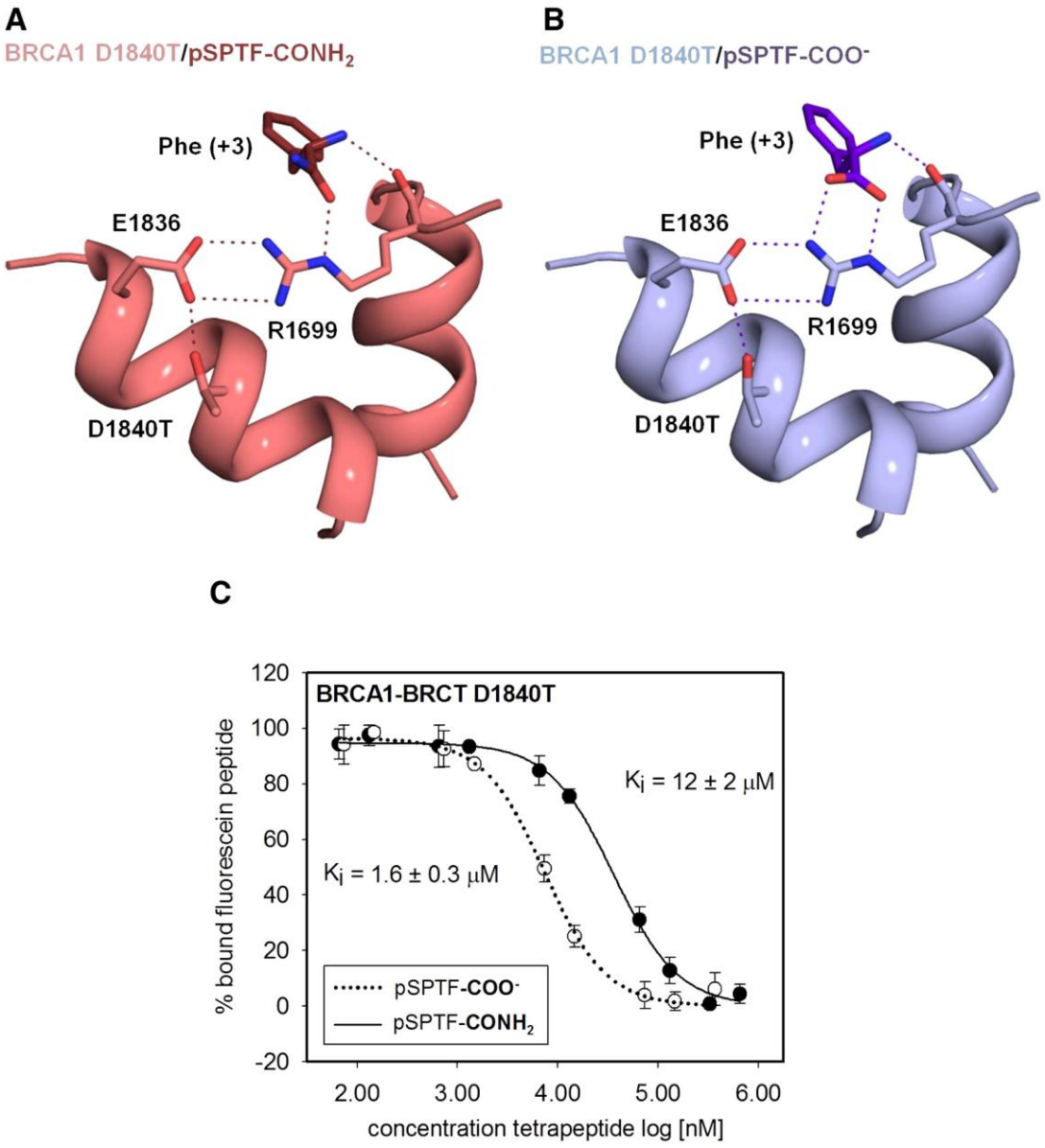


Figure 3. Binding and structural characteristics of BRCA1 BRCT D1840T in complex with tetrapeptides containing amidated or free C-termini.

A Structural features of BRCA1 D1840T bound to pSPTF-$CONH_2$. Dashed lines indicate hydrogen bonds.

B Structural features of BRCA1 D1840T bound to pSPTF-COO^-. Dashed lines indicate hydrogen bonds.

C Competition assay of free and C-terminally amidated tetrapeptides. The solid line indicates the tetrapeptide with an amidated C-terminus, and the dashed line indicates the tetrapeptide with a free carboxylate C-terminus. K_i is calculated from the competition curve as described in Experimental Procedures.

inhibitor of MDC1 T2067D than wild type MDC1, with a nearly 4-fold enhanced K_i compared to the wild type. Thus unlike BRCA1 D1840T, MDC1 T2067D induces a significant change in the specificity of the BRCT domain for the C-terminus of the peptide tail. While wild-type MDC1 shows a 50-fold preference for the γH2AX tail with a free C-terminal carboxylate versus an amidated end, the T2067D mutant shows only a 10-fold preference, very similar to that observed for BRCA1.

To understand the structural basis for this change in specificity, we crystallized and determined the structure of MDC1 T2067D bound to the γH2AX tetrapeptide pSQEY-$CONH_2$ at 1.33 Å resolution. This crystal form is isomorphous with that of the unliganded MDC1 BRCT domain previously determined [25] and contains two BRCT-tetrapeptide complexes in the asymmetric unit (Table 1). One complex in the asymmetric unit reveals an interaction between Arg1933 and Asp2067 that is similar to the salt bridging pattern of the analogous residues in BRCA1 (Figure 4B). However, unlike BRCA1, Glu2063 adopts a rotamer not observed in either wild type MDC1 or BRCA1 such that it no longer directly contacts Arg1933. Instead, a pair of water molecules bind in the space

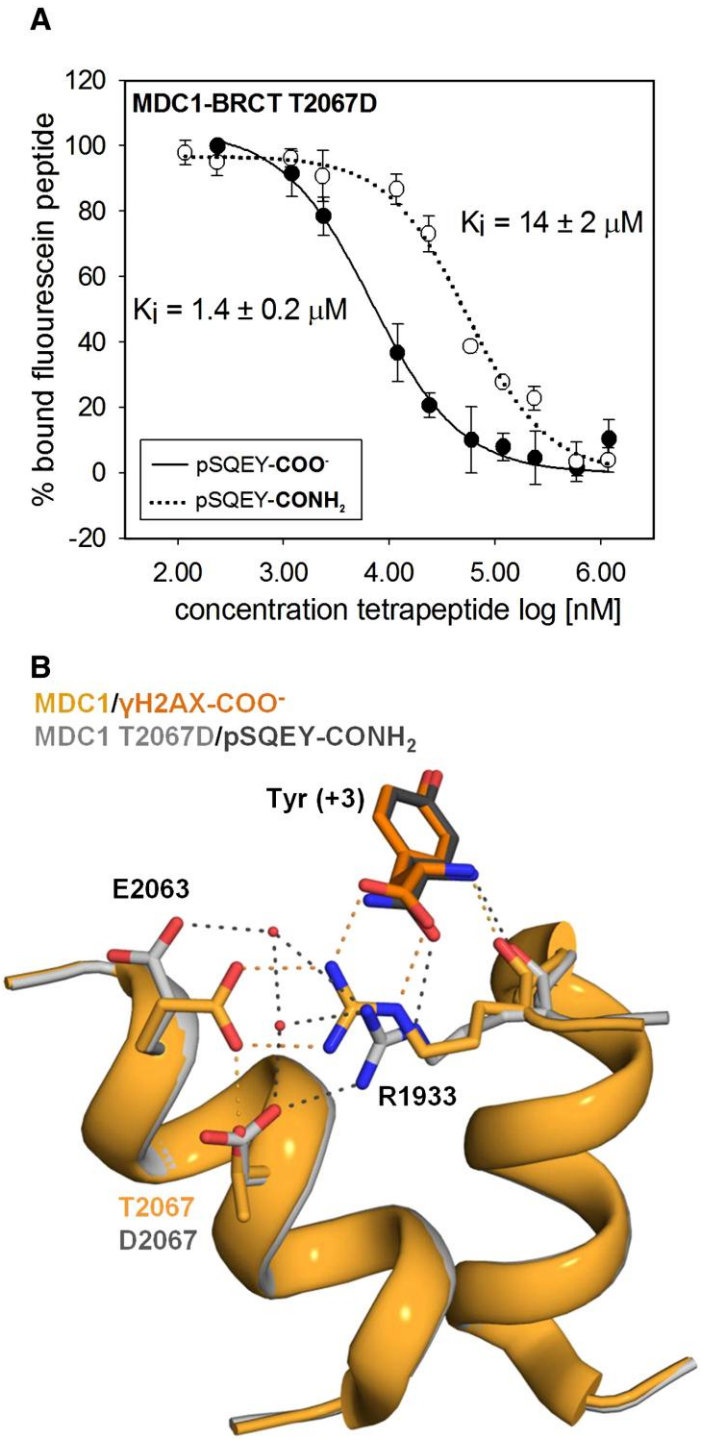


Figure 4. Binding and structural characteristics of MDC1 BRCT T2067D in complex with γH2AX tetrapeptides containing amidated or free C-termini.

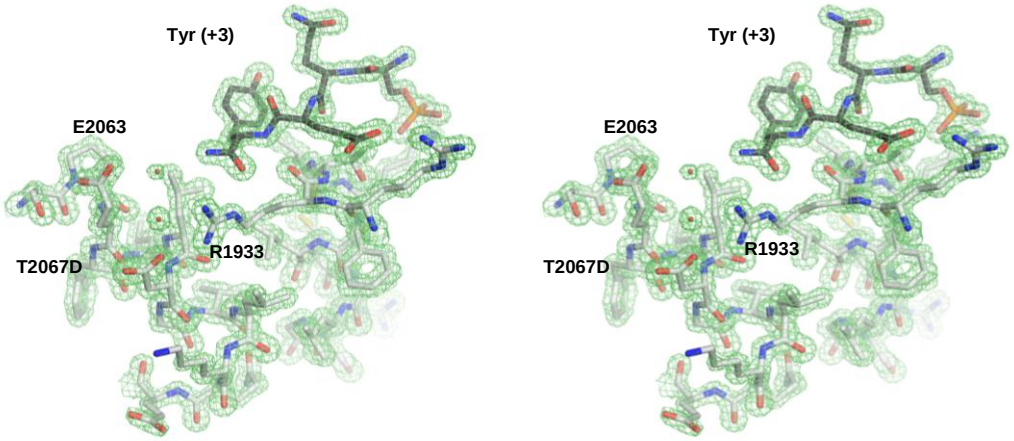
A Competition assay of T2067D in complex with free and amidated tetrapeptides. Dashed line indicates an amidated tetrapeptide, and the solid line indicates a free carboxylate tetrapeptide. The K_i is calculated from the competition binding curve as described in Experimental Procedures.

B Structural features and water network of T2067D bound to pSQEY-$CONH_2$ as shown in chain A (gray) aligned with WT MDC1 (orange). Dashed lines in orange and gray indicate hydrogen bonds. Distance shown in red indicates the increased distance between Arg1933 NH_2 and the tetrapeptide C-terminal amide observed in the mutant structure.

created between Arg1933, Glu2063, Asp2067 and the peptide C-terminus. This arrangement allows the Arg to shift lower in the binding pocket compared to the wild type, increasing the distance between the amidated C-terminus of the peptide and the Arg1933 guanidium, while maintaining hydrogen bonding interactions with the peptide main chain carbonyl group. The water molecules may also stabilize peptide binding through bridging hydrogen bonds between the terminal amide of the peptide and Glu2063. In the other complex in the asymmetric unit, a second set of alternative conformations are observed for Glu2063 and Asp2067. In this alternative conformation, it is Glu2063 that forms a direct salt bridge contact to Arg1933 in a manner that is similar to wild type BRCA1, while the Asp2067 swings away from the Arg1933 and contacts Lys2071 (Figure 5B). Thus, the T2067D mutation in MDC1 does induce a more BRCA1-like recognition of phospho-peptides, through a relaxation of the requirement

A

MDC1 T2067D/pSQEY-CONH$_2$ (Chain A)



B

MDC1 T2067D/pSQEY-CONH$_2$ (Chain B)

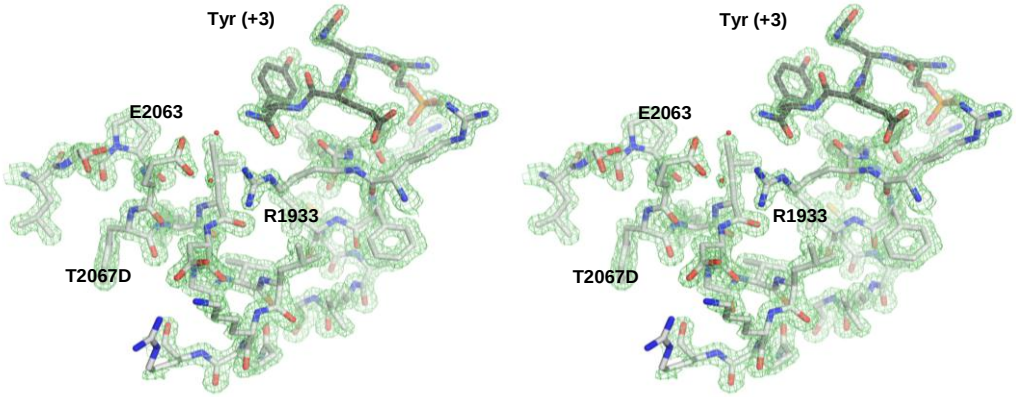


Figure 5 $2F_o$-F_c electron density of both molecules in the asymmetric unit corresponding to the relevant regions of MDC1. Stick representation of the +3 binding region. The electron density map is shown as a green mesh contoured at 2 σ.

A MDC1 T2076D bound to pSQEY-CONH$_2$ as seen in chain A (grey).

B MDC1 T2067D bound to pSQEY-CONH$_2$ as seen in chain B (grey).

for a C-terminal main chain carboxylate at the peptide +3 position. This is accomplished through a shift to a more BRCA1-like conformation in the mutant, as well as an increase in the flexibility and hydration of the recognition surface, which results in a significant enhancement in the affinity of the mutant MDC1 for the C-terminally amidated phospho-peptide.

3.2f: The MDC1 BRCT is highly selective for Ser139-phosphorylated γH2AX and does not bind alternative phospho-peptide targets. Recent studies have demonstrated that the +3 C-terminal tyrosine residue of H2AX (Tyr142) is phosphorylated in a manner that is coordinated with the phosphorylation status of Ser139 [28, 29]. In undamaged chromatin, Tyr142 is phosphorylated by the WSTF component of the WSTF–SNF2H (WICH) chromatin remodeling complex [28]. In response to a DNA double strand break, this phosphorylation is lost due to the action of the EYA phosphatase, and H2AX is subsequently phosphorylated at Ser139 to allow the subsequent binding of MDC1. The balance of these two phosphorylation events may help direct the damaged cell to either an MDC1-dependent DNA repair/cell survival pathway, or an apoptotic pathway [29].

We used fluorescence polarization spectroscopy to quantitate the effects of Tyr142 phosphorylation on interactions with the MDC1 BRCT domain. Our results show a profound, >300-fold reduction in the affinity of MDC1 BRCT – γH2AX interactions upon phosphorylation of Tyr142

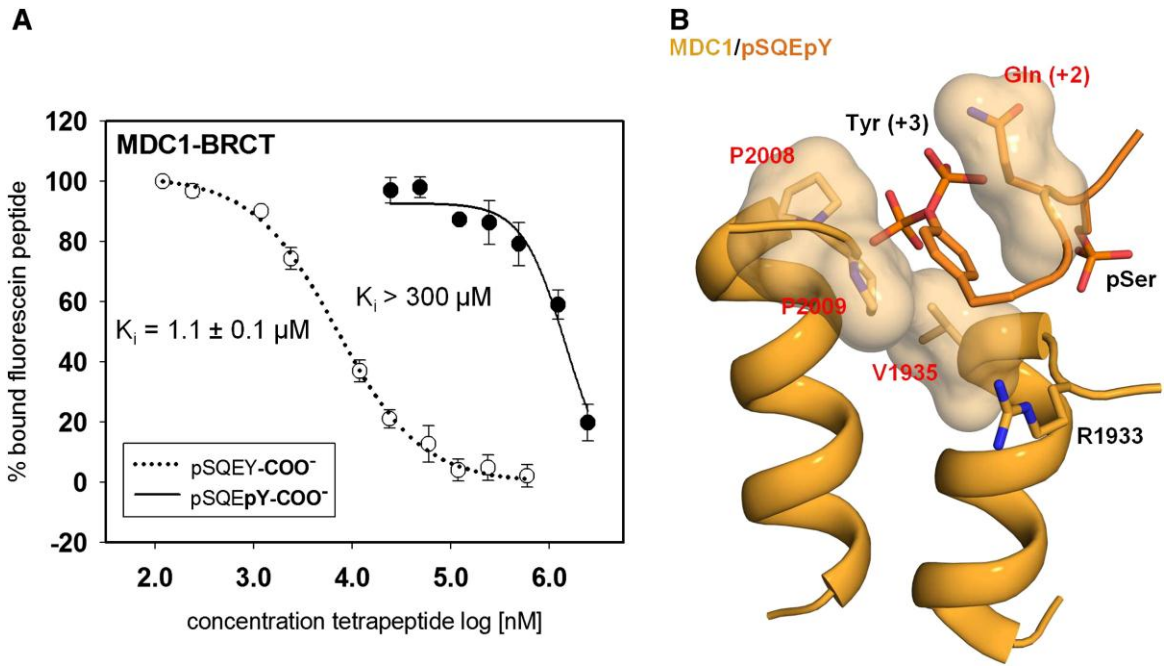


Figure 6 Binding data and model of γH2AX peptide containing phosphorylated Tyr142 bound to MDC1 BRCT.

A Competition assay of WT MDC1 BRCT with γH2AX tetrapeptide with and without phosphorylation of Tyr142. Dashed line indicates the □H2AX tetrapeptide (pSQEY-COO$^-$) and the solid line indicates the tetrapeptide with phosphorylated Tyr142 (pSQEpY-COO$^-$).

B Structural model showing the implications of a phosphorylated Tyr142 in the +3 binding pocket. Two rotomers of the phosphate group of pTyr142 are shown, with a surface representation of nearby residues that would potentially clash with the phosphate group labelled in red.

(Figure 6A), which agrees with the results of peptide pull down experiments [28]. To understand the structural basis for this discrimination, we modeled the pSer/pTyr tetrapeptide (pSQEpY-COO$^-$) into the MDC1 binding pocket (Figure 6B). The modeling suggests that the phosphate of

pTyr142 would clash with either the MDC1 specificity loop or the absolutely conserved Gln140 depending on the pTyr rotamer.

The MDC1 BRCT domain has also been shown to specifically bind the C-terminally phosphorylated tail of topoisomerase IIα, an interaction which specifically regulates the DNA decatenation cell cycle checkpoint [30]. The sequence of the topoisomerase IIα peptide (pSDEDDLF-COO^-) is unrelated to that of γH2AX. We used fluorescence polarization to test for interactions of this unusual target with the MDC1 BRCT but were unable to detect any binding activity (data not shown). The previously reported binding activity may therefore represent an interaction that requires additional contacts, perhaps involving other regions of the proteins or additional protein cofactors.

