

**Structural and Functional Characterization of BRCTs  
Domain**

*By*

**Dilip C. Badgujar**

[LIFE09200704005]

**Tata Memorial Centre**

**Mumbai**

*A thesis submitted to the*

*Board of Studies in Life Sciences*

*In partial fulfilment of requirements*

*For the Degree of*

**DOCTOR OF PHILOSOPHY**

*Of*

**HOMI BHABHA NATIONAL INSTITUTE**



**JUNE 2014**

# HOMI BHABHA NATIONAL INSTITUTE

## Recommendations of the Viva Voce Board

As members of the Viva Voce Board, we certify that we have read the dissertation prepared by Dilip C. Badgujar entitled “**Structural and Functional Characterization of BRCTs domain**” and recommend that it may be accepted as fulfilling the thesis requirements for the award of Degree of Doctor of Philosophy.

Chairperson - Dr. Vinay Kumar

Date: 30/06/2014

Convener – Dr. Ashok Varma

Date:

Member 1- Dr. Prasanna Venkatraman

Date:

Member 2- Dr. Kakoli Bose

Date:

External Examiner – Dr. T.P. Singh

Date:

Final approval and acceptance of this thesis is contingent upon the candidate's submission of the final copies of the thesis to HBNI.

I hereby certify that I have read this thesis prepared under my direction and recommend that it may be accepted as fulfilling the thesis requirement.

Date: 30/06/2014

Place: Kharagpur

  
(Ashok K. Varma)  
Guide

# STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at Homi Bhabha National Institute (HBNI) and is deposited in the Library to be made available to borrowers under rules of the HBNI.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgement of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the Competent Authority of HBNI when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

Navi Mumbai,

Dilip C. Badgujar

Date: 30.6.14

# **DECLARATION**

I, hereby declare that the investigation presented in the thesis has been carried out by me. This work is original and has not been submitted earlier as a whole or in part for a degree / diploma at this or any other Institution / University.

Navi Mumbai,

Dilip C. Badgujar

Date: 30.6.14

## List of Publications arising from the thesis

### Journal

1. **Badgujar DC**, Sawant U, Mahadik H, Gadewal N, Varma AK (2012) Pathogenicity of Mutations Discovered in BRCA1 BRCT Domains is Characterized by Destabilizing the Hydrophobic Interactions. *Journal of Cancer Science & Therapy* **4**: 386-393.
2. **Badgujar DC**, Sawant U, Yadav L, Hosur M, Varma AK (2013). Preliminary crystallographic studies of BRCA1 BRCT-ABRAXAS complex. *Acta Crystallographica Section F: Structural Biology and Crystallization Communications* **69**: 1401-1404.

Navi Mumbai,

Dilip C. Badgujar

June, 2014

*Dedicated to*  
*My Beloved Parents*

## ACKNOWLEDGEMENTS

*First and foremost, a sincere and heart filled gratitude to my mentor Dr. Ashok Varma for being such a wonderful and motivating guide. His stimulating suggestions, experience and encouragement accompanied by the freedom of thought that he granted, has not only helped me to discover my potential but also kept me sparked throughout my Ph.D tenure. I am also thankful to him as he has given me lot of opportunities to improve. I will never forget the generosity that he offered on several occasions. His everlasting moral support and kindness made way for the successful completion of this endeavour. I am indebted to him more than he knows. Sir, I thank you for everything and will always be proud to have been mentored by you.*

*I am thankful to Dr. Shubhada Chiplunkar (Director, ACTREC) for providing me an opportunity to work in this institution and the excellent infrastructure. I also want to thank Dr. Rajiv Sarin (Ex-Director, ACTREC) and Dr. Surekha Zingde (Ex-Deputy Director, ACTREC) for their help and support directly or indirectly.*

*I have been most fortunate to have eminent DC members on my panel. Firstly, I wish to thank Prof. Hosur, the Chairperson of the committee, for all his efforts towards motivating me and developing my scientific attitude. I never met such a simple man with clear thinking, honest attitude and so much enthusiasm towards science. The best part of Prof. Hosur is that he never puts you down and at the same time he conveys what he wants to say. He has taken lots of efforts in correcting my thesis. He is a star for me, to whom I can look up to and follow him to achieve dedication towards science. I am highly indebted to Prof. Vinay Kumar, for helping me in critically analysing the data. His suggestions about understanding basic crystallography and interaction with my guide really helped me. His comments were invaluable and had a major impact in shaping up my project. I am also thankful to Dr. Prasanna and Dr. Kakoli Bose for analysing the data and for giving moral support. I want to specially thank Dr. Prasanna because she motivated me to get inspired by interacting and listening to well known scientists.*

*I am extremely thankful to common instrument facility for their help and support specially Mr. Dandekar who is a very dedicated and resourceful person and always kin to help everyone. I am also thankful to administration and accounts department for their constant help and support. Special thanks to Miss Seema and Miss Sujata for helping me and my 007 batch. I also want to thank photography department for their help especially Mr. Shyamrao, he is the person who kept me grounded and taught me some important lessons of life.*

*I want to express my warmest thanks to all Varma lab members: Ulka Sawant (UU), Sopan (Sopu), Vikrant (Professional), Lumbini (complicated), Bhanu (Hyderabad Prince), Rajan (complaint box) and Quadir (cool). I cannot thank UU enough for all the efforts she had put in crystallization and protein expression. I also want to thank UU for being so kind and helpful throughout my tenure. I don't think, I could have finished my*

*PhD without her moral support and help both scientifically as well as non-scientifically. I think every Varma lab member is unique in one way or the other and I want to thank all of them for making a positive working environment in lab. I want to thank Hafiza for being a good friend who constantly helped me during my tenure. I also want to thank Neha for being a good friend as her moral support helped me achieve this endeavour. I also want to thank my all the trainees Mandar, Varsha, Medha and Polomy for their hard work and support. Although I am the first student of Lab, I want to thank all my seniors Amit Gautam, Amit Fuljale, Atul, Ajit, Manoj Baba and Sapna for their guidance and help.*

*My special thanks to Hemant for being wonderful friend from last 12 years and also inspiring me to do photography. I also want to thank my weekend photography group members Manohar and Bhairav for making creative, humorous and passionate environment during photography outings.*

*I am really thankful to be a part of batch 007: Akhil (Mr. senti), Bihari (Mr. Cool), Vinayak (Superman), Dimpu (Mr. Chinkey), Hemant (Mr. Artist), Gaurav (Mr. Crow), Ajit (Shri. Scientist), Sameer (Mr. Model), Harsh (Mr. Gaddu) and one and only “ladki” of our batch Ratika (Miss. Gorgeous) for maintaining cheerful, fun loving and caring atmosphere. I specially want to thank Akhil and Bihari, who made my tough and tense days of PhD as joyous as possible. Akhil, you deserve a special gratitude for being a wonderful friend. You are the most honest and genuine person I have ever met. Bihari, is a good listener and also a good human being; he taught me how to keep cool in tremendous pressure or tension. I really want to thank all my batch mates for taking my care when it was utterly required. The best part was not only to be a part of this batch but also to enjoy their company in hostel parties. I also want to thank pg room partners Rupa, Mellissa, Kaushal, Sapna, Richa, Cris and Manohar for making this time more joyful and happening. Special thanks to Rupa and Mellisa guiding me difficult time of my life and keeping pg room as funny place. I would also like to thank my MSc friends Sanket, Vishaldeep, Rahul, Deepak and Vipul.*

*I am thankful to GOD who gave me strength during struggling times of my life and making me realize that GOD never leaves you alone during hard time of your life. I am thankful to my family; especially I say thanks and sorry to my parents for not taking their responsibility in these past years as a son in my selfish pursuit to finish Ph.D. I have no word that can suffice for their constant sacrifices in making my life beautiful and easy. I know that I have reached so far only because you all believed in me. You all have been my source of strength and inspiration throughout my life.*

*Dilip C. Badgujar*

# CONTENTS

	Page No.
<b>Synopsis</b> .....	i-xviii
<b>List of Figures</b> .....	xix-xxii
<b>List of Tables</b> .....	xxiii
<b>Abbreviations</b> .....	xxiv-xxvi
<b>Chapter-1: Introduction and Review of Literature</b> .....	<b>1-31</b>
<b>1.1. Cancer</b> .....	<b>1-5</b>
1.1.1. Breast Cancer	
1.1.1.1. Sporadic Breast cancer	
1.1.1.2. Hereditary Breast Cancer	
<b>1.2. BRCA1</b> .....	<b>5-24</b>
1.2.1. BRCA1 functions	
1.2.1.1. Transcriptional activation by BRCA1	
1.2.1.2. Role of BRCA1 in DNA repair	
1.2.1.3. BRCA1 and Chromatin Remodelling	
1.2.1.4. BRCA1 in Cell- Cycle Check Point Control	
1.2.1.5. Apoptosis and BRCA1	
1.2.2. BRCA1 Domain organization	
1.2.2.1. Ring Finger Domain	
1.2.2.2. Central DNA Binding Domain	
1.2.2.3. BRCT Domain	
1.2.3. Three dimensional structure of different BRCA1 domains	
1.2.3.1. Structure of RING Domain of BRCA1 and BARD1 Complex	
1.2.3.2. Crystal Structure of BRCT Domain	
1.2.3.3. BRCT Domain as a Phosphopeptide Interaction Module	
A. Crystal Structure of BRCA1 BRCT-BACH1	
B. Crystal Structure of BRCA1 BRCT-CtIP	
C. Crystal Structure of BRCA1 BRCT-ACC1	
D. Crystal Structure of BRCA1 BRCT-ATRIP	
E. Crystal Structure of BRCA1 BRCT-BAAT	
<b>1.3. BRCA1 BRCT Mis-sense Mutations: Application     in Clinical Management</b> .....	<b>24-30</b>
<b>1.4. Conclusion</b> .....	<b>30-31</b>

<b>Chapter-2: Materials and Methods.....</b>	<b>32-61</b>
<b>Source of reagents and instruments.....</b>	<b>32</b>
<b>Materials.....</b>	<b>32-33</b>
<b>2.1. Gene Cloning.....</b>	<b>34-42</b>
2.1.1. Selection of target protein/domain	
2.1.2. Restriction enzyme based cloning	
2.1.2.1. Preparation of Expression vector	
2.1.2.2. Preparation of Target gene	
2.1.2.3. Annealing and Ligation	
2.1.2.4. Screening of potential clones	
2.1.3. Site Directed mutagenesis	
2.1.3.1. PCR of Mutant Strand	
2.1.3.2. DpnI Digestion	
2.1.3.3. Transformation of SDM product and Screening	
<b>2.2. Protein expression and Purification.....</b>	<b>42-46</b>
2.2.1. Preparation of seed culture	
2.2.2. Scale up and induction	
2.2.3. Protein Purification	
2.2.3.1. Affinity Chromatography	
2.2.3.2. Ion Exchange Chromatography	
2.2.3.3. Gel Filtration Chromatography (GFC)	
2.2.3.4. Hydrophobic Interaction Chromatography (HIC)	
<b>2.3. Protein Characterization.....</b>	<b>46-50</b>
2.3.1. Circular Dichroism (CD)	
2.3.2. Fluorescence spectroscopy	
2.3.3. Mass spectrometry	
2.3.3.1. MALDI-TOF	
2.3.3.2. ESI	
2.3.4. Isothermal Titration Calorimetry (ITC)	
<b>2.4. Protein crystallization.....</b>	<b>50-54</b>
2.4.1. Vapour Diffusion method	
2.4.2. Micro batch method	
2.4.3. Dialysis method	
2.4.4. Liquid – liquid diffusion method	
2.4.5. Micro-fluidics technique	
<b>2.5. X-ray diffraction.....</b>	<b>54-55</b>
<b>2.6. Diffraction data processing.....</b>	<b>55-56</b>

<b>2.7. Structure solution</b> .....	56-58
2.7.1. Molecular replacement method	
2.7.2. Multiple Isomorphous Replacement method	
2.7.3. Anomalous Scattering based method	
<b>2.8. Electron density map interpretation</b> .....	58
<b>2.9. Crystallographic refinement</b> .....	58-59
<b>2.10 Structure validation</b> .....	60
<b>2.11 Conclusion</b> .....	61
<b>Chapter-3 Structural studies of BRCA1 BRCT-NCoA2 complex</b> .....	<b>62-84</b>
<b>3.1. Introduction</b> .....	62-64
3.1.1. BRCA1 and Estrogen receptors	
3.1.2. BRCA1 and NCoA2	
<b>3.2. Materials and methods</b> .....	64-72
3.2.1. Cloning of BRCA1 BRCT domain in pGEX- KT	
3.2.1.1. Preparation of Insert (BRCA1 BRCT domain)	
3.2.1.2. Preparation of Expression vector pGEX-KT	
3.2.1.3. Ligation	
3.2.1.4. Screening and DNA sequencing	
3.2.2. BRCA1 BRCT domain protein expression and purification	
3.2.2.1. Starter culture and scale up	
3.2.2.2. Protein purification	
3.2.2.3. Gel filtration chromatography	
3.2.2.4. SDS-PAGE analysis	
3.2.3. Biophysical characterization	
3.2.3.1. Mass Spectrometry	
3.2.3.2. Interaction analysis using ITC	
3.2.4. Protein structure determination	
3.2.4.1. Protein crystallization	
3.2.4.2. Diffraction data collection and processing	
3.2.4.3. Structure solution by molecular replacement	
<b>3.3. Results and discussion</b> .....	72-84
3.3.1. Cloning of BRCA1 BRCT domain	
3.3.2. Expression and purification of BRCA1 BRCT domain	
3.3.3. NCoA2 peptide interacts with BRCA1 BRCT domain	
3.3.4. Structure of BRCA1 BRCT - NCoA2 peptide complex	
3.3.5. Interactions analysis of BRCA1 BRCT and NCoA2	
3.3.5.1. Weak intermolecular interactions around pSer <sub>(0)</sub> residue	
3.3.5.2. Molecular geometry around Phe <sub>(+3)</sub> residues	
3.3.5.3. Residues other than pSer <sub>(0)</sub> and Phe <sub>(+3)</sub>	
3.3.6. Comparison of oligopeptides conformation	

3.4. Conclusion.....	84
----------------------	----

**Chapter-4: Characterisation of Nup153 and RBP-12 as interacting partners of BRCA1 BRCT.....85-98**

4.1. Introduction.....	85-86
4.2. Materials and Methods.....	86-89
4.2.1. Expression and purification BRCT domain protein of BRCA1	
4.2.2. Binding affinities	
4.2.3. Crystallization	
4.2.4. Crystal screening and data processing	
4.2.5. Molecular dynamics simulation	
4.3. Results and Discussion.....	89-97
4.3.1. Interaction analysis	
4.3.1.1. Nup153 peptide interacts with BRCA1 BRCT domain	
4.3.1.2. RBP-12 interacts with BRCA1 BRCT domain	
4.3.2. Crystallization of BRCA1 BRCT domain with Nup153 and RBP-12	
4.3.3. Data collection and data processing	
4.3.4. Model building and molecular dynamics simulations	
4.3.4.1. BRCA1 BRCT-Nup153 complex	
4.3.4.2. BRCA1 BRCT-RBP-12 complex	
4.3.4.3. Structural comparison	
4.4. Conclusion.....	97-98

**Chapter-5: Structural studies of BRCA1 BRCT-Abraxas complex.....99-120**

5.1. Introduction.....	99-101
5.2. Materials and methods.....	101-104
5.2.1. Expression and purification of BRCA1 BRCT domain	
5.2.2. Interaction analysis using ITC	
5.2.3. Protein structure determination	
5.2.3.1. Protein crystallization	
5.2.3.2. Diffraction data collection and processing	
5.2.3.3. Structure determination and refinement	
5.3. Results and Discussion.....	104-119
5.3.1. Abraxas peptides interacts with BRCA1 BRCT domain	
5.3.2. Structure determination of BRCA1 BRCT domain with Abraxas peptides	
5.3.2.1. Crystallization and structure determination	
5.3.2.1.1. BRCA1 BRCT-Abraxas peptide structures	
5.3.2.1.2. BRCA1 BRCT-A1 complex (Singly phospho)	
5.3.2.1.2.1. Interaction analysis	
A) pSer-406 residue	
B) Phe-409 binding	

5.3.2.1.3.	BRCA1 BRCT-A2 complex (Double phospho)	
5.3.2.1.3.1.	Interaction analysis	
	A) pSer -404 and pSer-406 binding	
	B) Phe+3 binding	
5.3.2.2.	Comparison of A1 and A2 complex structures	
5.4.	<b>Conclusion</b> .....	119-120

**Chapter-6: Structural basis and pathogenicity of mis-sense mutations discovered in BRCA1 BRCT domain.....121-137**

<b>6.1.</b>	<b>Introduction</b> .....	121-122
<b>6.2.</b>	<b>Materials and Methods</b> .....	122-126
6.2.1.	Gene cloning and site -directed- mutagenesis	
6.2.2.	Protein expression and purification	
6.2.3.	Protein characterization	
6.2.3.1.	Circular Dichroism	
6.2.3.2.	Fluorescence spectroscopy	
6.2.4.	Prediction of pathogenicity using online servers	
6.2.5.	In silico modelling	
<b>6.3.</b>	<b>Results and Discussion</b> .....	126-136
6.3.1.	Cloning of BRCA1 BRCT domain in pET3a and SDM	
6.3.2.	BRCA1 BRCT domain mis-sense mutant expression and purification	
6.3.3.	Characterization of BRCA1 H1686Q variant	
6.3.3.1	Circular Dichroism	
6.3.3.2	Fluorescence Spectroscopy	
6.3.4.	Prediction of pathogenicity of BRCA1 BRCT mutants	
6.3.5.	Intramolecular interactions involving pathogenic mutations	
<b>6.4.</b>	<b>Conclusion</b> .....	136-137

**Chapter-7: Analysis of BRCT domain containing proteins.....138-167**

<b>7.1.</b>	<b>Introduction</b> .....	138-140
<b>7.2.</b>	<b>Materials and Methods</b> .....	140-142
7.2.1.	Amino acid sequence alignment	
7.2.2.	Cloning of BARD1 ARD-BRCT domain in pGEX-Kt vector	
7.2.3.	Expression and purification of BARD1 ARD-BRCT	
<b>7.3.</b>	<b>Results and Discussion</b> .....	142-159
7.3.1.	In-silico analysis	
7.3.1.1.	Structural basis of single BRCT domain	
7.3.1.2.	Tandem BRCT domain	
7.3.1.2.1.	BRCA1 BRCT	
7.3.2.	Cloning of BARD1 ARD-BRCT in pGEX-KT	
7.3.3.	Expression and purification of BARD1 BRCT-ARD	

7.3.4. Crystallization of BARD1 ARD-BRCT	
7.4. <b>Conclusion</b> .....	159-160
<b>Chapter-8: Summary and Future Directions</b> .....	<b>161-166</b>
<b>References</b> .....	<b>167-180</b>
<b>Published Manuscripts</b> .....	<b>181</b>

*Synopsis*

---

## **SYNOPSIS**



**Homi Bhabha National Institute**

**Ph. D. PROGRAMME**

- |  |   |
|--|---|
| <b>1. Name of the Student:</b>                 | Dilip C. Badgular   |
| <b>2. Name of the Constituent Institution:</b> | Tata Memorial Centre, Advanced Centre for<br>Treatment Research and Education in Cancer |
| <b>3. Enrolment No. :</b>                      | LIFE09200704005   |
| <b>4. Title of the Thesis:</b>                 | “Structural and Functional Characterization of<br>BRCTs domain”                         |

Hereditary breast cancer has been observed to be caused by mutations in a gene located on the 17q21 chromosome, [1] and intensive efforts have been carried out to isolate this gene, which is also associated with the development of ovarian cancer. Miki *et al.* discovered this breast cancer associated gene 1 (BRCA1) using the positional cloning method [2]. BRCA1 accounts for 5-10% of all breast cancers and 40-45% of hereditary breast cancers. BRCA1 is recognized for its multi functionality, including DNA damage repair, chromosome remodeling [3], centrosome duplication [4], cytokinesis, and transcription activation [5]. This raises the question of how a single molecule can perform such diversified functions. One possible clue on this topic comes from the growing literature on BRCA1, which points to a long list of BRCA1 interactions with its cellular partners. Using macromolecular crystallography, structural biology, bioinformatics, and the biophysical approaches this study aims to address the following issues: (1) determination of the structure of singly and doubly phosphorylated BRCA1 BRCT domain, (2) assessment of whether carriers of the BRCA1 gene mutation are predisposed to a high risk of cancer; and (3) studies of other BRCT containing proteins. Such information would help in clinical management or allow BRCA1 gene to be used as a preventive/diagnostic biomarker. The present research showed that many of the BRCA1 mutant proteins do not fold properly, thereby resulting in tumor formation. A small molecular inhibitor could be designed to ensure proper folding in such an event [6]. The thesis is organized into eight chapters.

**Chapter 1** is a brief introduction to BRCA1 protein and its functions. The BRCA1 protein comprises of 1863 amino acids, and possesses three conserved domains, specifically the N-terminal ring domain, the central DNA-binding domain, and the C-terminal BRCT domain. The BRCA1 ring finger domain interacts with BRCA1 Associated Ring Domain (BARD1) forming a heterodimer, which acts as an E3 ubiquitin ligase [7]. This complex is responsible for ubiquitination of proteins, which ultimately leads to the degradation of proteins by the 26S proteasome. The BRCA1 ring domain also helps in DNA repair and transcription regulation. Mutations in the ring domain of BRCA1 have been found in some breast and ovarian cancer

cases, which specifically target the metal binding residues of the BRCA1 ring domain [8]. The integrity of the ring domain is essential for the proper functioning of BRCA1.

BRCA1 has tandem repeats of ~ 90-100 amino acids at the C-terminus; together these are referred to as the BRCA1 C-Terminal BRCT domain [9]. BRCT domain is organized in diverse ways, as a single unit in some proteins, and as multiple tandem repeats in others. The crystal structure of the BRCA1 BRCT (1646-1859) domain has been determined [10], and the molecule is composed of two BRCT tandem repeats (N-terminal BRCT and C-terminal BRCT). Each BRCT repeat has a conserved hydrophobic core and a large interfacial hydrophobic area. The overall domain organization for a single BRCT repeat is  $\beta 1-\alpha 1-\beta 2-\beta 3-\alpha 2-\beta 4-\alpha 3$ . The BRCT repeats are arranged in a head-to-tail manner, and are separated by a linker of few amino acids (~ 24 amino acids), with a larger linker resulting in a parallel juxtaposition of BRCT repeats and a shorter linker generating twisted BRCT domains.

The BRCT domain of BRCA1 interacts with numerous proteins in a phosphorylation dependent manner [11]. Different research groups have attempted to map the BRCT domain-interacting motif and found that the proteins containing singly phosphorylated serine (pSer) at position (0), aromatic amino acids at position (+1), aromatic or hydrophobic amino acids at position (+2), and phenylalanine (Phe) at the position (+3) were the best binding partners. An oriented peptide library-based approach also confirmed that the consensus motif pS-X-X-F (pS-phosphorylated serine, X-any amino acid, F- phenylalanine) containing proteins bind to the BRCA1 BRCT. A number of crystal structures resolved recently show that the phosphopeptide binds with the BRCA1 BRCT in a bipartite manner in which the phosphoserine holds the N-terminal BRCT and the phenylalanine interacts with the C-terminal BRCT. Two residues from the N-terminal BRCT (Ser-1655 and Lys-1702) form hydrogen bonds with the phosphoserine, and the three C-terminal residues (Phe-1704, Met-1775, and Leu-1839) contribute to the hydrophobic interaction with phenylalanine. The sequential analysis of the BRCT-domain-containing proteins showed that the residues contributing to the interactions were conserved, and this may be an

evolutionarily conserved function of the BRCT domain, It has been observed that doubly phosphorylated Abraxas, which is a DNA repair protein known to be phosphorylated by ATM (Ataxia Telangiectasia Mutated) and ATR (Ataxia Telangiectasia and Rad3-related protein) at the 404 and 406 positions, is also a binding partner of BRCT. However, there has been no structural report on how the doubly phosphorylated protein binds to BRCA1 BRCT.

**Chapter 2** of the thesis is a brief description of the various methodologies and techniques used in this investigation.

## **2.1 Gene Cloning**

The BRCA1 BRCT domain is cloned in pGEX-KT and pET3a vectors by sticky end based directional cloning. The BRCA1 BRCT domains (1646-1859) are specifically amplified by specific forward and reverse primers. In sticky end based directional cloning method both the insert and the vector are digested by identical set of restriction enzymes leading to cohesive ends which are then ligated to get the clone. The clone is further confirmed by DNA sequencing method.

## **2.2 Protein Expression and Purification**

### ***2.2.1 Expression and purification of BRCA1 BRCT***

The BRCA1 BRCT was expressed in BL21 (DE3) bacterial strain. Starter culture was raised in 100 mL Luria broth, and diluted in 8 litre culture to scale up the protein purification. The culture was induced by 0.4 mM IPTG at 24°C for 16 hours. The cells were pelleted by centrifugation and bacterial pellet was re-suspended in buffer (50 mM Tris pH- 7.5, 300 mM NaCl), further sonicated to lyse the cells. The proteins in the soluble fraction of the cell lysate were bound to GST sepharose 4B (GE Healthcare). The fusion protein tag was cleared by TEV protease. The native protein was further purified using FPLC superdex-200 gel filtration column.

### ***2.2.2 Expression and purification of BRCA1 variants***

The BRCA1 BRCT variants (H1686Q, P1749R, S1715R and M1775R) were expressed in BL21 (pLysS) bacterial strain individually. Cells were grown in LB medium containing 100µg/ml Ampicillin and 34 µg/ml Chloramphenicol at 37°C till it reaches to 0.6 OD. The culture was induced by 0.4 mM IPTG at 18°C for 16 hours. Further protein purification steps were carried out at 4°C. The IPTG induced bacterial cell pellet was re-suspended in buffer (20 mM Sodium phosphate pH- 6.0, 50 mM NaCl, 1 mM EDTA, 1 mM β-ME, 1 mM PMSF and 0.5% Triton X-100), and then sonicated to lyse the cells. The soluble proteins were bound to SP sepharose resin equilibrated in buffer (20 mM Sodium phosphate pH- 5.8, 50 mM NaCl, 1 mM EDTA, 1 mM βME, 1 mM PMSF, 5% Glycerol). Bound proteins were eluted with a gradient of NaCl (100–800 mM). The protein in the elution fraction contains some high molecular weight protein. Further protein purification was through ion-exchange chromatography using Q sepharose resins. Bound proteins were eluted with a gradient of NaCl (100– 800 mM). All the proteins were FPLC purified.

## **2.3 Biophysical Characterization**

### ***2.3.1 Circular Dichroism (CD) spectroscopy***

Circular dichroism (CD) is defined as the unequal absorption of left-handed and right-handed circularly polarized light. It is given in terms of molar ellipticity, which is a measure of the degree of rotation of plane polarised light on passage through the sample. CD spectroscopy is an excellent method of rapidly determining the secondary structure of proteins, using samples as less as 20 micrograms. Proteins rich in α-helical segments have negative bands at 222 nm and 208 nm and a positive band at 193 nm. Proteins with well-defined antiparallel β-pleated sheets have negative bands at 218 nm and positive bands at

195 nm, while disordered proteins have very low ellipticity above 210 nm and negative bands near 195 nm. CD can be used also to study protein interactions. CD measurements were carried out using Jasco J-815 CD spectrophotometer. Wavelength scan from 200 nm to 260 nm were conducted in Quartz cuvette of 0.1mm path length. Each spectrum was taken at a scan rate of 20 nm/min with 0.2 nm wavelength step.

### ***2.3.2 Fluorescence spectroscopy***

Fluorescence is the emission of radiation that occurs when a molecule in an excited electronic state returns to the ground state. The intrinsic fluorescence in proteins is due to aromatic amino acids Phe, Tyr and Trp. Fluorescence emission is always at a wavelength longer than the wavelength of radiation used for excitation. The fluorescence maximum is dependent on the environment around the aromatics in the protein. For example, if the environment around a Trp is completely hydrophobic, as happens in a well folded protein, the fluorescence maximum is at a wavelength shorter compared to when the Trp residue is exposed, as happens during denaturation. Thus fluorescence spectroscopy is useful to study alterations in protein conformation, either due to ligand binding or due to denaturation. Emission spectra were collected on HORIBA FL3-21 Spectrofluorometer. The variant and wild type BRCA1 BRCT were excited at 280 nm and 295nm, and the emission maxima were monitored.