3.3 Discussion

Our results provide structural insight into the central role played by a conserved arginine (BRCA1 Arg1699, MDC1 Arg1933) in the recognition of the phospho-peptide +3 residue by tandem BRCT repeats. We show that the details of salt bridging interactions between this arginine and acidic residues that bridge the interface between the BRCT repeats modulate the orientation of this residue and its interactions with the phospho-peptide backbone.

Arg1699 likely plays an especially critical role in BRCA1 function as three distinct missense variants (R1699W, R1699L, R1699Q) have been

uncovered at this position through breast cancer screening programs (see http://research.nhgri.nih.gov/bic/ for a listing of BRCA1 variants). Pedigree analysis suggests both R1699W and R1699Q are associated with an increased cancer risk, and these variants are functionally defective in a BRCA1 transcription assay [31], and in their ability to bind a pSer-x-x-Phe peptide in a pull-down assay [15]. The R1699Q mutation is particularly interesting, since it should be able to participate in hydrogen bonding interactions with the peptide +3 backbone that could mimic the salt bridging observed in the wild type. The fact that this mutant shows peptide binding defects indicates that the ability to form salt bridges, either with the phospho-peptide or with Glu1836 and Asp1840, and not simply hydrogen bonding, is critical for its function. Interestingly, an E1836K variant has also been found through breast cancer screening. We suggest that this variant could perturb the orientation of Arg1699 through electrostatic repulsion with the Lys1836, and thereby impact phospho-peptide recognition.

Arg1933 in MDC1, like Arg1699 in BRCA1, is also necessary for recognition of the phospho-peptide main chain at the +3 position. In spite of the structural similarity of the BRCA1- and MDC1-phospho-peptide complexes, MDC1 shows a dramatic, 50-fold selectivity for peptides with a terminal main chain carboxylate at the +3 position compared to peptides with an amidated C-terminus. In contrast, BRCA1 shows a much reduced, ~7-fold selectivity for a main chain carboxylate at the peptide +3 position

(Figure 7A). Introduction of a T2067D substitution in MDC1 induces a significant increase in the affinity of a C-terminally amidated γH2AX phospho-peptide such that the binding preference for a C-terminal carboxylate vs an amide drops to 10-fold (Figure 7A). The high resolution structure of this variant in two different crystal environments reveals an overall more BRCA1-like conformation and additional flexibility near Arg1933. This relaxation of the MDC1 structure allows a shift in the position of the Arg1933 guanidium group that relieves the electrostatic repulsion with the terminating amide group at the phospho-peptide C-terminus that would otherwise occur.

Interestingly, the complementary substitution in BRCA1, D1840T, does not induce an MDC1-like enhancement in the preference for a main chain carboxylate at the phospho-peptide +3 position. The reason for this likely is explained by differences in the way in which the +3 tyrosine is positioned in the specificity pocket of BRCA1 and MDC1 (Figure 7B). In both proteins, key interactions with the +3 side chain are made by residues from the β1'-α1' loop. Met1775 in BRCA1 makes the closest contact to the phenylalanine side chain at +3, while Pro2009 contacts the +3 tyrosine side chain in MDC1. In MDC1, these interactions place the +3 Tyr carboxylate such that it contacts the Arg1933 guanidinium group in a nearly co-planar orientation. In the BRCA1 complex, the +3 Phe sits lower in its binding pocket and its main chain approaches the Arg1699 guanidinium in a more orthogonal manner, ~20° offset from the angle of

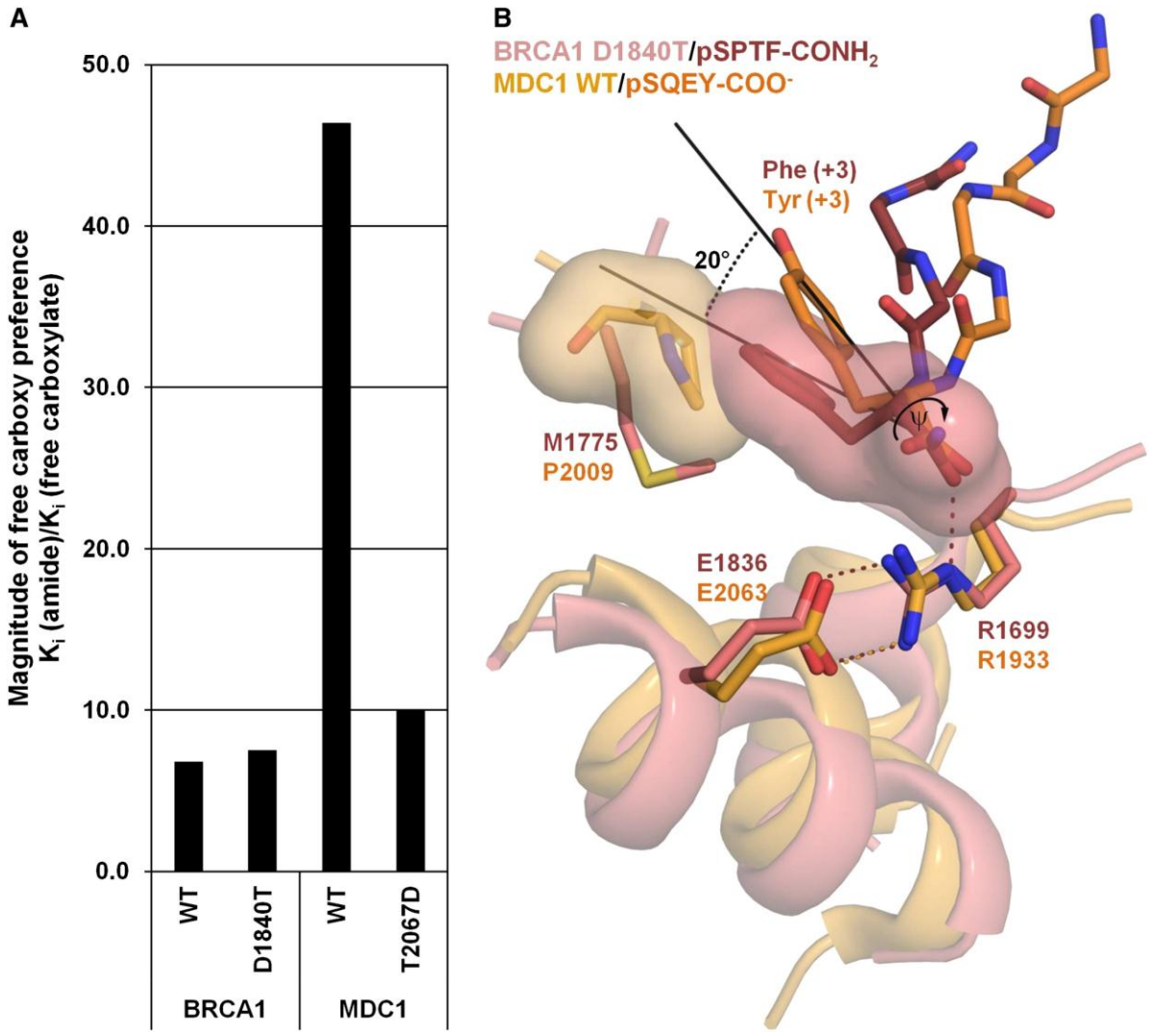


Figure 7 The threonine/aspartic acid mutation affects the specificity of MDC1 but not BRCA1.

A A graphical representation showing the magnitude of the preference of each of the recombinant BRCT domains for a free main chain carboxylate at the +3 position of the phospho-peptide.

B The alignment shows MDC1 BRCT in orange, and the BRCA1 BRCT D1840T in salmon. Solid black lines indicate the angular change in orientation of the side chains, and rotation about the ψ-angle is shown by a black arrow. Only the backbone is shown for peptide residues N-terminal to the Phe/Tyr. Surface representation is shown for Phe (+3) of

the BRCA1/pSPTF-$CONH_2$ structure and Pro2009 of the MDC1/H2AX structure. Salt bridging interactions of Arg1699/1933 are shown with dashed lines.

approach observed in MDC1. The more favourable hydrogen bonding geometry associated with the in-line approach of the C-terminal peptide carboxylate on the Arg1933 guanidinium group in MDC1 explains the enhanced affinity of MDC1 for the carboxylated chain terminus compared to BRCA1. In contrast, the BRCA1 orientation provides more flexibility, allowing for binding of extended peptide chains. In the BRCA1 orientation, a small rotation of the ψ torsion angle of the C-terminal residue is sufficient to shift the amidated C-terminus of the peptide away from the Arg1699 guanidinium group. However, in the MDC1 orientation a similar rotation is not sufficient to minimize the clash between the backbone and the Arg1933 guanidinium without a more significant reorganization of the interface. This likely explains MDC1's strong discrimination against binding peptide targets with extensions beyond the +3 position [24, 25].

The precise and rather rigid recognition of the peptide C-terminal carboxylate is central to the ability of MDC1 to mediate interactions with H2AX, and for the modulation of these interactions through phosphophorylation of H2AX at Tyr142. pTyr142 is an important H2AX mark in undamaged chromatin, where one of its roles is to likely block binding of MDC1 (Xiao 2009, Cook 2009). The anchoring of the Tyr142

through strong interactions with Arg1933, as well as van der Waals contacts with Pro2009, fixes the tyrosine on MDC1 such that its phosphorylation cannot be accommodated in the binding pocket.

The recognition of the C-terminally phosphorylated tail of H2A/H2AX by BRCT proteins in the DNA damage response is conserved throughout eukaryotes. In both *S. cerevisiae* and *S. pombe*, 53BP1 orthologs specifically bind the γH2A tail (...ApSQEL-COO^-) via their tandem BRCT repeats [32, 33]. The fact that the position of the H2A C-terminus is precisely conserved suggests that this may be a dominant specificity determinant in the yeast systems. However, the crystal structure of the BRCT domain of the *S. pombe* 53BP1 ortholog, Crb2, bound to the γH2A tail indicates that the γH2A C-terminal carboxylate does not contact Crb2 Arg616 (analogous to Arg1699 in BRCA1). It has been suggested that in this case chromatin recognition may not only involve BRCT – γH2A interactions, but may also rely on interactions of the adjacent tandem Tudor domain with lysine-methylated histones [32]. Intriguingly, the 53BP1 tandem BRCT domain also binds γH2AX *in vitro*, raising the possibility that 53BP1 may recognize nucleosomes presenting methyl-lysine and γH2AX histone tails during the DNA damage response.

Given the conformationally restrained yet structurally diverse nature of the BRCA1 and MDC1 peptide-binding grooves, it may be possible to develop inhibitors that are specific for a particular BRCT domain. Indeed, preliminary studies have identified both peptide and small molecule

inhibitors of the BRCA1 BRCT domain [27, 34, 35]. Here we have defined an important determinant of specificity between BRCA1 and MDC1 that could be utilized in the rational design of specific BRCT inhibitors. Such inhibitors could provide useful tools for the study of the role of specific BRCT domains, and potentially could provide new routes to modulate the DNA damage response in the treatment of cancer.

3.4 Materials and Methods

BRCT plasmids and purification. BRCA1 BRCT WT (residues 1646-1859), used in both FP and crystallization experiments, was expressed as an untagged recombinant protein in the plasmid pLM1-CD6. This plasmid was used as a template to clone BRCA1 BRCT D1840T using PCR mutagenesis, which was then ligated into the same plasmid. Both proteins were expressed in *E. coli* BL21 Gold (Invitrogen) for 15 hours at 22°C and purified as previously described [7]. Purified BRCA1 BRCT was buffer exchanged into crystallization buffer (5 mM Tris HCl pH 7.5, 400 mM NaCl) using spin concentration and quantified using a BCA assay.

MDC1 WT (1891-2089) was expressed as a Maltose Binding Protein (MBP) fusion protein using pKM596 (NEB). This construct was then used as a template to clone MDC1 BRCT T2067D using PCR mutagenesis and Gateway (Invitrogen). The proteins were over expressed in *E. coli* BL21 Gold at 20°C for 16 hours, and the cell pellets were then resuspended and sonicated. The lysate was cleared by ultra-centrifugation at 40000 × g for

one hour. The cleared lysate was further purified with an amylose resin affinity column (NEB). MDC1 used for crystallization was then cleaved using Actev protease (Invitrogen) and the MBP tag was removed using anion exchange chromatography. Purified MDC1 BRCT was buffer exchanged into crystallization buffer (10 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM DTT) and quantified using a BCA assay.

Fluorescence Polarization. Fluorescein-labelled peptides were obtained from the Alberta Peptide Institute (API) and are as follows: BACH1 peptide – Fluorescein-GGSRSTpSPTFNK-$CONH_2$; γH2AX peptide – Fluorescein-KKATQApSQEY-COO^-. Fluorescein peptide binding to BRCA1 and MDC1 BRCT domains was verified by titrating an increasing concentration of protein into a constant concentration of labelled peptide. Fluorescein fluorescence was excited at a wavelength of 485 nm and the emission was measured at 538 nm on a Perkin Elmer Envision plate reader. The change in polarization was graphed as a function of the log of the protein concentration, and the dissociation constant (K_d) was obtained from the resulting sigmoidal curve. The fluorescein labelled peptide was shown to have a similar K_d as previously reported values using both FP and direct binding methods (Figure 8) [24, 27]. Binding constants of the tetrapeptides were then determined using a competition assay, where the ability of the tetrapeptide to compete off the labelled peptide was measured. The concentration of protein used in the competition assay

was determined based on the K_d of the labelled peptide to its corresponding protein. The IC_{50} was obtained from the competition assay, and the K_i was then calculated using the Coleska-Wang equation [36]. For all assays, a concentration of 100 nM of labelled peptide was used in a reaction volume of 20 µL. Accurate concentrations of the tetrapeptides were determined by amino acid analysis.

Crystallization and Data Collection. The BRCA1 BRCT WT in complex with pSPTF-$CONH_2$ was crystallized using the sitting drop vapour diffusion method. BRCA1 BRCT WT was incubated on ice with an equimolar amount of pSPTF-$CONH_2$ for approximately three hours. Crystals were grown by mixing 2 µL of BRCA1 BRCT WT (10 mg mL^{-1} protein, 5 mM Tris-HCl pH 7.5, 400 mM NaCl) with 1 µL of well solution (0.1 M Tris-HCl pH 7.3, 0.4 M Li_2SO_4 and 10 mM $NiCl_2$). BRCA1 BRCT WT in complex with pSPTF-COO^- grew in identical conditions. Crystals of BRCA1 BRCT D1840T in complex with both tetrapeptides grew in very similar conditions to the WT protein, but preferred a pH between 8.3 and 8.5. Crystals were present in the drops after several days, and were large enough to loop after approximately two weeks. Crystals were cryoprotected in well solution supplemented with 26% glycerol and approximately 250 µM tetrapeptide, then looped and flash frozen using liquid nitrogen. Data were collected at beamline CMCF 08ID-1 at the Canadian Light Source (CLS).

The MDC1 BRCT T2067D in complex with pSQEY-$CONH_2$ was crystallized using the hanging drop vapour diffusion method. MDC1 BRCT T2067D was incubated on ice with an equimolar concentration of pSQEY-$CONH_2$ for approximately three hours. Crystals were grown by adding 1 µL of MDC1 BRCT T2067D (7.5 mg mL^{-1}, 10 mM Tris-HCl pH 7.5, 100 mM NaCl, 1 mM DTT) and 1 µL well solution (24% PEG 8000 and 0.1 M HEPES pH 8.0). Plates were present after approximately one week, and were looped using Mitegen MicroMesh. Crystals were cryoprotected in well solution supplemented with 26% glycerol and approximately 250 µM pSQEY-$CONH_2$. Data were collected at beamline 8.3.1 at the Advanced Light Source.

Model Building and Refinement. BRCA1 wild type and D1840T (1846-1859) complex diffraction data were indexed and scaled using HKL2000 [37]. Crystals grew in the space group $P6_122$ with one molecule in the asymmetric unit. The BRCA1 BRCT structures were solved with molecular replacement (Phaser) [38] using a BRCA1/BACH1 peptide complex (PDB ID: 1T15) as starting model. The first three residues (1846-1848) were not visible in the electron density. Electron density for the tetrapeptides was clear, and the ligand was modelled in to the phosphate binding pocket. The occupancies of the tetrapeptides in each model were adjusted to 70-80% to give reasonable B-factors. Structures were refined using TLS and restrained refinement in REFMAC [39-41]. The TLS

parameters were defined using chain A as the single TLS group. Electron densities of the +3 binding region shows a change in side chain positions as a result of the corresponding mutation previously described (D1840T) (Figure 9). Geometric restraints were tightened from program defaults to maintain correct geometry.

A strong peak located above the side chains of His1805 and His1673 of a symmetry mate was modelled as a nickel cation coordinating to the imidazole groups as observed in the previous unbound BRCA1 WT structure [7]. Although three other waters were expected to be coordinated to the nickel, it was not possible to satisfactorily model them into the BRCA1 WT complex structures. However, in the BRCA1 D1840T complex structures, two waters coordinated to the nickel were modelled with reasonable geometry. In addition to the nickel peak, another strong peak was located near the N-terminus of the tetrapeptide. Since both sulphate and chloride were present in the crystallization conditions, it was likely that an anion was bound to the ammonium cation at the N-terminus of the tetra peptide. Chloride was the only ion that resulted in satisfactory geometry, and was refined at the same occupancy as the tetrapeptide. Model building was done using the program Coot [42].

MDC1 T2067D (1891-2089) was indexed and scaled using HKL2000. Crystals grew in the $P2_12_12_1$ space group, and phases were solved by molecular replacement using a previous MDC1 BRCT structure (PDB ID: 2ADO). The electron density for pSQEY-$CONH_2$ was clear, and

the ligand was modelled in easily. The orientation of Glu2063 and Asp2067 in chain A was clear at 1.35 Å (Figure 5A). The electron density around chain B indicated an alternate conformation, in addition to the one in chain A, where Glu2067 interacted with Arg1933 while Asp2067 rotates away (figure 5B). However, given the two conformations present in chain B, the electron density was not quite as clear.

The structure was refined with TLS and restrained refinement using anisotropic B-factors in REFMAC. For TLS, the main chain of each molecule in the asymmetric unit and each tetrapeptide were defined as individual TLS groups. Residues 1891-2089 were modelled in to chain B, but the final six residues did not have clear electron density in chain A, and could not be modelled.

Accession Numbers. Atomic coordinates and structure factors have been deposited into the Protein Data bank (PDB) with the ID codes 3K05 for MDC1 BRCT T2067D/pSQEY-$CONH_2$, 3K0H for BRCA1 BRCT/pSPTF-$CONH_2$, 3K0K for BRCA1 BRCT/pSPTF-COO^-, 3K15 for BRCA1 BRCT D1840T/pSPTF-$CONH_2$ and 3K16 for BRCA1 BRCT D1840T/pSPTF-COO^-.

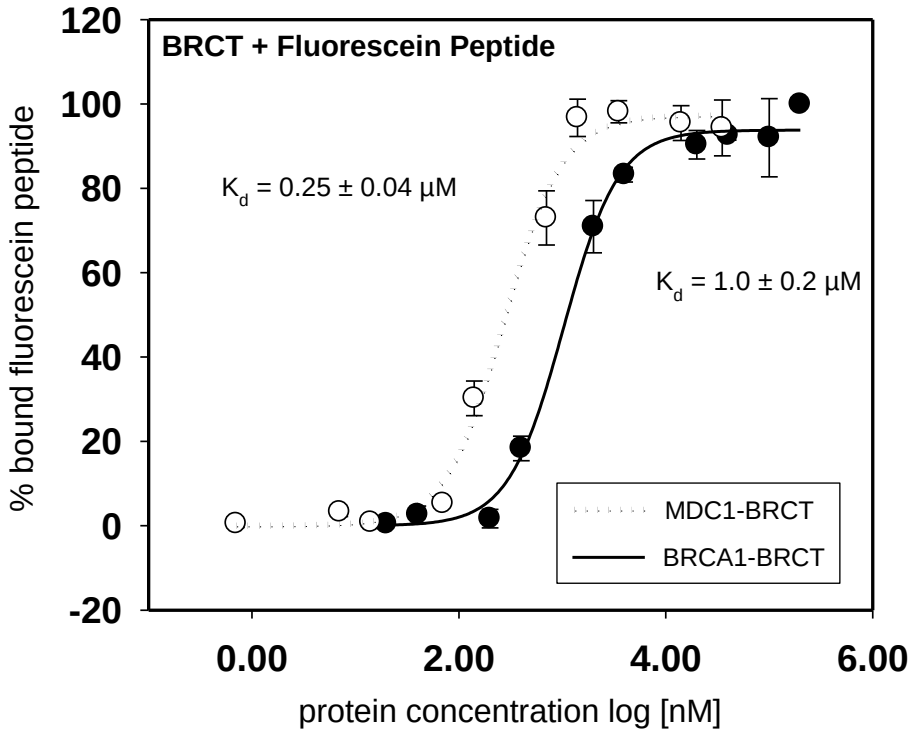
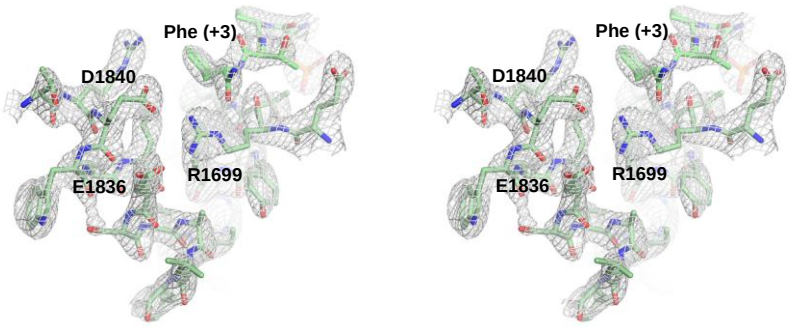


Figure 8. Direct binding of a fluorescein labelled peptide to the BRCT domain of BRCA1 and MDC1. Peptides sequences are as follows: BACH1: Fluorescein-GGSRST**pSPTF**NK-CONH2, H2AX: Fluorescein-AcP-KKATQAS**pSQEY**-COO^-. An increasing concentration of protein was titrated into a constant concentration of fluorescein labelled peptide, the K_d is defined as the inflection point of the resulting curve.

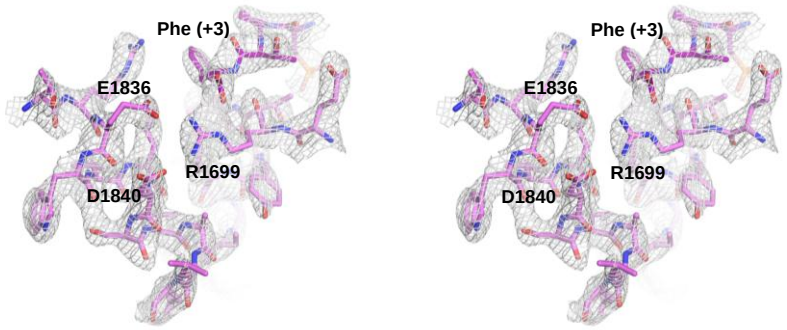
A

BRCA1/pSPTF-CONH$_2$



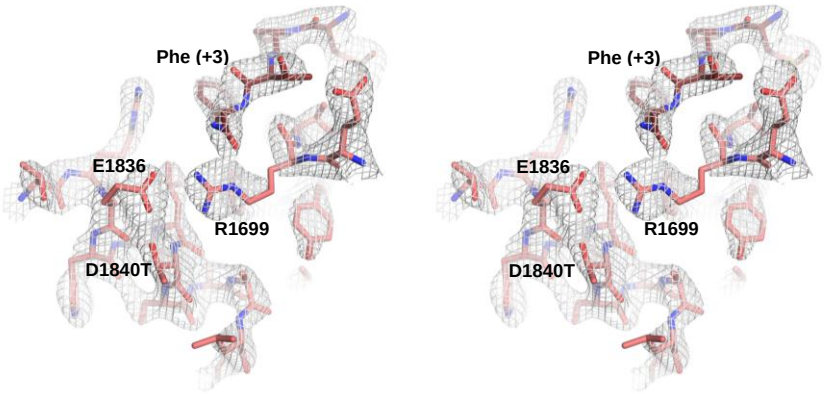
B

BRCA1/pSPTF-COO$^-$



C

BRCA1 D1840T/pSPTF-CONH$_2$



D

BRCA1 D1840T/pSPTF-COO$^-$

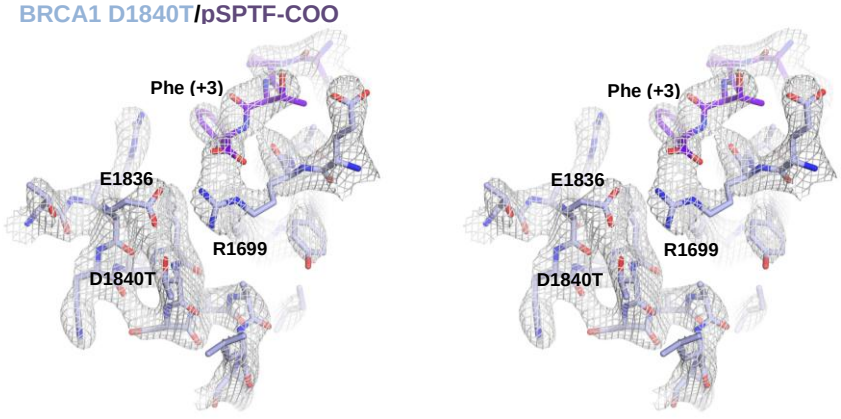


Figure 9 $2F_o$-F_c electron density corresponding to the relevant regions of BRCA1. Stick representation of the +3 binding region. The electron density map is shown as a gray mesh contoured at 2 σ.

A BRCA1 WT bound to pSPTF-$CONH_2$ (green).

B BRCA1 WT bound to pSPTF-COO^- (magenta).

C BRCA1 D1840T bound to pSPTF-$CONH_2$ (salmon).

D BRCA1 D1840T bound to pSPTF-COO^- (blue).

Table 1. Data Collection and refinement statistics

	WT BRCA1-BRCT		D1840T BRCA1-BRCT		T2067D MDC1-BRCT
	H_2N-pSPTF-**$CONH_2$**	H_2N-pSPTF-**COO-**	H_2N-pSPTF-**$CONH_2$**	H_2N-pSPTF-**COO-**	H_2N-pSQEY-**$CONH_2$**
Data Collection					
Space Group	$p6_122$	$p6_122$	$p6_122$	$p6_122$	$p2_12_12_1$
Cell Dimensions					
a,b,c (Å)	115.07,115.07,123.25	114.42, 114.42, 123.00	114.57, 114.57,123.19	114.34, 114.34, 123.87	55.64, 74.58, 103.41
α,β,γ (°)	90,90,120	90,90,120	90,90,120	90,90,120	90,90,90
Wavelength (Å)	0.97934	0.97934	0.97934	0.97934	1.11587
Resolution (Å)	50-2.70	50-2.70	50-2.80	50-3.0	50-1.33
$^2R_{sym}$	6.3 (51.3)*	6.1 (55.7)*	6.4 (57.1)*	8.3 (56.1)*	5.0 (48.9)*
I/σI	24.3 (2.68)*	49.7 (2.35)*	50.1 (2.55)*	37.8 (3.37)*	23.7 (2.7)*
Completeness (%)	99.1 (99.4)*	99.3 (94.0)*	99.5 (96.3)*	99.9 (99.4)*	98.3 (95.7)*
Redundancy	6.8 (5.6)*	19.2(7.7)*	19.5 (9.1)*	20.2 (14.1)*	3.8 (3.7)*
Refinement					
Resolution (Å)	29.44-2.70	35.8-2.70	37.9-2.80	38.7-3.0	37.9-1.33
No. Reflections	13,654	13,511	12,190	10,133	97,828
R_{work}/R_{free}	24.7/29.6	25.9/29.9	25.0/28.0	22.4/28.8	18.2/20.2
No. Atoms					
Protein	1731	1742	1730	1730	3212
Water	18	12	19	6	469
Nickel	1	1	1	1	0
Chloride	1	1	1	1	0
Overall B-factor ($Å^2$)	50	49	49	57	15
R.M.S. Deviations					
Bond Lengths	0.008	0.006	0.007	0.007	0.008
Bond Angles	1.11	1.03	1.02	1.06	1.30
Ramachandran					
Preferred (%)	90.5	86.7	86.7	85.2	99.7
Allowed (%)	9.5	13.3	13.3	14.8	0.3
Disallowed (%)	0	0	0	0	0

*Values in parentheses' are for highest resolution shell

$^2Rsym= \sum|(I_{hkl}) - <I>| / \sum(I_{hkl})$, where I_{hkl} is the integrated intensity of a given reflection.

3.5 References

1. Mohammad, D.H. and M.B. Yaffe, *14-3-3 proteins, FHA domains and BRCT domains in the DNA damage response.* DNA Repair (Amst), 2009.
2. Miki, Y., et al., *A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1.* Science, 1994. **266**(5182): p. 66-71.
3. Koonin, E.V., S.F. Altschul, and P. Bork, *BRCA1 protein products ... Functional motifs.* Nat Genet, 1996. **13**(3): p. 266-8.
4. Callebaut, I. and J.P. Mornon, *From BRCA1 to RAP1: a widespread BRCT module closely associated with DNA repair.* FEBS Lett, 1997. **400**(1): p. 25-30.
5. Futreal, P.A., et al., *BRCA1 mutations in primary breast and ovarian carcinomas.* Science, 1994. **266**(5182): p. 120-2.
6. Easton, D.F., et al., *A systematic genetic assessment of 1,433 sequence variants of unknown clinical significance in the BRCA1 and BRCA2 breast cancer-predisposition genes.* Am J Hum Genet, 2007. **81**(5): p. 873-83.
7. Williams, R.S., R. Green, and J.N. Glover, *Crystal structure of the BRCT repeat region from the breast cancer-associated protein BRCA1.* Nat Struct Biol, 2001. **8**(10): p. 838-42.
8. Manke, I.A., et al., *BRCT repeats as phosphopeptide-binding modules involved in protein targeting.* Science, 2003. **302**(5645): p. 636-9.
9. Yu, X., et al., *The BRCT domain is a phospho-protein binding domain.* Science, 2003. **302**(5645): p. 639-42.
10. Yu, X. and J. Chen, *DNA damage-induced cell cycle checkpoint control requires CtIP, a phosphorylation-dependent binding partner of BRCA1 C-terminal domains.* Mol Cell Biol, 2004. **24**(21): p. 9478-86.

11. Wang, B., et al., *Abraxas and RAP80 form a BRCA1 protein complex required for the DNA damage response.* Science, 2007. **316**(5828): p. 1194-8.
12. Kim, H., J. Huang, and J. Chen, *CCDC98 is a BRCA1-BRCT domain-binding protein involved in the DNA damage response.* Nat Struct Mol Biol, 2007. **14**(8): p. 710-5.
13. Liu, Z., J. Wu, and X. Yu, *CCDC98 targets BRCA1 to DNA damage sites.* Nat Struct Mol Biol, 2007. **14**(8): p. 716-20.
14. Glover, J.N., R.S. Williams, and M.S. Lee, *Interactions between BRCT repeats and phosphoproteins: tangled up in two.* Trends Biochem Sci, 2004. **29**(11): p. 579-85.
15. Williams, R.S., et al., *Structural basis of phosphopeptide recognition by the BRCT domain of BRCA1.* Nat Struct Mol Biol, 2004. **11**(6): p. 519-25.
16. Clapperton, J.A., et al., *Structure and mechanism of BRCA1 BRCT domain recognition of phosphorylated BACH1 with implications for cancer.* Nat Struct Mol Biol, 2004. **11**(6): p. 512-8.
17. Shiozaki, E.N., et al., *Structure of the BRCT repeats of BRCA1 bound to a BACH1 phosphopeptide: implications for signaling.* Mol Cell, 2004. **14**(3): p. 405-12.
18. Williams, R.S. and J.N. Glover, *Structural consequences of a cancer-causing BRCA1-BRCT missense mutation.* J Biol Chem, 2003. **278**(4): p. 2630-5.
19. Stewart, G.S., et al., *MDC1 is a mediator of the mammalian DNA damage checkpoint.* Nature, 2003. **421**(6926): p. 961-6.
20. Goldberg, M., et al., *MDC1 is required for the intra-S-phase DNA damage checkpoint.* Nature, 2003. **421**(6926): p. 952-6.
21. Lou, Z., et al., *MDC1 is coupled to activated CHK2 in mammalian DNA damage response pathways.* Nature, 2003. **421**(6926): p. 957-61.

22. Sedelnikova, O.A., et al., *Histone H2AX in DNA damage and repair.* Cancer Biol Ther, 2003. **2**(3): p. 233-5.
23. Stucki, M. and S.P. Jackson, *gammaH2AX and MDC1: anchoring the DNA-damage-response machinery to broken chromosomes.* DNA Repair (Amst), 2006. **5**(5): p. 534-43.
24. Stucki, M., et al., *MDC1 directly binds phosphorylated histone H2AX to regulate cellular responses to DNA double-strand breaks.* Cell, 2005. **123**(7): p. 1213-26.
25. Lee, M.S., et al., *Structure of the BRCT repeat domain of MDC1 and its specificity for the free COOH-terminal end of the gamma-H2AX histone tail.* J Biol Chem, 2005. **280**(37): p. 32053-6.
26. Yan, J. and A.M. Jetten, *RAP80 and RNF8, key players in the recruitment of repair proteins to DNA damage sites.* Cancer Lett, 2008. **271**(2): p. 179-90.
27. Lokesh, G.L., et al., *High-throughput fluorescence polarization assay to identify small molecule inhibitors of BRCT domains of breast cancer gene 1.* Anal Biochem, 2006. **352**(1): p. 135-41.
28. Xiao, A., et al., *WSTF regulates the H2A.X DNA damage response via a novel tyrosine kinase activity.* Nature, 2009. **457**(7225): p. 57-62.
29. Cook, P.J., et al., *Tyrosine dephosphorylation of H2AX modulates apoptosis and survival decisions.* Nature, 2009. **458**(7238): p. 591-6.
30. Luo, K., et al., *Topoisomerase IIalpha controls the decatenation checkpoint.* Nat Cell Biol, 2009. **11**(2): p. 204-10.
31. Vallon-Christersson, J., et al., *Functional analysis of BRCA1 C-terminal missense mutations identified in breast and ovarian cancer families.* Hum Mol Genet, 2001. **10**(4): p. 353-60.
32. Kilkenny, M.L., et al., *Structural and functional analysis of the Crb2-BRCT2 domain reveals distinct roles in checkpoint signaling and DNA damage repair.* Genes Dev, 2008. **22**(15): p. 2034-47.

33. Hammet, A., et al., *Rad9 BRCT domain interaction with phosphorylated H2AX regulates the G1 checkpoint in budding yeast.* EMBO Rep, 2007. **8**(9): p. 851-7.
34. Simeonov, A., et al., *Dual-fluorophore quantitative high-throughput screen for inhibitors of BRCT-phosphoprotein interaction.* Anal Biochem, 2008. **375**(1): p. 60-70.
35. Lokesh, G.L., et al., *Thermodynamics of phosphopeptide tethering to BRCT: the structural minima for inhibitor design.* J Am Chem Soc, 2007. **129**(35): p. 10658-9.
36. Nikolovska-Coleska, Z., et al., *Development and optimization of a binding assay for the XIAP BIR3 domain using fluorescence polarization.* Anal Biochem, 2004. **332**(2): p. 261-73.
37. Otwinowski, Z. and W. Minor, *Processing of X-ray Diffraction Data Collected in Oscillation Mode, Methods in Enzymology.* Methods Enzymol, 1997. **276**: p. 307-326.
38. McCoy, A.J., *Solving structures of protein complexes by molecular replacement with Phaser.* Acta Crystallogr D Biol Crystallogr, 2007. **63**(Pt 1): p. 32-41.
39. Murshudov, G.N., A.A. Vagin, and E.J. Dodson, *Refinement of macromolecular structures by the maximum-likelihood method.* Acta Crystallogr D Biol Crystallogr, 1997. **53**(Pt 3): p. 240-55.
40. Winn, M.D., M.N. Isupov, and G.N. Murshudov, *Use of TLS parameters to model anisotropic displacements in macromolecular refinement.* Acta Crystallogr D Biol Crystallogr, 2001. **57**(Pt 1): p. 122-33.
41. Winn, M.D., G.N. Murshudov, and M.Z. Papiz, *Macromolecular TLS refinement in REFMAC at moderate resolutions.* Methods Enzymol, 2003. **374**: p. 300-21.
42. Emsley, P. and K. Cowtan, *Coot: model-building tools for molecular graphics.* Acta Crystallogr D Biol Crystallogr, 2004. **60**(Pt 12 Pt 1): p. 2126-32.

Chapter 4

[1]BRCA1 as a therapeutic target: Preliminary inhibitor screening and design

[1]A version of part of this chapter has been published in Yuan, Z., [¥]Kumar, E.A., [¥]Campbell, S.J., [¥]Palermo, N.Y., Kizhake, S., Glover, J.N.M., Natarajan, A. **Exploiting the P-1 pocket of BRCT domains toward a structure guided inhibitor design**. *ACS Med Chem Lett*, 2011. 2(10): p. 764-767. [¥]Equal Contribution

4.1 Introduction

Many cancer therapies function by breaking the covalent bonds of DNA in tumour cells, resulting in the activation of apoptosis or, if the damage is extensive enough, necrosis. While dangerous for healthy tissues, the fact that these therapies have a more detrimental effect on the rapidly proliferating cancer cells makes them therapeutically effective. In addition, tissue specificity and methods of drug administration can help to decrease off target side effects. Nevertheless, the damage incurred to normal tissues can be very severe, with side effects ranging from hair loss and radiation burns to death.