### ***2.3.3 Peptide binding analysis using ITC***

Isothermal titration calorimetry (ITC) measures directly the energy associated with a chemical reaction triggered by the mixing of two components through injection of one of the components. Modern ITC instruments operate on the heat compensation principle, and the measured signal is the amount of power (microcalories per second) necessary to

maintain constant the temperature difference between the reaction and reference cells. Isothermal titration calorimetry is the only method that measures not only the magnitude of the binding affinity but also the magnitude of the two thermodynamic terms that define the binding affinity: the enthalpy ( $\Delta H$ ) and entropy ( $\Delta S$ ) changes. ITC was performed using MicroCal ITC-200 from GE Healthcare. BRCA1 BRCT wild- type and BRCA1 H1686Q mutant were titrated with synthetic phospho peptides of NcoA2, Nup153 and Abraxas proteins. Total injections were 15 with 1 $\mu$ l volume each with stirring speed 1000 rpm. Spacing between two injections was 100 sec with 40 sec as initial delay. Curve fitting and calculation of  $K_d$  value was by using Origin software (Origin Lab).

## **2.4 Crystallization**

### ***2.4.1 BRCA1 BRCT: oligopeptide complexes***

BRCA1 BRCT protein was concentrated to a concentration of 25 mg/mL using MWCO (Molecular Weight Cut off, Millipore) filtration device. The protein concentration was measured by Bradford protein estimation method, and also by measurement of UV (280 nm) absorption using nanodrop spectrophotometer. The BRCA1 BRCT protein and different oligopeptides were mixed in 1:1.5 molar ratio. Crystals of the complex were grown at 22°C in MES buffer of pH 6.5 by the hanging drop vapour diffusion method, with 30% PEG MME 5K containing various amounts of ammonium sulphate as the precipitant.

### **2.4 X-ray diffraction data collection and processing**

The crystals of BRCA1 BRCT: NCoA2 complex were initially screened for diffraction at the home X-ray source and MAR-DTB Image Plate diffractometer. X-ray diffraction intensity data were collected by the oscillation method, using the BM-14 beamline on ESRF, Grenoble, France. Freshly grown crystals in 30% glycerol were flash frozen directly in liquid nitrogen and shipped

to Grenoble, France. All the data were indexed, scaled and reduced using Collaborative Computational Project No.4 (*CCP4*) software suite. The BRCA1 BRCT: (A2) Abraxas complex crystals were diffracted at home source (ACTREC, Mumbai). Diffraction data oscillation frames were processed using *iMosflm* and were scaled using *SCALEPACK* [12].

### 2.6 Structure solution and Refinement

The crystal structures of BRCA1 BRCT: NCoA2 complex and BRCA1 BRCT: (A2) Abraxas complex were solved by the molecular replacement method using atomic coordinates from PDB entry-1T29 as the search model. Subsequent structure refinement was done by *refmac-5* and several cycles of manual model building using *oot*[13].

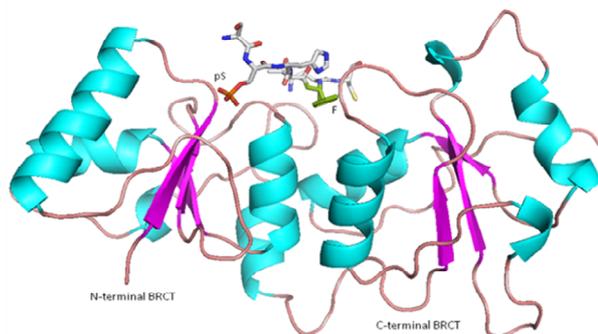
### 2.7 In-silico analysis

Sequence alignment were carried out using Clustal W [14]. All the structures were superimposed using Pymol ([www.pymol.com](http://www.pymol.com)) and root mean square deviation were calculated. The **Protein Interfaces, Surfaces and Assemblies** (PISA) software from EBI were used to analyse the dimer interface[15].

In **chapter 3** of the thesis molecular association between BRCA1 BRCT and NCoA2 peptide has been investigated using biophysical and crystallographic techniques.

The nuclear receptor co-activator NCoA2 is a potential substrate of ATM (Ataxia Telangiectasia Mutated) and ATR (Ataxia Telangiectasia and Rad3-related protein), and BRCA1 is known to interact with NCoA2, but the molecular mechanism is poorly understood. BRCA1 BRCT is a phosphopeptide recognition module, and the NCoA2 sequence was analyzed for the presence of a conserved motif. It was found to be present in the N-terminus region of the protein. In order to confirm the interaction between the NCoA2 and BRCA1 BRCT, the synthetic peptide representing the conserved motif of NCoA2 (NH<sub>2</sub>-P-R-R-N-pS-H-T-F-N-C-COOH), has been titrated with purified BRCA1

BRCT (1646-1859) using isothermal titration calorimetry. The binding affinity was  $K_d = 0.08\mu\text{M}$ . The BRCA1 BRCT oligopeptide complex was crystallized, and diffraction quality of the crystals was found to be very good. These crystals were cryo-protected with 30% glycerol in mother liquor and flash frozen in liquid nitrogen. The diffraction data were collected at the BM14 beam line in ESRF, Grenoble, France.

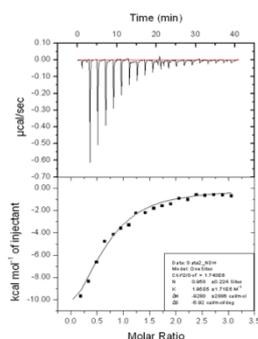


**Figure-1:** BRCA1 BRCT-NCOA2 Complex

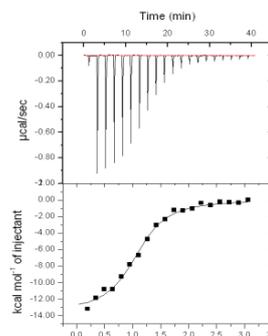
The data to the resolution of  $1.9\text{\AA}$  showed 87.1% completeness with  $R_{\text{sym}}$  value of 7.7%. The structure of the BRCA1 BRCT-NCOA2 complex was solved by using the molecular replacement method, and was refined using *refmac-5* to conventional  $R_{\text{factor}}$  of 22% ( $R_{\text{free}}$  24%). The complex structure consisted of two BRCT repeats each with one  $\beta$ -sheet and three  $\alpha$ -helices, arranged in a head-to-tail manner. The phosphorylated serine and phenylalanine of the peptide hold BRCA1 BRCT in a two branched manner. Phosphoserine holds the N-terminal BRCT and phenylalanine binds in the C-terminal hydrophobic pocket. The phosphate group of serine forms hydrogen bonds with Gly-1656, Lys-1702, and Ser-1655, and phenylalanine forms hydrophobic interactions with Met-1775, Leu-1701, Phe-1704, and Val-1741 (**Figure 1**). Most of the residues responsible for the interactions are conserved in the BRCT domain family, which indicates an evolutionary function of this folding. Superimposition with the earlier reported structures indicate that the backbone atoms of the four residues from the phosphorylated serine to phenylalanine overlap to within  $0.4\text{\AA}$ ; however there are significant changes at both of the terminal residues. This indicates that the modes of interaction for all of the single phosphoproteins are similar, but differences may lie in the extent of their affinities and mutual exclusions with other partners.

**Chapter 4** is a description of biophysical characterization of BRCA1 BRCT binding with Nup153 and RNA binding domain peptides.

Nup153 is a nuclear transport protein and part of the nuclear pore complex. It has been reported that Nup153 can be a potential interacting partner of the BRCA1 BRCT domain [16]. However, another protein-RNA binding domain has also been reported to be a potential interacting partner of the BRCA1 BRCT [17]. The RNA binding domain helps in the transcription activation function of the BRCA1 BRCT. In order to understand the molecular associations with BRCA1, the synthetic peptides from Nup153 (NH<sub>2</sub>-S-A-G-S-pS-F-V-F-G-T-COOH) and the RNA binding domain (NH<sub>2</sub>-A-S-F-G-pS-T-F-S-S) were examined for their binding with BRCA1 BRCT. The binding affinity using isothermal titration calorimetry estimates the BRCA1 BRCT-Nup153 affinity with dissociation constant,  $K_d$ , of 0.1 $\mu$ M (**Figure 2.1**) and BRCA1 BRCT- RNA binding domain affinity with  $K_d$  value of 5.1 $\mu$ M (**Figure 2.2**).



**Figure 2.1:** BRCA1 BRCT - Nup153

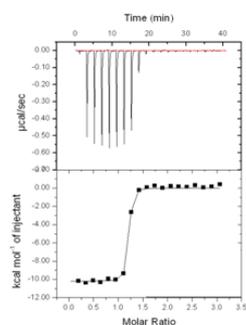


**Figure 2.2:** BRCA1 BRCT - RNA binding domain

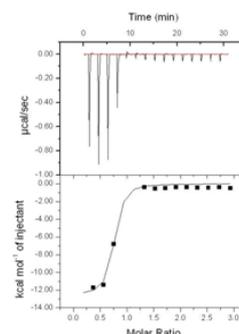
Further, attempts were made to unravel these interactions using crystallographic analysis. The cell parameters of the BRCA1: Nup 153 complex were observed to be,  $a=85.49$ ,  $b=178.72$ ,  $c=194.36\text{\AA}$  and  $\alpha=\beta=\gamma=90^\circ$ . Crystals of the complex however diffracted poorly to the resolution of  $\sim 6.0\text{\AA}$ . We are attempting to improve the diffraction quality of crystals.

Investigation of binding of BRCA1 BRCT with Abraxas is described in **chapter 5** of the thesis.

BRCA1 can perform a DNA repair function by interacting with Abraxas,



**Figure 3.1** BRCA1 BRCT - A1 peptide



**Figure 3.2:**BRCA1 BRCT- A2 peptide

Abraxas is known to be doubly phosphorylated at the Ser-404 and Ser-406 residues [18].

In order to assess the implication of the doubly phosphorylated Abraxas, synthetic peptides containing sequences of Abraxas **A1**=NH<sub>2</sub>-G-F-G-E-Y-S-R-pS<sub>406</sub>-P-T-F-COOH

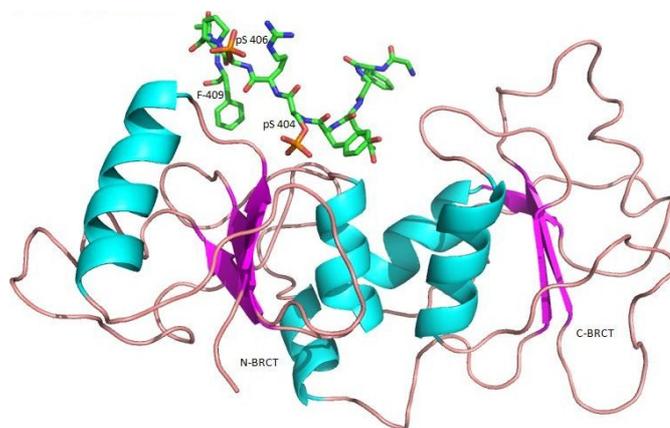
**(Figure 3.1)** and **A2**=NH<sub>2</sub>-G-F-G-E-Y-

pS<sub>404</sub>-R-pS<sub>406</sub>-P-T-F-COOH **(Figure 3.2)**

were procured and isothermal titration calorimetric study has been performed.

We observe that doubly phosphorylated Abraxas interacts more strongly ( $K_d=0.2$   $\mu$ M) with BRCA1 BRCT than singly

phosphorylated. Abraxas ( $K_d=1.2$   $\mu$ M).



**Figure-4:** BRCA1 BRCT: Abraxas complex

After observing the strong association of Abraxas with BRCA1 BRCT, it was decided to get complex crystals with BRCA1 BRCT. These complex crystals were screened for diffraction analysis at a home source and it was found that the BRCA1 BRCT: **A1** complex diffracted poorly to the resolution of  $\sim 6.0$   $\text{\AA}$  and thus further studies were not pursued. The BRCA1 BRCT: **A2** peptide complex diffracted at  $3.5$   $\text{\AA}$ . The intensity data

were processed with completeness of 99% and  $R_{\text{merge}}$  value of 37%. The structure was solved by molecular replacement method and further refined using *refmac-5* ( $R_{\text{factor}}$ -29.83%,  $R_{\text{free}}$ -32.72%). The complex structure of BRCA1 BRCT: Abraxas (**A2**) was arranged in a head-to-tail manner, and the doubly phosphorylated Abraxas peptide held the N-terminal BRCT. The phosphoserine at 404 position forms the main interacting center by creating not only hydrogen bonds with Ser-1655 and Gly-1656, but also hydrophobic interactions with Lys-1702. This is a unique phosphopeptide binding motif which possesses two phosphoserines in the structure (**Figure 4**). Comparing the structure of the BRCA1 BRCT-Abraxas complexes with the earlier reported structures, it was observed that the pSer-404 and pSer-406 of the Abraxas bound to the BRCT with the mutual exclusion of the earlier reported structures. In the reported complex structures the phosphoserine peptides held the N-terminal BRCT, while the Phe held the C-terminal BRCT, however, in the case of doubly phosphorylated peptide (A2) both phosphoserines (pS404 and pS406) were found to interact with the N-terminal BRCT. The atomic coordinates of the refined structure have been deposited in the PDB (ID: 4JLU).

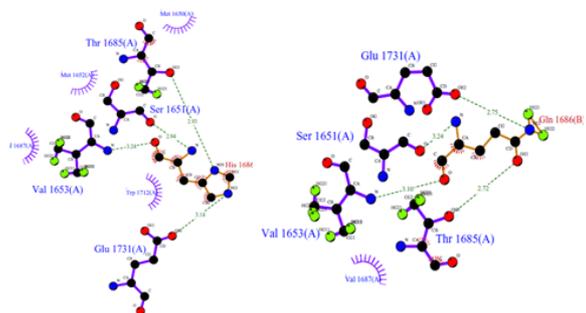
In **chapter 6** of the thesis the observed mis sense mutations have been interpreted in the light of crystallographic and structural biology results, with a view to characterizing genetic biomarkers for genetic counseling.

The BRCA1 gene mutates frequently, leading to the development of hereditary breast and ovarian cancer. All BRCA1 mutations have been documented in the BIC database; these are composed of mostly mis-sense mutations or premature terminations. Such mutations have been discovered throughout the BRCA1 gene's 1863 amino acids [19]. The clinical management of the mutations discovered in the BRCA1 gene across the world is a major challenge. Due to the absence of specific functional assays for BRCA1 and lack of evidence to be segregated with the disease. Taking into account the

importance of the mutations to genetic counseling and structurally guided drug design, all of the possible locations of the reported pathogenic mutations in the BRCA1 BRCT structure were explored and the grade of pathogenicity was evaluated using in-vitro, in-silico, and biophysical approaches [20]. Online servers such as the *Align-GVGD* [21], *Polyphen* [22], *Mutpred* [23] are also useful to predict the pathogenicity of a particular mutation. The major questions are whether these mutations are the cause of breast cancer and how to develop a possible genome-

based diagnostic marker for the disease.

In the C-terminus of the BRCA1 alone, approximately 420 distinct variants/mutations are reported in the BIC's



database. Some of these are predicted to be pathogenic mutations and others have an unknown clinical significance. A few reported pathogenic mis-sense substitutions, such as

BRCA1 H1686Q, P1749R, S1715R, and C1697R have been selected for study here.

These mutant genes have been cloned, and attempts were made to express the mutant proteins in bacterial cells. Some of these mutants were found to be insoluble when

overexpressed in *E. coli* cells and were difficult to purify. However, BRCA1 H1686Q

was purified and biophysically characterized using size exclusion chromatography, circular dichroism (CD) and fluorescence spectroscopy. The gel filtration profile of this

mutation indicates that it is present in the monomeric population, while the CD analysis shows little change in the secondary structure as compared to the wild-type protein.

Fluorescence spectroscopy also shows that the Trp is buried in the hydrophobic pocket. The biophysical data thus suggests that there is no alteration in the tertiary structure of the mutant protein. However, the BRCA1 H1686Q variant showed a loss of transcription

activation function *in vitro*. The protein folding of BRCT domain for each mutation was different from the wild-type. Looking at the three-dimensional structural changes and weak intermolecular interactions it has been found that most of the pathogenic mutations are destabilizing the hydrophobic core of BRCT domains.

**Chapter 7** describes structural characterization of different BRCT domain-containing proteins using in-silico and biophysical approach.

The BRCT domain is present in many DNA repair proteins, including BARD1[24], MDC1[25], and 53BP1 (p53 binding protein) [26]. The BRCT domain in MDC1 acts as a mediator in DNA damage signaling, which is responsible for the activation of the intra-S phase and G2/M checkpoints. It has been well-reported that most of the BRCT repeats are packed in a head-to-tail manner, and that the BRCA1 BRCT domain acts as a phosphopeptide recognition module. However, no reported complex structure is present in the protein database for BARD1 BRCT to identify the residues involved in phosphopeptide binding. The sequence similarity among these three proteins is also not significant. Our sequence analysis shows that BRCA1BRCT has 22.28%, and 11.73% sequence homology with BARD1 BRCT and MDC1 BRCT respectively. BARD1 BRCT and MDC1 BRCT share 10-20% sequence similarity. Interestingly, structural comparison shows significant similarity among the three BRCT domains. Currently, there are 134 redundant structures for BRCT in the protein structural database (PDB). BRCA1 BRCT (PDB ID-1JNX) domain overlaps with BARD1 BRCT (PDB ID-2NTE) and MDC1 BRCT (PDB ID-2ADO) domains with the rmsd values of 1.4Å and 2.8Å respectively. This result further suggests that either properly folded BARD1 dimer interface inhibits the molecular association with binding sequences reported for BRCA1 BRCT or the BARD1 may adopts different binding sequences[20].These variations

highlight the fact that the BRCTs of different proteins may have different binding motifs, unlike BRCA1 BRCT.

**Chapter 8** of the thesis is a concise Summary along with a possible future course of investigation of the problem pursued in the thesis.

The BRCT domain was initially identified as the product of the breast and ovarian cancer gene BRCA1. In the human genome, approximately 23 genes code for BRCT domains, ranging from single to multiple modules. BRCA1 BRCT acts as a phosphoprotein interaction module for signal transduction and DNA double strand repair (DDR), however the molecular association of the BRCT has not been well explored. This study aimed to characterize the BRCA1 BRCT domain with singly phospho-interacting partners, such as NCoA2, NUP153, the RNA binding domain, and doubly phosphorylated Abraxas. The crystal structures of protein-oligopeptide complexes showed that the phosphorylated serine and phenylalanine hold repeats of the BRCT, and the structural details provide insights into why the BRCT prefers pSer over pThr (because the pSer interacting pocket is too shallow to hold pThr).

The results presented here help to elucidate the molecular interactions between BRCA1 BRCT and phosphoproteins, and it was found that singly phospho-NCoA2 interacting residues were frequently mutated in most cases. The Nup153 and RNA binding domains were also characterized as interacting partners of the BRCA1 BRCT domain, and were crystallized with the BRCA1 BRCT domain, but crystals diffracted poorly at the home X-ray source. Abraxas is a novel doubly phospho-interacting partner of BRCA1, and it was found that the doubly phosphorylated Abraxas binds more strongly than the singly phosphorylated Abraxas. The enhanced binding may be due to the extra phosphate group present in the Abraxas, and its crystal structure revealed that both

phosphate groups of the Abraxas interacted with the N-terminal of the BRCT domain. This is the first structure reported to have a doubly phosphorylated association with BRCA1 BRCT.

BRCA1 is found to be mutated in breast and ovarian cancer cases and most of the mutations are of unknown clinical significance. In this study an attempt was made to characterize some of the pathogenic mutations, including BRCA1 C1697R, H1686Q, P1749R, and S1715R. The *in-silico* analysis revealed that the mis-sense mutations (discovered to be pathogenic) present in BRCT modules result in the loss of intramolecular hydrogen bonding and hydrophobic interactions, which ultimately leads to destabilization of the hydrophobic pocket. In conclusion, it was observed that all reported pathogenic mutations were located in the three-dimensional hydrophobic core of the BRCT. Most of the mutations were insoluble when expressed in *E. coli*; however the BRCA1 H1686Q mutation could be purified showing slight alterations in the structure and transcription activation function. Using the multimodal approach it can be concluded that a particular mutation discovered in BRCA1 BRCT is pathogenic, and this information will help clinicians in to implement genetic counseling and clinical management.

### **References:**

1. Hall, J.M., et al., *Linkage of early-onset familial breast cancer to chromosome 17q21*. Science, 1990. **250**(4988): p. 1684-9.
2. Futreal, P.A., et al., *BRCA1 mutations in primary breast and ovarian carcinomas*. Science, 1994. **266**(5182): p. 120-2.
3. Hu, Y.-F., Z.L. Hao, and R. Li, *Chromatin remodeling and activation of chromosomal DNA replication by an acidic transcriptional activation domain from BRCA1*. Genes & development, 1999. **13**(6): p. 637-642.
4. Deng, C.-X., *Roles of BRCA1 in centrosome duplication*. Oncogene, 2002. **21**(40): p. 6222.
5. Monteiro, A.N., A. August, and H. Hanafusa, *Evidence for a transcriptional activation function of BRCA1 C-terminal region*. Proceedings of the National Academy of Sciences, 1996. **93**(24): p. 13595-13599.

6. Lokesh, G., et al., *Thermodynamics of phosphopeptide tethering to BRCT: the structural minima for inhibitor design*. Journal of the American Chemical Society, 2007. **129**(35): p. 10658-10659.
7. Brzovic, P.S., et al., *Binding and recognition in the assembly of an active BRCA1/BARD1 ubiquitin-ligase complex*. Proceedings of the National Academy of Sciences, 2003. **100**(10): p. 5646-5651.
8. Roehm, P.C. and J.M. Berg, *Sequential metal binding by the RING finger domain of BRCA1*. Biochemistry, 1997. **36**(33): p. 10240-10245.
9. Manke, I.A., et al., *BRCT repeats as phosphopeptide-binding modules involved in protein targeting*. Science Signaling, 2003. **302**(5645): p. 636.
10. Williams, R.S., R. Green, and J.M. Glover, *Crystal structure of the BRCT repeat region from the breast cancer-associated protein BRCA1*. Nature Structural & Molecular Biology, 2001. **8**(10): p. 838-842.
11. Yu, X., et al., *The BRCT domain is a phospho-protein binding domain*. Science Signaling, 2003. **302**(5645): p. 639.
12. Bailey, S., *The CCP4 suite: programs for protein crystallography*. 1993: Daresbury Laboratory.
13. Emsley, P., et al., *Features and development of Coot*. Acta Crystallographica Section D: Biological Crystallography, 2010. **66**(4): p. 486-501.
14. Thompson, J.D., D.G. Higgins, and T.J. Gibson, *CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice*. Nucleic acids research, 1994. **22**(22): p. 4673-4680.
15. Krissinel, E. and K. Henrick, *Inference of macromolecular assemblies from crystalline state*. Journal of molecular biology, 2007. **372**(3): p. 774-797.
16. Lemaitre, C., et al., *The nucleoporin 153, a novel factor in double-strand break repair and DNA damage response*. Oncogene, 2012. **31**(45): p. 4803-4809.
17. Anderson, S.F., et al., *BRCA1 protein is linked to the RNA polymerase II holoenzyme complex via RNA helicase A*. Nature genetics, 1998. **19**(3): p. 254-256.
18. Wang, B., et al., *Abraxas and RAP80 form a BRCA1 protein complex required for the DNA damage response*. Science Signaling, 2007. **316**(5828): p. 1194.
19. Bic, L. and J.P. Gilbert, *Learning from AL: new trends in database technology*. Computer;(United States), 1986. **19**(3).
20. Ashok, K., *Pathogenicity of Mutations Discovered in BRCA1 BRCT Domains is Characterized by Destabilizing the Hydrophobic Interactions*. Journal of Cancer Science & Therapy, 2012.
21. Tavtigian, S.V., et al., *Comprehensive statistical study of 452 BRCA1 missense substitutions with classification of eight recurrent substitutions as neutral*. Journal of medical genetics, 2006. **43**(4): p. 295-305.
22. Adzhubei, I.A., et al., *A method and server for predicting damaging missense mutations*. Nature methods, 2010. **7**(4): p. 248-249.
23. Li, B., et al., *Automated inference of molecular mechanisms of disease from amino acid substitutions*. Bioinformatics, 2009. **25**(21): p. 2744-2750.
24. Birrane, G., et al., *Crystal structure of the BARD1 BRCT domains*. Biochemistry, 2007. **46**(26): p. 7706-7712.
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*. Journal of Biological Chemistry, 2005. **280**(37): p. 32053-32056.
26. Joo, W.S., et al., *Structure of the 53BP1 BRCT region bound to p53 and its comparison to the Brca1 BRCT structure*. Genes & development, 2002. **16**(5): p. 583-593.

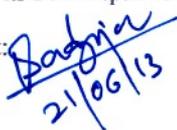
**Publications from thesis**

- 1) **Badgujar D C**, Sawant U, Mahadik H, Gadewal N, Varma AK (2012) Pathogenicity of Mutations Discovered in BRCA1 BRCT Domains is Characterized by Destabilizing the Hydrophobic Interactions. **J Cancer Sci Ther** 4: 386-393. doi:10.4172/1948-5956.1000172.