Resistance to DNA damaging agents can occur when the cancer cell's DNA damage response (DDR) is up-regulated to combat the extensive damage caused by the therapy [1-4]. As a result, larger dosage and duration is required to have the same effects as previous treatments. Homologous recombination (HR) plays a key role in the DDR, repairing double strand breaks in a pathway that is dependent on the recruitment and proper function of the tumour suppressor BRCA1. After histone phosphorylation and ubiquitylation by numerous repair factors comprising the IR induced foci, BRCA1 is recruited through its BRCT domains through a phosphorylation dependant interaction with the adapter protein Abraxas [5-7] (Figure 1A). The BRCT domain of BRCA1 interacts with the phosphorylated tail of Abraxas, an interaction that is essential to the function of BRCA1 (Figure 1B). Mutations in this domain

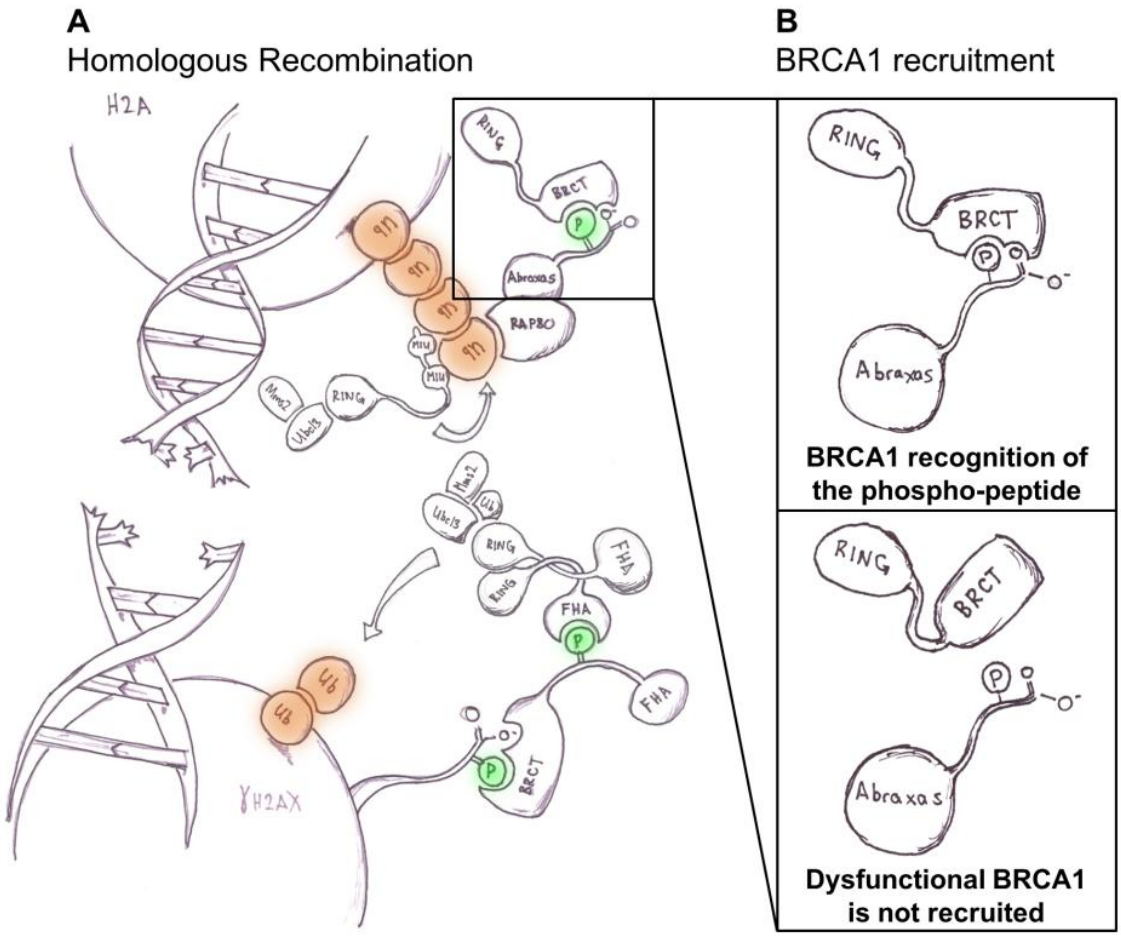


Figure 1 BRCA1 recruitment in Homologous Recombination. A Cartoon illustrating the initiation process of Homologous Recombination. In the event of a double strand bread, the histone H2AX is phosphorylated. MDC1 BRCT binds to the phosphorylated tail of γH2AX, and recruits RNF8. In concert with RNF168, RNF8 ubiquitylates adjacent histones, recruiting RAP80/Abraxas/BRCA1. B In order to be recruited, BRCA1 BRCT interacts with the phosphorylated tail of Abraxas. Mutations that disrupt the binding surface of BRCA1 prevent its recruitment.

of BRCA1 correlate to a roughly 85% lifetime risk of acquiring breast cancer, and accounts for about 50% of hereditary cancer cases [8].

Although BRCA1 is universally expressed in all tissues, dysfunction seems only to effect breast and ovarian tissues. Interestingly, women with BRCA1 mutation positive breast cancers respond very well to DNA damaging therapies, presumably due to an inherent dysfunction of the HR pathway [9, 10]. However, resistance to these agents can be conferred, and in some cases has been traced to a back mutation in the BRCA1 gene, which may actually restore function of the BRCA1 protein [11]. As a result, the HR pathway can work to counter the effects of the DNA damage therapies.

Given the essential role that BRCA1 plays in genome maintenance, the question arises about whether or not BRCA1 might be a good therapeutic target. While hereditary mutations in BRCA1 can result in tumourigenesis, the majority of breast cancers are sporadic, and a significant proportion retain functional BRCA1 [10, 12]. If the Protein Protein Interactions (PPIs) of BRCA1 could be inhibited in these tumours, the resulting effect could be to sensitize these tumours to current forms of cancer therapies. Recently, inhibitors of Poly-ADP-Ribose 1 (PARP1), a protein involved in Base Excision Repair (BER), have been shown to be very effective against BRCA1 mutation positive breast cancers [13]. This is likely because upon inhibition of PARP1, single strand breaks (SSBs) are not repaired effectively and a double strand break is formed when the replication machinery reaches the SSB. Usually, redundant repair pathways would be available to repair these DSBs in the case of BER

inhibition, but since HR is also dysfunctional due to the BRCA1 mutation, the cell activates apoptosis. We would like to mimic this synthetic lethality by inhibiting the phospho-peptide binding activity of the BRCT domain of BRCA1, and propose that the synthetic inhibition of BRCA1 in combination with a DNA breaking agent would have similar effects. This sensitization of tumour cells to DNA damaging agents may result in increased efficacy and shorter treatment times of toxic chemo and radiotherapies (Figure 2).

The BRCT domain has been revealed to be a very important structural module in the DNA damage response. As mentioned in the previous chapter, a typical BRCT domain is made up of a three or four stranded parallel β-sheet flanked on one side by two α-helices, and one on the other [14, 15]. Two of these domains are typically packed in a head to tail manner to form a phospho-peptide binding motif (Figure 3A) . The BRCA1 BRCT has been shown to interact with an array of different phosphorylated partners involved in the DDR (including BACH1 and CtIP), all of which contain the pSPTF motif that binds in a conserved manner in the phospho-peptide binding pocket [16-19]. In particular, the interaction between the BRCA1 BRCT domain and Abraxas has been very well characterized both biochemically and structurally. The BRCA1 BRCT binds to the extreme C-terminus of the Abraxas peptide, with the phospho-serine binding in a conserved pocket in the N-terminal BRCT repeat made up of S1655, G1656 and K1702. The Phe(+3) binds in the interface

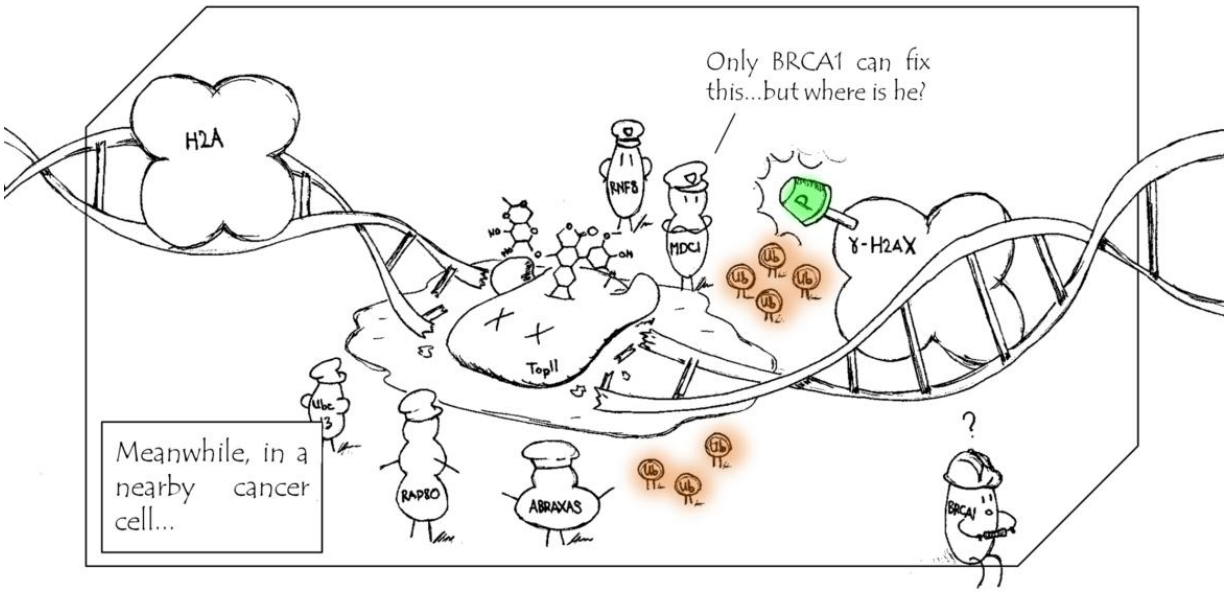


Figure 2 Consequences of BRCA1 inhibition. After double strand breaks caused by DNA damaging agents such as etoposide or radiation, inhibition of BRCA1 by a Chinese finger trap (or other effective inhibitors) would prevent the repair of the lesion.

between the two BRCT repeats via an essential interaction with M1775 [15, 19, 20]. In the case of Abraxas, the free carboxy C-terminal tail makes an important electrostatic interaction with the guanidinium group of Arg1699 (Figure 3B) [21].

In order to identify preliminary inhibitors of this BRCA1/pSPTF interaction, we employed high through-put screening (HTS), virtual screening, and peptidomimetics. Initially, screening began with the NCI diversity set (http://dtp.nci.nih.gov/index.html). Later, we obtained libraries of marine animal extracts (MAE) from the Andersen Lab (University of British Columbia) supplied through the Holmes lab (University of Alberta).

In addition, the Weinfeld lab and the Hall lab (University of Alberta) provided a library of polycyclic compounds previously used to identify inhibitors of the DNA repair protein Polynucleotide Kinase Phosphatase (PNKP) [22]. After an unsuccessful attempt at HTS, we ran virtual screening on a clean fragment library from ZINC, comprised of about 130,000 drug like, soluble compounds (http://zinc.docking.org/subsets/clean-fragments). Finally, in collaboration with the Natarajan lab (University of Nebraska), we used peptidomimetics and took advantage of the wealth of structural data available on the BRCA1/pSPTF interaction to design high affinity inhibitory peptides.

While HTS does not reveal any *bona fide* lead inhibitors of BRCA1, virtual screening identifies a single weak inhibitor of the BRCT domain. However, the chemical properties of the compound make it difficult to further characterize, and it is likely not a viable inhibitor. Using peptidomimetics, however, we have been more successful in the development of synthetic tight binding ligands of BRCA1.

4.2 Results

4.2a: The phospho-peptide binding pocket of BRCA1 BRCT provides a hotspot for inhibitor binding. Before attempting inhibitor screening of BRCA1, we addressed the concern of potential binding surfaces on the BRCT domain of BRCA1. Previous reports from our lab and others have calculated the binding affinities and solved the crystal structure of the

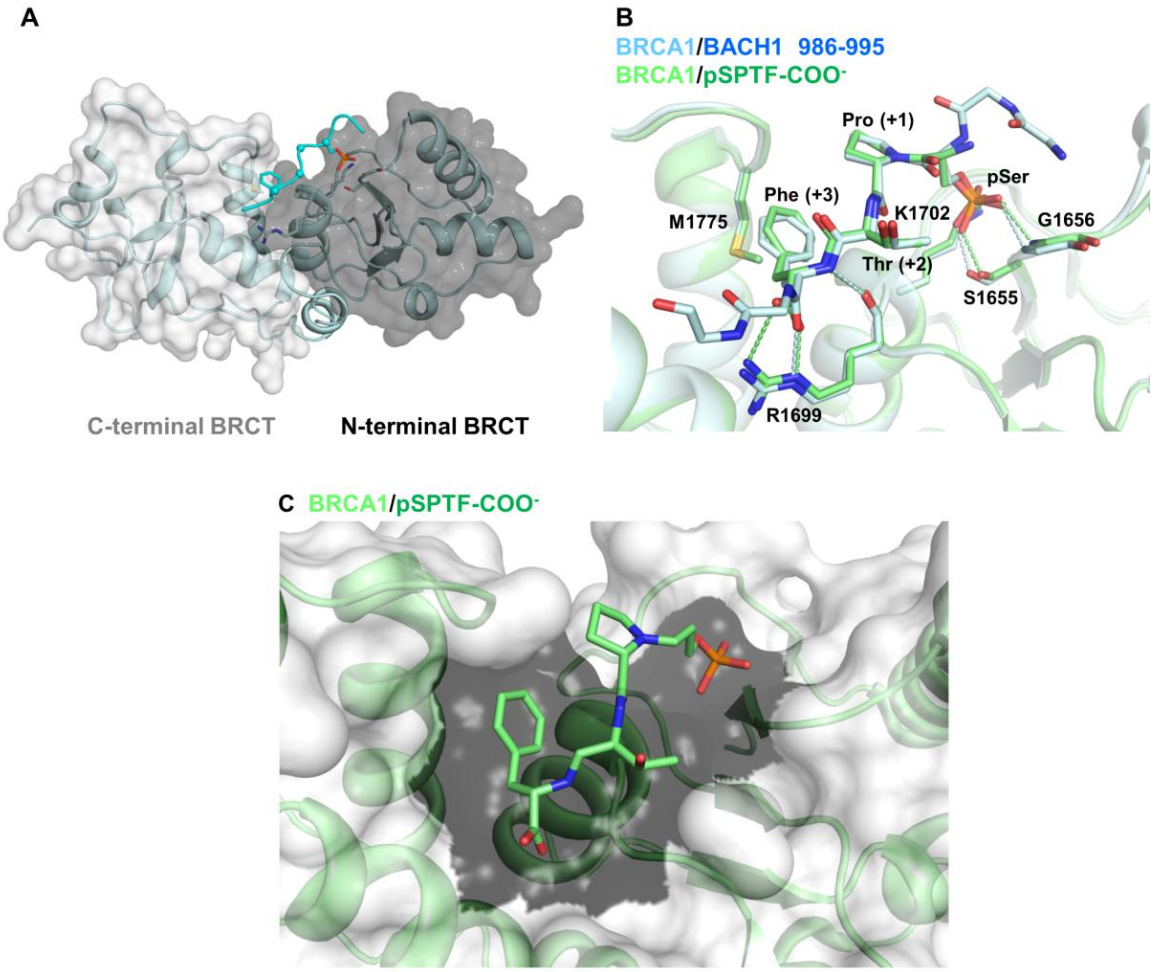


Figure 3 Phospho-peptide binding site provides a possible hotspot.

A Crystal structure of BRCA1 BRCT. The BRCT domains are indicated by surface representation – the C-terminal BRCT in white, and the N-terminal BCT in gray. The phospho-serine recognition peptide is shown in cyan.

B A crystal structure alignment of BRCA1 bound to its minimal recognition sequence (pSPTF) and bound to a longer consensus deca-peptide corresponding to the BACH1 sequence.

C Surface representation indicating the binding surface of the minimal tetrapeptide. Binding surface is shown in black, and has a total surface area of about 1000 $Å^2$

BRCA1 BRCT bound to an array of peptide targets. The minimal target sequence is comprised of the motif pSXXF with a free carboxy tail on the Phe(+3) (Figure 3B) [21, 23, 24]. The tetrapeptide binds identically as compared to longer peptides, indicating that the conformation of binding is fixed in the BRCT pocket [21]. In addition, previous reports have used ITC to show that the pSXXF motif is responsible for the majority of interaction with BRCA1 [23]. The heat of enthalpy is reduced very little in the tetrapeptide as compared to a longer decapeptide, indicating that the tetrapeptide alone makes the majority of interactions with the BRCA1 BRCT. Finally, the binding interface between the BRCT/tetrapeptide complex comprises approximately 1000 $Å^2$ on the BRCT domain (Figure 3C). This interface, including both a positive patch and an hydrophobic patch, is small but sufficient to provide a hotspot to screen for inhibitors of protein-protein interactions. Previous reports have suggested that a minimal binding interface of about 800-1100 $Å^2$ is required to serve as a viable target for inhibitors of PPIs [25, 26].

4.2b: Preliminary screening of the NCI diversity set reveals false positives. In order to identify inhibitors of BRCA1, the Natarajan Lab developed a Fluorescence Polarization (FP) assay to screen quickly for inhibitors of the BRCA1/Abraxas interaction. They measured the ability of a small molecule to displace a fluorescently labelled phospho-peptide from the BRCT domain of BRCA1. When bound to the BRCT, the labelled

phospho-peptide tumbles more slowly than when it is free in solution. As a result, the polarization of the emitted light from the FITC-labelled deca-peptide after excitation by polarized light will be maintained. If an inhibitor succeeds in competing the phospho-peptide out of the binding pocket, the polarization of the emitted light will be decreased. The Natarajan lab used this technique to identify lead inhibitors of BRCA1 in the NCI diversity set, a consolidated library of compounds shown to have favourable pharmacological properties, and the results were published in 2006. We were able to obtain many of these lead compounds, and reproduced the IC_{50} values measured for the inhibitors with BRCA1 as well as obtain IC_{50} values with the BRCT domain of MDC1 for comparison (Figure 4A).