Other publications

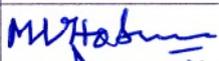
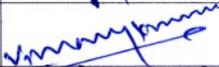
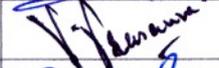
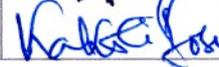
- 2) **Badgujar D C**, Varma A K(2009) Structural Biology : A Novel Tool for Drug Discovery. **Indian Association of Cancer Research, Newsletter.**
- 3) Vikrant, Shah Disha, Nakhwa Pallvi, **Badgujar DC**, Varma AK (2013) Characterization of the MERIT40 for its DNA repair function. **Manuscript - Under Revision- PLoS ONE.**

Signature of Student:


  
21/06/13

Date:

**Doctoral Committee:**

S. No.	Name	Designation	Signature	Date
1	Dr. M. V. Hosur	Chairperson		18/6/13
2	Dr. Ashok Varma	Convener		18/6/13
3	Dr. Vinay Kumar	Member		18/6/13
4	Dr. V. Prasanna	Member		21/6/13
5	Dr. Kakoli Bose	Member		21/6/13


  
26.6.13

Dr. S.V. Chiplunkar  
Chairman Academics and  
Training Program  
ACTREC



Dr. R. Sarin  
Director  
ACTREC


  
26.6.13

Dr. S.V. Chiplunkar  
Dy. Director  
ACTREC



Prof. K. Sharma  
Director, Academics  
Tata Memorial Centre

## List of Figures

Figure No.	Figure title	Page No.
1.1	Tumor spectrum in adults by cell type.	2
1.2	Estimated cancer cases in India	3
1.3	Contribution of different breast cancer susceptibility genes towards development of breast cancer	4
1.4	Role of BRCA1 transcriptional regulation	6
1.5	BRCA1 participates in NHEJ type of DNA repair.	8
1.6	BRCA1 participates in DNA damage signaling.	9
1.7	Domain organizations and interacting partners of BRCA1	11
1.8	Ribbon representation of BRCA1-BARD1 ring domain heterodimer (PDB ID-1JM7)	14
1.9	PDBsum analysis BRCA1-BARD1 RING domains complex (PDB ID-2JM7)	15
1.10	Ribbon representation of the BRCA1 BRCT domain (PDB ID- 1JNX).	16
1.11	Ligplot analysis of BRCA1 BRCT- BACH1 complex (PDB ID: 1T29)	18
1.12	Ligplot analysis of BRCA1 BRCT-CtIP complex (PDB ID-1Y98).	20
1.13	Ligplot analysis of BRCA1 BRCT: ACC1 (PDB ID: 3COJ).	21
1.14	Ligplot analysis of BRCA1BRCT-ATRIP (PDB ID: 4IGK).	23
1.15	Ligplot analysis of BRCA1BRCT-BAAT (PDB ID: 4IFI)	24
1.16	Superposition of BRCA1 BRCT domain complex structures	24
1.17	Multiple sequence alignment of the BRCA1 central region (exon 11-13)	27
1.18	Multiple sequence alignment of the BRCA1 BRCT domain.	28
1.19	The positions of selected mis-sense mutations in BRCA1 BRCT domain structure	29
1.20	The surface occlusion by Lys-1775 mutant.	30
2.1	pGEX-kT vector map	35
2.2	Typical circular Dichroism spectra of protein.	48
2.3	Phase diagram of protein crystallization mediated by precipitant.	51
2.4	Pictorial representation crystallization methods.	52
2.5	Dialysis method in protein crystallization	53
2.6	Schematic representation of liquid-liquid diffusion method	53
3.1	The mechanism of transcription co-activation using NCoA2 protein.	64
3.2	Domain organization of NCoA2 protein	69
3.3	PCR amplification of BRCA1 BRCT (A) and screening of potential clones of BRCA1 BRCT (B).	73
3.4	Protein expression profile of BRCA1 BRCT	74
3.5	FPLC chromatogram of BRCA1 BRCT domain.	74
3.6	SDS-PAGE profile of FPLC purified fractions of BRCA1 BRCT domain	75
3.7	The mass spectrometry profile of BRCA1 BRCT.	75
3.8	Representative ITC result obtained for the interaction of BRCA1 BRCT with NCoA2 peptide	76
3.9	The final atomic model and 2Fo-Fc electron density map for the NCoA2 peptide at 1.7 Å resolution.	78
3.10	NCoA2 peptide recognition by BRCA1 BRCT domain.	79
3.11	Electrostatic surface representation of BRCA1 BRCT-NCoA2 complex complex structure.	80
3.12	The phosphate group of peptides from NCoA2 (A) and BACH1 (B) forms three hydrogen bonds with BRCA BRCT domain	80
3.13	The hydrogen bond network formed by Arg-1699 with NcoA2 peptide.	81

3.14	Space filling model representing environment around Phe+3 of NCoA2 peptide	81
3.15	Contribution of individual residues in the phospho-peptide to the surface area burial in the BRCT/NCoA2 complexes	82
3.16	Two-dimensional representation of the interactions between BRCA1 BRCT (purple) and NCoA2 residues (purple).	83
3.17	Structural superposition of oligopeptides represented in stick model, where BACH1 (Cyan) CtIP (red), ACC1 (green) and NCoA2 (yellow),	83
4.1	Domain organisation of Nup153	86
4.2	Domain organisation of RBP-12	86
4.3	Representation of ITC result obtained from interaction of BRCA1 BRCT domain with Nup153 peptide.	90
4.4	Representation of ITC result obtained from interaction of BRCA1 BRCT domain with Nup153 peptide.	90
4.5	Crystals of BRCA1 BRCT domain complex with Nup-153 and RBP-13	91
4.6	Diffraction image of BRCA1 BRCT-Nup153 complex crystal	92
4.7	BRCA1 BRCT-Nup153 model with electron density map	93
4.8	Diffraction image of BRCA1 BRCT-RBP-12 complex crystal	93
4.9	RMSD of BRCA1 BRCT-Nup modelled structure over 2000 pico second.	94
4.10	Ribbon representation of BRCA1 BRCT-Nup153 model	95
4.11	RMSD of BRCA1 BRCT-RBP-12 modelled structure over 2000 pico seconds.	96
4.12	Ribbon representation of BRCA1 BRCT-RBP-12 model.	96
4.13	Structural comparison of Nup153, RBP and CtIP peptides bound to BRCA1 BRCT domain.	97
5.1	Homologous recombination repair by BRCA1 A DNA repair complex	101
5.2	Representation of ITC results obtained from interaction between BRCA1 BRCT domain and A1 peptide.	105
5.3	Representation of ITC results obtained from interaction between BRCA1 BRCT domain and A2 peptide.	105
5.4	Representation of ITC results obtained from interaction between BRCA1 BRCT domain and A3 peptide.	106
5.5	Electron density map (2fo-fc) covering the A1 Abraxas peptide	110
5.6	Superposition of three BRCA1 BRCT-A1 complex peptides	110
5.7	A1 peptide recognition by BRCA1 BRCT domain.	111
5.8	Electrostatic surface representation of BRCA1 BRCT-A1 complex structure	111
5.9	The phosphorylated serine present at 406 position in A1 peptide form three hydrogen bonds BRCA1 BRCT domain.	112
5.10	Space filling model representing environment around Phe (409) of A1 peptide.	113
5.11	Schematic representation of weak intermolecular interaction between Abraxas A1 peptide and BRCA1 BRCT domain	113
5.12	Electron density map (2fo-fc) covering the A2 Abraxas peptide	114
5.13	A2 peptide recognition by BRCA1 BRCT domain.	115
5.14	Phosphate group of Serine residue present at 404 position forms two hydrogen bonds with BRCA1 BRCT domain.	116
5.15	The phosphoserine at 406 positions is stabilized by three hydrogen bonds with the symmetry related molecule Thr-1677.	116
5.16	Schematic representation of weak intermolecular interaction between Abraxas A2 peptide and BRCA1 BRCT domain,	117
5.17	The phenyl ring is stabilized by hydrogen bonds from Glu-1698 and Thr-1692	127

5.18	Superposition of stick model of A1 peptide (green) with A2 peptide (magenta).	118
6.1	Schematic representation of purification of BRCA1 BRCT mis-sense mutants.	124
6.2	PCR amplification of BRCA1 BRCT domain (~650bp) A), The restriction digestion of potential clones B).	127
6.3	Cation exchange chromatographic profile of BRCA1 H1696Q mutant	128
6.4	Anion exchange chromatographic profile BRCA1 H1696Q mutant,	128
6.5	Gel filtration chromatogram of BRCA1 H 1686 Q mutant protein.	129
6.6	The SDS-PAGE profile of FPLC purified BRCA 1 H 1686 Q protein,	129
6.7	CD spectra for wild- type BRCA1 BRCT (blue) and BRCA1 H 1686Q mutant (red) in near UV range.	131
6.8	Fluorescence emission spectra of BRCA1 BRCT and BRCA1 H 1686 Q mutant excited at 280 nm.	130
6.9	The superposition of wild type BRCA1 BRCT (green) with modelled mutant BRCA1 P1749R (yellow),	131
6.10	Schematic representations of weak intramolecular interactions between BRCA1 BRCT wt (A) and pathogenic mutations P1749R (B),	132
6.11	The superposition of wild type BRCA1 BRCT (cyan) with modelled mutant BRCA1 H1686Q (green)	132
6.12	Schematic representations of weak intramolecular interactions between BRCA1 BRCT wild-type (A) and pathogenic mutation H 1686 Q (B)	133
6.13	The superposition of wild type BRCA1 BRCT (green) with modelled mutant BRCA1 S1715 (cyan),	134
6.14	Schematic representations of weak intramolecular interactions between BRCA1 BRCT wild-type (A) and pathogenic mutations S1715R (B)	134
6.15	The superposition of wild type BRCA1 BRCT (green) with modelled mutant BRCA1 C1697 (cyan),	135
6.16	Schematic representations of weak intramolecular interactions between BRCA1 BRCT wild-type (A) and BRCA1 C1697R (B),	135
7.1	Schematic representation of different BRCT domain family members	139
7.2	The multiple sequence alignment of single BRCT repeat containing proteins,	143
7.3	Mapping of conserved residues on XRCC1 BRCT domain	144
7.4	Pair wise secondary structure alignment of RFC1 BRCT with N-terminal BRCT	145
7.5	A-Structural superposition of RFC1 BRCT (PDB ID:-2EBU) (magenta) with N-terminal BRCT (PDB ID:1JNX)	145
7.6	Superposition of different crystal structures of single BRCT repeat containing proteins.	145
7.7	Superposition of different crystal structures of single BRCT repeat containing proteins	146
7.8	Multiple sequence alignment of tandem BRCT domain containing proteins	149
7.9	Structural superposition of BRCT domain containing proteins,	150
7.10A	Structural superposition of BRCT domains of BRCA1	151
7.10B	Structural superposition of BRCT domains of TopBp1 protein	151
7.10C	The structural superposition of MDC1 N-terminal BRCT domains	152
7.10D	The structural superposition of N and C-terminal BRCT of Brc1	152
7.10E	The structural superposition of N and C-terminal BARD1	153
7.11	Structural superposition of BRCT domains of BARD1 and BRCA1	154
7.12	Pair wise alignment of BARD1 BRCT and BRCA1 BRCT domain	155
7.13	Functional domains BARD1 protein.	156
7.14	Schematic representation of DNA repair function of BRCA1 and	156

	<b>BARD1</b>	
7.15	A) PCR amplification of BARD1 ARD-BRCT domain, B) Restriction digestion of potential clone of BARD1 BRCT.	157
7.16	Protein expression profile of BARD1 ARD-BRCT Protein	158
7.17	SDS-PAGE profile of BARD1 ARD-BRCT domain FPLC fractions	158

---



---

## List of Tables

<b>Table No.</b>	<b>Table title</b>	<b>Page No.</b>
1.1	List of BRCA1 BRCT domain interacting partners	17
1.2	Summary of BRCA1 BRCT domain mis-sense mutant structures.	30
2.1.	Characteristics of expression vectors	36
2.2	Typical 20 $\mu$ l restriction digestion reaction	38
2.3	Typical 50 $\mu$ l PCR reaction	39
2.4	Program for PCR amplification	39
2.5	List of sequencing primers	41
2.6	Characteristics of different superdex chromatographic media used in GFC	45
3.1	ITC experimental and injection parameters.	69
3.2	Summary of data collection, processing and refinement statistics	71
3.3	RMSD values for protein C $\alpha$ atoms obtained upon superposition of different BRCA1 BRCT domain complex structures	78
4.1	ITC experimental and injection parameters	87
5.1	ITC experimental and injection parameters.	102
5.2	Initial crystallization condition for BRCA1:A1 and BRCA1:A2 complex	107
5.3	Data collection and refinement statistics	108
5.4	The phi and psi angle for the residues of A1 and A2 peptide.	119
6.1	List of BRCA1 BRCT domain mis-sense mutations selected for study	122
6.2	List of site- directed- mutagenesis primers for BRCA1 mis-sense mutants	123
7.1	List of non-redundant BRCT domain containing protein structures from PDB.	140
7.2	Comparison of two different BRCT domain containing protein crystal structures	146
7.3	RMSD in $\text{\AA}$ upon comparison of two different single BRCT domain containing solution structures	147
7.4	RMSD values expressed in $\text{\AA}$ upon comparison of two different BRCT domain containing proteins. In parenthesis, the percentage identified of superposed amino acids is given	147
7.5	The structural superposition of N-terminal BRCT domains from BRCA1, MDC1, BARD1, TopBp1 and Brc2 protein. RMSD values in $\text{\AA}$ and percent identity given in parenthesis.	148
7.6	The structural superposition of C-terminal BRCT domains from BRCA1, MDC1, BARD1, TopBp1 and Brc2, protein. RMSD values in $\text{\AA}$ and percent identity given in parenthesis.	148
7.7	The sequence similarity and rmsd between N-terminal and C-terminal BRCT repeat for respective proteins.	153
7.8	List of BRCT domain containing proteins with preferred consensus sequence of interaction	154

---

---

## Abbreviations

BC	: Breast cancer
HBC	: Hereditary Breast Cancer
BRCA1	: BRCA1 Associated Ring Domain-1
DNA	: DeoxyriboNucleic Acid
RNA	: RiboNucleic acid
CtIP	: CtBP Interacting Protein
RING	: Really Interesting New Gene
BARD1	: BRCA1 Associated Ring Domain-1
TCR	: Transcription-Coupled DNA Repair
HR	: Homologous Recombination
NHEJ	: Non-Homologous End Joining
ATM	: Ataxia Talangiectasia Mutated
ATR	: ATM and RAD3 related
MDC1	: Mediator of DNA Damage Checkpoint -1
RNF	: RING finger containing protein
UBC	: Ubiquitin Conjugating Enzyme
MRE11	: Meiotic Recombination 11 homolog
MRN	: Mre11-Rad50-Nbs1 complex
RAP80	: Receptor Associated Protein-80
HDAC	: Histone DeAcetylases
RB	: Retinoblastoma protein
Chk1	: Checkpoint kinase 1
BRCT	: BRCA1 C-Terminal domain
BAP1	: BRCA1 Associated Protein 1
DB-I	: DNA Binding domain-I
PARP1	: Poly ADP Ribose Polymerase
BIC	: Breast cancer Information Core
BACH1	: BTB and CNC homology 1

ACC1	: Acetyl Coenzyme A Carboxylase
ATRIP	: ATR Interacting Protein
$\mu\text{M}$	: micro Molar
GST	: Glutathione S-Transferase
FAS	: Fatty Acyl Synthase complex
BIC	: Breast Cancer Information Core
LB	: Luria Bertani
SDS PAGE	: Sodium Dodecyl Sulfate Poly-Acrylamide Gel Electrophoresis
PCR	: Polymerase Chain Reaction
UV	: Ultra Violet
IPTG	: Isopropyl $\beta$ -D-1 thiogalactopyranoside
GFC	: Gel Filtration Chromatography
HIC	: Hydrophobic Interaction Chromatography
CD	: Circular Dichroism
MALDI-TOF	: Matrix-Assisted Laser Desorption and Ionisation-Time Of Flight
ESI	: Electron Spray Ionisation
ITC	: Isothermal Titration Calorimetry
MR	: Molecular Replacement
COOT	: Crystallographic Object-Oriented Toolkit
NCoA2	: Nuclear receptor CoActivator 2
AR	: Androgen Receptor
ER	: Estrogen Receptors
TEV	: Tobacco Etch Virus protease
CCP4	: Collaborative Computational Project No.4
FPLC	: Fast Protein Liquid Chromatography
LLG	: Log Likelihood Gain
RBP-12	: RNA binding proteins 12
MDS	: Molecular Dynamics Simulations
Align GV-GD	: Align Grantham Variation Grantham Deviation

OPLS	: Optimized Potentials for Liquid Simulations
NMR	: Nuclear Magnetic Resonance
RMSD	: Root Mean Square Deviation
Top BP1	: Topoisomerase (DNA) II Binding Protein 1
MCPH1	: Microcephalin1
FCP1	: F-cell Production 1
DNA Pol $\lambda$	: DNA polymerase $\lambda$
53BP1	: p53 Binding Protein
Lig3	: DNA ligase 3
PARP1	: Poly (ADP-ribose) Polymerase 1
RFC	: Replication Factor C
ARD	: Ankyrin Repeat Domain
SPR	: Surface Plasmon Resonance

---

*Introduction and review of literature*

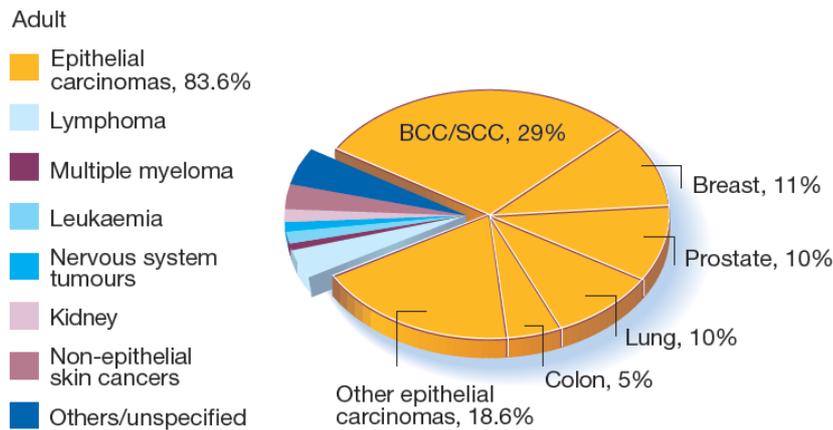
---

*Chapter 1*

## 1.1. Cancer

Cancer is a major cause of death in humans, and has existed since the first humans walked the earth. Archaeologists have found evidence of tumor in a 2700 year old human skeleton [27]. The World Health Organization (WHO) has reported 14.1 million new cases and 8.2 million deaths in 2012 due to cancer. Cancer has been recorded in a variety of biological species such as molluscs, arthropods, jawless fish, cartilaginous and bony fish, amphibians, reptiles and mammals [28, 29]. Most of the cancers originate from an individual cell. Normally, cells in the organisms grow and divide in a controlled way. However, when this cell-cycle regulation is lost, the property of contact inhibition is also lost, and the resulting uncontrolled growth leads to the formation of primary tumor. Cancer cells can detach from the primary tumor and circulate through the blood and the lymphatic system and invade other parts of the body to form a new tumor. The formation of a new tumor from a primary tumor is called metastasis [30]. Therefore, cancers are broadly categorised into two types: primary cancers (in the organ or tissue where cancer originated) and secondary cancers (metastatic tumor). Depending upon what types of normal cells are converted into cancerous cells, cancers may also be classified into the following different types:

- a) **Carcinoma**: cancer that begins in the skin cell or in the outer lining of organs such as the liver or kidney
- b) **Sarcoma**: cancer that arises from altered cells of mesenchymal origin
- c) **Leukemia**: cancer of the blood or bone marrow
- d) **Lymphoma**: cancer of the lymphocytes
- e) **Myeloma**: cancer of plasma cells from the bone marrow
- f) **Mixed types**: include the tumors developed from different tissues.



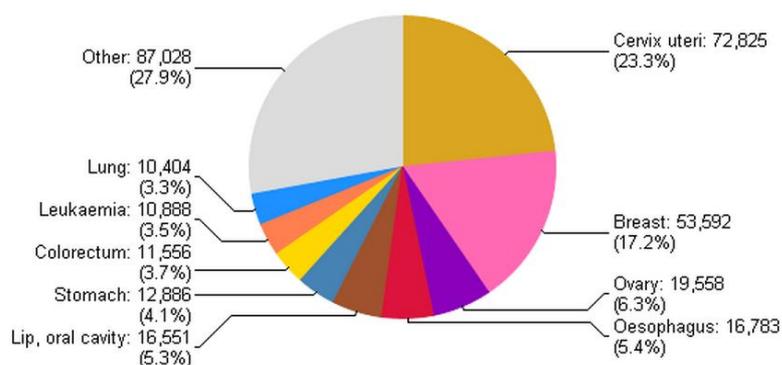
**Figure-1.1: Tumor spectrum in adults by cell type.** The frequency of Basal cell carcinoma (BCC) and squamous cell carcinoma (SCC) account for 29% and rank highest among all cell types.

Cancer can be in the form of either a solid tumor or a liquid / “soft” tumor. The genetic material of cancerous cells is altered in specific regions, compared to normal cells, leading to the theory that “cancer is due to genetic changes”. The genetic alterations may be due to external factors such as radiation, carcinogens, or due to internal factors such as deficiencies in the DNA damage repair mechanisms within the cells [31]. The genes which are altered or mutated in cancer are described as cancer-susceptibility genes. These genes can be divided into three major categories: gatekeeper genes, caretaker genes and landscaper genes [31, 32]. Gatekeeper genes are responsible for regulation of cell growth and differentiation, and these include tumor suppressor genes and oncogenes [33]. Caretaker genes maintain the genomic integrity of the cell, and when affected, they are indirectly responsible for cancer progression [34]. Landscaper genes indirectly affect cancer progression by controlling cellular environment around the tumor.

### 1.1.1. Breast Cancer

As shown in **Figure-1.1**, breast cancer (BC) is the third largest public health problem worldwide. According to Global Cancer Statistics 2011 [35], the BC incidence rate in India is around (17.2%) of all the cancers, and stands second in the number of cancer

cases overall (**Figure-1.2**). In developed countries, one in ten women is diagnosed with BC, while in developing countries, the rates are slightly lower, but the incidences are increasing day by day.



**Figure-1.2: Estimated cancer cases in India.** The incidence of breast cancer in India is about 17% which stands second when compared to other cancer types.

The most prominent and well established risk factors for BC are early menarche (beginning of menstrual function) and late age at first childbirth. Breast cancer (BC) can be divided into two categories, sporadic BC and hereditary BC. The genetic changes in sporadic BC are due to external factors while they are inherited in hereditary BC. The inherited mutations are mostly in breast cancer susceptibility genes, specifically, *BRCA1* and *BRCA2* [36-38].

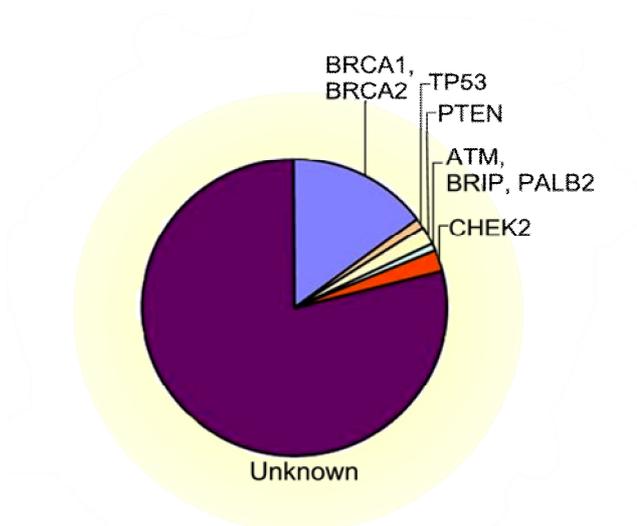
#### 1.1.1.1. Sporadic Breast cancer

Though sporadic breast cancer (SBC) accounts for more than 80% of all BC, very limited information exists about the definite causes. One of the major causes is assumed to be the protein AKT1 serine/threonine kinase, which is resident in the cellular membrane. *AKT1* is an oncogene that is regulated by various extracellular factors [39]. In 50% of SBC cases, there is an over-expression of AKT1 protein [40]. AKT1 is associated with *BRCA1*, and is responsible for the sequestration of *BRCA1* in the cytoplasm. This leads

to the non-availability of BRCA1 in the nucleus to carry out the DNA repair mechanism [40]. However, its genesis and progression is poorly understood.

### 1.1.1.2. Hereditary Breast Cancer

Paul Broca in 1866 was the first person to describe the high prevalence of breast cancer among the members of one family, suggesting linkage to heredity [41]. Hereditary breast cancer (HBC) can be caused by mutations in the group of genes involved in DNA damage repair pathways. These genetic variants are passed from generation to generation. HBC appears to be autosomal dominant in character. Approximately 10-30% of BC is considered to be HBC, whereas only 5-10% of BC cases are discovered with strong hereditary component [42]. The key evidence came after years of efforts made by Hall *et al* (1990) to identify, by linkage analysis, the breast cancer susceptibility gene. The gene was found on the long arm of human chromosome 17 [1]; this gene was subsequently named *BRCA1* [43]. *BRCA1* is known to act as a tumor suppressor gene because in more than 90% of breast cancer cases, this gene is found to be mutated [30, 31].



**Figure-1.3: Contribution of different breast cancer susceptibility genes towards development of breast cancer.** *BRCA1 and BRCA2 are found to be involved more in development of breast cancer compared to others.*

Also transfection of wild-type *BRCA1* into human tumor cell lines results in growth inhibition through cell-cycle arrest. Other known breast cancer susceptibility genes are shown in **Figure-1.3**, but *BRCA1* and *BRCA2* are the most frequently mutated in BC.

## **1.2. BRCA1 (breast cancer 1, early onset)**

The gene *BRCA1* is about 100 kb long and comprises of 24 exons, of which exon one is non-coding while exon four has Alu repeats (transposable element) [44]. *BRCA1* encodes a 7.8 kb mRNA that codes for a BRCA1 protein having 1,863 amino acids and a molecular weight of 220 kDa [37]. BRCA1 possesses two nuclear localization signals which help the protein to cross the nuclear membrane easily. BRCA1 is conserved in mammals but not in lower animals [45]. BRCA1- knockout mice die between embryonic day 7.5 and day 13, suggesting that BRCA1 is essential for embryonic development [46, 47] .

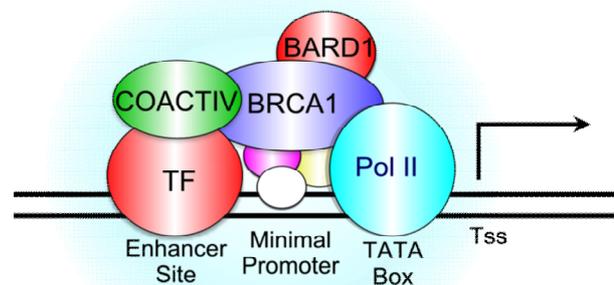
### **1.2.1. BRCA1 functions**

BRCA1 interacts with several proteins and performs, directly or indirectly, various functions such as transcriptional activation [48], chromosomal remodelling [3], DNA repair signalling [49], cell -cycle control [50], and apoptosis [51]. The most important involvement of BRCA1, though, is to maintain the genomic stability by performing DNA repair through homologous recombination [52] and non-homologous end joining [53].

#### **1.2.1.1. Transcriptional activation by BRCA1**

Mammalian gene transcription is a complicated process driven by the molecular complex containing RNA polymerase II, which is a multi-subunit enzyme [54]. RNA polymerase II, along with some general transcriptional factors, assembles at a promoter site (TATA box) and initiates the transcription **Figure-1.4** [17]. The transcription process can be stimulated or inhibited by interaction with co-activators or co-repressors respectively. The

function of BRCA1 as a transcriptional activator was discovered when the C-terminal of BRCA1 fused with a GAL4 DNA binding domain lead to transcriptional activation in both yeast and mammalian systems [55]. Further, BRCA1-null mice or BRCA1-deficient cells are unable to execute a transcription-coupled DNA repair (TCR) mechanism [56]. The genes that are actively transcribed undergo TCR. Though there is no direct evidence that BRCA1 can perform TCR, a possible clue is that BRCA1 is co-purified with RNA polymerase II [57]. BRCA1 does not bind to DNA in a sequence-specific manner but, rather binds to abnormal or damaged DNA. The proteins p21 and GADD45 are cell-cycle regulators, and the transcriptional activation of their genes is performed by BRCA1 with the help of protein p53, which is also a tumor suppressor gene product [58]. The combination of BRCA1/p53 is also responsible for the activation of cell survival and repair pathways [59].



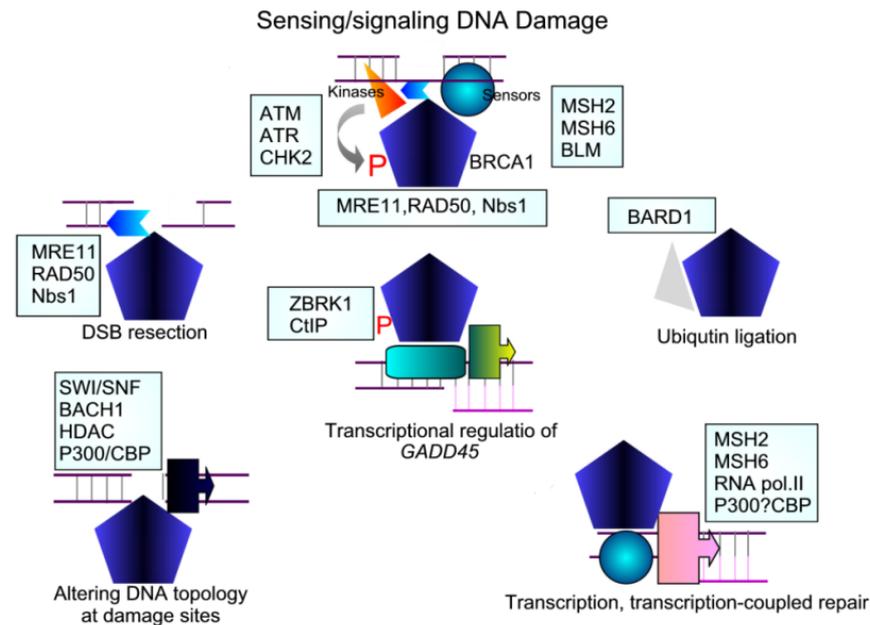
**Figure-1.4: Role of BRCA1 transcriptional regulation.** *BRCA1 interacts with basic transcription machinery components such as RNA pol-II and RNA helicase A, which ultimately leads to transcriptional activation of downstream genes.*

The BRCT domain of BRCA1 interacts with CtIP (CtBP Interacting Protein) and facilitates transcriptional repression [60]. CtIP was originally identified as a binding partner of CtBP that ultimately suppresses transcription by interacting with pRB and p300 proteins [61]. The transcriptional regulation function of BRCA1 depends upon the

presence or absence of damaged DNA. In the absence of damaged DNA, BRCA1 interacts with RNA helicase A, which is a part of the RNA polymerase II complex, and helps in transcriptional regulation [57, 62-64]. In the presence of damaged DNA, the two RING domains of BRCA1-BARD1 complex are known to associate with RNA polymerase II at the transcriptional site blocked due to DNA damage. The exact function of BRCA1-BARD1 complex with respect to RNA polymerase II is unclear, but it may help in the ubiquitination and degradation of RNA polymerase II components. This degradation facilitates the binding of BRCA1 with DNA and also further recruitment of DNA repair proteins [65]. It is believed that TF II, a component of RNA polymerase II complex, can shift to repair mode in stalled RNA polymerase [66].

#### **1.2.1.2. Role of BRCA1 in DNA repair**

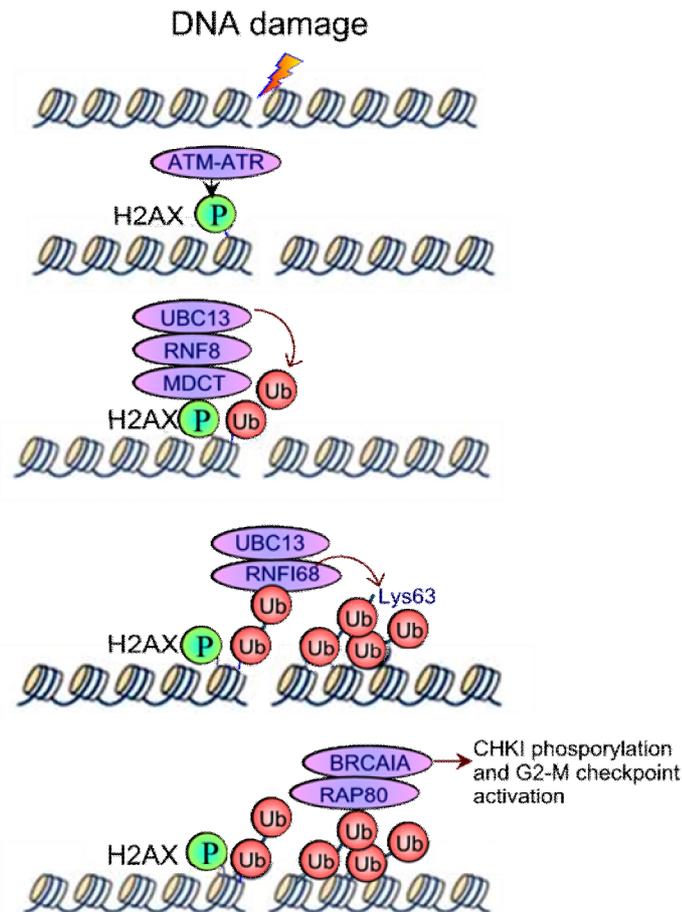
Various *in-vitro* binding assays demonstrate that BRCA1 can be associated with proteins involved in different DNA repair pathways, such as homologous recombination (HR) and non-homologous end joining (NHEJ) [53]. HR is the most accurate double-strand repair mechanism, and the absence of this pathway leads to genomic instability [52]. The first observation that BRCA1 is involved in DNA repair came when Scully *et al* in 1997 showed that BRCA1 at the DNA damage foci, is associated with RAD51, a DNA repair protein [67]. The RAD51 is a central player in homologous recombination and therefore BRCA1-RAD51 complex repairs the DNA damage by homologous recombination [68] [69]. The MRN complex formed by MRE1, RAD50 and NBS1 is found to interact with BRCA1 and help in HR in DT40 cells [70]. However, the molecular mechanisms by which BRCA1 is associated with RAD50, and how it affects the function of the MRN complex is still not clear **Figure-1.5**.



**Figure-1.5: BRCA1 participates in NHEJ type of DNA repair.** The ATM and ATR phosphorylate BRCA1; in turn BRCA1 recruits RAD50, MRE11 and Nbs1 complex to the site of DNA damage. Different proteins which are also recruited to damaged site such as ZBRK1 and CtIP ultimately help in transcription coupled DNA repair and NHEJ.

Upon the occurrence of DNA damage, the ATM and ATR get activated and phosphorylate the histone H2AX. This leads to further recruitment of mediator proteins such as MDC1 (mediator of DNA damage checkpoint-1) at the damage site. MDC1 promotes the accumulation of E3 ubiquitin ligase complex, comprising of RING Finger domain protein-8 (RNF8) and the ubiquitin conjugating enzyme UBC-13, and mediates the ubiquitinylation of H2AX. RNF-168 is another E3 ubiquitin ligase that amplifies the damage signal by forming lys-63 linked chains on H2AX. These chains are recognised by the ubiquitin interacting motif of RAP80 (receptor associated protein-80) **Figure-1.6** [71].

RAP-80 protein is part of the BRCA1-A complex, whose other components are BRCA1, BRCC35, MERIT-40 and Abraxas [72]. The coiled coil domain protein CCDC-98/ Abraxas, on phosphorylation at the Ser-406 residue, is known to interact with BRCA1 BRCT domain [73]. The ubiquitin interacting motif of RAP80 is essential for the retention of the BRCA1-A complex at the DNA damage site [74].



**Figure-1.6: BRCA1 participates in DNA damage signaling.** Upon DNA damage, ATM and ATR phosphorylate histone H2A, which leads to activation of MDC1. Activated MDC1 recruits ubiquitin conjugase and ligase at damage site resulting into the Lys-63 linked ubiquitination on histone. RAP-80 protein recognizes these ubiquitin and forms BRCA1-A complex with the help of Abraxas.

### 1.2.1.3. BRCA1 and Chromatin Remodelling

The alteration of physical states of chromosomes through post-translational modifications, such as phosphorylation, acetylation and methylation, is known as chromosome remodelling [75]. Several protein complexes such as BASC (BRCA1-associated genome-surveillance complex) are involved in this process. BRCA1 is a member of BASC complex, and is also known to interact with the SWI-SNF proteins, which form chromosomal remodelling complexes [76-78]. There is evidence that BRCA1 can interact with histone deacetylases such as HDAC1 and HDAC2, and may be

responsible for the tighter binding of histone with the DNA [79]. Upon chromatin remodelling due to BRCA1, genes such as KU and GADD45 are activated [78]. It is also a known fact that BRCA1 BRCT can interact with proteins like BACH1 that are involved in chromatin remodelling [77].

#### **1.2.1.4. BRCA1 in Cell- Cycle Check Point Control**

The ability to control the precise timing of cell cycle events is essential to repair the damaged DNA prior to cell division [80]. The cellular machinery responsible for cell cycle arrest is called cell- cycle checkpoint. Exposure of cells to ionising radiation results in G2/M checkpoint activation, which allows the cell to repair damaged DNA [80]. Chk1 is a G2/M checkpoint regulator protein and BRCA1 is known to regulate the expression and phosphorylation of Chk1 [81]. The experiment performed by Xu *et al* [82] has shown that cells that are deficient in the wild-type *BRCA1* exhibit a higher mitotic index, compared to the cells that contain wild-type *BRCA1*. BRCA1 is phosphorylated in the G1 and S phases of the cell cycle. It is also found that cells containing wild-type retinoblastoma (RB) are sensitive for BRCA1-dependent G1 arrest, while RB<sup>-/-</sup> (null) cells do not show activation [83]. Further evidence shows that BRCA1 interacts not only with hypo-phosphorylated RB, but also with the RbAp46 and RbAp48 proteins and controls the cell-cycle check point [79].

#### **1.2.1.5. BRCA1 and Apoptosis**

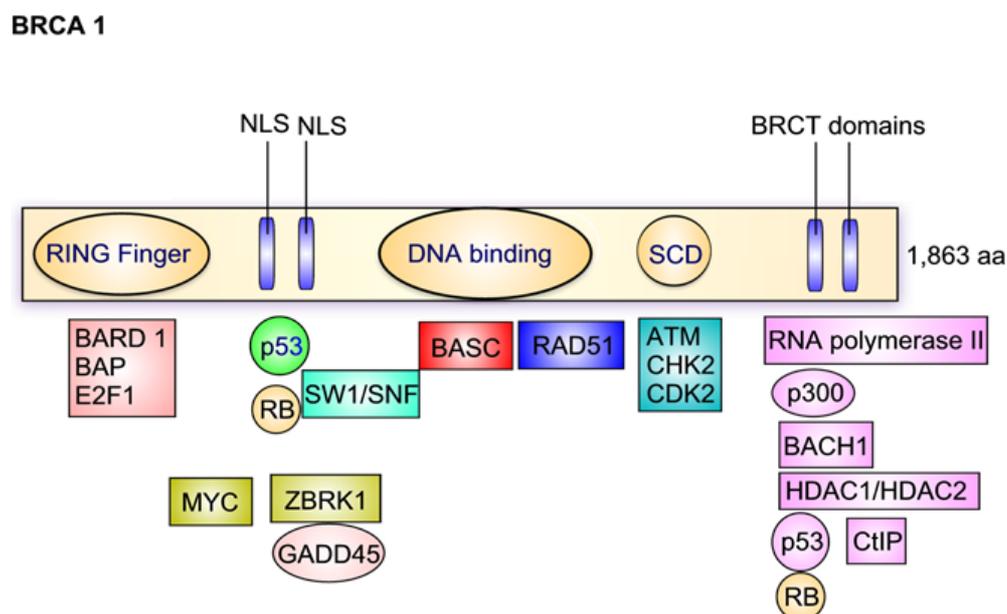
When BRCA1 expression is compromised, cells are more sensitive to apoptosis. But when BRCA1 is over-expressed, the *PIG3*, *Bax* and *PERP*, the p53 dependent apoptosis activating genes are down-regulated [59]. BRCA1 does not always inhibit apoptosis; it is

also responsible for the activation of interferon  $\gamma$  pathway, which ultimately leads to the up- regulation of genes associated with apoptosis [51].

The multifunctional *BRCA1* gene is a subject of extensive studies because mutations in any domains may lead to the formation of breast and ovarian cancer. Studies of genetic signature of *BRCA1* domains might foster the way for targeted therapy and genetic counselling.

### 1.2.2. BRCA1 Domain organization

The BRCA1 protein is comprised of three distinct known domains: the N-terminal ring finger RING domain (~ first 100 amino acids) [84], the central DNA binding domain (452-1079) [85], and the C-terminal BRCT domain (1646-1859) [86, 87]. BRCA1 uses these domains to perform different functions (**Figure-1.7**).



**Figure-1.7: Domain organizations and interacting partners of BRCA1.** Functional domains of *BRCA1* are shown to interact with a number of different proteins

#### 1.2.2.1. Ring Finger Domain

Sequence homology analysis has revealed the presence of a *Really Interesting New Gene* (RING) at the N-terminus of BRCA1 [84]. This domain contains cysteine and histidine residues known to be conserved among Zn-binding proteins of diverse origins and functions. Therefore RING domain of BRCA1 is described as a Zn-binding domain. The ring finger domain is inferred to be crucial for BRCA1 function for the following reasons: i) several clinically important mutations have been found in this region [88], ii) yeast two-hybrid analysis shows that the ring finger domain is involved in the interactions with the BARD1 and BAP1 proteins, which are crucial for the ubiquitin activity of BRCA1 [89, 90].

#### **1.2.2.2. Central DNA Binding Domain**

The capability of BRCA1 to perform DNA-dependent functions like transcriptional activation and DNA damage signaling is due to its ability to interact with DNA. Paull *et al* suggest that the central region of BRCA1 (residues 452–1079) is involved in binding with DNA in a non-specific manner [91]. This suggestion was further supported by the identification of two separate regions, DB-I and DB-II, that bind to DNA with low micromolar affinities [85]. There are few regions at the centre of BRCA1 which are biophysically characterized to be protease-resistant. However, these regions do not form well folded conserved domains. The central regions of BRCA1 are important for several reasons: (i) several clinically relevant mutations have been discovered in the central domain, (ii) the central region has binding affinity to several proteins involved in DNA damage repair, and (iii) multiple phosphorylation sites that are involved in the signalling and repair pathways are present in this region.

### 1.2.2.3. BRCT Domain

The BRca1 C-Terminal (BRCT) domain comprises of about 100 amino acids [92], and BRCT occurs in proteins either as a single unit (TopBP1) [93] or as tandemly linked multiple units separated by a linker of about 25-60 amino acids [94]. BRCA1 contains two tandemly organised BRCT domains. Interestingly, BRCT modules are found in many different proteins [87]. The BRCT domain has been reported in other DNA repair proteins such as 53BP1 [86], MDC1 [95], PARP1 [86], DNA ligase IV [86] etc. The tandem repeats of BRCA1 BRCT are known to interact with specifically phosphorylated binding partners.

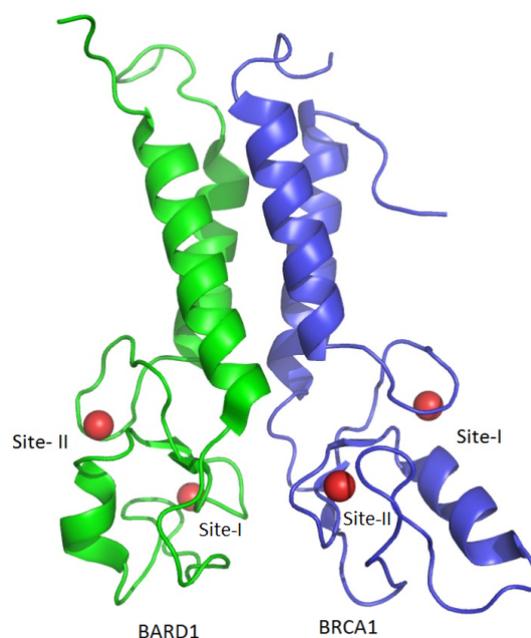
### 1.2.3. Three dimensional structures of functional domains of BRCA1

Full-length BRCA1 is a 220 kDa protein, and has not been purified in a bacterial system. There are no reported structures for full-length BRCA1; however, the individual domains have been expressed, purified and structurally characterised by different investigators [10, 96]. The structures of N-terminal RING and C-terminal BRCT domains have been determined using NMR and X-ray diffraction techniques [10, 97].

#### 1.2.3.1. Structure of RING Domain

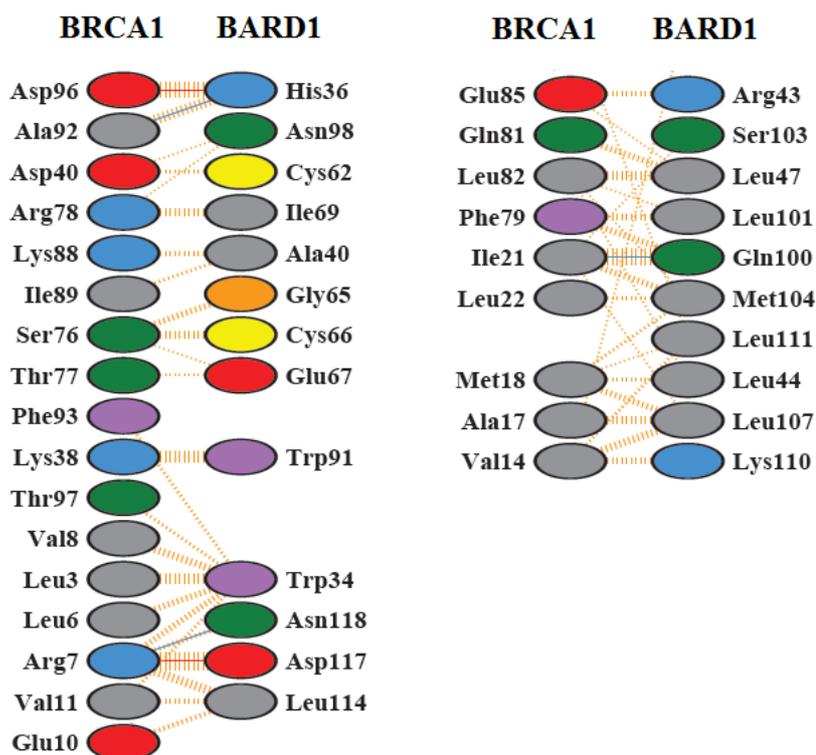
The ring finger domain of BRCA1 forms a homodimer in solution, and this homodimer displays ubiquitin-ligase activity. Similarly the RING domain of BARD1 also dimerises and displays ubiquitin-ligase activity. BRCA1 and BARD1 proteins form heterodimers through interactions between their RING domains. Interestingly, the heterodimer displays several fold higher ubiquitin-ligase activity, compared to that of either homodimer. No crystal structure is available for the RING-RING domain of BRCA1 and BARD1 complex. Klevit *et al* have used NMR to determine the structure (**Figure 1.8**) of the heterodimer in solution [96]. The complex adopts a fold that is highly conserved in other

ring-finger-domain-containing proteins [98, 99]. The ring finger is characterised by eight conserved Cys and His residues arranged in a manner to form two separate  $Zn^{2+}$  binding sites (site-I and site-II).



**Figure-1.8: Ribbon representation of BRCA1-BARD1 ring domain heterodimer (PDB ID-1JM7).** *BRCA1:BARD1 ring domain heterodimer structure shows the presence of two  $Zn^{2+}$  binding sites on each protein.*

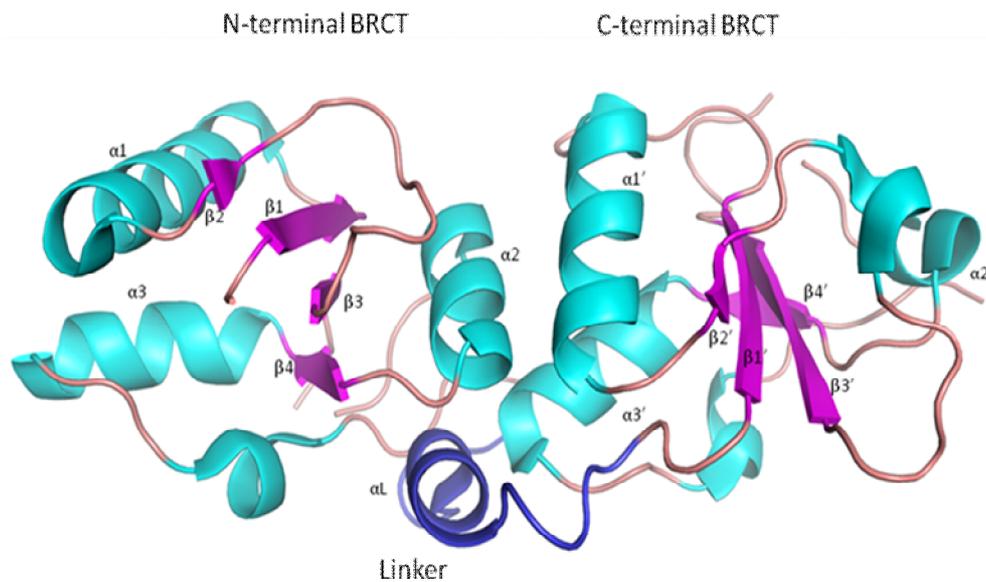
The BRCA1 RING motif comprises of a short anti-parallel three-stranded  $\beta$ -sheet and two Zn-binding loops along with a central  $\alpha$  helix. The BARD1 ring domain structure is similar to BRCA1 ring but lacks the central  $\alpha$  helix between the two Zn-binding loops. Hetero-dimerization is a consequence of interactions at the hydrophobic core of the four helix bundle [96]. The amino acid residues involved in these interactions are shown in **Figure-1.9**. The complex structure provides a platform to study the effect of cancer-predisposing mutations on the BRCA1 ring domain. The mutations reported on BIC (<http://www.nhgri.nih.gov>) which occur in Cys residues of  $Zn^{2+}$  binding sites I or II are known to be deleterious and predispose individuals to high risk of breast and ovarian cancers.



**Figure-1.9:** PDBsum analysis BRCA1-BARD1 RING domains complex (PDB ID-2JM7). Interactions were drawn by using PDBsum database [42].

### 1.2.3.2. Crystal Structure of BRCT Domain

Koonin *et al* have identified the presence of the BRCA1 C-terminal (BRCT) domain by hydrophobic cluster analysis [86]. Further, using limited proteolysis approach, Williams *et al* found two BRCT domains at the C-terminus (1646-1863) of BRCA1. BRCT domain of BRCA1 is essential for the tumor suppressor functionality [10]. The first crystal structure of the BRCT domain was reported by Zhang *et al*, who crystallized the BRCT domain of XRCC1, and solved the crystal structure at 2.5 Å resolution [100]. The overall shape of BRCA1 BRCT repeat is like that of a cylinder about 70 Å long and 30-35 Å wide (**Figure-1.10**) [10].



**Figure-1.10: Ribbon representation of the BRCA1 BRCT domain (PDB ID- 1JNX).** The BRCA1 is comprised of two tandem BRCT domains such as N-terminal (1646-1736) and C-terminal BRCT domain (1760-1863) and a linker domain (1737-1759).

Each BRCT domain is comprised of three  $\alpha$  helices and a single four-stranded  $\beta$  sheet. The two BRCT repeats, arranged in a head- to- tail manner are separated by a linker of 23 amino acids. The central  $\beta$  sheet is flanked by a pair of  $\alpha$  helices ( $\alpha 1$  and  $\alpha 3$ ) on one side and  $\alpha 2$  helix on the other side. The overall topology of a single BRCT repeat is  $\beta 1$ – $\alpha 1$ – $\beta 2$ – $\beta 3$ – $\alpha 2$ – $\beta 4$ – $\alpha 3$  [10]. The ligand binding is stabilised by hydrophobic amino acids from  $\alpha 2$  of the N-terminal repeat, and residues from  $\alpha 1'$  and  $\alpha 3'$  of the C-terminal repeat [101–103]. In-vitro and structural studies have revealed that most of the cancer-causing mutations discovered in the BRCT domains destabilize the protein fold [104, 105].

### 1.2.3.3. BRCT Domain as a Phosphopeptide Interaction Module

The BRCT domain of BRCA1 is known to perform diverse functions by interacting with several phosphorylated proteins. The chemical interactions are with a contiguous stretch of amino acids of consensus sequence  $pS_0$ - $X_1$ - $X_2$ - $F_{+3}$  (the phosphorylated serine at 0

position, X can be any amino acid and phenylalanine residue at +3 position) [106-109]. Different binding partners of the BRCA1 BRCT domain have different amino acids at the +2 position [102, 110], while the residue at +1 position is mostly proline. It has been observed that phosphopeptide interacts with BRCA1-BRCT in a “two-knob” manner. The two “knobs” of interaction are the pS and F residues. The phosphate group of the phosphopeptide interacts with residues at the N-terminal BRCT via salt bridge and hydrogen bonding interactions. The phenylalanine (+3) residue interacts at the hydrophobic pocket in the interface region of N-terminal and C-terminal BRCT repeats. The conserved residue Arg-1699 plays an important role in recognizing the backbone of Phe (+3) residue through hydrogen bonding. These interactions appear to exist across species as the interacting residues from BRCT domain are conserved on different species.

Till date, crystal structures of five complexes between interacting partners peptides and BRCA1 BRCT have been reported (**Table-1.1**). From each structure, it has been observed that the association between pSer<sub>(0)</sub>-X-X Phe<sub>(+3)</sub> and BRCA1- BRCT is very much conserved [111].

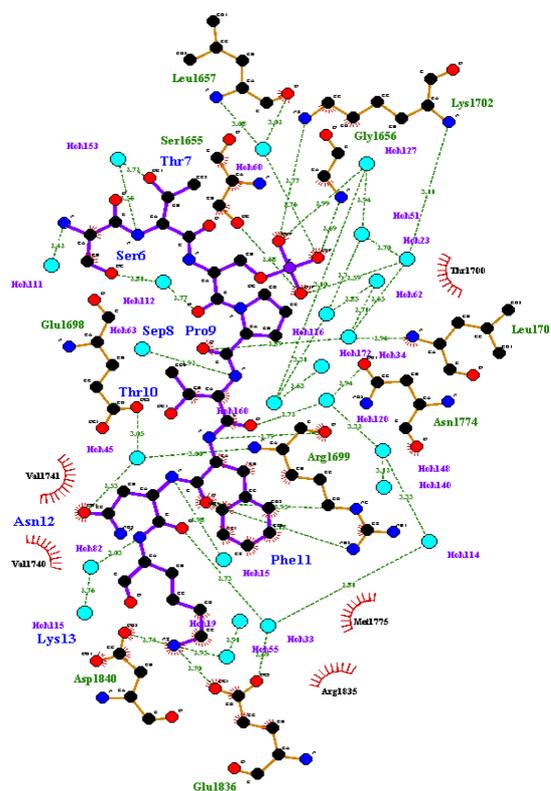
**Table-1.1:** List of BRCA1 BRCT domain interacting partners. Where pS indicate phosphorylated serine

Sr.No.	BRCA1 BRCT Interacting Partners	Peptide Sequence	Affinity ( $\mu\text{M}$ )	PDB-ID and [Reference]
1	BACH1	ISRSTpSPTFNKQTK	0.9	1T29 [112]
2	CtIP	PTRVSpSPVFGAT	3.7	1Y98 [102]
3	ACC1	DSPPQ-pS-PTFPEAGH	5.2	3COJ [113]
4	ATRIP	PEACpSPQFG	28.2	4IGK [110]
5	BAAT	VARpSPVFSS	3.34	4IFI [110]

#### A) Crystal Structure of BRCA1 BRCT-BACH1

BACH1 is one of the important binding partners of BRCA1 in the cells [114]. The association between phosphorylated BACH1 and the BRCA1-BRCT domain was

analysed through GST-BRCT pull-down assay, mass spectrometry and western blotting. Phosphorylated BACH1 was shown to be directly associated with the BRCT domain of BRCA1 [115]. The phosphorylation of BACH1 at the Ser-990 position is essential for interaction with the BRCT domain. The absence of BACH1 in cells leads to the loss of DNA repair function and ultimately to tumor progression [116]. The BACH1 oligopeptide of sequence (NH<sub>2</sub>-ISRSTpSPTFNKQTK-COOH) was found to interact with the BRCT domain, and the binding affinity of the interaction was estimated to be 0.9  $\mu$ M (Kd). The details of the molecular interactions are revealed by the complex crystal structure **Figure-1.11** [112]. The BRCT domain structures in the liganded and unliganded [10] structure are almost identical with a RMSD of about 0.5 Å for 213 C- $\alpha$  atom pairs.



**Figure-1.11: Ligplot analysis of BRCA1 BRCT- BACH1 complex (PDB ID: 1T29).** The pSer-990 and Phe-993 residues of peptide (pink bonds) forms network of hydrogen bonds and hydrophobic interactions with BRCA1 BRCT (orange bonds). On each Ligplot figure, the hydrogen bonds are shown in dashed line and hydrophobic interactions in radial arcs.

The BACH1 phosphopeptide binds to the BRCT domain in a “two knob model”. In this model the amino acids, pSer-990, and Phe-993, bind respectively to the N-terminal BRCT and interface region between N and C-terminal BRCT. The pSer-990 forms three intermolecular hydrogen bonds with residues Ser-1655, Gly-1656 and Lys-1702 from the N-terminal BRCT, while the Phe-993 forms hydrophobic interactions within the hydrophobic pocket created by residues from the interface and C-terminal BRCT.

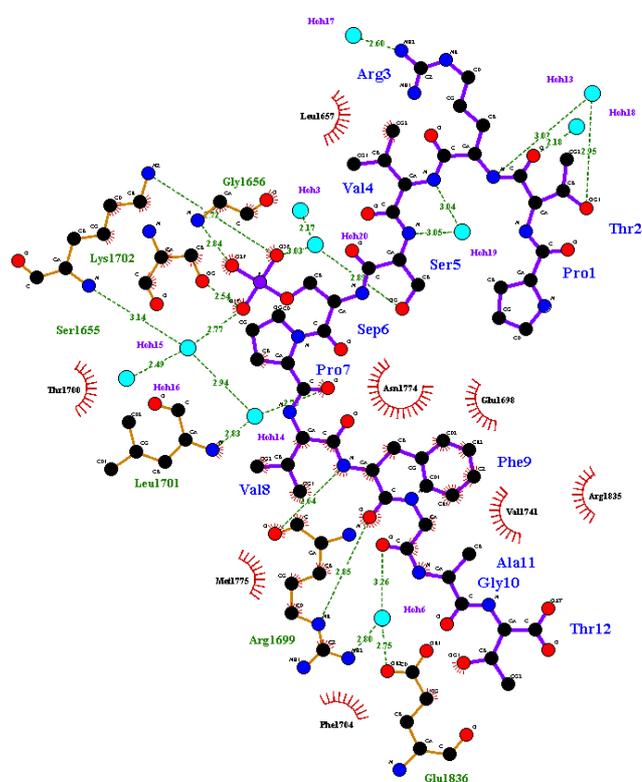
The residues of the BRCT domain that contributed to the interaction with phosphopeptide are conserved among BRCA1 orthologs [111, 112]. As shown in **Figure-1.11**, the amino acids Arg-1699 and Met-1775 are involved in the interactions with BACH1 phosphopeptide. However, BRCA1 M1775R [36] and M1775K [117] mutations found in the high risk families are impairing the interactions between BRCT and pBACH1, this indicates the importance of BACH1 association in transcriptional activation function BRCA1 BRCT domain [112] [113]

#### **B) Crystal Structure of BRCA1 BRCT-CtIP**

CtIP is a phosphoprotein that consists of 897 amino acids and binds to CtBP, which is a transcriptional co-repressor [118]. There are no reported conserved domains on CtIP, but the protein possesses a few motifs that are essential for interaction with the CtBP and pRB family of proteins, such as pRB, p107 and p300v [119]. CtIP is known to be a nuclear protein with four reported putative nuclear localization signals [120]. The occurrence of DNA damage by ionizing radiation activates ATM kinase that phosphorylates two Ser residues of CtIP. CtIP was identified as an interacting partner of the BRCA1 -BRCT domain using yeast two hybrid screening method [121] [122]. The crystal structure of BRCT with a

pCtIP peptide of amino acid sequence  $\text{NH}_2\text{-}_{323}\text{-PTRVSpSPVFGAT-}_{333}\text{-COOH}$  has been determined [102]. The phospho CtIP peptide was titrated with highly purified BRCA1-BRCT using ITC, and it was found that BRCA1- BRCT bound to the CtIP peptide with a dissociation constant of  $\sim 3.4 \mu\text{M}$ . The association of pCtIP with BRCT is  $\sim 5$  times weaker than that of pBACH1, which is supported by loss in weak intermolecular interactions observed in the crystal structure. Furthermore, tumor-associated mutations in BRCA1 leads to abrogation of BRCT-CtIP interaction, which indicates the importance of this complex for the tumor suppression function of BRCA1.