In order to characterize these inhibitors further, we attempted to solve the crystal structure of the lead compounds from the Natarajan Lab bound to the BRCT domain of BRCA1. After many attempts at co-crystallization and inhibitor soaking, we were unable to locate the electron density of an inhibitor bound to BRCA1 (Figure 4A). To test the validity of the inhibitors identified in Lokesh et al. (2006), we developed a robust, non-fluorescence based assay. We tested the ability of a biotin-labelled, phosphorylated Abraxas peptide (pSPTF) immobilized on streptavidin agarose beads to pulldown the BRCA1 BRCT both with and without the presence of inhibitor. The BRCA1 BRCT (1646-1859) was expressed by in vitro transcription translation in the presence of 35-S labelled methionine, and resolved on a 15% SDS-PAGE gel. The gel was imaged

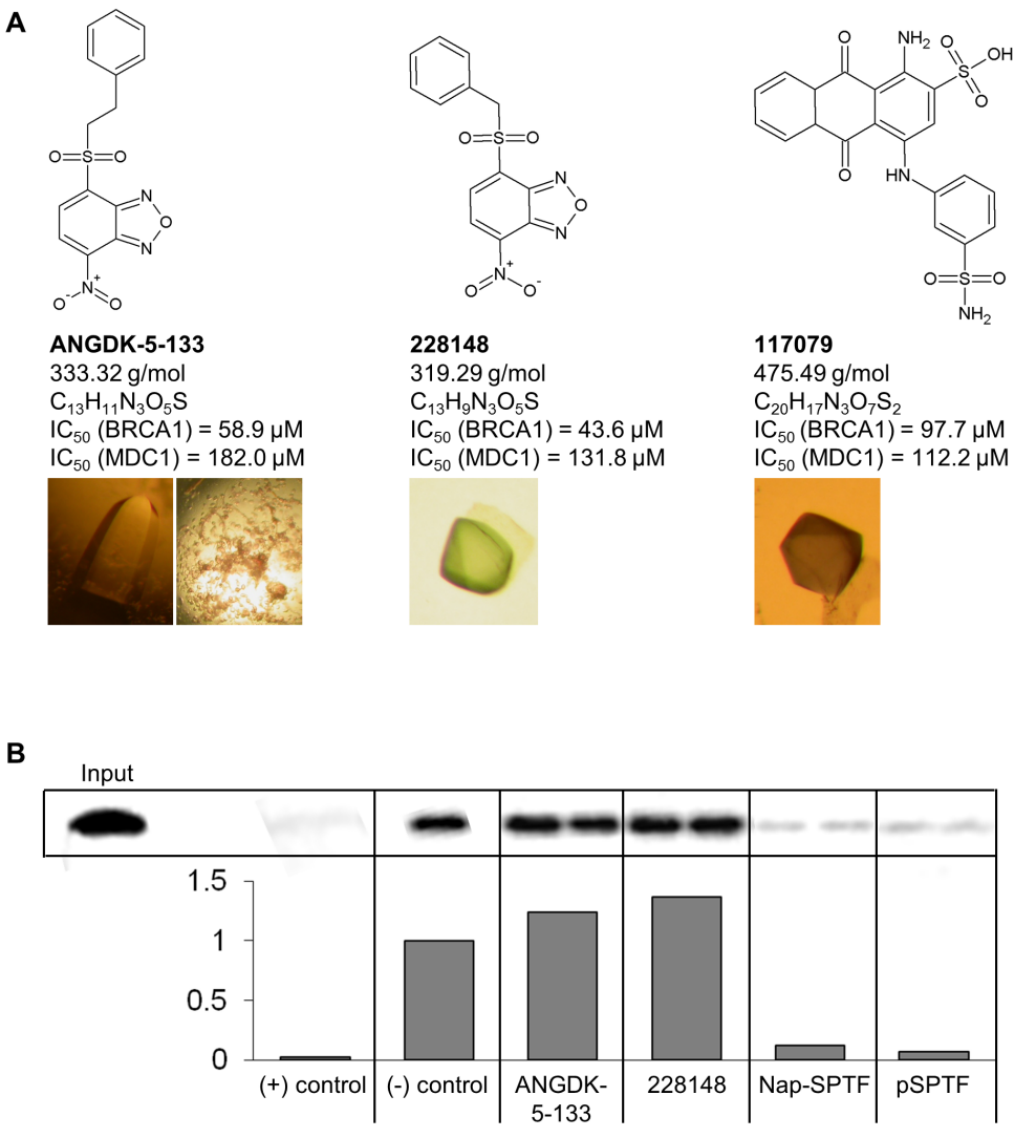


Figure 4 Initial leads from the NCI Diversity Set are not verified by a streptavidin pulldown.

A Initial leads identified from the NCI diversity set. IC_{50} values determined by fluorescence polarization competition assays. Crystals of BRCA1 BRCT/inhibitor complexes were promising for inhibitor binding, but inhibitor was not visible upon structure solution.

B Streptavidin pulldown implicating the initial lead compounds as false positives. The (+) control (SPTF biotin peptide) is analogous to inhibitor binding, and the negative control (pSPTF biotin peptide) indicates peptide binding. Known peptide inhibitors of BRCA1 (Nap-SPTF and pSPTF)

were used as additional (+) controls. Band intensities were normalized and represented as a percentage of binding as compared to the (-) control.

by exposure to a phosphor screen for 48 hours. In the absence of inhibitor ((-) control), BRCA1 binds strongly to the beads, and is eluted by adding SDS-PAGE buffer and boiling for two minutes and resolved on an SDS-PAGE gel (Figure 4B). When the Abraxas peptide is dephosphorylated, no binding is detected ((+) control). None of the lead compounds were effective at inhibiting the pulldown of BRCA1. Two known peptide inhibitors were used a positive controls to verify the assay including the minimal tetrapeptide (pSPTF), and a modified tetrapeptide with a napthalated amino acid N-terminal to the phospho-serine (Nap-pSPTF). Both of these peptides effectively inhibit the pulldown of BRCA1 with the beads. These results indicate that in the NCI diversity set, there are in fact no lead inhibitors of BRCA1.

The FP assay still serves a purpose for initial HTS and has been proven previously to be accurate, sensitive and time effective. The exact cause of false positives in the FP assay is currently unknown. We were, however, able to show that the lead inhibitors from the Natarajan Lab are active with no substantial change in the total fluorescence, and that the majority of them do not alone cause a decrease in polarization (data not shown). Additional hits identified by FP HTS can be removed as false positives based on total fluorescence measurements. A non-specific

interaction between the inhibitor and the fluorescein-labelled peptide would certainly cause the observed artifact, but is difficult to observe given the nature of the assay. As a result, leads from the FP assays are now verified using the streptavidin pulldown.

4.2c: HTS using additional libraries fails to yield lead compounds. Marine animal extracts (MAEs) are an excellent source of large, chemically diverse bioactive toxins with potential activity towards therapeutic targets. Many of these toxins have evolved to target aspects of phosphorylation pathways, given its importance in cell signalling. MAEs may contain an array of small molecule toxins, making each extract a potential library in itself. Specifically, marine animal extracts have previously been used to identify inhibitors of Ser-Thr protein phosphatases PP-1, PP-2A and calcineurin [27]. We use them here in an attempt to identify inhibitors of the phospho-peptide binding pocket of BRCA1. Marine animal extracts were obtained as evaporated solids on 96 well plates, and were re-suspended in methanol. The methanol suspension was typically at a 10 mg/ml concentration, and was added directly to the high throughput FP assay containing BRCA1 bound to fluorescent Abraxas peptide. MAEs that caused a decrease in polarization approximately two standard deviations below the mean were taken as hits (Figure 5A). Approximately 5% of the extracts in the initial screen show competition activity in the assay. These extracts were further tested in the

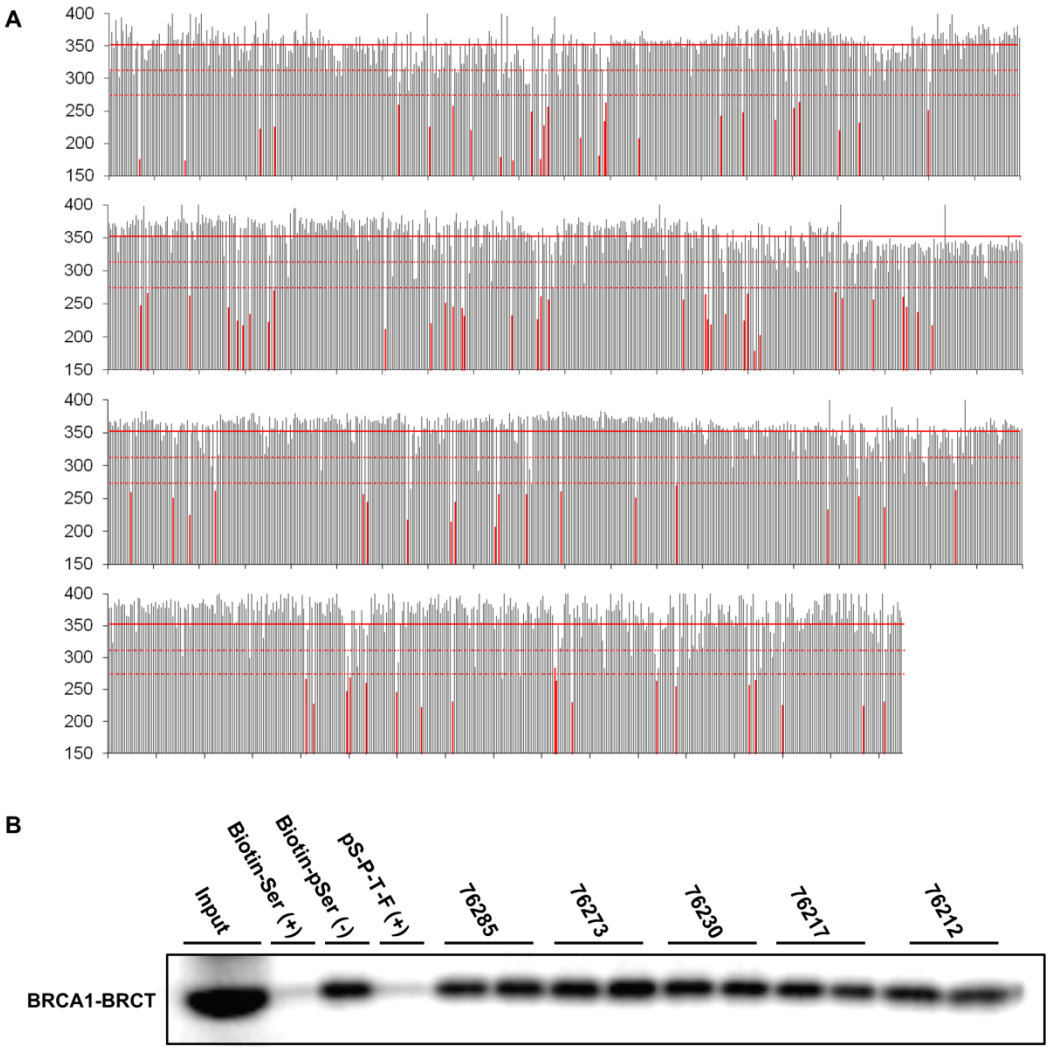


Figure 5 Marine Animal Extract (MAE) screening fails to yield any lead compounds.

A High through-put FP screen of MAEs. A hit is identified as any sample where the mP reading is 2 standard deviations below the mean (indicated by dashed lines). Each of the leads is marked as a red column.

B A sample gel of a streptavidin pulldown on leads identified using the FP assay. No hits from the initial FP assay were active in the pulldown, as all were unable to inhibit the pulldown of GST-BRCA1.

streptavidin pulldown, and do not show any inhibition whatsoever (Figure 5B). Approximately 2,000 marine animal extracts were screened, and no identifiable inhibitors were found.

In addition to the MAE, a small library of polysubstituted piperidines was supplied by the Hall lab (University of Alberta), which was used by the Weinfeld library to identify inhibitors of PNKP [22, 28]. The initial FP assay identified two possible inhibitors of BRCA1, but were not verified by the streptavidin pulldown (data not shown). Approximately 400 compounds were screened in this library, and no inhibitor of BRCA1 was found.

4.2d: In silico screening reveals a single low affinity inhibitor of BRCA1. In the wake of unsuccessful HTS for inhibitors of BRCA1, we attempted a paradigm shift that would use an entirely different approach at inhibitor screening. To screen larger libraries faster, we performed virtual screening using the docking program Autodock 4. After defining a search area on a macromolecule, Autodock uses an automated approach to place the inhibitor coordinates in a manner that minimizes the binding energy, parameterized by known protein/inhibitor complexes. We obtained a large, clean fragment library from ZINC comprised of about 130,000 compounds. The clean fragment library is described as drug fragments that are relatively soluble in water. This is important given that biochemical assays are required to test the positive hits produced by

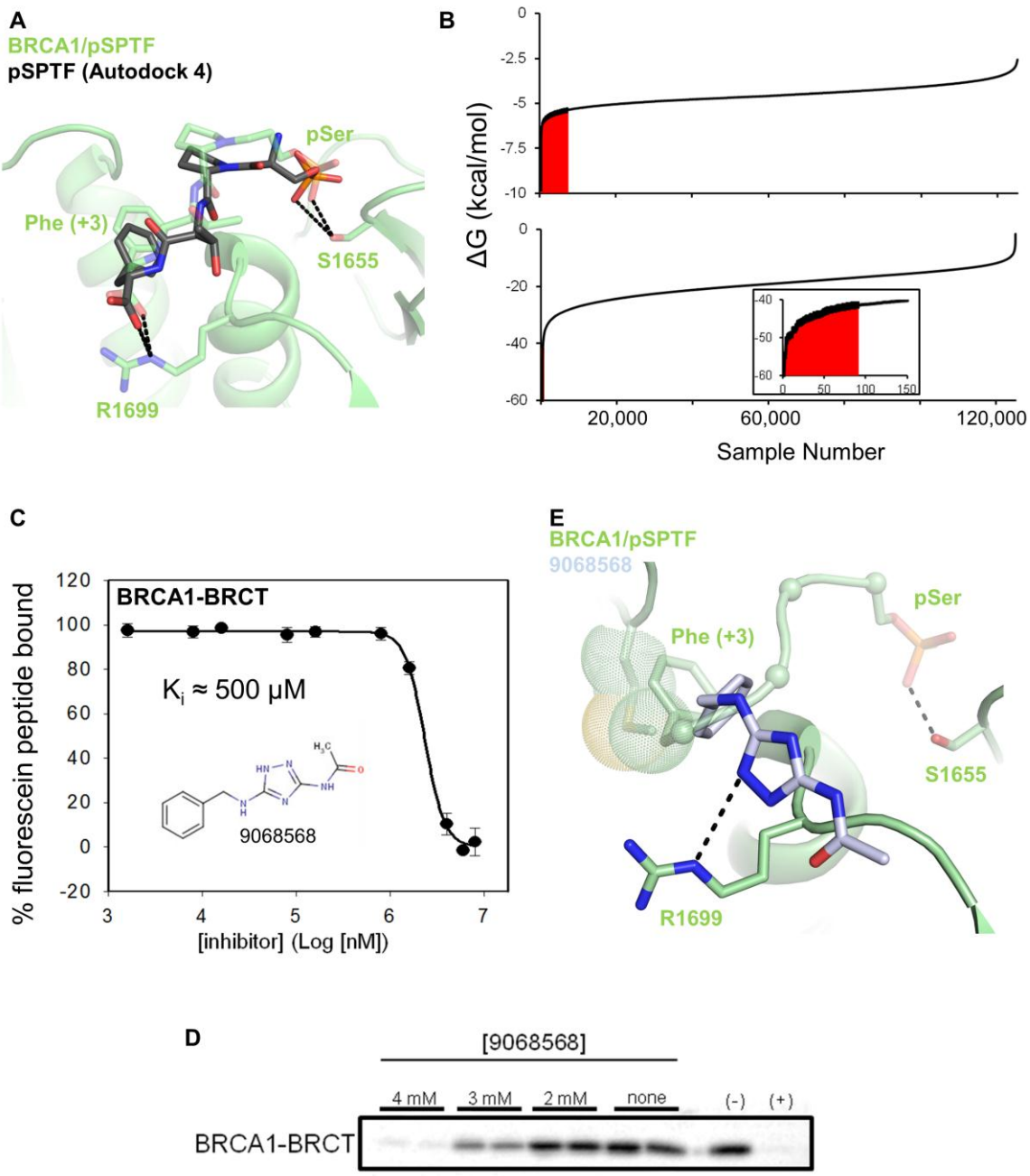


Figure 6 Virtual screening reveals a single low affinity inhibitor.

A Alignment of BRCA1 bound to its minimal tetrapeptide (PDB ID: 3K0K) in green and the placement of the tetrapeptide by Autodock 4 in black. Important residues involved in binding are labelled, and hydrogen bonds are shown as dashes.

B Graphical representation of the binding energy distribution of small molecules after Autodock 4 (above) and BEAR (below). Area under the curve is shaded up to the positive control tetrapeptide (pSPTF).
C FP competition assay titrating the lead compound 9068568 into a constant concentration of BRCA1 BRCT.
D Steptavidin pulldown showing the ability of the lead compound 9068568 inhibit the precipitation of GST-BRCA1.
E Computational model of the possible mode of binding of 9068568 predicted by Autodock 4. R1699 is shown as sticks, and possible hydrogen bonds are shown as dashed lines. The phenyl ring of 9068568 binds in a similar position as Phe (+3).

Autodock. In addition, the phospho-peptide binding pocket of BRCA1 is comprised of two well-defined pockets, each of which may bind a different type of fragment. The selection of this library made it possible to identify multiple fragments that interact with different sites on BRCA1 BRCT, with the possibility of designing a tight binding inhibitor by combining the individual components. Although fragments tend to bind with a much weaker affinity than larger compounds, our assays can detect binding into the mM range, provided the inhibitor stays in solution at these very high concentrations.

In order to follow the accuracy of inhibitor placement and binding constant prediction of Autodock, the tetrapeptide pSPTF-COO^- was used as a positive control. The search box encompassed roughly half of the BRCA1 BRCT, centred on the phospho-peptide binding site. The majority

of the top hits were placed either in the phospho-serine binding pocket or the phenylalanine binding pocket. While the positive control tetrapeptide is placed correctly in the phospho-peptide binding pocket, the predicted energy of binding is quite high, placing the tetrapeptide in the top 7%, or within the top 10,000 of the ~130,000 compounds tested (Figure 6A). In order to increase the accuracy of the predictions, we used Binding Estimation After Refinement, or BEAR, a MD based procedure that re-scores the Autodock 4 results following a structural minimization, a molecular dynamics (MD) simulation, and a second structural minimization [29]. In the first step, the Autodock 4 hits are structurally minimized, accounting for flexibility both in the inhibitor and the protein. Then, MD is used to simulate the likelihood of this interaction over a psec timescale by giving flexibility to the inhibitor. False positives will not likely stay in a tight, bound conformation, and will be eliminated following a second structural minimization. Following the BEAR simulation, the tetrapeptide was re-ranked into approximately the top 0.1%, or top 100, suggesting that the BEAR procedure is able to better rank the binding affinities of the compound library, compared to Autodock alone (Figure 6B). From this, we identified the top 50 hits, and where possible, obtained a sample of the inhibitor. One compound, 9068568 (Chembridge), showed inhibition activity in both the FP assay and the streptavidin pulldown (Figure 6C). In the FP assay, inhibition was detected at a K_i of approximately 500 μM, with similar activity in the streptavidin pulldown. Autodock 4 as well as MD

places this inhibitor in the Phe(+3) binding pocket, with potential hydrogen bonding to R1699 in BRCA1 (Figure 6D). Attempts to solve the crystal structure of the inhibitor bound to BRCA1 were unsuccessful.