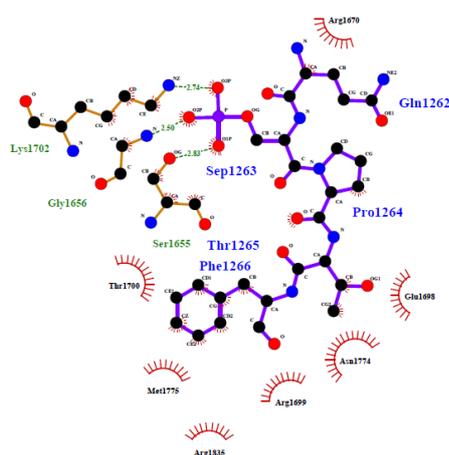
Crystal structure of BRCA1 BRCT-CtIP complex revealed conserved binding pocket for pSer(0) –Pro-Val- Phe (+3) sequences with the BRCT domain. The pSer (0) and Phe (+3) form the main interactions, where pSer (0) interacts with BRCT1 and Phe (+3) binds to hydrophobic pocket present between BRCT1 and BRCT2 [123] (**Figure-1.12**).



**Figure-1.12: Ligplot analysis of BRCA1 BRCT-CtIP complex (PDB ID-1Y98).** The pSer and Phe+3 residues of CtIP peptide (purple bonds) form main interacting centre in the complex.

### C) Crystal Structure of BRCA1 BRCT-ACC1

ACC1 is an ATP dependent carboxylase that converts acetyl Co-A to malonyl Co-A for the synthesis of long-chain fatty acids through fatty acid synthase (FAS) [124, 125]. Different investigators have identified ACC1 as a BRCA1 BRCT domain interacting partner [126-128]. There are two isozymes of ACC1 i.e. ACC1- $\alpha$  and ACC1- $\beta$ . ACC1- $\alpha$  is essential for embryonic development because an ACC1-deficient mouse is embryonically lethal. It is found that ACC1- $\alpha$  and FAS levels are up-regulated in many cancers [129] [130]. ACC1 possesses a short-term regulatory mechanism in which active ACC1 can be converted into inactive ACC1 upon phosphorylation at Ser-79 residue. BRCA1 BRCT is known to interact with phosphorylated ACC1. ACC1 residues (1258-1271) of sequence NH<sub>2</sub>-1258-DSPPQpSPTFPEAGH-1271-COOH bind to BRCA1 BRCT with ~3.7  $\mu$ M binding affinity. The ACC1 has five times less affinity towards BRCT domain compared to BACH1 peptide. The three dimensional structure of the BRCT-ACC1 oligopeptide complex provides atomic-level details of the interaction.

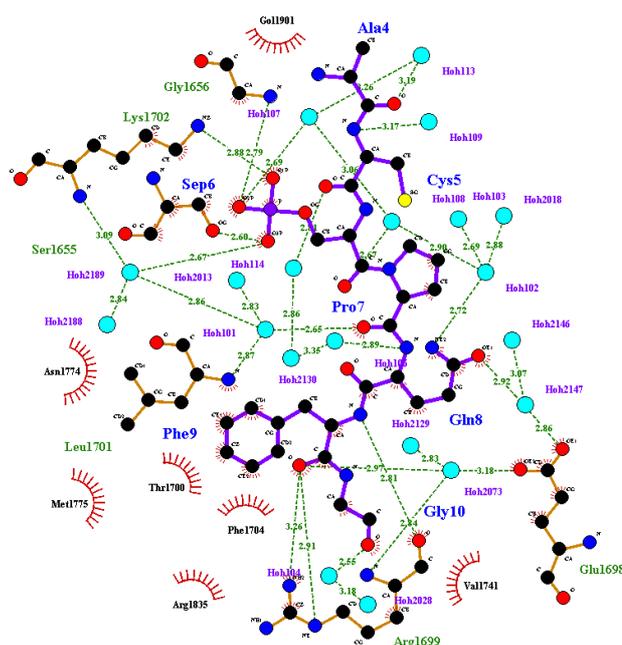


**Figure-1.13: Ligplot analysis of BRCA1 BRCT: ACC1 (PDB ID: 3COJ).** The ACC1 peptide (purple bonds) form less number of hydrogen bonds with BRCA1 BRCT domain (orange bonds) compared to BACH1 and CtIP peptides.

The pSer(0) and Phe (+3) binding sites are quite similar to those of BRCA1-BACH1 and BRCA1-CtIP complexes. **Figure-1.13** shows the details of interactions. The ACC1 residues from Pro-1261 to pSer1263 interact with the N-terminal of BRCT while residues from Pro-1264 to Gly-1270 bind to C-terminal of BRCT. The phosphorylated serine interacts with positively charged region of N-terminal, whereas the Phe-1266 of ACC1 forms the hydrophobic interactions with the C-terminal BRCT repeat [131].

#### **D) Crystal Structure of BRCA1 BRCT-ATRIP**

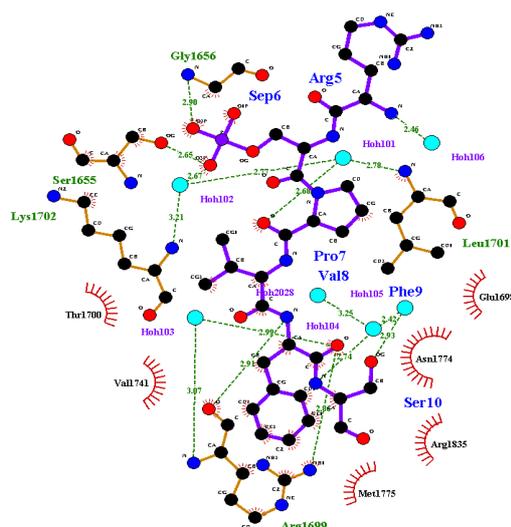
The BRCA1- BRCT interacts with ATRIP and BAAT proteins [132, 133]. Phosphorylation at Ser-239 position in human ATRIP is crucial for interaction with BRCA1 BRCT. When Ser-239 of ATRIP is mutated to Ala, there is abrogation in the complex formation, which results in G2-M checkpoint [132]. The BAAT1 protein is co-localized with ATM and is responsible for auto-phosphorylation of ATM at Ser-1981. The assembly of BRCA1, ATM and BAAT is essential for DNA double-strand break-induced ATM activation [133]. Recently, the BRCA1 BRCT domain has been co-crystallized with the ATRIP peptide (NH<sub>2</sub>-<sub>235</sub>PEACpSPQFG<sub>243</sub>-COOH), and the crystal structure has been determined to the high resolution of 1.75Å. In the complex structure, the pSer-239 interacts with N-terminal BRCT pocket while Phe-243 is inserted in the interface pocket (**Figure-1.14**). The pSer forms hydrogen bonding with Ser-1656 and Gly-1655, and the Phe-243 forms hydrophobic interactions with Arg-1699, Leu-1701, Phe-1704, Asn-1774, Met-1775, Arg-1835 and Leu-1839.



**Figure-1.14: Ligplot analysis of BRCA1BRCT-ATRIP (PDB ID: 4IGK).** The pSer (0) residue (purple bonds) forms hydrogen bonds to N-terminal BRCT (orange bonds) while Phe+3 forms hydrophobic interactions with C-terminal BRCT.

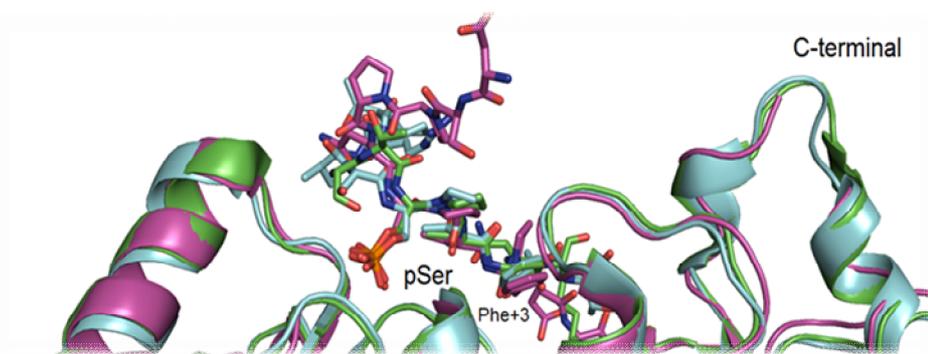
### E) Crystal Structure of BRCA1 BRCT-BAAT

It has been observed that the BAAT peptide of sequence NH<sub>2</sub>-<sub>266</sub>-VARpSPVFSS-<sub>274</sub>-COOH has a 3.3  $\mu$ M binding affinity with BRCA1 BRCT domain. The crystal structure has been solved to 2.2 Å resolution to unravel the conserved binding mode of pS-X-X-F motif [110]. The pSer-269 and Phe-272 are the key amino acids holding BRCT1 and BRCT2 respectively. The pSer forms hydrogen bonds with Ser-1655 and Gly-1655 (**Figure-1.15**). However, one water-mediated hydrogen bond was also observed with Lys-1702. The phenyl ring is inserted into BRCT2 pocket and its carbonyl group forms hydrogen bond with Arg-1699. Even though pSer (0) and Phe (+3) binding is conserved, the whole BRCT-BAAT complex is stabilized additionally by van der Waals interactions and water mediated hydrogen bonding.



**Figure-1.15: Ligplot analysis of BRCA1BRCT-BAAT (PDB ID: 4IFI).** The pSer residue of BAAT peptide (purple bonds) forms only two hydrogen bonds with Ser-1655 and Gly-1656 residues of BRCA1 BRCT domain (orange bonds).

**Figure-1.16** shows superposition of all the ligands-bound to BRCT domain structures.



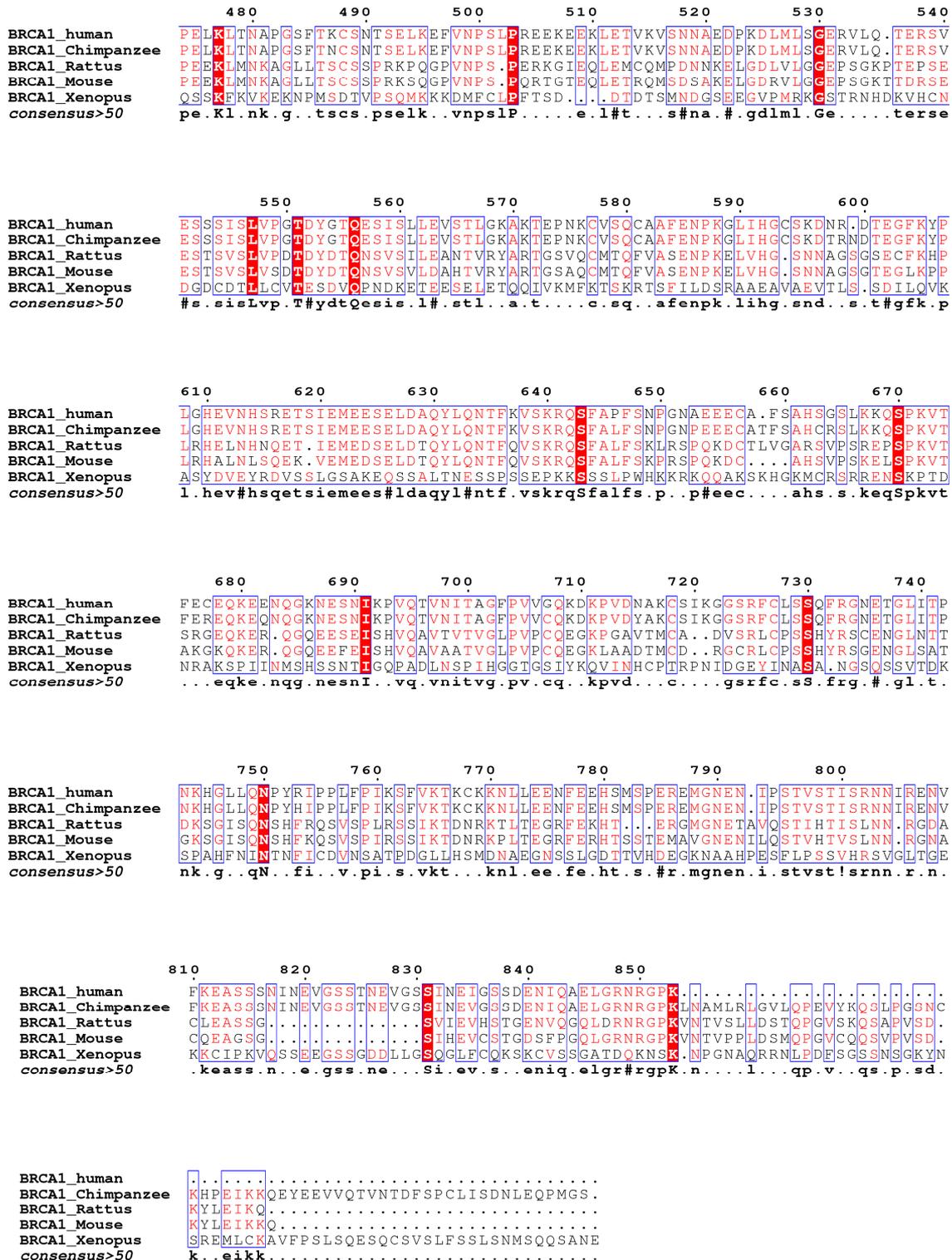
**Figure-1.16: Superposition of BRCA1 BRCT domain complex structures.** The cartoon representation of BRCA1 BRCT-BACH1 (PDB ID-1T29) (green), BRCA1 BRCT-CtIP (PDB ID-1y98) (cyan) and BRCA1 BRCT-ACC1 (PDB ID -3COJ) (pink). The peptide of each structure is shown in stick model

### 1.3 BRCA1 BRCT Mis-sense Mutations: Application in Clinical Management.

Around 5-10% of breast cancers are hereditary in nature and approximately 90% of hereditary breast and ovarian cancers are caused by mutations in *BRCA1* and *BRCA2* genes [1, 92, 134]. An individual who is heterozygous for *BRCA1* mutations, possesses an increased risk of breast cancer development compared to normal population [135].

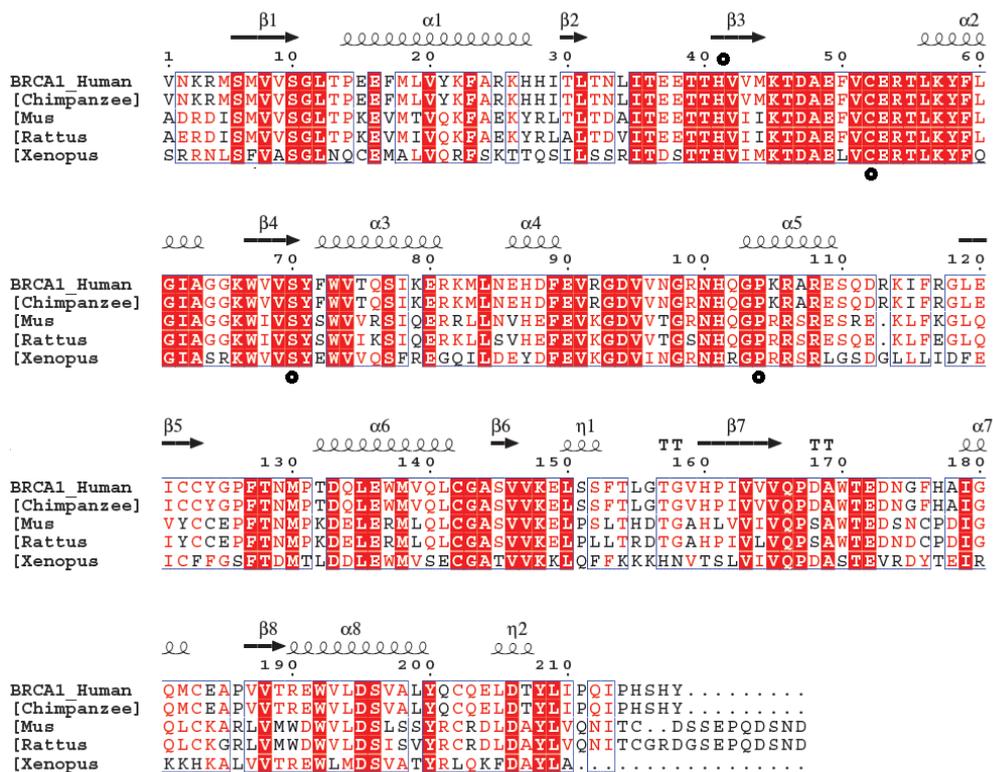
Several databases are available to serve cancer populations, but the Breast Cancer Information Core (BIC) data base is a repository for *BRCA1* and *BRCA2* mutations, which was established in 1995. Since its inception, more than 1700 member research communities from 48 different countries have registered to use the database. The BIC data are derived from online published literature and online depositions of mutations [136]. Although the cancer-causing mutations are distributed throughout the 1,863 amino acids of *BRCA1*, they are concentrated more in three regions: i) highly conserved RING domain ii) the region encoded by exons 11-13 (65% of sequence) and iii) the BRCT domain [137, 138] [103]. Over 2,000 mutations discovered in *BRCA1/2* have been reported in the BIC database. These mutations may be classified as either deleterious mutations or polymorphisms or mutations with unknown clinical significance. The risk associated with mis-sense mutations is difficult to predict. Presently, testing of samples of family members is being done to establish linkage between particular mutation and increased cancer risks. However, this is not satisfactory as genetic evidence is often unavailable, and moreover the substantial percentages of breast cancers are non-hereditary. Further, increasing evidence suggests that hereditary and sporadic *BRCA1*-associated tumors respond differently to treatment [139]. In view of the multifaceted role of *BRCA1*, development of simple biochemical assay is very challenging. Thus a test capable of distinguishing a cancer predisposing mutant from a polymorphism would be of great value in clinical management of BC. In the absence of genetic and biochemical means of classification, prediction about the functional effects of mis-sense mutants of *BRCA1* have been attempted through methods of amino acid sequence comparison and structural modelling.





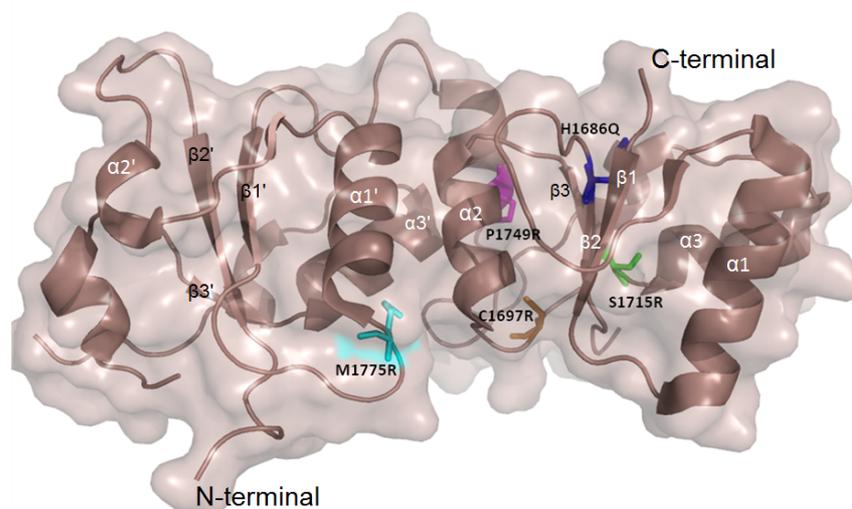
**Figure-1.17: Multiple sequence alignment of the BRCA1 central region (exon 11-13).** Amino acid sequence from species *Mus musculus*, *Rattus norvegicus*, *Xenopus tropicalis*, *Human sapiens* and *Chimpanzee* (*Pan troglodytes*) *The figure was prepared in ESPript [140].*

**Figure-1.17** shows alignment of exon 11-13 of BRCA1 from vertebrate species covering a wide spectrum of evolutionary distance. This is the region of BRCA1 which is involved in binding to various proteins during its proper functioning.



**Figure-1.18: Multiple sequence alignment of the BRCA1 BRCT domain.** BRCT domain is selected from from *Mus musculus*, *Rattus norvegicus*, *Xenopus tropicalis*, *Human sapiens* and *Chimpanzee (Pan troglodytes)*; The mutations of BRCT domain selected for study are shown in black circle. The figure was prepared in ESPript [140].

It is seen that disease-associated mis-sense mutations occur at highly conserved residues, whereas, polymorphisms are in regions of lower conservation. Similarly, experimentally established disease causing mutations of BRCA1 BRCT are also found in the highly conserved region as shown in **Figure-1.18**.



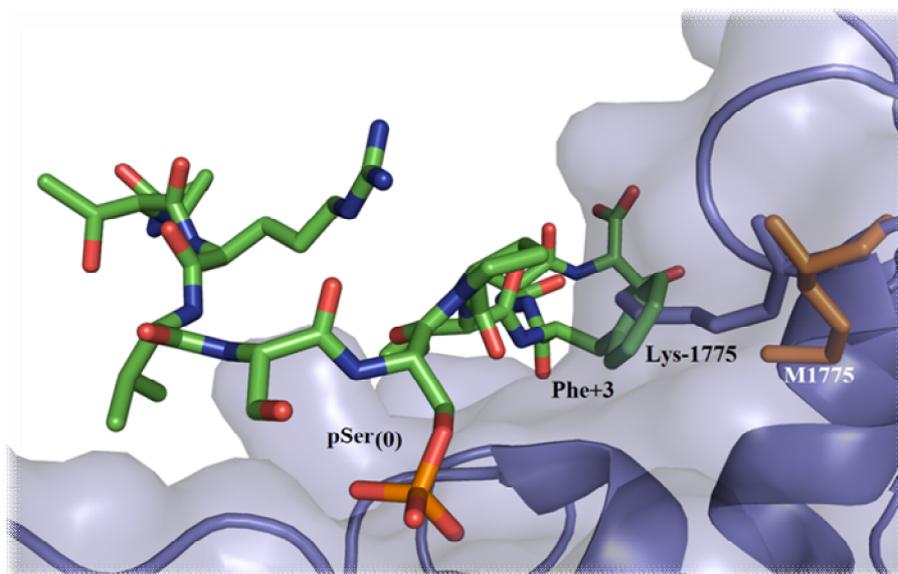
**Figure-1.19:** The positions of selected mis-sense mutations in BRCA1 BRCT domain structure. BRCA1 BRCT domain shown as surface as well as cartoon representation, where the selected mis-sense mutations are shown in stick model.

**Figure-1.19** shows locations of the few disease-associated mutations that have been selected for study in the present thesis. To get insights into the mechanisms by which disease associated mutations exert their influence, crystal structures of BRCT mutants with phosphopeptides have been determined by investigators, and these are listed in **Table-1.2**.

Mis-sense mutations in BRCA1 listed in the **Table-1.2** are reported to impair the phospho-protein binding [10, 141, 142]. For example, **Figure-1.20** shows the M1775K mutant structure overlaid on the structure of wild-type BRCT complexed with the phosphopeptide. As can be seen, the side-chain of the lysine-1175 residue is in steric contact with the Phe +3 residue from the phosphopeptide, and this unacceptable steric contact would prevent phosphopeptide binding to the mutant BRCT, thereby adversely affecting the function of BRCA1. Such structural information would help in clinical management of the disease.

**Table-1.2:** Summary of BRCA1 BRCT domain mis-sense mutant structures.

Sr. No.	BRCA1Mis-sense	PDB ID	References
1	BRCA1 M1775R	1N5O	[105]
2	BRCA1 M1775K	2ING	[117]
3	BRCA1 V1809F	1T2U	[143]
4	BRCA1 D1840T	3K15,	[144]
5	BRCA1 D1840T	3K16	[144]
6	BRCA1 G1656D	3PXA	[145]
7	BRCA1 T1700A	3PXB	[145]
8	BRCA1 R1699Q	3PXC	[145]
9	BRCA1R1835P	3PXD	[145]
10	BRCA1 E1836K	3PXE	[145]



**Figure-1.20: The surface occlusion by Lys-1775 mutant.** The BRCA1 BRCT-CtIP structure (PDB ID: 1Y98) superposed with BRCA1 M1775K mutant structure (PDB ID: 2ING), CtIP peptide, M1775 of complex structure and K1775 of mutant structure are shown in stick model.

#### 1.4 Conclusion

Breast cancer is one of the most common causes of death among women across the globe and in India. Breast cancer can be broadly divided as either sporadic cancer or hereditary breast cancer. The mutations in the breast cancer susceptibility genes, in sporadic breast cancer, are caused by environmental factors, while they are inherited in hereditary breast

cancer. *BRCA1/2* genes are two of the most important breast cancer susceptibility genes. *BRCA1* has two distinct domains in addition to a large unstructured DNA binding central region. These domains are N-terminal RING finger domain and C-terminal BRCT domain. The structural analysis revealed that BRCT domain of *BRCA1* acts as a phosphopeptide recognition module. Interactions between a few binding partners and *BRCA1*-BRCT domain have been structurally characterised, but the possibility that the *BRCA1*-BRCT domain may have more binding partners is to be expected. The BRCT domain recognizes phosphopeptides that possess the pS-X-X-F motif. The literature indicates that doubly-phosphorylated proteins can also interact with *BRCA1* BRCT domain. There is a long list of BRCT domain- containing proteins. The basic function of the BRCT domain is to act as a protein-protein interaction module, but there are a few proteins that do not have reported interacting partners, such as *BARD1* BRCT domain. It will be interesting to find potential reasons for the lack of in-vivo binding partners. Most of the breast and ovarian cancer mutations are present in the BRCT and RING finger domains of *BRCA1*. It is very difficult to distinguish between clinically significant mutations from mere polymorphisms. Mutations occurring in highly conserved region are being suggested as causing. Development of additional tools to reliably distinguish between disease-causing and polymorphic mutations is urgently required in clinical management. Structural studies are also required to understand disease-causing mechanism of *BRCA1* BRCT domain mis-sense mutations.

*Materials and Methods*

---

*Chapter 2*

### Source of reagents and instruments

1. Bacterial culture – The LB medium, Ampicillin and IPTG – Himedia, India
2. Common salts, buffers, precipitant, detergents and organic solvents- Himedia (India), Sigma (USA), Merck (Germany), Fluka (Germany), SRL, Qualigens (India).
3. Crystallization screen- Hampton research (USA)
4. Synthetic peptides – USV Biotech (India).
5. Crystallization buffers, salts, precipitants, organic solvents- Hampton research (USA)
6. MWCO (Molecular Weight Cut Off) filter unit – Millipore (USA)
7. Pre-packed gel filtration column- Superdex-200 and 75 – 16/60- GE Healthcare (Sweden)
8. DNA/ protein electrophoresis- Agarose (Himedia, India), EtBr, Bromophenol blue, Acrylamide , Bis-acrylamide, Bradford reagent, Protein ladder- Sigma (USA), Merck (Germany), Fluka (Germany); SDA-PAGE apparatus- Bio-rad (USA).
9. Restriction and modification enzymes and DNA isolation kit- Fermentas (USA), NEB (USA) , Qiagen (Germany), Sigma (USA)
10. Plasmids or DNA ladder- Clone JET PCR cloning kit, DNA ladder - Fermentas (USA); pGEX-KT and pET-TEV from Dr. John Ladias, pET3a- Dr. Sanjay Gupta
11. FPLC system– ACTA purifier – GE healthcare –(Sweden)
12. Vibration free cooling incubator- Sanyo (Japan)
13. X-ray source- Rigaku (Japan)
14. Stereo microscope –Olympus (Japan)
15. Detector- MAR-345 (Germany)
16. -20 incubator- Sanyo (Japan)
17. -80 deep freezer – Thermo Fisher (USA)
18. ITC-200- GE Healthcare (Sweden)
19. CD spectrophotometer – JASCO (Japan)
20. Spectrofluorometer- Horiba (Japan)

### Materials

**Luria Bertani (LB) medium-** For bacterial growth LB medium is used. Powdered Luria broth powder 20g was dissolved in 900 ml of distilled water (D/W) and the volume is adjusted to 1litre with D/W. The LB agar plates were prepared by adding 35 g of agar powder per litre of LB broth. LB broth and LB agar is sterilised by autoclaving. LB agar is poured in 90 mm sterile plates.

**Antibiotics and IPTG-** Ampicillin and Chloramphenicol were used for selection and propagation of plasmids carrying respective markers. Ampicillin stock solution (100 mg/ml) was prepared using D/W, while chloramphenicol stock solution (34 mg/ml) was prepared using 100% ethanol. Both the antibiotics were sterilised by filtering through 0.5 micron filter aseptically. The ampicillin and chloramphenicol is added in LB broth or in LB agar to a final concentration of 100 $\mu$ g/ml and 34 $\mu$ g/ml respectively. 1M IPTG stock solution was prepared by dissolving 2.38 g IPTG powder in 7 ml autoclaved D/W and finally volume made up to 10 ml D/W.

**Agarose gel electrophoresis -** Tris Borate EDTA (TBE) buffer: 0.9 M Tris base, 0.9 M Boric acid, 0.02 M EDTA (10X buffer stock was made and diluted to 0.5X for use); 6X gel loading dye: 0.25% bromophenol blue, 30% glycerol; EtBr:0.5  $\mu$ g/ml.

**Protein estimation –** Bradford reagent

**SDS PAGE:**

30 % Acrylamide - 28.8 g Acrylamide and 0.2 g Bis-acrylamide were dissolved in 60 ml D/W on a magnetic stirrer and the volume was made up to 100ml by D/W.10% ammonium per sulphate (APS);Tetramethylethylenediamine (TEMED). 1.25 M Tris buffer – pH 6.8 and 1.25 M Tris buffer – pH 8.8; 4 X sample loading buffer: 250 mM Tris buffer pH 6.8, 20% glycerol, 8% SDS, 8%  $\beta$  mercapto-ethanol (BME), 0.04 % bromophenol blue; electrophoresis buffer: 25mM Tris base, 250 mM Glycine and 0.1% SDS. Staining solution – 0.5 % bromophenol blue prepared n D/W. Destaining solution - 45% Methanol and 10% glacial acetic acid.

## **Methods**

The quantity of pure protein required for biophysical and crystallographic studies is rather high, and often, it is difficult to isolate and purify such large quantities from natural sources. However, it is possible to exploit the advances of molecular biology methods to produce the protein of interest in bacterial expression system. The gene coding for the heterologous protein is inserted either in bacterial chromosome or in a plasmid/ vector, which is a small, circular and double stranded DNA that can replicate independently of the chromosomal DNA inside the cell. The procedures of gene cloning, site-directed-mutagenesis, protein expression, purification, crystallization, structure determination and biophysical techniques used in the present thesis are described briefly in this chapter.

### **2.1. Gene Cloning**

#### **2.1.1. Selection of target protein/domain**

The target protein for cloning is either a full length protein or the functional domain. In spite of extensive knowledge of molecular biology, it is not possible to express every gene in the bacterial system [146]. Identification of a functional domain is really important, because small errors can affect expression level greatly [147]. There are various servers that can be used to predict secondary structural elements and disorder region in the protein [148, 149], and this information is helpful in functional domain prediction. The targets which are difficult to express can sometime be efficiently expressed by changing the starting and the end point of the protein sequences. After selecting a target protein/domain of a gene, corresponding gene sequence can be incorporated into an expression vector. There are different kinds of gene cloning methods like homology based cloning [150], ligation-independent cloning [151], and restriction

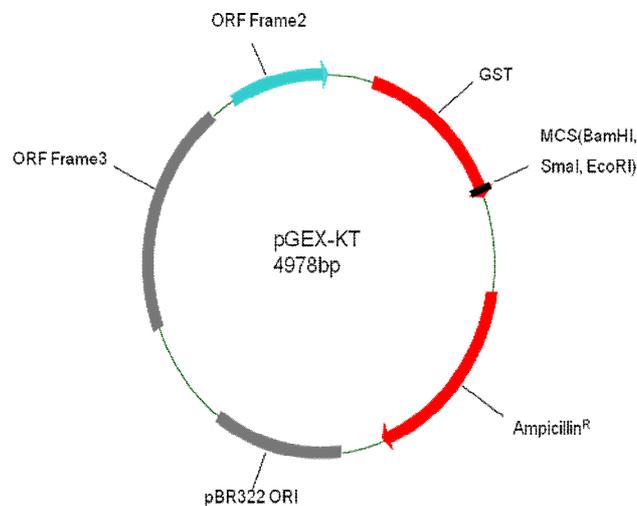
enzyme based cloning [152]. For the work reported in this thesis, the restriction enzyme based cloning was used and therefore, this method is described in detail.

### 2.1.2. Restriction enzyme based cloning

In this method, both expression vector and target genes are restriction digested with the same set of restriction enzymes. This process provides sticky ends that enable insertion of foreign gene into the plasmid/vector

#### 2.1.2.1. Preparation of expression vector

The protein of interest can be expressed either as a fusion protein with affinity tag or as a native protein. The choice of the expression vector is made, considering the advantage gained in the subsequent purification of expressed protein. The affinity tags like GST, 6HIS, MBP etc are helpful in protein expression, purification, and rarely interfere with the activity of target protein [153, 154] .



**Figure-2.1: pGEX-kT vector map.** The pGEX-kT vector mainly comprised of four regions such as two open reading frames, origin of replication, ampicillin resistance gene and GST tag. (MCS is multiple cloning site).

We have cloned different regions of BRCA1 using pGEX-kT and pET3a vectors and characteristics of these vectors are reported in **Table-2.1**. The fusion protein expressed in

pGEX-kT vector has N-terminal GST tag, while protein expressed in pET3a vector doesn't have tag. Restriction sites for the pGEX-kT vector are shown in **Figure-2.1**.

**Table-2.1:** Characteristics of expression vectors

Plasmids	Replicon	Affinity Tag	Resistance marker
pGEX-kT	pMB1	GST	Ampicillin
pET3a	ColE1	No tag (Native)	Ampicillin

The plasmid DNA can be incorporated into the DH5 $\alpha$  cells in the process of bacterial transformation as follows:

➤ **Bacterial transformation**

- Uptake of foreign DNA by bacteria is referred to as transformation [155]. The bacterial cells are made competent to uptake foreign DNA by treatment with transformation buffer, which contains calcium chloride (CaCl<sub>2</sub>) or Rubidium chloride (RbCl).
- The competent cells (DH5 $\alpha$ ) are thawed on ice, and incubated for 45 minutes with approx 50-100 ng of DNA.
- Heat- shock is given to mixture for about 2 minutes by dipping into a water bath kept at 42°C.
- The mixture is put back on ice for four or five minutes.
- Further, 800  $\mu$ l of sterile Luria broth is added to the mixture aseptically and the mixture is incubated on shaker incubator (or water bath) at 37°C for 45 minutes. The mixture is harvested for one minute on table-top centrifuge.
- Re-suspend the pellet with 200  $\mu$ l of LB medium,
- Spread the mixture evenly all over the surface on LB agar plate containing appropriate selection marker.

- Incubate the plate at 37°C for overnight.

For cloning purpose, the plasmid DNA is required in  $\mu\text{g}$  quantities with appropriate restriction enzymes, and following is the protocol used for plasmid DNA extraction from bacterial cells.

➤ **Plasmid DNA isolation**

The bacterial cells taken from a single colony are seeded in 10 ml of LB medium, further grow overnight at 37°C with shaking at 300 rpm.

- Centrifuge the overnight grown culture at 6000 rpm for 2 minutes at 4°C.
- Resuspend the bacterial pellet in 250  $\mu\text{l}$  (depending upon the weight of pallet) of buffer P1 containing 50 mM Tris HCl (pH 8.0), 10mM EDTA, 100  $\mu\text{g/ml}$  RNase-A and transfer into new microfuge tube.
- Add 250  $\mu\text{l}$  buffer P-2 containing 200 mM NaOH, 1% SDS, and mix it gently 4-6 times by inverting the tube.
- Add 350  $\mu\text{l}$  buffer N-3 having 3.0 M Potassium acetate (pH-5.5) and invert the tube gently 4-6 times.
- Clear the cell debris by centrifugation at 13000 rpm for 10 minutes at 4°C.
- Collect the supernatant and apply on the spin column (silica gel). Centrifuge at 13000 rpm for 2 min and discard the flow-through.
- Wash the spin column with 0.75 ml buffer PE of 75% ethanol, 25mM NaCl, 5 mM Tris-HCl, (pH7.5) and centrifuge for approx 60 sec, and decant the flow through.
- Give an extra empty spin to decant the residual wash buffer.

- Transfer the spin column on new microfuge tube and add 30  $\mu$ l of elution buffer (10 mM Tris HCl, pH 8.5) at the centre of the column. Centrifuge for 13000 rpm for 2 min, and this results in elution of DNA of interest.

The isolated plasmid DNA is subjected to restriction digestion and a composition of reaction mixture is shown in **Table-2.2**.

**Table-2.2:** Typical 20  $\mu$ l restriction digestion reaction

Plasmid DNA (1 $\mu$ g)	15 $\mu$ l
Restriction Enzyme-A (1unit/ $\mu$ l)	1 $\mu$ l
Restriction Enzyme-B (1unit/ $\mu$ l)	1 $\mu$ l
Buffer (10x)	2 $\mu$ l
Distilled water	1 $\mu$ l

The restriction digestion is carried out at 37°C for 2 hours using fast- digest restriction - enzymes. However, normal restriction enzymes take over-night at 37°C to digest the required region. The digested plasmid can be purified by resolving it on 1% agarose gel. The method of separation of DNA by agarose gel electrophoresis is explained below.

#### ➤ **Agarose gel electrophoresis**

This is a commonly used method for analytical and preparative separation of nucleic acids. The agarose gel of concentration ranging from 0.2 to 2% can be used for the separation of DNA fragments of different length. The DNA fragments are separated on the basis of charge and mass [156]. The standard DNA marker loaded on the same gel can be used to determine sizes of DNA fragments. The DNA can be visualized by adding a fluorescent dye like ethidium bromide (EtBr), which binds to DNA through intercalation. The EtBr stained DNA fragments are visualized using UV light and documented by Gel documentation system (Fisher Scientific, UK). DNA, encoding the gene of interest can be

purified, and quantified by measuring absorbance ( $A_{260}$ ) on NanoDrop spectrophotometer (Thermo Scientific, USA).

### 2.1.2.2. Preparation of target gene

The deoxyribonucleic acid encoding the gene of interest can be amplified using Polymerase Chain Reaction (PCR) [157]. Two gene specific PCR primers are required to amplify the gene of interest. The forward primer will have a restriction site, nucleotide sequence for protease cleavage site (at the N-terminal tag) and 15-20 bases from the sense strand of gene of interest. The reverse primer comprises restriction enzyme site, one or two stop codons and 15-20 bases complementary to the antisense strand of the gene. A complete reaction mixture is of 20 to 50  $\mu$ l volume and the typical PCR composition and reaction program are listed in **Table-2.3** and **Table-2.4**, respectively.

**Table-2.3:** Typical 50 $\mu$ l PCR reaction

Chemicals	Amount ( $\mu$ l)
5xGC Buffer	10
dNTPs (10 mM)	2
Template (100 ng/ $\mu$ l)	1
Forward primer (10 picomole/ $\mu$ l)	2
Reverse primer (10 picomole/ $\mu$ l)	2
Phusion polymerase (1 unit/ $\mu$ l)	1
Deionised water	32

**Table-2.4:** Program for PCR amplification

Step	Temperature ( $^{\circ}$ C)	Duration (min:sec)
1) Initial denaturation	94	15:00
2) Denaturation	94	00:45
3) Annealing	55	00:30
4) Elongation	72	01:00
Run steps 2-4 for 32 more cycles		
Final Elongation	72	10:00

The amplified PCR product can be resolved and extracted from the agarose gel, the gel extraction technique is described below.

➤ **Extraction of DNA fragments from gel**

- Excise the gel slice containing desired DNA fragment under UV light.
- Weigh the gel slice, add approximately three volumes of gel solubilisation solution which contains 6 M Guanidine thiocyanate, 50 mM Tris-HCl (pH 7.5) and 20 mM EDTA (pH 8.0) and incubate at 55°C for 10 min or till it dissolve the gel pieces.
- Apply the solubilised gel pieces on the spin column (made up of silica gel)
- Centrifuge for 13000 rpm for two min and discard the flow through.
- Wash the spin column using 0.7 ml wash buffer (75% Ethanol, 25 mM NaCl, 5 mM Tris-HCl, pH7.5).
- Give an extra empty spin to decant the residual wash buffer.
- Elute the DNA fragment by adding prerequisite amount of elution buffer (10 mM Tris HCl, pH 8.5) on to the spin column.

The purified PCR product is subjected to restriction digestion by the same set of restriction enzymes which were used for digestion of vector described in **Table-2.1**. The digested PCR product is purified by gel extraction technique. Now we have PCR product with sticky ends complementary to the vector.

### **2.1.2.3. Annealing and Ligation**

The digested vector and PCR product are then mixed in 1:3 molar ratio and allowed to anneal via their sticky ends. The vector and the PCR product are ligated using Quick T4 DNA ligase (Fermentas, USA). The ligation reaction is carried out at 22°C for ~10 minutes. The ligation mixture is then transformed into DH5( $\alpha$ ) cells to yield colonies in a LB plate for overnight incubation.

#### 2.1.2.4. Screening of potential clones

To check the success of ligation reaction, these colonies are generally screened. . This screening can be performed in different ways, like restriction method, colony PCR and selection of blue- white colony. The work reported in this thesis was carried out using restriction digestion method. Few colonies of are seeded from the ampicillin plate and grown overnight for DNA isolation. The isolated DNA was set for restriction digestion using same set of enzymes that was used while preparing vector and target gene. This digested product was resolved on agarose gel. The positive clones are those that show the band for vector and target gene on agarose gel. Finally the positive clones were confirmed by DNA sequencing using vector specific sequencing primers listed in **Table-2.5**.

**Table-2.5:** List of sequencing primers

Name	Primers
pGEX 5'	5'-[GGG-CTGGCAAGCCACGTTTGGTG]-3'
pGEX 3'	5'-[CCG-GGAGCTGCATGTGTCAGAGG]-3
T7 promoter	5'-[TAATACGACTCACTATAGGG]-3
T7 terminal	5'-[CTAGTTATTGCTCAGCGGTG]-3

#### 2.1.3. Site-Directed-Mutagenesis

Site-directed-mutagenesis (SDM) implies point mutation at a particular position in the gene [158, 159], and involves following three major steps.

##### 2.1.3.1. PCR of Mutant Strand

This procedure needs two site-specific primers, and both the primers should have the desired mutation and should be able to anneal to complementary DNA strands of the plasmid. The desired mutation should be present at the centre of the primer sequences of about 25-45 nucleotides in length. The amplification is carried out by *Pfu* DNA polymerase (New England Biolabs, USA), because this polymerase replicates strands

with high fidelity. The PCR reaction can be set and run according to **Table-2.3** and **Table-2.4** and care must be taken to increase the time for polymerisation.

### **2.1.3.2. DpnI Digestio**

The DpnI is an endonuclease which specifically targets the methylated DNA strands [159]. Most of the plasmid DNA synthesised in *E. coli* is dam methylated and therefore, is susceptible to DpnI digestion. DpnI is added to the 50µl of reaction mixture and incubated for about one hour at 37°C.

### **2.1.3.3. Transformation of SDM product and Screening**

The DpnI digested product is transformed into competent DH5α cells which are plated on LB plate with appropriate selection marker. The single colony is inoculated into 10 ml LB broth with appropriate antibiotic and grown overnight to get enough cells for plasmid DNA isolation. The plasmid DNA isolated from these colonies are subjected to DNA sequencing with primers listed in **Table-2.5**.

## **2.2. Protein expression and Purification**

There are different bacterial strains having specific properties that can be utilised to check the protein expression. The choice of host bacterial strain depends on level of gene expression. For routine protein expression purpose *E. coli* BL21 (DE3) strain is used because it lacks various proteases. The most used strains are:

- **BL21 (DE3)** - This strain is deficient in *lon* and *omp-t* proteases and has DE3 lysogen with T7 polymerase. It is most commonly used strain for expression of heterologous proteins.
- **Rosetta (DE3)** – This is the modified version of BL21 (DE3) strain which can be used for expression of eukaryotic protein having some rare codons. The cells are equipped with additional tRNAs that can recognise following codons: AUA, AGG, AGA, CUA, CCC, and GGA.

- **BL21 DE3 (pLysS)** – This strain possesses DE3 lysogen and expresses T7 polymerase upon IPTG induction. The pLysS plasmid produces T7 lysozyme which controls the leaky expression or basal level expression. This strain is appropriate for expression of toxic genes

### **2.2.1. Preparation of seed culture**

The single colony from ampicillin resistance plate is inoculated aseptically in 100 ml of autoclaved LB broth containing 100µg/ml ampicillin or appropriate selection marker. The culture is then allowed to incubate overnight on shaker incubator at 37°C with continuous agitation at 300 rpm.

### **2.2.2. Scale up and induction**

Scaling up of protein production was done by inoculating 1%, by volume, seed culture in sterilized LB broth containing ampicillin. This culture was grown on shaker incubator till the OD at  $\lambda = 600$  nm reaches between 0.6-0.8. The over expression of recombinant protein is then induced by adding isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG) to a concentration of 0.4 mM in each flask. The culture flasks are then incubated in the shaker incubator at 24°C for 14-16 hours under agitation at 300 rpm. Bacterial culture is then pelleted by centrifugation at 6000 rpm. The supernatant is discarded and bacterial pellet is stored at -80°C until further use.

### **2.2.3. Protein Purification**

Differences in the physical properties of the bio-molecules are used for their separation and purification. Chromatographic methods that exploit different physical properties individually are listed below.

- Affinity chromatography- Ligand specificity or bio recognition
- Ion exchange chromatography- Overall electric charge

- Gel Filtration chromatography- Molecular size/mass
- Hydrophobic interaction chromatography- Hydrophobicity

Each separation techniques have its own advantages and disadvantages, very often two or more purification techniques are used in tandem. .

### **2.2.3.1. Affinity Chromatography**

Affinity chromatography exploits biological recognition phenomenon, in which one bio-molecule has strong affinity toward another particular bio-molecule (for instance antigen-antibody). Different affinity tags have been developed to help in protein purification procedure. The GST tag has affinity towards the glutathione molecules which can be immobilized on a resin. Poly HIS tag (6-8 HIS residues) [160] specifically binds to Ni-NTA/ Ni-IDA resins and facilitates one step protein purification. The affinity tags MBP (Maltose Binding Protein) [161], intein [162] and calmodulin [163] etc. are used for protein purification using affinity chromatography. Protein of interest is expressed as a fusion protein with the affinity tag along with a protease cleavage site that helps in removal of tag from fusion protein.

### **2.2.3.2. Ion Exchange Chromatography**

In this method, different proteins are separated on the basis of their intrinsic electrostatic charges. The pI (Isoelectric point) of a protein is the pH at which the net charge on the protein is zero. The charge of a target protein depends on the pH of buffer used for purification [164]. If the pH of the purification buffer is lower than the isoelectric point (pI) of a protein then the protein will have a net positive charge. The positively charged protein molecules can bind to cation exchangers e.g Sulfopropyl (S), Sulfomethyl, Sulfoethyl (SE) sepharose and carboxy methyl cellulose, and can be eluted by salt

gradient. This process is referred as cation exchange chromatography. However, when the pH of the purification buffer is higher than the pI of a given protein, then the protein has net negative charge and binds to the anion exchange materials. The materials Trimethyl aminomethyl (Q), Trimethylamino-hydroxypropyl (QA) and Diaethyl-2(-hydroxypropyl) aminoethyl (QAE) sepharose act as strong anion exchangers while DEAE (Diethyl aminoethyl) and DMAE (Dimethyl aminoethyl) act as weak anion exchangers [165]. Again the bound proteins are eluted by salt gradient. The choice of chromatographic methods majorly depends upon the kind of polar species (positive or negative) dominating during the purification process. At isoelectric point, the electrophoretic mobility of protein is zero. This property of each protein can be used for separation of proteins by the method of isoelectric focusing. This technique can resolve the proteins according to differences in their pI [166].

### 2.2.3.3. Gel Filtration Chromatography (GFC)

The Gel Filtration Chromatography (GFC) separates bio molecules according to differences in their size. This is a very old and widely used technique for separation of protein from mixtures. The GFC can also be referred to as size exclusion chromatography. The molecules of different sizes are passed through chromatographic medium e.g. Superdex, Sephadex, sephacryl, sepharose and superpose that possesses pores of different sizes.

**Table-2.6:** Characteristics of different superdex chromatographic media used in GFC.

Type medium	Fractionation range in KDa kDa
Superdex-30	> 10
Superdex-75	10-70
Superdex-100	10-600

The sephadex is prepared by cross linking dextran with epichlorohydran; the degree of cross linking can be varied to purify different sizes of protein, as shown in **Table-2.6**. The medium is equilibrated with buffer (25 mM Tris HCl pH-7.5, 150 mM NaCl), which leads to formation of continuous channel by linking these variable pores. The gel filtration can be performed using fast protein liquid chromatography (FPLC). The elution of the proteins can be monitored by measuring absorbance at wavelength of 280 nm. The method can be also used to perform buffer exchange along with separation.

#### **2.2.3.4. Hydrophobic Interaction Chromatography (HIC)**

HIC separates bio molecules as per differences in their hydrophobicity. The ammonium sulphate precipitation and ion exchange chromatography techniques are used to capture and concentrate the protein molecule from the cell lysate. HIC is the method of choice to be followed by ion exchange chromatography technique, because in both the cases, protein is in high salt buffer. High salt concentration enhances the interaction between hydrophobic moieties of the protein sample and chromatographic media. There are different types of chromatographic media that can be used in HIC such as SOURCE 15, sepharose fast flow and sepharose 6 fast flow etc. The protein can be eluted by lowering the salt concentration [167]. There are various factors like pH, temperature and ionic strength which can also affect the structure and solubility of protein, thereby affecting the interaction with hydrophobic surfaces of separation such as media.

### **2.3. Protein Characterization**

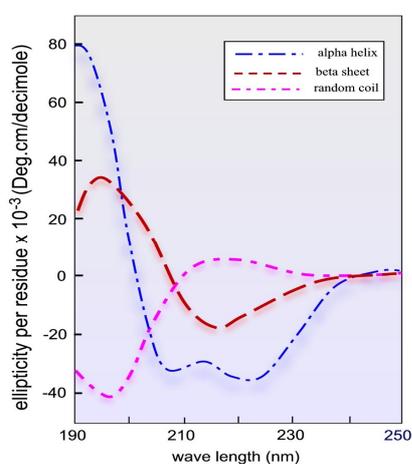
A number of biophysical techniques are used to structurally characterize a given protein in solution. Some of these methods used in the present work are described below.

#### **2.3.1. Circular Dichroism (CD)**

The CD is a technique which rapidly determines the overall folding and secondary structural elements of the protein. The CD is a measure of the degree of unequal

absorption of left- and right handed circularly polarised light [168]. Once the protein molecules differentially absorb the left and right handed light, then the emerging light is elliptically polarised. The CD signal/data can be expressed in terms of molar ellipticity (in degree, units- $^{\circ}\text{cm}^2 \text{dmol}^{-1}$ ).

The CD measurements can be divided into different categories depending upon the energy of electromagnetic radiation: the far UV range, which is below 250 nm, where peptide bond contribution is more, or the near UV range (250-300 nm) where aromatic side chains contribute more. The far-UV CD result gives characteristic spectra for secondary structure of protein. The  $\alpha$ -helices containing protein gives negative peak at  $\lambda=220$  nm and 208 nm and a positive peak at  $\lambda=193$  nm. Proteins possessing  $\beta$  sheets have a negative peak at  $\lambda=218$  nm and a positive peak at  $\lambda=195$  nm [168]. The disordered secondary element/ random coil protein are characterised by a low ellipticity at 210 nm and negative band near  $\lambda=195$  nm [169]. Different software/tools like SELCON [170], K2D2 [171] and CONTIN [172] are available that can estimate the secondary structure content of the protein [173]. The near-UV spectra can give information about tertiary structure. The CD signal at near-UV is very weak and therefore longer path length and higher sample concentrations are required. Typically 0.25-2 mg/ml, 0.5ml in volume and for far-UV CD 0.1 mg/ml and 200  $\mu\text{l}$  in volume.



**Figure-2.2: Typical circular Dichroism spectra of protein.**

The CD technique is used: **a)** to determine the secondary structural elements in protein **b)** to monitor the changes in conformation of protein for example due to mutations and also **c)** to determine the melting temperature of the macromolecules (thermal stability of proteins).

**2.3.2. Fluorescence spectroscopy**

Fluorescence is the phenomenon in which the protein sample absorbs a lower wavelength photon, undergoes electronic excitation, then emits longer wavelength radiation [174]. The difference in the measurement is called Stokes shift, and it make accurate measurement possible. The class of molecules which are capable of undergoing electronic transition are called as fluorophore. The fluorophore can be either intrinsic or extrinsic. The fluorescence signal intensity depends upon the path- length and concentration of solutes. The fluorescence spectroscopy of protein molecules are dependent upon three important characteristics **a)** the dynamic nature of fluorescence signal, **b)** localized nature and **c)** its redundancy [175].

Out of 20 amino acids, the aromatic amino acids (Phe, Trp and Tyr) act as intrinsic fluorophores. The quantum yield of Phe residue is however too small. The protein sample under study is excited at fixed wavelength such as  $\lambda=280$  nm for Trp and Tyr both, whereas  $\lambda=295$  nm specifically for Trp. When the Trp and Tyr are present in hydrophobic environment, high quantum yield is obtained that indicate proper folding of protein [176]. However, unfolded/partially folded protein or Trp/Tyr residues in a hydrophilic environment result low fluorescence intensity. Some of the extrinsic fluorophores used are ANS [177], Bis-ANS [178]. The fluorescence intensity and  $\lambda_{\max}$  are monitored to study the protein-folding or ligand-binding.

**2.3.3. Mass spectrometry**

Mass spectrometry is the technique that requires 0.1-10 pmol sample to quickly identify the protein mass with an accuracy of 0.01%. Two types of mass spectrometers, differing in the method of molecular ion generation, are commonly used for characterization of bio molecules: a) Matrix-Assisted Laser Desorption and Ionisation-Time of Flight mass spectrometer (MALDI-TOF) [179] and b) Electron Spray Ionisation (ESI) mass spectrometer [180].

#### **2.3.3.1. MALDI-TOF**

The given bio molecule is co-crystallized with an organic matrix, and then with the help of a laser beam these crystalline particles are converted to different ionic species in the gaseous phase. The time taken by different ionic species to travel through a known distance enables one to estimate the charge to mass ratio. From these measurements one can arrive at the exact molecular weight of the molecule of interest. This method is useful for a molecular weight range of 700-200,000 Da.

#### **2.3.3.2. ESI**

The sample in liquid form is converted into droplets by spraying through the tip of a capillary kept at high voltage (0.5–4 Kv). The fine droplets are allowed to pass through a desolvation capillary, which is continuously supplied by higher voltage. The solvent evaporates due to heat and dry gas present in the chamber. The particles move towards mass analyser under higher potential differences. Electron spray ionization process leads to creation of multiple protonated ions that lower the mass to charge ratio. The charged species are mass analysed based on their mass to charge ratio. The ESI technique is more sensitive compared to MALDI-TOF.

#### **2.3.4. Isothermal Titration Calorimetry (ITC)**

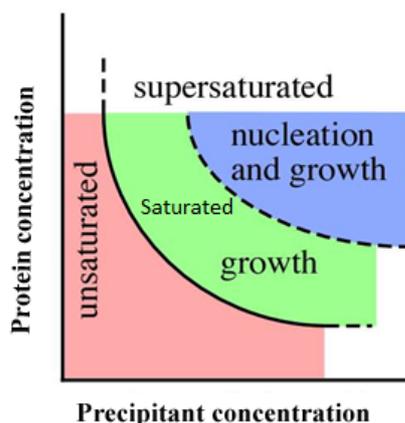
When bio-molecules interact with one another, either release of heat into the surroundings or absorption of heat from the surroundings is observed. ITC is the technique that

measures this heat produced (exothermic) or heat absorbed (endothermic) to determine binding affinities [181]. Isothermal Titration Calorimeter is a very sensitive instrument, and it is used even at nano molar concentrations of reactants. In a typical ITC experiment, the protein sample is placed into a cell (sample cell) and the ligand molecule whose interaction with the protein is being investigated is loaded in the syringe. Both the protein sample and the ligand are taken in the same buffer solution. The ligand molecule is then, injected repeatedly into the cell, leading to generation or absorption of heat. The generation of heat ceases when all the protein molecules are saturated by bound ligand. Once the titration is completed all  $\Delta H$ , dissociation constant  $K_a$ , stoichiometry  $N$  and entropy change  $\Delta S$  are obtained.

#### **2.4. Protein crystallization**

Crystallization is the process of orderly precipitation of solutes from solution. Protein crystallization is one of the bottle necks for x-ray structure determination. For successful crystallization of a protein molecule, the protein should be functionally active, highly purified and homogenous. The process of crystallization can be divided into two steps: 1) nucleation, where protein molecules are sequestered to form clusters and 2) growth, the small nuclei grow into few larger crystals. **Figure-2.3** is the general phase diagram for protein crystallization. The diagram consists of three regions; a) unsaturated region, b) saturated region, and c) supersaturated region. Growth can happen in regions b) and c) while nucleation can happen only in c) region. To find out best crystallization condition is like searching for a needle in a haystack.