4.2e: Chemical properties of the weak inhibitor make further characterization difficult. Since the structural characterization of 9068568 was difficult, we attempted to further characterize the binding of the inhibitor to BRCA1 using biochemical means. To define the binding site of 9068568, we obtained several peptide binding deficient mutants of BRCA1 including G1656D, T1700A, R1699Q and E1836K (Figure 7A). Each of these mutants show decreased, but substantial binding to the BRCA1 recognition phopsho-peptide [30-32]. We predicted that if any of these residues were essential for the interaction of 9068568, we would cease to see inhibition. However, inhibition was detected in all of the mutants, suggesting that either these residues are not involved, or that the interaction of 9068568 with BRCA1 is non-specific (Figure 7A). The possibility of a non-specific effect was supported by the fact that by the time inhibition is observed, the inhibitor is severely insoluble in aqueous solution. Upon spinning down of the FP assay to remove precipitate, inhibition is no longer observed (data not shown).

In order to characterize 9068568 binding, we attempted to order chemically similar compounds that are more soluble in our biochemical assays. Using SciFinder, we identified 5529374, 5128664 and 4022738

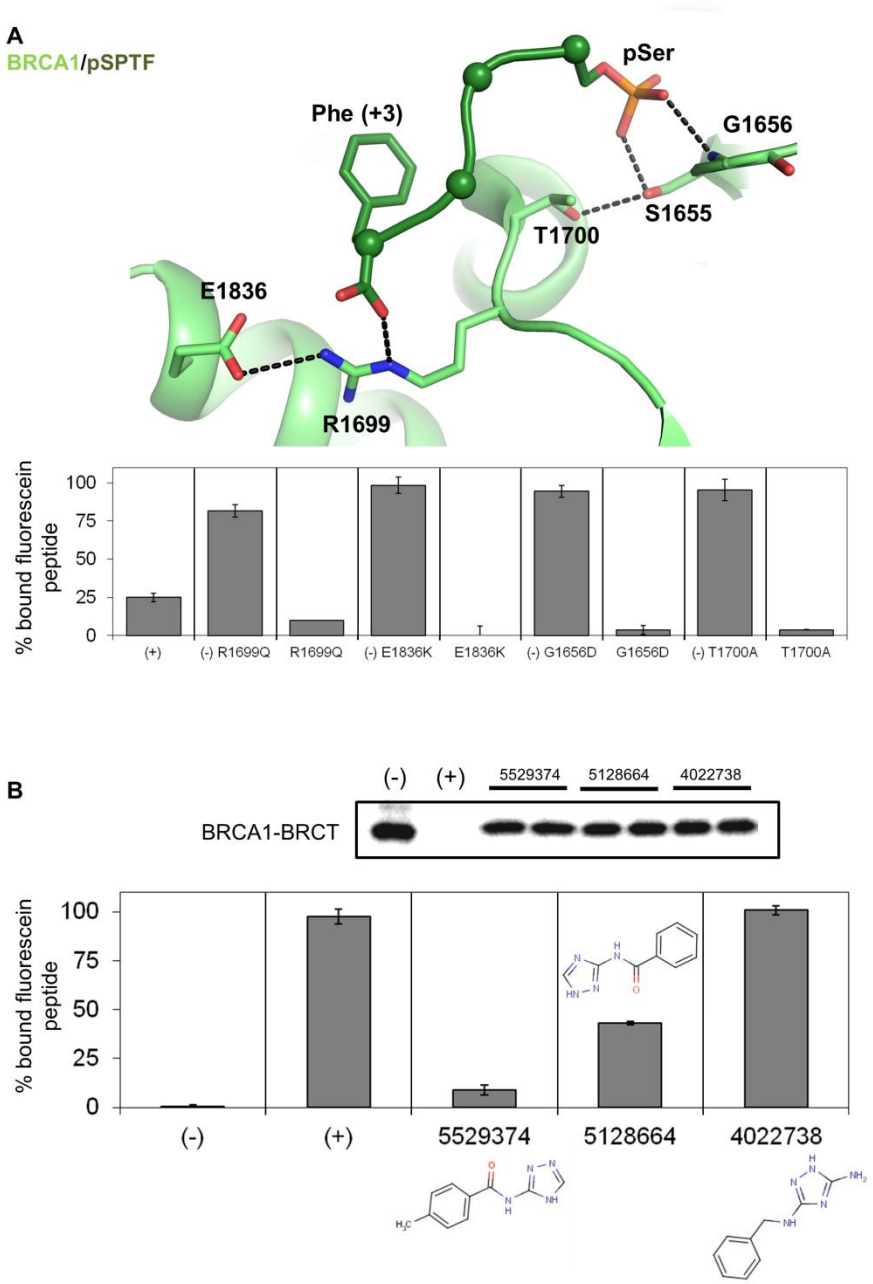


Figure 7 9068568 is likely interacting with the BRCA1 BRCT non-specifically.

A Crystal structure of BRCA1 BRCT bound to the minimal tetrapeptide (pSPTF) (PBD ID: 3K0K) showing the positions of residues mutated to map binding of 9068568. Mutation of residues in both the phospho-serine

binding pocket and the Phe (+3) binding pocket fail to abrogate binding of 9068568, suggesting non-specific binding.

B Streptavidin pulldown and FP assay of compounds homologous to 9068568. Two compounds show activity in the FP assay, but fail to inhibit precipitation of GST-BRCA1 BRCA1 in the pulldown.

which are all very similar in structure with 9068568, and have the functional groups important for binding based on the Autodock 4 prediction (Figure 7B). These compounds are significantly more soluble in our reaction conditions compared to 9068568, but only 5529374 shows significant activity in the FP assay. Unfortunately, none of the homologous compounds showed any activity in the strepravidin pulldown indicating that perhaps the interaction shown by 9068568 works via a non-specific precipitation effect.

4.2f: Peptidomimetics reveals additional binding pocket in BRCA1 that can be exploited to increase peptide affinity [33]. After a significant (non-automated) attempt at HTS, we concluded that perhaps manual random screening was not the most ideal way to identify inhibitors of BRCA1. In light of this, our collaborators set out to design tight binding peptides based on the original tetrapeptide. They had previously confirmed our findings that the tetrapeptide with a free carboxy C-terminus binds significantly tighter than the same peptide with an amidated terminus, based on an electrostatic interaction with R1699 [24]. They showed in

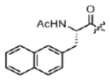
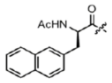
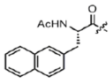
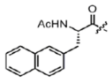
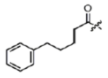
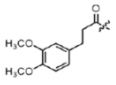
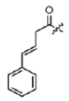
addition to this that if the N-terminal phospho-serine is acetylated removing the positive charge on the phospho-serine, binding affinity is increased further.

In 2006, Lokesh et al. predicted a possible additional binding pocket adjacent to the phospho-serine binding pocket formed by Pro1659, Val1654, Leu1657 and Phe1662 (PVLF) (Figure 8A) [23]. They suggested that if a peptide could be designed with a hydrophobic side chain that could interact with that pocket, affinity could potentially be increased significantly. As a result, they designed many different peptides based on the pSPTF motif, with various residues N-terminal to the phospho-serine. We used both FP and the streptavidin pulldown to characterize the binding of these modified peptides.

We measured the ability of the modified tetrapeptide to compete off a BRCT bound TMR-long probe labelled decapeptide using FP and used the Coleska-Wang equation to calculate the K_i based on the K_d of the labelled peptide (Table 1). Peptides 1 and 2 are variations of the Abraxas peptide, verifying the importance of the free carboxylate group on the Phe(+3) C-terminus. Peptides 3 and 4 show that a valine can be placed in the +2 position with no significant effect on binding (pSPVF).

Peptides 5-8 have an unnatural amino acid (Nap = napthyl side chain) at the P-1 position and were generated to explore the PVLF patch. Peptides 5 and 6 examine the effect of the stereocenter (R and S) at the P-1 position and peptides 7 and 8 determine if the C-terminus of the

Table 1

	R-Group	Peptide	K_i (μM)
1		pSPTF-$CONH_2$	1.85 ± 0.06
2		pSPTF-COOH	0.27 ± 0.01
3		pSPVF-$CONH_2$	1.96 ± 0.09
4		pSPVF-COOH	0.20 ± 0.01
5		pSPTF-$CONH_2$	1.24 ± 0.04
6		pSPTF-$CONH_2$	19.44 ± 1.91
7		pSPVF-$CONH_2$	0.54 ± 0.03
8		pSPVF-COOH	0.13 ± 0.03
9		pSPVF-COOH	0.10 ± 0.02
10		pSPVF-COOH	0.50 ± 0.05
11		pSPVF-COOH	12.05 ± 0.29
12		pSPVF-COOH	15.67 ± 0.36
13		pSPVF-COOH	4.04 ± 0.07
14		pSPVF-COOH	0.08 ± 0.01
15		pSPVF-COOH	0.04 ± 0.01

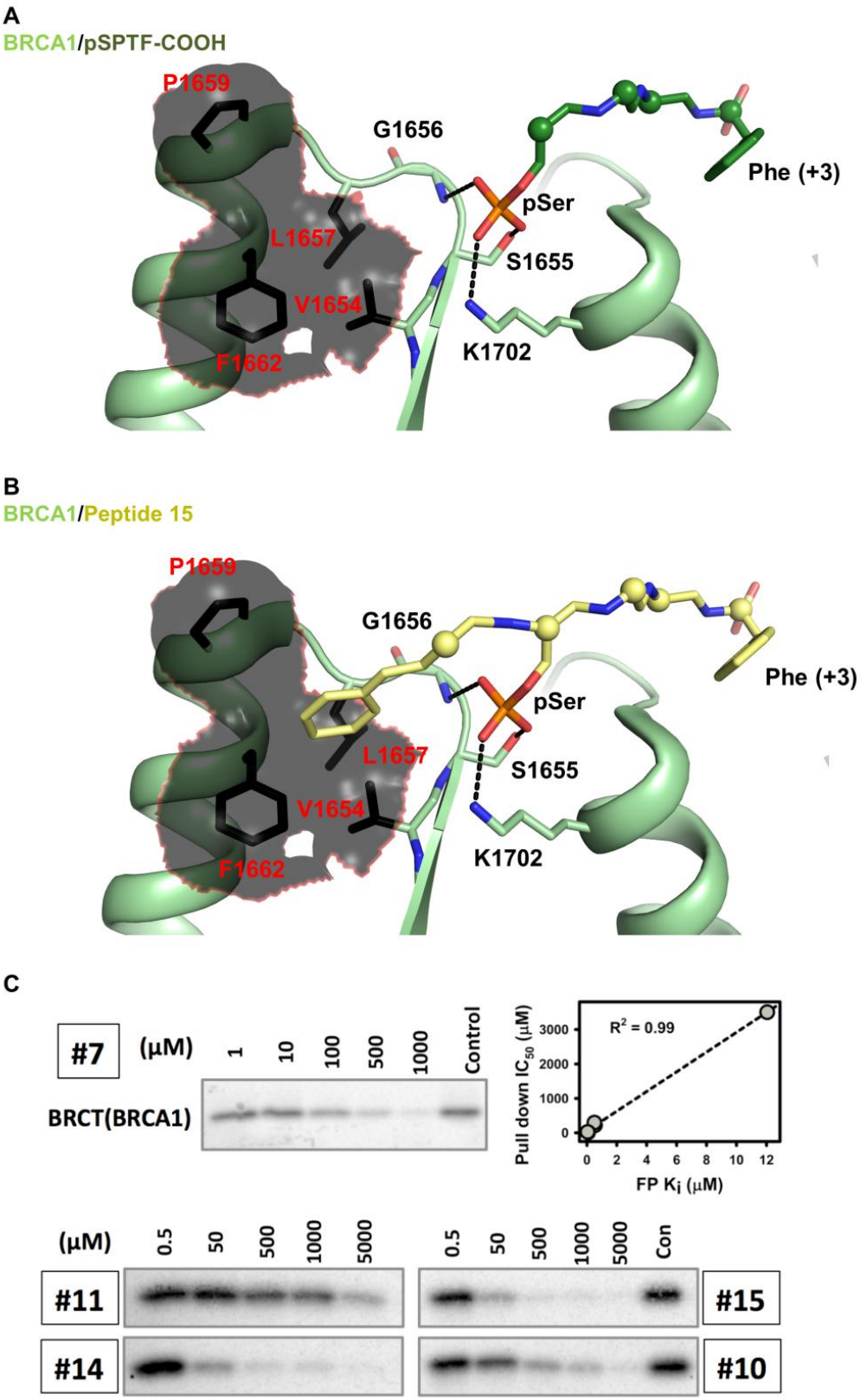

A
BRCA1/pSPTF-COOH
P1659
G1656
Phe (+3)
pSer
L1657
S1655
V1654
K1702
F1662
B
BRCA1/Peptide 15
P1659
G1656
Phe (+3)
pSer
L1657
S1655
V1654
K1702
F1662
C
#7
(μM)
1
10
100
500
1000
Control
BRCT(BRCA1)
Pull down IC50 (μM)
3000
2000
1000
0
R2 = 0.99
0
2
4
6
8
10
12
FP Ki (μM)
(μM)
0.5
50
500
1000
5000
0.5
50
500
1000
5000
Con
#11
#15
#14
#10

Figure 8 Peptidomimetics is used to design higher affinity inhibitors of the BRCA1 BRCT.

A Crystal structure of BRCA1 BRCT bound to the minimal tetrapeptide (PDB ID: 3K0K) showing the position of the PVLF hydrophobic pocket adjacent to the phospho-serine binding pocket. PVLF pocket is shown as black surface representation. The tetrapeptide is shown as green sticks, and BRCA1 is shown as a green cartoon.

B Placement of modified peptide 15, showing the possible position of the phenyl ring upon binding of the peptide. Modified peptide is shown as yellow sticks, and BRCA1 is shown as green cartoon.

C Streptavidin pulldowns of high affinity modified peptides. An increasing concentration of modified peptide was titrated into a constant concentration of GST-BRCA1 BRCT.

peptide has an effect on the occupancy of the PVLF pocket. Conjugating the Nap amino acid to the N-terminus of pSPVF peptides results an increase in the K_i values (μM) when compared to the parent peptides (5 vs 1, 7 vs 3 and 8 vs 4). Reversal of the stereochemistry from S to R by incorporating the unnatural Nap amino acid results in ~15-fold loss of activity (6 vs 5). These results suggest that incorporating a hydrophobic group at the P-1 position to occupy the PVLF cluster is a viable approach to design BRCA1 BRCT inhibitor peptides with increased affinity.

Next we generated a set of seven peptide mimics (9-15) with varying functionalities at the P-1 position to determine the optimal linker length and flexibility, as well as the optimal size of the hydrophobic group. We observed a decrease in activity with increase in the chain lengths

through peptide mimics 9-11, with mimic 9 showing a 2-fold higher activity than the parent peptide 4. We next explored the size of the PVLF hydrophobic cluster using a napthyl ring and a 3,4-dimethoxysubstituted phenyl ring instead of the phenyl ring in mimics 12 and 13. With these mimics we observed ~40 and ~150-fold loss of activity compared to 9 suggesting that the phenyl ring has the optimal size to occupy the PVLF patch. Although we observed ~2.5-fold loss of activity in 10 compared to the parent peptide 4, we next explored if linker flexibility as opposed to linker length is responsible for this loss of activity. Mimics 14 and 15 were generated with different levels of linker flexibility compared to 4. Indeed, we observed increased activity (> 5-fold) in 14 and 15 with restricted linkers compared to 10 (Figure 8C). This systematic approach resulted in the identification of mimic 15 with a K_i value of 40 nM for the BRCA1 BRCT. Using the crystal structure of BRCA1 BRCT bound to pSPTF (PDC ID: 3K0K), we modelled the modified peptide 15 into the binding pocket of BRCA1, using the tetrapeptide crystal structure as a fixed anchor using pymol (Figure 8B). In the model, the phenyl ring on the -1 residue can orient itself to sit the PVLF hydrophobic pocket.

To determine the relative affinities of the peptide mimics for BRCA1 BRCT in using an alternate assay, we conducted a competitive streptavidin pulldown on five of the modified tetrapeptides (7, 10, 11, 14 and 15) (Figure 8C). A mixture of phospho-BACH1 peptide immobilized on to streptavidin-agarose beads and peptide mimics were incubated with the

BRCA1 BRCT. The relative band intensity was quantified and the pulldown IC_{50} values were determined. A high correlation (R^2 = 0.99) between the FP derived K_i values and pull-down IC_{50} values is observed. The most potent peptide mimics 14 and 15 identified from the FP assay were also the most potent in the in vitro transcription translation assay. We attempted to crystallize the peptide/BRCT complex, but no crystals of diffraction quality were obtained (data not shown).

4.3 Discussion

Given the results described above, without automated methods, it is unlikely that HTS screening is the most appropriate method to indentify inhibitors of BRCA1. While it may still be applicable in certain circumstances, our data suggest that the phospho-peptide binding pocket is not as ideal a target for randomized screening as initially thought. Small molecule inhibitors of enzymes typically recognize a single, very well defined structural motif. Conversely, the binding surface of a protein-protein interaction is often much more expansive, with multiple recognition pockets. While hotspots of protein-protein interactions show a region that could be inhibited by randomized screening, that fact still remains that this binding surface is much more specific, and not as well defined as that of an enzyme. In BRCA1 in particular, the three binding pockets that can be exploited for synthetic inhibition are relatively shallow, and inhibitors would require very specific structural characteristics in order to interact with all

three. In addition, known peptide ligands of BRCA1 absolutely require both the phospho-serine and the (+3) Phe in order to show any binding whatsoever [34]. It is therefore likely that the peptide binding pocket of BRCA1 BRCT is one such surface, where the required interactions are too structurally defined to identify an inhibitor using randomized, non-automated methods.

Alternatively, analysis of the BRCT-BRCA1/pSPTF complex structures suggested the presence of a hydrophobic cluster (PVLF) at the P-1 position of pSXXF binding site. A systematic structure guided iterative synthesis and screening of peptide mimics resulted in the identification of BRCA1 BRCT peptide inhibitors with higher affinity than the minimal tetrapeptide pSPTF. A constrained 3-carbon linker with a phenyl ring conjugated to the N-terminus of pSPVF peptide results in mimics with low nanomolar K_i values. These results indicate that the most successful approach to high affinity inhibitors of BRCA1 is likely peptidomimetics. By exploiting the PVLF binding pocket on BRCA1, we are able to make tight binding, specific artificial inhibitors of BRCA1. In addition, since BRCT domains are already quite specific to their corresponding consensus sequences, using the minimal tetrapeptide is likely a good approach to ensure specificity, and prevent binding to other BRCT domains.

Currently, much progress has been made to the idea of peptides as drug molecules. Various types of administration and peptide designs have made it feasible to administer peptides therapeutically. Peptoids are

variations on protein structure where the side chain is moved to the backbone nitrogen instead of the Cα, and are highly proteolytically stable [35]. Cyclic peptides have been shown to increase serum stability, in addition to improving affinities for targets, and could potentially be of use here [36]. In order to increase cellular uptake of peptide mimetics, poly-arginine tags can be added to the peptide, although this has not been attempted extensively [37, 38]. In light of these recent advances, it is possible that the high affinity peptide inhibitors described here could be modified appropriately. Eventually, they could provide leads towards the development of viable candidates for therapeutics, making the inhibition of BRCA1 function as a potential cancer treatment a real possibility.