The best way to get the leads or hits is to systematically expose the protein dissolved in varieties of buffer solutions to different combinations of precipitants such as salts, poly

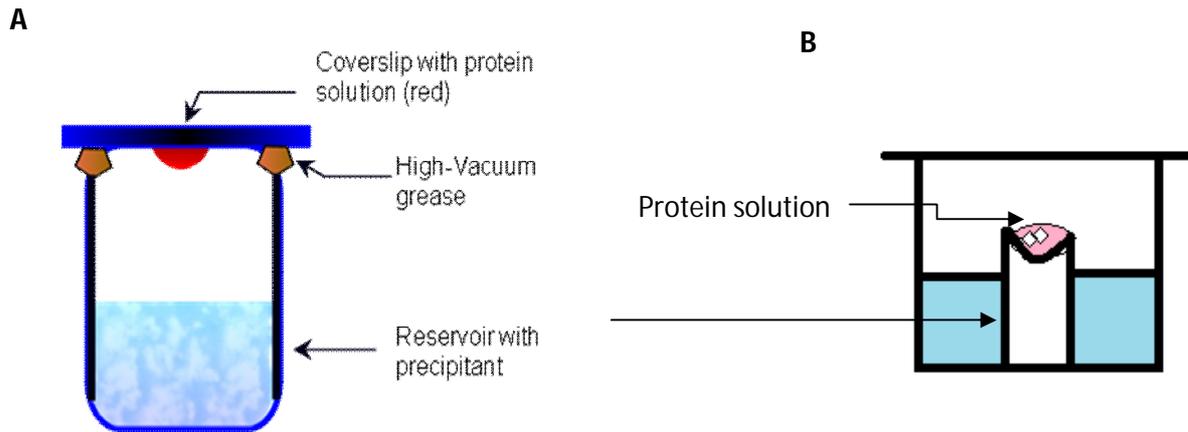


**Figure-2.3: Phase diagram of protein crystallization mediated by precipitant.**

ethylene glycols, of different molecular weights. The sparse matrix screens from M/S Hampton research, like crystal screen 1 and crystal screen 2, are the most popular. The protein crystallization trials can be performed in different ways as described below.

#### **2.4.1. Vapour Diffusion method**

This is the most widely used protein crystallization method. A mixture of protein and precipitant solutions is allowed to equilibrate, through the vapour phase, with a reservoir solution containing higher concentration of the precipitant. The best example is the hanging drop technique in which a protein solution (2-5 $\mu$ l) is mixed with equal volume of reservoir solution, and then the mixture equilibrates with the reservoir solution in a sealed chamber as shown in **Figure-2.4A**. After mixing protein with reservoir, the precipitant concentration in the drop becomes half compared to reservoir concentration. Due to vapour diffusion the water in protein drop transfers to reservoir till equilibrium is reached. At equilibration the net transfer of water ceases which leads to transfer of protein concentration into the nucleation zone. In the nucleation zone few nuclei are formed, which decreases the protein concentration and pushes the nuclei into growth phase. This vapour diffusion method can also be used in the sitting drop mode **Figure-2.4B**.



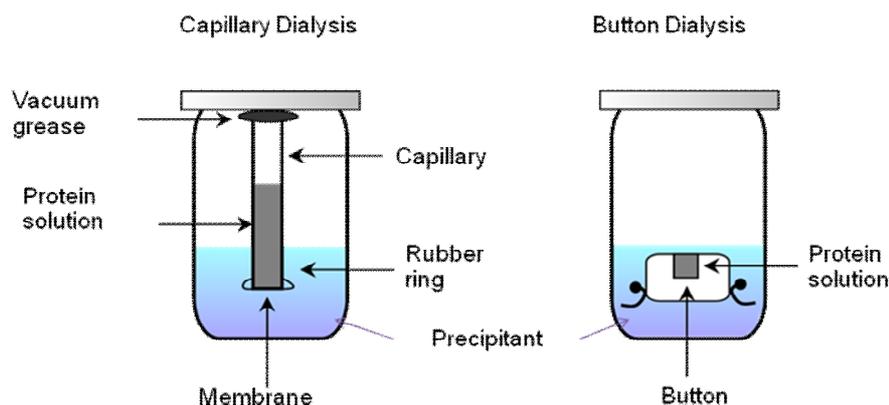
**Figure-2.4: Pictorial representation crystallization methods.** *The hanging drop method is shown in (A) and sitting drop method is shown in (B).*

### 2.4.2. Micro batch method

The vapour diffusion technique sometimes yields too many small crystals in the drop. To overcome the problem, the rate of evaporation is slowed down by using low density paraffin oil (0.87 mg/ml) [182]. The aqueous proteins solution is denser than the oil, and therefore protein remains under the oil, and the rate of evaporation is lowered. It is the method of choice while fine tuning the crystallization condition arrived by screening, and it is very easy to shift from vapour diffusion to micro batch and vice versa [183].

### 2.4.3. Dialysis method

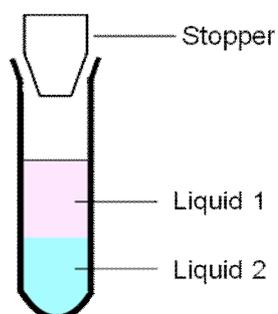
The dialysis method uses slow diffusion and maintains equilibration of precipitant with solute molecules through a semi permeable membrane. The dialysis membrane is used to cover the capillary or the dialysis button which allows the movement of surrounding salt molecules into the solute. As the diffusion occurs through the membrane the system changes from saturated to supersaturated state [184]. This method has its own importance because it allows us to shift the dialysis button/capillary from one condition to another until the appropriate condition is found.



**Figure-2.5: Dialysis method in protein crystallization.** *Capillary Dialysis (a) and Button Dialysis method using dialysis button (b).*

#### 2.4.4. Liquid–liquid diffusion method

This is the most successful method of obtaining crystals, and is sometimes referred to as ‘free-interface diffusion’ method. The denser precipitant solution is layered first, and then the protein solution is added carefully from the side of the tube. The tube is sealed by cork and left undisturbed for 24 hours. Over a period of time, precipitant moves up, and eventually that may lead to crystal formation as shown in **Figure-2.6**. If the crystal does not appear on the interface region, the experiment is again set with higher concentration of precipitant. The commonly used solvent/precipitant pairs for this method are: water/acetone; chloroform or methylene chloride/petroleum ether or cyclohexane; and formic acid/diisopropyl ether. This method can also be useful when very small amount of protein sample (0.1 ml) is available.



**Figure-2.6: Schematic representation of liquid-liquid diffusion method.**

#### **2.4.5. Micro-fluidics technique**

The micro-fluidics technique is just a modification and miniaturization of liquid-liquid diffusion method of protein crystallization, and involves the liquids and small devices arranged on micrometer scale chip [185]. It is also known as “lab on chip” which means the multiple lab operations are designed on the same chip. Using micro fluidics technique the mixture containing protein, precipitant and additives (which has higher diffusion rate) can be mixed completely upon delivery, which cannot be achieved by manual method or using current robotic system. This technique also allows the user to have fine control over transport phenomena, which is useful in protein crystal growth. This technique also increases the surface-to-volume ratio of protein solution, and this has a positive effect on crystallisation kinetics [185]. Recently the micro-fluidics system has been used in the successful batch crystallization of proteins and subsequent in-situ data collection [186].

#### **2.5. X-ray diffraction**

The high energy x-rays of wave length in the range of 0.3-2 Å are used in the determination of three dimensional structures of proteins by the diffraction method. There are three different types of x-ray sources: sealed tube x-ray generators, rotating anode x-ray generator and the synchrotron. In the in-house laboratory, sealed tubes and rotating anodes are used as x-ray source, and the wavelength of x-ray generated is fixed, and depends upon the choice of the anode material. While in the case of synchrotron, wavelength is tuneable, and also the intensity of x-ray beam produced is much higher than in a laboratory source.

When a collimated monochromatic x-ray beam hits a protein single crystal, diffracted x-rays emerge along specific directions leading to formation of discrete spots on the two-dimensional planar detector placed perpendicular to the incident x-ray beam direction.

The directions of the diffracted rays are governed by Bragg's law:

$$\lambda = 2d_{hkl} \sin \theta_{hkl} \quad (\text{Bragg's Law}) \quad \text{-----} \quad (1.0)$$

where  $\lambda$  is the wavelength of the incident x-ray beam,  $2\theta_{hkl}$  is the deviation of the diffracted beam,  $hkl$ , from the incident beam, and  $d_{hkl}$  is the spacing between lattice planes with Miller indices  $h, k, l$ . Each diffraction spot ( $h, k, l$ ) is also described as Bragg reflection ( $h, k, l$ ). The corrected intensity of Bragg reflection ( $h, k, l$ ) is equal to square of the structure factor  $F_{hkl}$ , which is calculated according to the following formula:

$$F_{hkl} = \sum_{A=1}^n f_A e^{2\pi i(hx_A + ky_A + lz_A)} \quad \text{-----} \quad 1.1 [187]$$

Where,  $f_A$  is the scattering factor of atom A and  $x_A, y_A$  and  $z_A$  are the co-ordinates of atom A in the unit cell, expressed as fractions of unit cell axes lengths; a, b, c and n is the number of atoms in the unit cell.

Miller indices of the reflections are co-ordinates of the lattice points in the three dimensional diffraction space known as reciprocal space. The x-ray diffraction data set is typically collected by the “rotation” or the “oscillation method”, in which the crystal is rotated through a solid angle and a contiguous series of images, is collected to get complete dataset [188].

## 2.6. Diffraction data processing

Several software packages such as *iMOSFLM* [189], *DPS* [190], *d\*TREK* [191], *XDS* [192], and *HKL* [193] are available to process the diffraction data collected by the oscillation method. The aims of data processing are to: i) assign Miller indices  $h, k$  and  $l$  to each reflection, and ii) estimate corrected relative intensity of each reflection. By examining systematic absences and by comparing reflection intensities space group symmetry of the crystal can be determined [194]. Finally the data quality is judged by the

values of  $R_{\text{merge}}$ ,  $I/\sigma$  and data completeness. Typically, the ideal dataset should have completeness  $\geq 95\%$ ,  $R_{\text{merge}} \leq 5\%$  and  $I(hkl) / \sigma I(hkl) \geq 2.0$  [195].

## 2.7. Structure solution

The aim of structure determination experiments is to get the atomic coordinates through interpretation of the experimental electron density map of the target molecule. The electron density  $\rho(x, y, z)$  is related to the structure factor  $\mathbf{F}_{hkl}$  by the following equation:

$$\rho(x, y, z) = \frac{1}{V} \sum_h \sum_k \sum_l F_{hkl} e^{-2\pi i(hx+ky+lz)} \quad \text{-----1.2 [187]}$$

where,  $V$  is the volume of unit cell. The structure factor  $\mathbf{F}_{hkl}$  is a complex number, and therefore, is characterised by magnitude  $|\mathbf{F}_{hkl}|$  and phases  $\varphi_{hkl}$ . The amplitudes  $|\mathbf{F}_{hkl}|$  can be deduced by taking square root after applying different (Lorenz and polarisation and other) corrections to the measured intensities  $I_{hkl}$  [196]. The phases cannot be measured directly during data collection. In order to determine the electron density  $\rho(x, y, z)$ , however, one requires phases. This is the famous crystallographic phase problem. There are various methods developed for determining phases, and some of these are:

- a) Molecular replacement method [197-199]
- b) Multiple Isomorphous Replacement method [200]
- c) Anomalous Scattering based methods [201].

In this thesis Molecular Replacement is used as the method of phasing, which is described next, followed by other phasing methods.

### 2.7.1. Molecular Replacement method

In case of Molecular Replacement method, the unknown structure of the protein is solved by making use of known structure of a related protein. The known protein structure is positioned in the unit cell of the unknown protein crystal, by performing rotational and translational searches. In essence, three rotational searches and three translational searches are performed in Patterson space, to achieve best agreement between observed and calculated structure factors. The following are the most recent software packages available for Molecular Replacement calculations: Molrep [202], Phaser [203], AMoRe [204], and CNS [205]. The orientation of the molecule is determined by rotating the Patterson map of homologous structure over the Patterson map of the unknown structure. Once MR solution is obtained, phases of the reflections are calculated by using equation (1.1).

### 2.7.2. Multiple Isomorphous Replacement method

In isomorphous replacement method the idea is to perturb the structure factor by isomorphous attachment of heavy atoms to the protein molecules in the crystal [206]. The heavy atoms such as Pt, Hg, As, I, W etc. contribute for diffraction more compared to the atoms typically present in proteins because of the large number of electrons present in these atoms. The perturbation in diffraction intensities due to addition of few heavy atoms to protein crystals is then easily measured. The positions of heavy atoms in protein can then be determined through analysis of difference Patterson maps [207] computed using  $|\mathbf{F}_{PH}| - |\mathbf{F}_P|$  as the Fourier coefficients, where  $\mathbf{F}_{PH}$  is the structure factor of heavy atom derivative crystal and  $\mathbf{F}_P$  is the structure factor of native crystal. By using Harker construction two possible values for the phase of each reflection are determined. This ambiguity arising out of bimodal probability distribution can be removed by preparing a second derivative crystal in which the heavy atoms bind at different sites on the protein

[208]. Thus to get the complete phase information one needs to prepare at least two derivative crystals.

### **2.7.3. Anomalous Scattering based method**

Whenever the wavelength of incident radiation matches with the absorption edge, the scattering from the atom is described as ‘anomalous’ scattering. This physical phenomenon is exploited to derive reflection phases in the diffraction experiment. By using properly selected wavelength from synchrotron radiation, anomalous scattering is observed in case of heavy atom containing protein crystals [209, 210]. The consequence of anomalous scattering is that intensities of Friedel and Bijvoet mates [211] are no longer equal, and this difference can be used for initial phase estimation. This method facilitates the phase estimation from a single derivative dataset as shown by Ramaseshan *et al.* in 1957 [209]. In protein crystallography, a Se atom is incorporated in the protein by growing bacteria with seleno-methionine instead of normal methionine, and the incorporated Se atom is used as the heavy atom for phasing. The phase information is then obtained by single wavelength anomalous diffraction method (SAD) [212, 213]. For Se-SAD method, the dataset can be collected at home source, and for successful phasing the dataset should be highly redundant [214].

### **2.8. Electron density map interpretation**

With phase information available, the electron density map of the target protein can be calculated using Fourier transform equation (1.2). The electron density map is visualised and atomic model can be manually fitted to the map by using the Crystallographic Object Oriented Tool kit (COOT) program. One normally needs to know the primary structure of the protein. The difference map which uses, as Fourier coefficients,  $(|F_o| - |F_c|)$  is used to locate ligand and water molecules in the unit cell.

### **2.9. Crystallographic refinement**

The initial model which is obtained through manual or semiautomatic interpretation of the electron density map may not be perfect due to insufficient phase information, and also due to errors in initial model building. This situation can be improved by the iterative process of crystallographic refinement and electron density interpretation. Prior knowledge about the stereochemistry of the chemical entities involved can be incorporated into the refinement process. The ultimate goal is to get a stereochemically acceptable atomic model, which can account for the experimentally collected diffraction data. The refinement process is to search for global minimization of a target function,  $E$ , which depends upon two components  $E_{\text{chem}}$  and  $E_{\text{data}}$ :

$$E = E_{\text{chem}} + w_{\text{data}} E_{\text{data}} \quad \text{----- (1.3)}$$

where,  $E_{\text{chem}}$  is the empirical chemical information which involves atomic positions and describes covalent and non-covalent interactions.  $E_{\text{data}}$  represents differences in observed and calculated diffraction data, and  $w_{\text{data}}$  specifies relative weights of the two terms. The target function is formulated either as a likelihood target function or as a least square target function. The parameters varied during the minimisation are atomic coordinates, atomic vibration parameters and atomic occupancies. Estimation of bulk solvent and its contribution to structure factor is also among the varied parameters. The parameter space is systematically searched via rigid body or TLS or molecular dynamics coupled simulated annealing [215] or Monte Carlo approaches [216]. When more than one molecule is present in the asymmetric unit, the non-crystallographic symmetry, if present, is often imposed during refinement. The over fitting of the data can be avoided by taking cross validation using reflections not included in the refinement process [217]. The following software packages for crystallographic refinement are widely used: REFMAC5 [218], CNS [205], BUSTER-TNT [219], SHELX [220] etc.

### 2.10. Structure validation

The refined protein structure needs appropriate validation. The structure can be validated by different quality indicators, such as stereochemistry of the model, Ramachandran plot,  $R_{\text{factor}}$  and  $R_{\text{free}}$ , which are defined in equation (1.3). During the process of structure refinement about 5-10% of diffraction data is removed and the remaining data is used to guide the refinement process. The  $R_{\text{free}}$  value is calculated by analysing, how well the model predicts the 5-10% observations that were not used during refinement. The use of  $R_{\text{free}}$  value avoids the bias in the validation of the model. The R factor measures the agreement between the diffraction pattern and the atomic model. The best refined structure will have  $R_{\text{factor}}$  0.0, which is not possible due to disordered solvent molecules in the crystal and errors in the data.  $R_{\text{factor}}$  close to about 20% is acceptable.

$$R = \frac{\sum || F_{\text{obs}} - k F_{\text{calc}} ||}{\sum |F_{\text{obs}}|} \text{-----} 1.3[42]$$

Where,  $F_{\text{obs}}$  and  $F_{\text{calc}}$  are the observed and calculated structure factors while  $k$  is the scale factor and  $R$  is  $R_{\text{factor}}$  calculated over all the reflections. The computer programs Procheck [221], Molprobity [18], ADIT [222] are used to validate the stereochemistry of the protein model. In a polypeptide, torsion angles phi and psi determine relative positions of neighbouring amino acids in space. In order to avoid steric clashes only certain combinations of phi and psi angles for any given residue are observed in protein structures. The favourable and unfavourable conformations for the polypeptide backbone are marked in a plot of phi and psi values of each amino acid in the polypeptide [223]. For acceptable protein structure, all amino acids should occupy allowed /favourable region in this plot.

**2.11 Conclusion:**

There are three different ways to clone heterologous gene in *E. Coli*. This facilitates over expression and purification of recombinant protein for biophysical and structural studies. The heterologus protein can be purified in single step by affinity chromatography making use of different affinity tags. The x-ray crystallography is one of the powerful tool to determine atomic level structure of biomolecules of any size. Along with x-ray crystallography, different biophysical techniques, as described in this chapter are used in the characterization of protein and protein-protein interaction.

*Structural studies of BRCA1 BRCT-NCOA2 complex*

---

### 3.1. Introduction

Hormones and hormone receptors play an important role in breast carcinogenesis [224]. Over expression of estrogen receptor, progesterone receptor or human epidermal growth factor receptor 2 (HER2) can provide the oncogenic signal [225]. Therefore breast cancers may be classified into following types:

- Endocrine (Estrogen and progesterone) receptor positive: these types grow in response to the hormone estrogen or progesterone.
- HER2 positive: this type is very aggressive and grows in response to growth hormone.
- Triple positive: (Presence of all three receptors).
- Triple negative: (All three receptors are absent).

In carriers of the triple negative BC, the *BRCA1* gene is found to be mutated [226]. Although *BRCA1* is widely expressed in different cells, and is responsible for maintaining the genomic integrity, mutations in *BRCA1* result in tumors only in breast and ovarian tissues by a mechanism which is still not understood. Since breast and ovarian tissues exclusively respond to the hormones, estrogen, androgen and progesterone [227], it is likely that mutant *BRCA1* in association with these hormones is responsible for breast and ovarian cancers. In fact, androgen receptor (AR), by interacting with *BRCA1*, regulates the signalling of prostate and mammary epithelial cell proliferation [228]. This suggests that *BRCA1* may have additional functions specific to mammary tissue, the growth of which is influenced by hormones like androgen, estrogen and progesterone. It has been reported that *BRCA1* in association with the exogenous co-activator NCoA2, enhances AR signalling in both prostate and breast cancer cell lines. To

understand this phenomenon at the atomic level, we have studied the protein-protein interactions between BRCA1 and NCoA2 using bio-physical tools.

### **3.1.1. BRCA1 and Estrogen receptors**

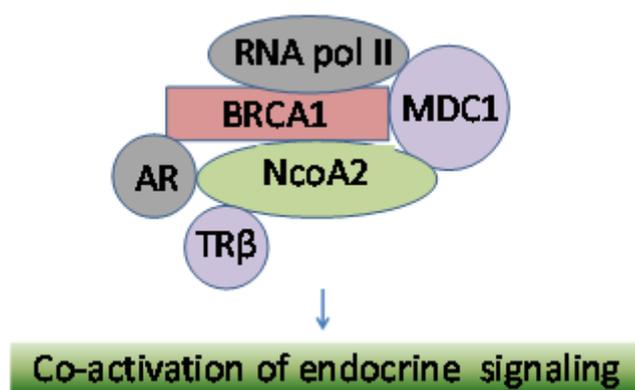
The estrogen receptor (ER) is a well-studied marker in breast cancers, and tumors positive for ER respond well to the hormonal treatment. Experimental evidence suggests that estrogen plays a role in the growth and development of breast cancers [229]. Estrogen secreted by ovaries binds to the estrogen receptors present on specific cell types (breast epithelial cells) leading to transcriptional activation of specific genes that are ultimately responsible for the growth of breast cells [230]. BRCA1 inhibits the signalling process by interacting with estrogen receptor [231]. However, the mutant BRCA1 loses the ability to interact with estrogen receptor, which may be one of the reasons for tumorigenesis [232]. There are two types of estrogen receptors:  $\alpha$  and  $\beta$ . The  $\alpha$  has higher affinity for BRCA1 and acts as a physiological regulator in the breast cells, while the  $\beta$  is anti-proliferative in nature and opposes the actions of estrogen receptor  $\alpha$  in reproductive tissues [233]. The drug tamoxifen, which is an antagonist of estrogen, is used in the breast cancer treatment, as it stops breast cell growth in ER positive tumors [234].

### **3.1.2. BRCA1 and NCoA2**

Fernand et al [235] demonstrated that androgen can inhibit the proliferation of human breast cancer cells (ZR-75-1). Recently it has been reported that under conditions where MAP kinase is already stimulated by EGFR, additional AR stimulation leads to cell cycle arrest through p21 activity [236]. It is known that BRCA1 interacts with androgen receptor (AR) and enhances the AR activity. Further, the ability of BRCA1 to stimulate AR activity was enhanced by several co-activators, including CBP [237], ARA70 [238]

and GRIP1/NCoA2 [239]. Direct interactions of BRCA1 with NCoA2 have been identified earlier [228].

The transcription activator NCoA2, enhances the transcriptional activity by binding to different nuclear receptors like class I (ER, AR, progesterone receptor) and class II (vitamin D receptor, retinoic acid receptor). **Figure-3.1** shows schematically the way in which NCoA2 protein interacts with BRCA1 and leads to transcriptional activation using RNA polymerase II.



**Figure-3.1: The mechanism of transcription co-activation using NCoA2 protein.** *The RNA pol II is known to interact with BRCA1 and also NCoA2 and AR proteins help BRCA1 transcription by interacting with BRCA1 protein.*

However, the exact domain of BRCA1 that is involved in the interaction with NCoA2 protein was not reported. NCoA2 in its N-terminal region contains the consensus amino acid sequence (S-X-X-F) required for interaction with BRCA1 BRCT domain. If the Serine-195 is phosphorylated in-vivo, the NCoA2 may act as an interacting partner of BRCA1 BRCT domain [108]. To explore this interaction, we have carried out biophysical and crystallographic studies described in this chapter.

## 3.2. Materials and Methods

### 3.2.1. Cloning of BRCA1 BRCT domain in pGEX- KT

The BRCA1 BRCT domain (1646-1859) was cloned into pGEX-KT vector by restriction enzyme based method, as described below.

### **3.2.2. Preparation of Insert (BRCA1 BRCT domain)**

The BRCA1 BRCT domain region from (1646-1859) was PCR amplified from full length cDNA of human-*BRCA1* (Kind Gift from Richard Baer) using forward (5'GTCGGA TCCGAGAACCTGTACTTTTCAGGGTGTGAACAAACGTATGTCCAT3) and reverse (5'-GTCGAATCCCTATTAGG GGATCTGGGGTATCAGTATGG-3') primer. The forward primer was designed to have a BamHI and TEV protease cleavage site, while EcoRI restriction site and a stop codon were included in the reverse primer. The PCR amplified *BRCA1 BRCT* region was purified and restriction digested using BamHI and EcoRI, as described in chapter 2.

#### **3.2.2.1. Preparation of Expression vector pGEX-KT**

The full length DNA of vector pGEX-KT was digested by the same set of restriction enzymes as that for the insert, and the linearised vector was purified as described in Chapter2. The purified vector was then quantified using Nanodrop spectrophotometer (Thermo Scientific, USA). This digested vector was further used for ligation with the digested insert of BRCA1 (1646-1859).

#### **3.2.2.2. Ligation**

The restriction enzyme digested insert and vector were mixed in 1:3 molar ratio, and further treated for ligation by Quick DNA ligase at 22°C for about 5 minutes. The ligation product was then transformed in *E.coli* (DH5 $\alpha$ ) competent cells, and the cells were plated on LB plate with ampicillin as selection marker.

#### **3.2.2.3. Screening and DNA sequencing**

The bacterial colonies obtained after the transformation were processed to isolate plasmid DNA for insert screening. These plasmid DNAs were digested by BamHI and EcoRI restriction enzymes and electrophoresed on 1 % agarose gel. The DNA samples which showed insert release of appropriate size, were subjected to DNA sequencing using pGEX-5' forward sequencing primer 5' GGG-CTGGCAAGCCACGTTTGGTG-3' and pGEX-3' reverse sequencing primer 5'-CCG-GGA-GCT-GCA-TGT-GTC-AGA-GG-3'.

### **3.2.3. BRCA1 BRCT domain protein expression and purification**

BRCA1 BRCT construct (1646-1859) cloned in pGEX-kT vector was expressed in *E. coli* BL21 (DE3) strain and the fusion protein was purified using GST affinity chromatography as described below.

#### **3.2.3.1. Starter culture and scale up**

The BRCA1 BRCT protein expression was optimised in a smaller culture by varying the IPTG concentration and also the induction temperature. The optimization results indicate that about 30% of protein was in the soluble form. To scale up the protein expression, a starter culture is prepared by inoculating BRCA1 BRCT domain expressing cells aseptically into 100 ml LB broth containing 100 µg/ml ampicillin. The LB culture was grown for 16 hours on shaker incubator at 37°C under agitation at 300 rpm/min. The 10 flasks containing 1 litre LB broth were seeded with 1% starter culture and incubated on shaker incubator at 37°C, under agitation at 300 rpm, until the culture OD<sub>600</sub> reached between 0.6-0.8. The culture was then induced by adding IPTG to a concentration of 0.4 mM to each flask under agitation at 300 rpm for 16 hours. The induced culture was harvested at 6000 rpm by centrifugation in Sorvall SLC-3000 rotor. The bacterial pellet was either used immediately for protein purification or was stored in -80 °C freezer until further use.

### **3.2.3.2. Protein purification**

The cell pellet was thawed on ice, and was re-suspended in buffer-1 (50 mM Tris, 300 mM NaCl, 0.1% Triton X-100, pH-7.5) of volume depending on the pellet size. The cells were disrupted by ultra-sonication under ice cold conditions. The sonication was done in cycles of 1 min vibration and 1 min rest; 8-10 such cycles were performed depending upon the volume. The cell debris was removed from cell lysate by centrifugation at 18,000 rpm for 50 minutes in SS-34 rotor. The cleared cell lysate contains the BRCA1 BRCT domain protein fused with GST. The column packed with GST sepharose 4B resin was equilibrated against buffer-1. The cleared cell lysate was passed through the column twice to enable the binding of fusion protein to the column. The column was washed with 5-6 column volumes of buffer-1. The BRCT domain protein was separated from GST beads by on column cleavage using TEV protease, which was incubated on the column for 3 hours at room temperature. The BRCT domain protein, now mixed with TEV protease, was further purified by passing through Ni-NTA resin, which would extract 6His tag TEV protease. The BRCT protein was further purified by gel filtration chromatography.

### **3.2.3.3. Gel filtration chromatography**

The BRCT protein was concentrated by centricon (10 kDA Molecular weight cut off) centrifugal filter units (Millipore, USA), and injected into AKTA purifier system (GE, USA) connected to Hiload superdex-200 column (GE, USA). The protein flow was monitored by measuring the absorbance at 280 nm. The purified protein fractions were collected and further analysed for purity on SDS-PAGE.

### **3.2.3.4. SDS-PAGE analysis**

The SDS-PAGE separates the biomolecules depending upon their molecular mass. The 12% gel was prepared and loaded with different samples to monitor the protein expression and purification process. Each sample was mixed with sample dye and denatured by heating at 100 °C for 5 min. The protein bands were visualised by staining with coomasie brilliant blue.

### **3.2.4. Biophysical characterisation**

The purified BRCT domain was further analysed by mass spectrometry to determine its exact molecular weight and identity.

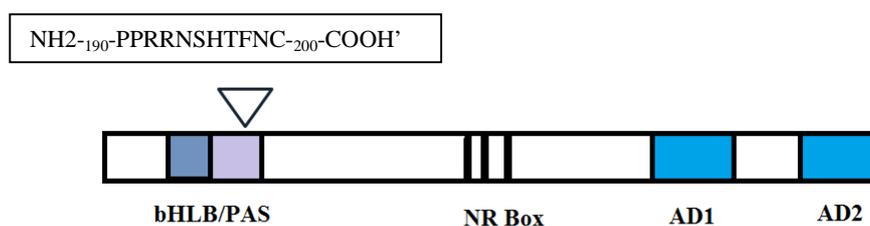
#### **3.2.4.1. Mass Spectrometry**

FPLC purified BRCT domain protein at a concentration of about 0.5 mg/ml was used for mass spectrometry analysis. The peptide mass finger printing was done by performing in solution trypsin digestion, followed by extraction of peptides, whose masses were then determined by mass spectrometry. The peptide peaks were then analysed in MASCOT software to determine protein identity.