4.4 Materials and methods

Protein cloning, expression and purification. BRCA1 and MDC1 were cloned expressed and purified as previously described in Chapter 3. BRCA1 mutations were made as described in chapter 3.

Fluorescence Polarization Assay. Purified BRCA1 was diluted to a concentration where the fluorescein labelled peptide (FITC-GGSRST**pS**PTFNK-$CONH_2$) was still 100% bound (Chapter 3, Figure 8). HTS was accomplished on a 384 well plate, and 5 µL of potential inhibitor has added to each well, bringing the total volume up to 20 µL. Plates were excited at 495 nm at emission was detected at 521 nm in a Perkin-Elmer

Envision plate reader. Potential hits were identified as having a mP value two standard deviations below the mean of all samples on each plate. Streptavidin pulldowns were used to verify hits obtained from FP.

FP for modified peptides was performed by our collaborators using a TMR-long probe instead of FITC. Titrations of increasing amount of modified peptide were added to a constant concentration of protein and labelled decapeptide and the change in polarization was graphed vs. the log of the peptide concentration. The K_i was calculated using the Coleska-Wang equation. A more detailed method can be found in Yuan et al. (2011).

Streptavidin pulldown. Streptavidin beads were washed with BRCA1 binding buffer (200 mM NaCl, 50 mM Tris pH 7.5, 0.1% (v/v) Triton X-100, 1 mM EDTA and 1 mM DTT). Biotin labelled BACH1 peptides containing the BRCA1 BRCT consensus sequence (biotin-SRST**pS**PTFNK-$CONH_2$ and biotin-SRST**S**PTFNK-$CONH_2$) were immobilized on the streptavidin beads by a 30 minute incubation at a 10:1 molar excess of peptide. The beads were then washed thoroughly to remove excess peptide. Hits obtained from the FP assay were added to 10 µL bed volume peptide/streptavidin beads and 1 µL of BRCA1 BRCT obtained from a S-35 supplemented transcription translation reaction. Beads were washed thoroughly, and BRCA1 BRCT binding was detected by resolving the band

on a 15% SDS-PAGE gel and exposure to a phosphor-screen. When appropriate, bands were quantified using ImageQuant.

Crystallization and data collection. Crystals of BRCA BRCT (1649-1859) were obtained using vapour diffusion, by mixing 2 µL of 10 mg/mL BRCA1 BRCT in storage buffer (400 mM NaCl and 50 mM Tris pH 7.5) with 1 µL of well solution (10 mM Tris pH 8.1, 0.5 M $LiSO_4$ and 0.5 mM $NiCl_2$). Crystals grew after 2 weeks, and were soaked gradually with a molar excess of inhibitor. Crystals were then backsoaked for 10 minutes in cryoprotectant lacking inhibitor, and frozen in liquid nitrogen. Data was collected at beamline CMCF 08ID-1 at the Canadian light source, and the structure was solved using molecular replacement (Phaser) with 1JNX as the search model. No significant positive density in the F_o-F_c was identified that could correspond to a bound inhibitor.

Computational screening. The Clean Fragment library was downloaded from ZINC Free Library (UCSF) and Autodock 4 was used initially to dock inhibitors into the phospho-peptide binding pocket of BRCA1 (see *Using Autodock 4 for Virtual Screening,* Online Tutorial, Lindstrom et al, The Scripps Research Institute, 2008). BEAR was set up as described by Degliesposi et al. (2010). Hits within the top 100 after BEAR were ordered depending on their price and availability. FP and streptavidin pulldowns were conducted on all hits, with only 9068568 showing activity in both.

Peptide synthesis. The peptides were synthesized using standard Fmoc-chemistry by our collaborators or by the Tufts peptides core as described in Yuan et al. (2011). The peptides were synthesized on Rink Amide NovaGelTM resin (0.25 mmol) (EMD) using N-α-Fmoc-protected amino acids (EMD) or unnatural N-α-Fmoc protected amino acids (3B Scientific Corporation or Fischer Scientific) and TBTU-HOBt coupling chemistry on a Focus XC synthesizer (Aapptec). A more detailed method can be found in Yuan et al. (2011).

1.5 References

1. Kastan, M.B. and J. Bartek, *Cell-cycle checkpoints and cancer.* Nature, 2004. **432**(7015): p. 316-23.
2. Roossink, F., et al., *The role of ATM and 53BP1 as predictive markers in cervical cancer.* Int J Cancer, 2012.
3. Squatrito, M. and E.C. Holland, *DNA damage response and growth factor signaling pathways in gliomagenesis and therapeutic resistance.* Cancer Res, 2011. **71**(18): p. 5945-9.
4. Al-Ejeh, F., et al., *Harnessing the complexity of DNA-damage response pathways to improve cancer treatment outcomes.* Oncogene, 2010. **29**(46): p. 6085-98.
5. Kim, H., J. Huang, and J. Chen, *CCDC98 is a BRCA1-BRCT domain-binding protein involved in the DNA damage response.* Nat Struct Mol Biol, 2007. **14**(8): p. 710-5.
6. Liu, Z., J. Wu, and X. Yu, *CCDC98 targets BRCA1 to DNA damage sites.* Nat Struct Mol Biol, 2007. **14**(8): p. 716-20.
7. Wang, B., et al., *Abraxas and RAP80 form a BRCA1 protein complex required for the DNA damage response.* Science, 2007. **316**(5828): p. 1194-8.

8. Narod, S.A., *Modifiers of risk of hereditary breast and ovarian cancer.* Nat Rev Cancer, 2002. **2**(2): p. 113-23.
9. Martin, S.A., C.J. Lord, and A. Ashworth, *DNA repair deficiency as a therapeutic target in cancer.* Curr Opin Genet Dev, 2008. **18**(1): p. 80-6.
10. Turner, N.C., et al., *BRCA1 dysfunction in sporadic basal-like breast cancer.* Oncogene, 2007. **26**(14): p. 2126-32.
11. Swisher, E.M., et al., *Secondary BRCA1 mutations in BRCA1-mutated ovarian carcinomas with platinum resistance.* Cancer Res, 2008. **68**(8): p. 2581-6.
12. Rakha, E.A., et al., *Expression of BRCA1 protein in breast cancer and its prognostic significance.* Hum Pathol, 2008. **39**(6): p. 857-65.
13. Farmer, H., et al., *Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy.* Nature, 2005. **434**(7035): p. 917-21.
14. Zhang, X., et al., *Structure of an XRCC1 BRCT domain: a new protein-protein interaction module.* EMBO J, 1998. **17**(21): p. 6404-11.
15. Williams, R.S., R. Green, and J.N. Glover, *Crystal structure of the BRCT repeat region from the breast cancer-associated protein BRCA1.* Nat Struct Biol, 2001. **8**(10): p. 838-42.
16. Cantor, S.B., et al., *BACH1, a novel helicase-like protein, interacts directly with BRCA1 and contributes to its DNA repair function.* Cell, 2001. **105**(1): p. 149-60.
17. Li, S., et al., *Binding of CtIP to the BRCT repeats of BRCA1 involved in the transcription regulation of p21 is disrupted upon DNA damage.* J Biol Chem, 1999. **274**(16): p. 11334-8.
18. Varma, A.K., et al., *Structural basis for cell cycle checkpoint control by the BRCA1-CtIP complex.* Biochemistry, 2005. **44**(33): p. 10941-6.

19. Clapperton, J.A., et al., *Structure and mechanism of BRCA1 BRCT domain recognition of phosphorylated BACH1 with implications for cancer.* Nat Struct Mol Biol, 2004. **11**(6): p. 512-8.
20. Shiozaki, E.N., et al., *Structure of the BRCT repeats of BRCA1 bound to a BACH1 phosphopeptide: implications for signaling.* Mol Cell, 2004. **14**(3): p. 405-12.
21. Campbell, S.J., R.A. Edwards, and J.N. Glover, *Comparison of the structures and peptide binding specificities of the BRCT domains of MDC1 and BRCA1.* Structure, 2010. **18**(2): p. 167-76.
22. Freschauf, G.K., et al., *Mechanism of action of an imidopiperidine inhibitor of human polynucleotide kinase/phosphatase.* J Biol Chem, 2010. **285**(4): p. 2351-60.
23. Lokesh, G.L., et al., *Thermodynamics of phosphopeptide tethering to BRCT: the structural minima for inhibitor design.* J Am Chem Soc, 2007. **129**(35): p. 10658-9.
24. Joseph, P.R., et al., *Structural characterization of BRCT-tetrapeptide binding interactions.* Biochem Biophys Res Commun, 2010. **393**(2): p. 207-10.
25. Wilson, A.J., *Inhibition of protein-protein interactions using designed molecules.* Chem Soc Rev, 2009. **38**(12): p. 3289-300.
26. Clackson, T. and J.A. Wells, *A hot spot of binding energy in a hormone-receptor interface.* Science, 1995. **267**(5196): p. 383-6.
27. Carr, G., et al., *Protein phosphatase inhibitors isolated from Spongia irregularis collected in Papua New Guinea.* J Nat Prod, 2007. **70**(11): p. 1812-5.
28. Freschauf, G.K., et al., *Identification of a small molecule inhibitor of the human DNA repair enzyme polynucleotide kinase/phosphatase.* Cancer Res, 2009. **69**(19): p. 7739-46.
29. Degliesposti, G., et al., *BEAR, a novel virtual screening methodology for drug discovery.* J Biomol Screen, 2011. **16**(1): p. 129-33.

30. Williams, R.S. and J.N. Glover, *Structural consequences of a cancer-causing BRCA1-BRCT missense mutation.* J Biol Chem, 2003. **278**(4): p. 2630-5.
31. Lee, M.S., et al., *Comprehensive analysis of missense variations in the BRCT domain of BRCA1 by structural and functional assays.* Cancer Res, 2010. **70**(12): p. 4880-90.
32. Coquelle, N., R. Green, and J.N. Glover, *Impact of BRCA1 BRCT domain missense substitutions on phosphopeptide recognition.* Biochemistry, 2011. **50**(21): p. 4579-89.
33. Yuan, Z., et al., *Exploiting the P-1 pocket of BRCT domains toward a structure guided inhibitor design.* ACS Med Chem Lett, 2011. **2**(10): p. 764-767.
34. Yu, X., et al., *The BRCT domain is a phospho-protein binding domain.* Science, 2003. **302**(5645): p. 639-42.
35. Simon, R.J., et al., *Peptoids: a modular approach to drug discovery.* Proc Natl Acad Sci U S A, 1992. **89**(20): p. 9367-71.
36. Schafmeister, C., Po, J, Verdine, GL, *An all-hydrocarbon cross-linking system for enhancing the helicity and metabolic stability of peptides.* J. Am. Chem. Soc., 2000. **122**: p. 5891-5892.
37. Mitchell, D.J., et al., *Polyarginine enters cells more efficiently than other polycationic homopolymers.* J Pept Res, 2000. **56**(5): p. 318-25.
38. Futaki, S., et al., *Arginine-rich peptides. An abundant source of membrane-permeable peptides having potential as carriers for intracellular protein delivery.* J Biol Chem, 2001. **276**(8): p. 5836-40.

Chapter 5

General Discussion

5.1 Overall Summary

5.1a: Overview. This work provided significant structural insight into the mechanisms of HR that function to recruit BRCA1 to sites of DSBs. This recruitment is essential to the tumour suppressor activity of BRCA1. In Chapter 2, we provided much needed structural detail of the relative functions of RNF8 and RNF168. Our data supports the concept that RNF8 initiates ubiquitylation while RNF168 serves to amplify these K63-ubiquitin chains. While further study is required, we also proposed a mechanism behind RING domain ubiquitylation based on the crystal structure of the RNF8/Ubc13/Mms2 complex. In Chapter 3, we provided a mechanism of BRCT substrate specificity, explaining how different BRCT domains can be recruited to different phosphorylated targets, despite the structural similarity in BRCT domains as well as their respective recognition sequences. Finally, in Chapter 4, we studied the possibility of BRCA1 as a therapeutic target. HTS failed to identify effective inhibitors, but significant progress was made in the design of high affinity peptide inhibitors of the BRCT domain of BRCA1. While we were unable to find a high affinity small molecule inhibitor of BRCA1 with our initial HTS, identifying an inhibitor by these traditional methods is an attainable goal. New approaches and brute force methods should be attempted before the BRCA1-BRCT is discarded as a potential target of high-throughput screening.

5.1b: Regarding RNF8/RNF168. We described in detail the crystal structure of RNF8 in complex with its E2 heterodimer Ubc13/Mms2, and showed that *in vitro*, RNF8 has significant inherent ubiquitylation activity. We showed that RNF8 forms a tight homo-dimer, mediated through a 50 residue coiled coil, which is somewhat important for RNF8 RING domain function. Based on our data and previous studies, RNF8 dimerization exists *in vivo* and plays an important role in its function. We were to provide insight into the biological relevance of the coiled coil, as its truncation weakens Ubc13 binding and reduces ubiquitylation activity.

The $RNF8_{345-485}$RING/Ubc13/Mms2 complex is likely symmetric, with two Ubc13/Mms2 bound to the RNF8 dimer. We confirmed that RNF8 binds to Ubc13/Mms2 in a canonical way, using its ZnFs to interact with the SPA motif on Ubc13. In addition to this, the complex structure reveals a possible mechanism of RING domain ubiquitylation through an interaction with the covalently bound ubiquitin of Ubc13/Mms2.

Moreover, we showed that RNF168 does not have the same innate ability to build K63-ubiquitin chains as RNF8, and likely requires a structural modification or additional factors in order to be fully active. Our results suggest that at least, the RNF168 RING domain alone is not sufficient to form ubiquitin chains. This certainly suggests that RNF8 initiates ubiquitylation in HR, given its inherent ability to build *de novo* chains. Since the function of RNF168 has been shown previously to be dependent on its MIU domains, they may work by binding RNF168 tightly

in proximity to the ubiquitin chains, and may even initiate a structural change [1, 2]. It is unclear in our biochemical data, however, whether the immediate extended C-terminal region of $RNF168_{1-113}$ is involved in inhibiting its own ubiquitin ligase activity, and if this is a possibility for the required conformational change. The theoretical consequences of our data on the HR pathway are highlighted in Figure 1.

While we suggested a mechanism for RING domain ubiquitylation, it remains to be tested and confirmed. Perhaps the clearest way to do this is to solve the structure of the $RNF8_{345-485}$/Ubc13~Ub/Mms2 complex, and observe any interactions that may exist between the RING domain and the covalently bound ubiquitin. If our prediction is accurate, the ubiquitin moiety should make significant contacts with the RING domain of RNF8 and have reduced flexibility. The ubiquitin may also be oriented differently by RNF8 than in the crystal structure of Ubc13~Ub/Mms2 (PDB ID: 2GMI) [3], and an alignment may indicate that the geometry for catalysis is preferable when RNF8 is bound. In addition, mutations could be made in the potential RNF8 binding sites of ubiquitin and the corresponding residues in RNF8, to see if the ubiquitylation activity is decreased. If these mutants impair ubiquitylation, these results would certainly confirm the hypothesis that there is a specific interaction of the RNF8 RING domain and ubiquitin, and that this interaction is important for catalysis. It is likely that these experiments will be performed in the near future by additional members of the Glover lab.

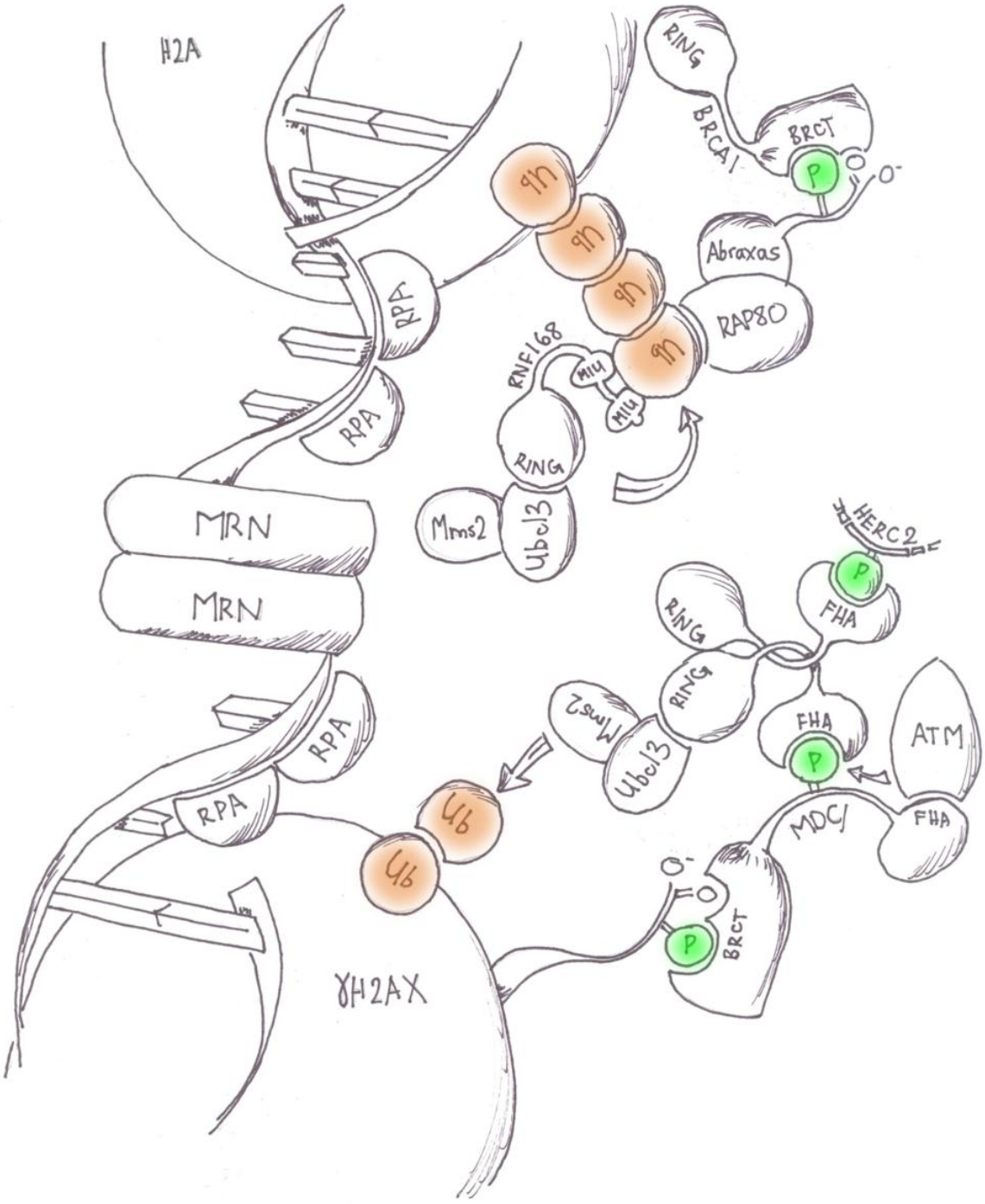


Figure 1. Structural insights into the function of RNF8. An RNF8 dimer forms an asymmetric complex with Ubc13/Mms2. It is likely that RNF8 initiates

ubiquitylation, and RNF168 is recruited to amplify the signal through its MIU domains.

5.1c: Regarding BRCT specificity. Here we described the mechanism of specificity at the +3 position of the phospho-peptide recognition sequence of BRCT domains. The structural characteristics of the specificity loops of BRCA1 and MDC1 dictate the position of the Phe/Tyr in BRCA1 and MDC1, respectively, and it is this positioning that gives MDC1 its sensitivity to an amide group at the C-terminus of the phospho-peptide.

Given the prevalence of the tandem BRCT repeat domains in the DDR, it is useful to know the mechanisms of binding of a domain that retains such structural homology. Other BRCT repeat containing proteins in the DDR, in addition to BRCA1 and MDC1, include TopBP1 and MCPH1. The details of the interactions of these BRCT domains with their target phospho-peptides have been thoroughly characterized [4-6]. The structures available not only indicate similar structural features, but also similarities in the target peptide – often, the phosphorylated residue is followed in the +3 position by a hydrophobic residue (Table 1). In the TopBP1 BRCT 7/8 crystal structure, the +3 residue is a Leucine and the MicroCePHalin 1 (MCPH1) BRCT recognizes both a Phenylalanine or a Tyrosine in this position (Figure 2A). While there are alternate phosphorylated recognition sequences for many BRCT domains that do not contain a hydrophobe in the +3 position, the structural features of

Table 1	Phosphorylated target	Recognition Sequence			
BRCA1	Abraxas/BACH1/CtIP	**pS**	P	T	**F**
MDC1	γH2AX	**pS**	Q	E	**Y**
TopBP1	BACH1	**pT**	P	E	**L**
MCPH1	Cdc27	**pS**	D	E	**F**

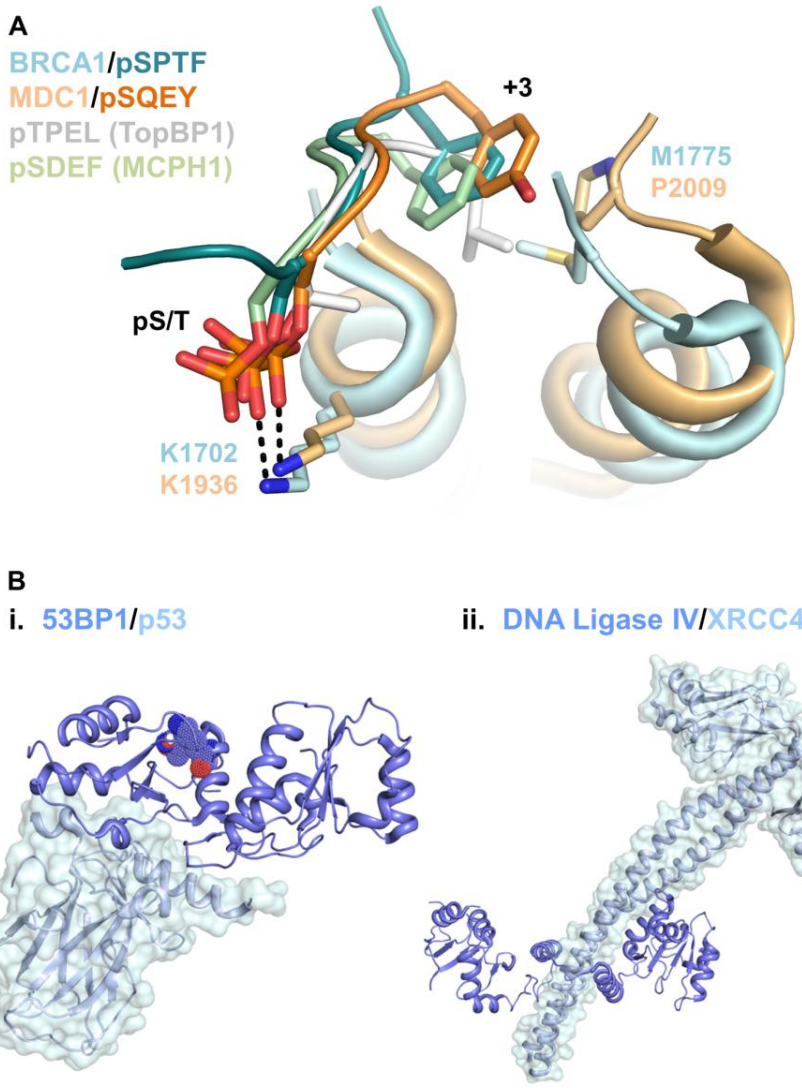


Figure 2. BRCT mode of binding.

A The +3 binding pocket formed by the interface of the N- and C-terminal BRCT domains is shown in cartoon. The conserved resiudes M1775 and P2009 that form the base of the pocket in BRCA1 and MDC1, respectively,

are shown as sticks. The pS/T and +3 residue on each peptide is shown as sticks.

B Non canonical BRCT binding. i. P53 interacts with the N-terminal BRCT repeat of 53BP1 in a phosphorylation independent manner (PDB ID: 1GZH). ii. DNA Ligase IV interacts with the coiled coil of XRCC4 mainly through its linker region (PDB ID: 3II6).

these interactions have either not been characterized, or they interact with a region completely outside of the phospho-peptide binding pocket [7-9]. In addition, many of these non-canonical interactions themselves have been difficult to observe *in vitro*. For example, as discussed in chapter 3, Lou et al. suggest that MDC1 binds to an alternate target in Top II α with the sequence pSXXD [10]. We were unable to show this interaction in vitro, and the structural basis of the interaction has yet to be identified. In addition, a non-canonical MDC1 di-phosphorylated peptide is thought to interact with BRCT domains 4/5 of TopBP1 [11] but again, unpublished work in our lab has shown that this interaction is difficult to reproduce *in vitro*. The crystal structure shows a very non-canonical mode of binding that does not involve the conserved pockets to the same extent as other well characterized phospho-peptide interactions.

Certainly, there are well defined cases of tandem BRCT repeats forming interactions with non-phosphorylated targets, including 53BP1 and DNA ligase IV. 53BP1 interacts with p53 via an interaction with its N-terminal repeat, and DNA Ligase IV interacts with the coiled coil of XRCC4

through the linker between the two BRCT domains (Figure 2B). The phospho-peptide binding pocket is well conserved in 53BP1, and it is not unlikely that it interacts with an as yet unidentified phosphorylated target *in vivo*. Conversely, the tandem BRCT repeats of DNA ligase IV are not packed canonically and the peptide binding pocket is not well defined. As our knowledge of BRCT domains expands, we are starting to discover previously thought unorthodox functions of the BRCT domain [12, 13]. This, however, does not diminish the the significance of the structural conservation observed in the phospho-peptide/BRCT interactions.

Since the target sequences of many BRCT repeats seem to be similar, the question of how the +3 binding pocket confers specificity to its target is of interest to us. The most well characterized phospho-peptide interactions of BRCT domains typically have the phosphorylated residue in the N-terminal binding pocket, and a +3 residue bound in the hydrophobic pocket made by the interface between the two BRCT repeats. The phosphorylated residue binding pocket is very well conserved, and is almost identical in phospho-peptide bound structures (Figure 2A). While the +3 residue is often hydrophobic, it is this essential interaction with the hydrophobic binding pocket that tends to determine specificity.

Given that many of these proteins function in the DDR, the idea of inhibiting any of them as a therapeutic target is quite possible. The mechanistic details of the interactions are therefore essential to understand. This similarity of the BRCT domains may prove to be a

hindrance while attempting to develop inhibitors of any of these proteins, thus it is beneficial to know in detail how these modules retain specificity to a very specific substrate.

5.1d: Regarding BRCA1 inhibition. While insight into the mechanisms of BRCT binding is valuable to the idea of inhibitor design, the fact remains that the inhibitor must eventaully be identified. This often entails the most time consuming processes, as well as the most clinically relevant. Unfortunately, the screening process can often be unsuccessful, especially in the case of Protein Protein Interactions (PPIs); given poorly defined binding surfaces relative to many enzymes [14]. Identifying inhibitors of the BRCA1 PPI likely requires a significantly more involved process than randomized screening. Our collaborators (the Natarajan Lab, University of Nebraska) have used structure aided drug design and peptidomimetics to develop high affinity inhibitors of the BRCA1 BRCT domain. It should be noted that inhibitor screening of BRCA1 is still a valid approach to identify drug like molecules, and automated approaches should be implemented and performed in parallel with peptidomimetics in an attempt to identify both small molecule and peptide based ligands of BRCA1.

As discussed briefly at the end of Chapter 4, the idea of peptides as potential inhibitors has come a long way in recent years. Some often used approaches to improving peptides as inhibitors, however, are not viable in

this situation. For example, to increase binding affinity and stability, the inherent secondary structure of inhibitory peptides can be stabilized by various covalent modifications such as peptides that have been 'stapled' by olefinic amino acids inserted into the primary sequence [15, 16]. However, the pSPTF recognition sequence of BRCA1 is too small, and lacks any sort of secondary structure upon binding. Certainly, the small size of the peptide would not be its limiting factor, as it is relatively similar in molecular weight and pharmacological properties to that of many small molecule drugs. As a result, the best way to improve the tetrapeptide as a drug would be an attempt to make the backbone less susceptible to proteases. Stability can be improved by replacing the peptide backbone with protease resistant peptidomimetic chains such as peptoids and oligopeptidosulfonamide (Figure 3A). Perhaps the greatest challenge to making the peptide a viable therapeutic agent would be to replace the phosphate group with a moiety of similar properties that is less susceptible to hydrolysis. The serum half-life of a phosphate group is very short, but is essential for the BRCA1/Abraxas interaction. While potentially difficult, there are many options for phospho-residue mimics, including replacement of the phospho-serine with glutamic acid. Other carboxylic groups can be substituted, as well as many other acidic functional groups. Although peptides have been considered poor options as therapeutics, due to their inherent physiological instability, the last 20 years have shown a marked increase in the use of biological molecules as drugs. While a

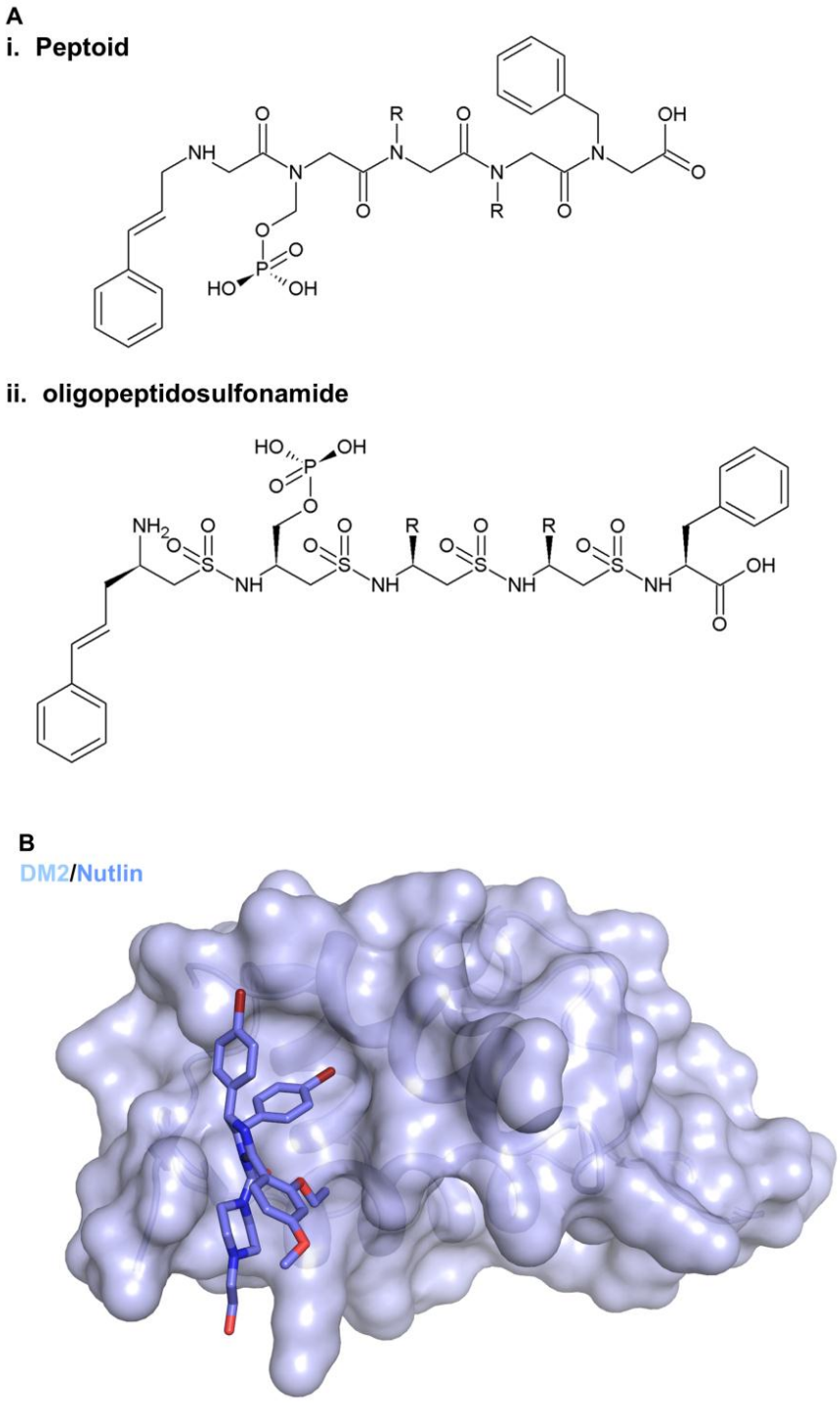


Figure 3. Inhibitor design for Protein Protein Interactions (PPIs).

A Examples of peptidomimetic oligomers with protease resistant backbones, using the BRCA1 peptide inhibitor 16 (see Chapter 4)

B HTS has been used successfully to inhibit the poorly defined surfaces of PPIs, as shown here by the DM2/nutlin interaction (PDB ID: 1RV1).

The nutlin is shown as sticks, and DM2 is shown with surface representaion.

significant endeavour, recent advances in peptidomimetics certainly make it possible to construct a stable, peptide based, biologically active inhibitor of BRCA1 based on the tetrapeptide pSPTF.

In general, conventional forms of inhibitor screening such as HTS and *in silico* screening have proven difficult in identifying inhibitors of PPIs. This is likely due to the fact that many of the current approaches have been optimized toward the discovery of inhibitors of extremely well-defined and reactive enzyme cavities. Nonetheless, HTS has been used previously to identify inhibitors of relatively poorly defined PPIs, particularly Nutlins, which were identified as inhibitors of the p53-hDM2 interaction by HTS (Figure 3B) [17]. In addition, virtual screening has been used successfully in the past to identify an inhibitor of the SH3 binding surface of the HIV type 1 Nef protein [18].

In order to obtain a small molecule inhibitor by randomized screening, new techniques must be implemented, and likely require a more automated approach. Robotic screening would be more appropriate in this situation, since much larger libraries of fragments can be screened much more quickly. An essential change would be the requirement of more samples, perhaps tens of thousands for biochemical screening, and libraries numbering in the millions of compounds for virtual screening. A

fragment-based method would likely be the best approach for HTS, since there are three well defined binding pockets to be utilized in the BRCA1 BRCT/pSPTF interaction. This process will be time and resource intensive, but inhibitors of BRCA1 may not only prove to be an essential therapeutic, but also extremely useful as a biochemical tool to probe the specific functions of BRCA1 and of the DDR in general. Given this, identification of BRCA1 inhibitors should be a priority, with the potential for major therapeutic and biochemical ramifications.

5.2 References

1. Doil, C., et al., *RNF168 binds and amplifies ubiquitin conjugates on damaged chromosomes to allow accumulation of repair proteins.* Cell, 2009. **136**(3): p. 435-46.
2. Stewart, G.S., et al., *The RIDDLE syndrome protein mediates a ubiquitin-dependent signaling cascade at sites of DNA damage.* Cell, 2009. **136**(3): p. 420-34.
3. Eddins, M.J., et al., *Mms2-Ubc13 covalently bound to ubiquitin reveals the structural basis of linkage-specific polyubiquitin chain formation.* Nat Struct Mol Biol, 2006. **13**(10): p. 915-20.
4. Singh, N., et al., *Molecular basis for the association of microcephalin (MCPH1) with the cell division cycle protein 27 (Cdc27) subunit of the anaphase-promoting complex.* J Biol Chem, 2011.
5. Shao, Z., et al., *Specific recognition of phosphorylated tail of H2AX by the tandem BRCT domains of MCPH1 revealed by complex structure.* J Struct Biol, 2011.

6. Leung, C.C., et al., *Molecular basis of BACH1/FANCJ recognition by TopBP1 in DNA replication checkpoint control.* J Biol Chem, 2011. **286**(6): p. 4292-301.
7. Derbyshire, D.J., et al., *Crystal structure of human 53BP1 BRCT domains bound to p53 tumour suppressor.* EMBO J, 2002. **21**(14): p. 3863-72.
8. Joo, W.S., et al., *Structure of the 53BP1 BRCT region bound to p53 and its comparison to the Brca1 BRCT structure.* Genes Dev, 2002. **16**(5): p. 583-93.
9. Wu, P.Y., et al., *Structural and functional interaction between the human DNA repair proteins DNA ligase IV and XRCC4.* Mol Cell Biol, 2009. **29**(11): p. 3163-72.
10. Luo, K., et al., *Topoisomerase IIalpha controls the decatenation checkpoint.* Nat Cell Biol, 2009. **11**(2): p. 204-10.
11. Wang, J., Z. Gong, and J. Chen, *MDC1 collaborates with TopBP1 in DNA replication checkpoint control.* J Cell Biol, 2011. **193**(2): p. 267-73.
12. Leung, C.C. and J.N. Glover, *BRCT domains: easy as one, two, three.* Cell Cycle, 2011. **10**(15): p. 2461-70.
13. Sheng, Z.Z., Y.Q. Zhao, and J.F. Huang, *Functional Evolution of BRCT Domains from Binding DNA to Protein.* Evol Bioinform Online, 2011. **7**: p. 87-97.
14. Clackson, T. and J.A. Wells, *A hot spot of binding energy in a hormone-receptor interface.* Science, 1995. **267**(5196): p. 383-6.
15. Wilson, A.J., *Inhibition of protein-protein interactions using designed molecules.* Chem Soc Rev, 2009. **38**(12): p. 3289-300.
16. Liskamp, R.M., et al., *Peptides and proteins as a continuing exciting source of inspiration for peptidomimetics.* Chembiochem, 2011. **12**(11): p. 1626-53.
17. Vassilev, L.T., et al., *In vivo activation of the p53 pathway by small-molecule antagonists of MDM2.* Science, 2004. **303**(5659): p. 844-8.

18. Betzi, S., et al., *Protein protein interaction inhibition (2P2I) combining high throughput and virtual screening: Application to the HIV-1 Nef protein.* Proc Natl Acad Sci U S A, 2007. **104**(49): p. 19256-61.