#### **3.2.4.2. Interaction analysis using ITC**

The ITC-200 instrument (GE, Sweden) was used to perform interaction analysis. The NCoA2 peptide is selected from N-terminal region of NCoA2 protein (**Figure-3.2**). The BRCA1 BRCT protein at a concentration of 0.02 mM and NCoA2 peptide at a concentration of 0.200 mM in the FPLC buffer (50mM Tris, pH-7.5 and 300mM NaCl) were used for ITC analysis. In a typical ITC experiment BRCA1 BRCT protein was kept in sample cell and peptide solution was injected as 2 µl aliquots under constant stirring at 1000 rpm. The ITC experiment consisted of a total 16 injections with a 210 sec gap between two successive injections, and was performed at 25°C. The experimental and injection parameters are recorded in **Table-3.1**. The ITC experimental data was fitted

using Origin software (Version 7.2) to calculate binding affinity and stoichiometry. The heat of dilution obtained by titrating the peptide with the buffer and was treated as blank isotherm. The base line was corrected by subtracting blank isotherm from test heat isotherm. The curve fitting is carried out by non-linear chi square analysis and  $K_a$ ,  $\Delta H$  and stoichiometry (N) values are calculated.



**Figure-3.2: Domain organization of NCoA2 protein.** It has N-terminal basic Helix-Loop-Helix domain (bHLH) and Per Arnt Sim domain (PAS). It also has three nuclear receptor box domains (NR Box) at central region and two Activation domains (AD1 and AD2) at the C-terminus. The NCoA2 peptide containing the BRCA1 binding sequence mapped at N-terminal domain.

**Table-3.1:** ITC experimental and injection parameters.

Experimental parameters	
Total injections	16
Cell Temperature (°C)	25
Ref power (μcal/sec.)	5
Initial Delay (sec.)	250
Syringe conc. (mM)	0.2
Cell conc. (mM)	0.02
Stirring speed (RPM)	1000
Injection parameters	
Injection volume (μl)	2
Duration (sec.)	40
Spacing (sec.)	210
Pipette volume (μl)	~38
Filter period (sec.)	5

### **3.2.5. Protein structure determination**

#### **3.2.5.1. Protein crystallization**

The FPLC purified BRCA1 BRCT protein was concentrated to 25 mg/ml as described in chapter 2. The concentrated protein was mixed with the peptide at a concentration of 10 mg/ml to obtain a 1:1.5 molar mixture, which was then incubated overnight at 4°C. The crystallization experiment was carried out at 295 K. Initial crystallization experiments were performed using Hampton crystallization kits (Hampton Research, Inc). The crystallization attempts used both sitting and hanging drop methods, where a 4 µl drop consisting of 2 µl complex solution and 2 µl precipitant was concentrated through vapour diffusion against 0.5 ml of precipitant solution. This crystallization condition was further optimised by systematically varying the concentration of precipitant and salt.

#### **3.2.5.2. Diffraction data collection and processing**

The complex crystals were cryo-protected using 30% (v/v) glycerol prepared in mother liquor solution, and then were snap frozen in liquid nitrogen before exposure to X-rays. The crystals diffracted to around 3 Å resolution on the rotating anode x-ray generator operated at 50 kV and 100 mA (National Chemical Laboratory, Pune). However, the crystals diffracted to around 1.7 Å resolution on the BM14 beam line at ESRF. The two diffraction datasets were collected at 100 K by the oscillation method. The exposure time and detector distance were set in such way as to record all possible reflections with minimum overloads. The diffraction data were processed using iMOSFLM [240] software, and scaled using SCALA program from CCP4 suite [241]. The complex crystals belong to P3<sub>2</sub>21 space group (no.154), and data intensity statistics is given in

#### **Table 3.2**

**Table-3.2-** Summary of data collection, processing and refinement statistics

No. of crystals	1
X-ray source	BM-14, ESRF
Wavelength (Å)	0.9000
Crystal to detector distance (mm)	156.04
Space group	P3 <sub>2</sub> 21
Unit cell parameters (Å)	$a = b = 65.8, c = 93.1$ $\alpha = \beta = 90.0^\circ, \gamma = 120.0^\circ$ .
Mosaicity (°)	0.6
Resolution limit (Å)	28.46-1.7 (1.79-1.7) <sup>a</sup>
Total No. Of reflections	304855
Unique reflections	25805
Redundancy	11.8(12.1)
$I/\sigma I$	10.0 (3.2)
Completeness (%)	100 (100)
$R_{\text{merge}}$ (%) <sup>b</sup>	25(239)
Wilson B factor	18.7
CC1/2	0.7 ( at 1.7 resolution)
<b>Refinement statistics</b>	
$R_{\text{work}}$ (%) <sup>c</sup>	21.57
$R_{\text{free}}$ (%) <sup>d</sup>	25.78
Total no. of residues	210
Total no. of water molecules	259
Overall B factor	23.07
B factor of ligand	35.5
RMSD bond length (Å)	0.008
RMSD bond angle (°)	1.3
<b>Ramachandran plot analysis</b>	
Most favoured (%)	96.28
Additionally allowed (%)	3.26
Disallowed region (%)	0.47

<sup>a</sup> Values in parentheses are for the highest-resolution shell (1.7-1.79)

<sup>b</sup>  $R_{merge} = \frac{\sum |I - \langle I \rangle|}{\sum I}$  where  $I$  is the observed integrated intensity,  $\langle I \rangle$  is the average integrated intensity obtained from multiple measurements, and the summation is over all observed reflections.

<sup>c</sup>  $R_{work} = \frac{\sum_{hkl} \| |F_{obs}(hkl)| - K |F_{calc}(hkl)| \|}{\sum_{hkl} |F_{obs}(hkl)|}$  Where  $F_{obs}$  and  $F_{calc}$  are observed and calculated structure factors respectively.

<sup>d</sup>  $R_{free}$  was calculated as for  $R_{work}$  but only 5% data left out of refinement procedure has been used in the calculations.

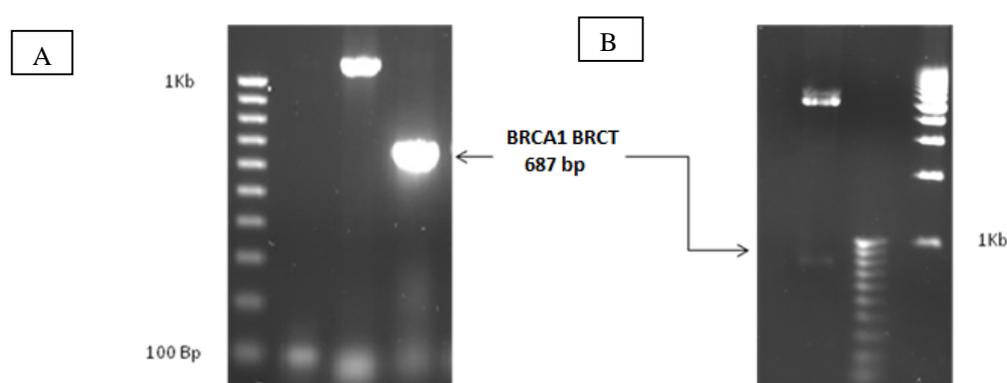
### 3.2.5.3. Structure solution by Molecular Replacement

The crystal structure of the complex was solved by the Molecular Replacement (MR) method [198] briefly described in chapter 2. The BRCA1 BRCT-BACH1 complex structure (PDB ID-1T15) [101] obtained from the protein data bank was used as the search model after the ligand and water molecules were removed from the coordinate file. The calculated  $V_M$  [242] value of  $2.32 \text{ \AA}^3 \text{ Da}^{-1}$  corresponds to a solvent content of 47%, and suggests presence of one protein molecule per asymmetric unit. The rotational and translation search for single molecule in the asymmetric unit was performed using PHASER software [203]. The CCP4 [241] and PHENIX [243] software suites were used for the refinement of the structure. Positive electron density was seen near Gly-1655 residue in the Fo-Fc difference map. The ligand containing the phosphorylated serine and remaining amino acid residues, was built into the electron density, and was subjected to several rounds of refinement. The appropriate LINK statement for modified serine residue was already available in the REFMAC dictionary [244]. In the final stages, few cycles of TLS refinement were also carried out. At the end of refinement most of the residues (96%) were in Ramachandran allowed region [223].

## 3.3. Results and discussion

### 3.3.1. Cloning of BRCA1 BRCT domain

**Figure-3.3A** shows DNA gel electrophoresis of the BRCA1 BRCT domain PCR product amplified from the cDNA of full length BRCA1. One can see a prominent band of approximate size 687 bp, which was inserted into the pGEX vector. **Figure-3.3B** shows the DNA insert release pattern from one potential clone of BRCA1 BRCT domain, after restriction digestion by BamHI and EcoRI enzymes. The clone was further confirmed by DNA sequencing.

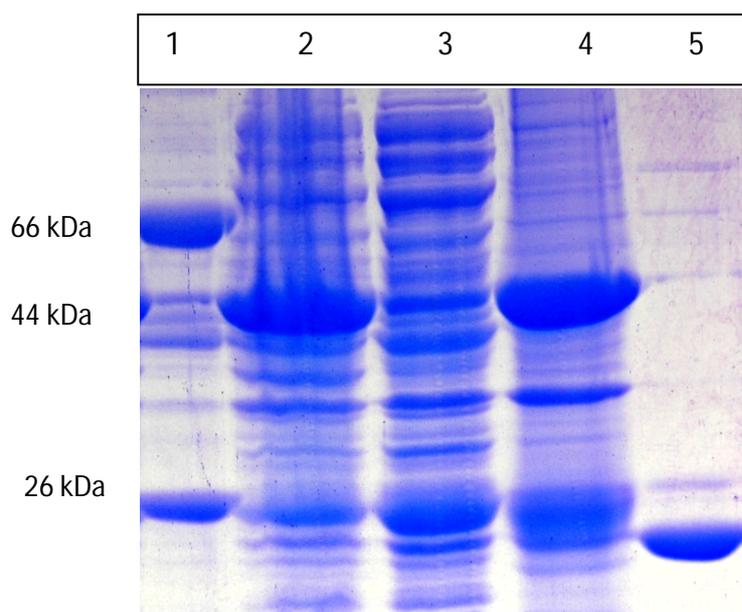


**Figure-3.3: PCR amplification of BRCA1 BRCT (A) and screening of potential clones of BRCA1 BRCT (B)**

### 3.3.2.Expression and purification of BRCA1 BRCT domain

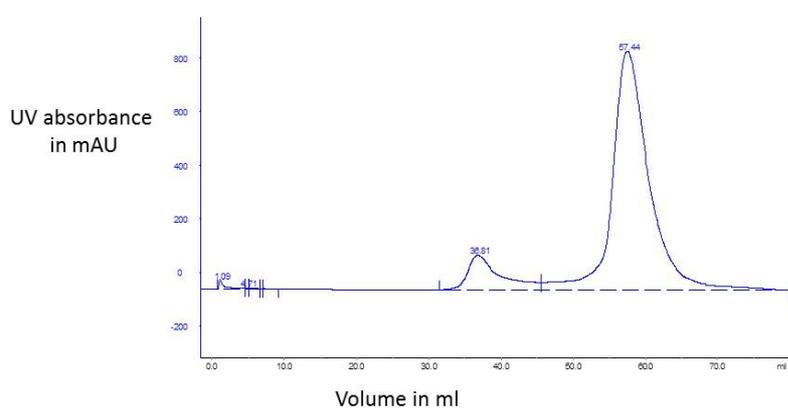
The BRCA1BRCT domain is cloned under *tac* promoter in pGEX vector and expressed in BL21 (DE3) through IPTG induction. The protein is purified by affinity chromatography using GST sepharose 4B resins. The sample purity is monitored by collecting protein samples at different stages of protein purification and loading them together on SDS-PAGE (**Figure-3.4**).

The partially purified protein was passed through a gel filtration column (volume 120 ml) in the fast protein liquid chromatography (FPLC) system, to separate oligomeric populations and contaminations. The protein elution fractions were monitored by measuring absorbance at 280 nm, and the elution profile is shown in **Figure-3.5**.

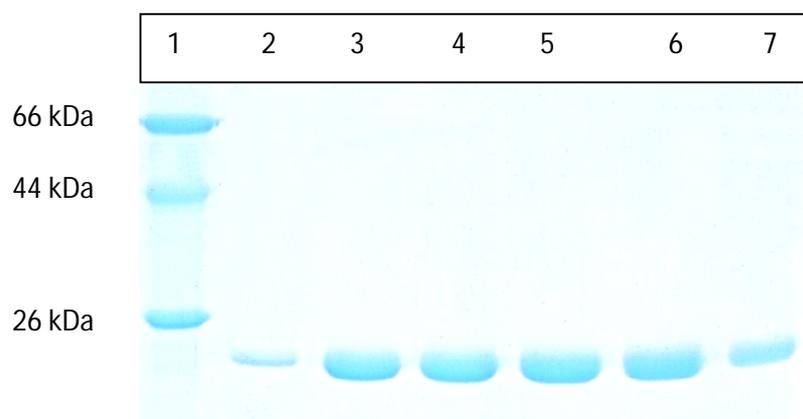


**Figure-3.4: Protein expression profile of BRCA1 BRCT.** Protein samples used to monitor the protein purification process are: **Lane 1:** the molecular weight ladder, **Lane 2:** induced whole cell extract, **Lane 3:** induced soluble fraction, **Lane 4:** fusion protein-bound beads and **Lane 5:** sample after TEV treatment.

The BRCA1 BRCT protein eluted at the expected elution volume along with a higher molecular weight species. On SDS-PAGE, the secondary peak corresponded to higher oligomer of BRCA1 BRCT. It can be seen that the protein purified from the FPLC was highly pure (**Figure-3.6**).

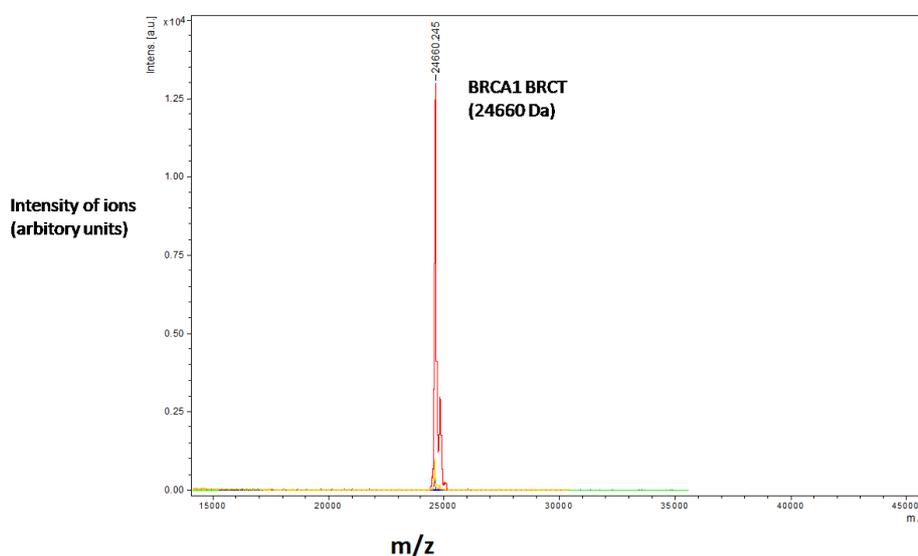


**Figure-3.5: FPLC chromatogram of BRCA1 BRCT domain.** Hiload Superdex 200 column is used for FPLC, where X axis represents the volume of buffer passed and Y axis represents UV280 absorption in arbitrary units.



**Figure-3.6:** SDS-PAGE profile of FPLC purified fractions of BRCA1 BRCT domain. Where Lane -1: molecular weight ladder, lane -1-6 different FPLC purified fractions.

**Figure-3.7** shows the mass spectrum of the purified sample. The mass profile clearly shows a single peak confirming the homogeneity of the sample. The experimental molecular weight is matching the expected value of 24660 Da.

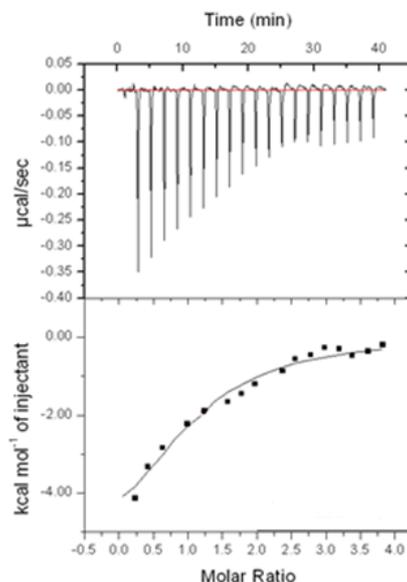


**Figure-3.7:** The mass spectrometry profile of BRCA1 BRCT. The X-axis represents molecular mass divided by ionic charge, and Y-axis represents the intensity of charged ions.

### 3.3.3. NCoA2 peptide interacts with BRCA1 BRCT domain

The enthalpic change on interaction of BRCA1 BRCT with NCoA2 peptide was explored using ITC as shown in **Figure-3.8**. The heat change shows that NCoA2 peptide

exothermically interacts with BRCA1 BRCT domain, and the binding affinity of this interaction was calculated to be 8.7  $\mu\text{M}$ .



**Figure-3.8: Representative ITC result obtained for the interaction of BRCA1 BRCT with NCoA2 peptide.**

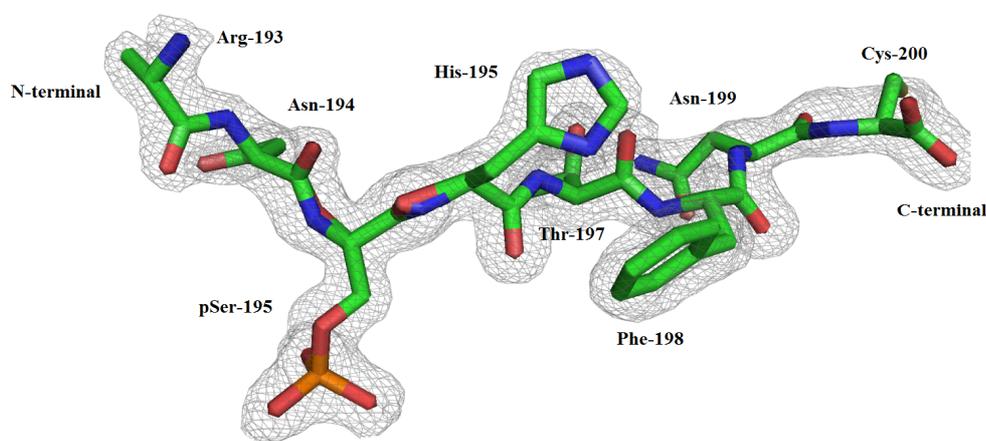
The binding affinity of NCoA2 with BRCA1 BRCT is weaker than those observed for other oligopeptides: BACH1 (0.9  $\mu\text{M}$ ) [21], CtIP (3.7  $\mu\text{M}$ ) [27] and ACC1 (3.4  $\mu\text{M}$ ) [28]. The lower binding affinity of NCoA2 peptide may be due to absence of polar residues at positions -2, +4 and +5. This indicates that the residues other than pS(0) and F(+3) also make significant contribution to the interaction with the BRCA1 BRCT domain. These differences in binding affinities may be helpful to BRCT domain in the selection of binding partner while performing different functions such as transcriptional activation, DNA repair and checkpoint activation.

#### **3.3.4. Crystal Structure of BRCA1 BRCT - NCoA2 peptide complex**

Single crystals of BRCA1 BRCT - NCoA2 peptide complex were obtained in 0.2M ammonium sulphate, 0.1M MES monohydrate pH6.5, and 30% poly ethylene glycol

mono methyl ether buffer condition. The crystals were confirmed as protein crystals by staining with IZIT dye from Hampton Research. Crystal of BRCA1 BRCT-NCOA2 peptide complex belongs to a trigonal system, and the space group was established as  $P3_221$ , based on intensity equivalents, systematic absences and MR calculations. Molecular Replacement calculations [198] yielded a single solution characterised by a TFZ score of 49 and a log likelihood gain (LLG) score of 3148.9, and thus indicated the presence of a single molecule in the asymmetric unit. Subsequently, the structure was refined by REFMAC5 [218] using standard protocols. The final refinement statistics are given in **Table-3.2**. The low crystallographic R-factor and the good stereochemistry indicate that the structure has been determined to good accuracy. Positive electron density is observed near Ser-1655 residue. Therefore the present structure is that of a molecular complex between BRCA1 BRCT and NCoA2 peptide. In the complex, eight of the eleven residues of the ligand could be built in the electron density map and refined satisfactorily. The three amino terminal residues Pro-191, Arg-192 and Arg-193, do not have clear electron density. Since for Arg-193 residue, the electron density was visible only up to the C  $\beta$  atom, this residue was modelled without side chain (**Figure-3.9**). Structural superposition with unliganded protein (BRCA1 BRCT PDB ID-1JNX) [10] yields rmsd of 0.48 Å over 213 C $\alpha$  atom pairs showing that the protein conformation is not altered much upon ligand binding. **Table-3.3** gives rmsd values for superposition of protein residues in complexes of BRCT with oligopeptides NCoA2, BACH1, CtIP and ACC1 [101, 102, 113]. The **Table-3.3** also gives the C $\alpha$  – C $\alpha$  separation of corresponding oligopeptide ligand residues when the protein residues alone were used to superpose the complex structures. It is interesting that the ligand residues, especially the central four residues, superpose much closer (0.36 Å) compared to all residues in peptide (0.68 Å). As expected, the

protein residues involved in binding the central ligand residues also overlap better on one another, suggesting that the ligand binding site is structurally conserve



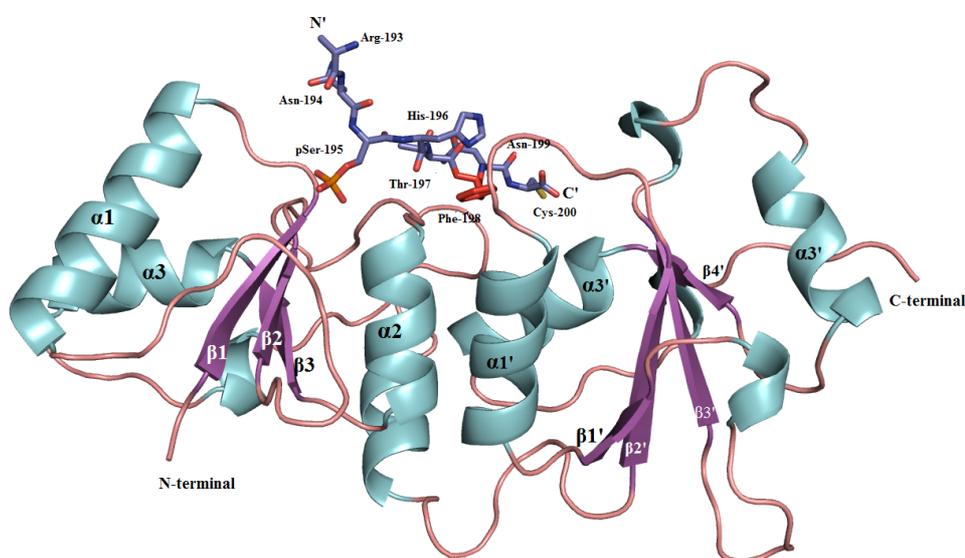
**Figure-3.9:** The final atomic model and 2Fo-Fc electron density map for the NCoA2 peptide at 1.7 Å resolution. The electron density map is contoured at  $1\sigma$  level.

**Table-3.3:** RMSD values for protein  $C\alpha$  atoms obtained upon superposition of different BRCA1 BRCT domain complex structures. The values in the parenthesis correspond to  $C\alpha$ - $C\alpha$  separation of oligopeptide ligands in the complexes.

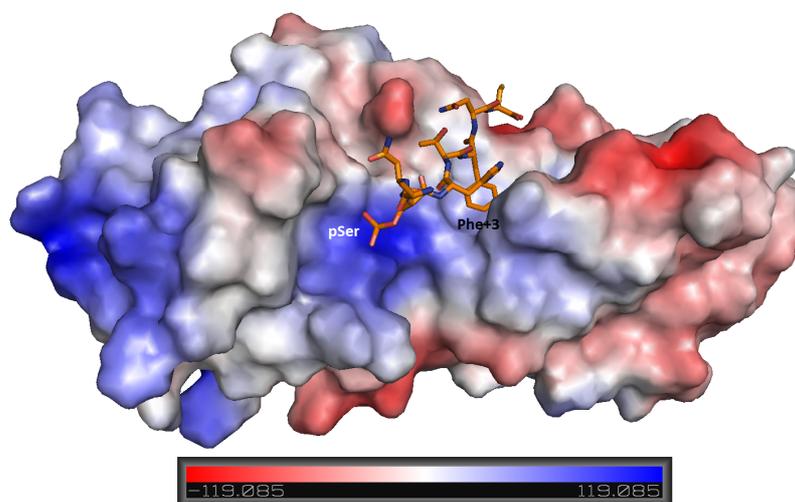
	BACH1	NCoA2	CtIP	ACC1	ATRIP
BAAT	0.28 (0.25)	0.61 (0.20)	0.86 (0.48)	0.9 (0.62)	0.45 (0.43)
ATRIP	0.29 (0.56)	0.66 (0.34)	0.79 (0.46)	0.58 (0.79)	
ACC1	1.0 (0.48)	0.43 (0.57)	0.84 (0.79)		
CtIP	1.0 (0.45)	0.67 (0.44)			
NCoA2	0.88 (0.27)				

### 3.3.5. Interaction of BRCA1 BRCT with NCoA2

The two BRCT domains in BRCA1 consist of three  $\alpha$ -helices and one  $\beta$ -sheet, with a topology of  $\beta 1-\alpha 1-\beta 2-\beta 3-\alpha 2-\beta 4-\alpha 3$ . The NCoA2 peptide is bound to the BRCT domain in a “two knob” manner, in which the pSer (0) residue from NCoA2 interacts exclusively with the N-terminal BRCT while the Phe (+3) residue, by being at the interface, interacts with both the N-terminal and C-terminal BRCT (**Figure-3.10**). The environment around the pSer (0) residue is electrostatically charged while that around Phe (+3) is hydrophobic, as may be seen from **Figure 3.11**.



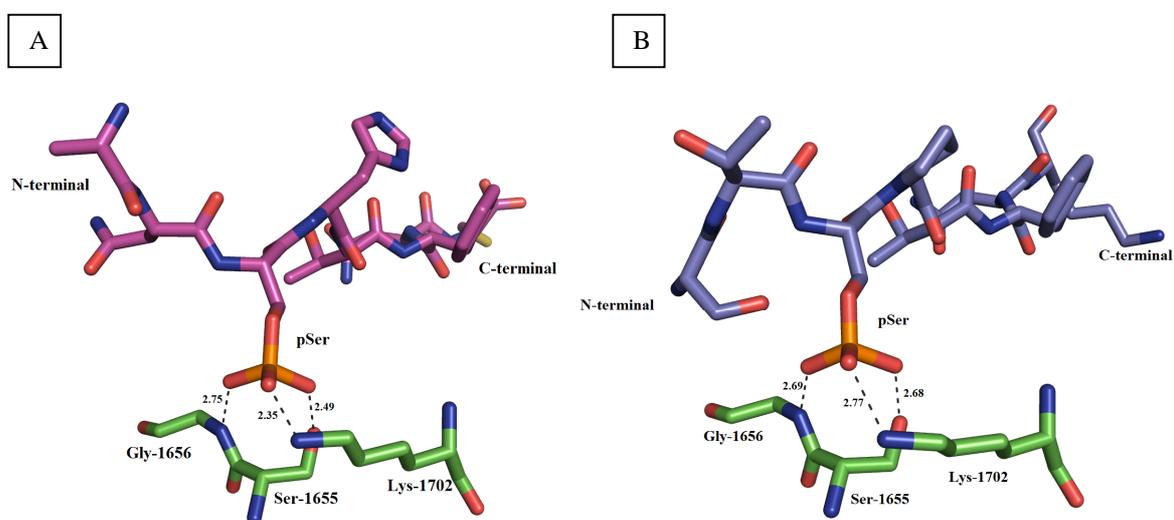
**Figure-3.10: NCoA2 peptide recognition by BRCA1 BRCT domain.** *Ribbon representation of BRCA1 BRCT bound to NCoA2 peptide (blue colour stick model). The  $\alpha$ -helices and  $\beta$ -sheets coloured cyan and magenta respectively. The secondary structure elements of C-terminal BRCT are labelled with primes. The critical residues of NCoA2 peptide denoted as pSer-195) and Phe-198). The figure was made using PyMOL [245].*



**Figure-3.11: Electrostatic surface representation of BRCA1 BRCT-NCoA2 complex structure.** The surface is colored by the electrostatic potential (red-negative, blue-positive). The figure was made using PyMOL [245].

### 3.3.5.1. Interactions around pSer<sub>(0)</sub> residue

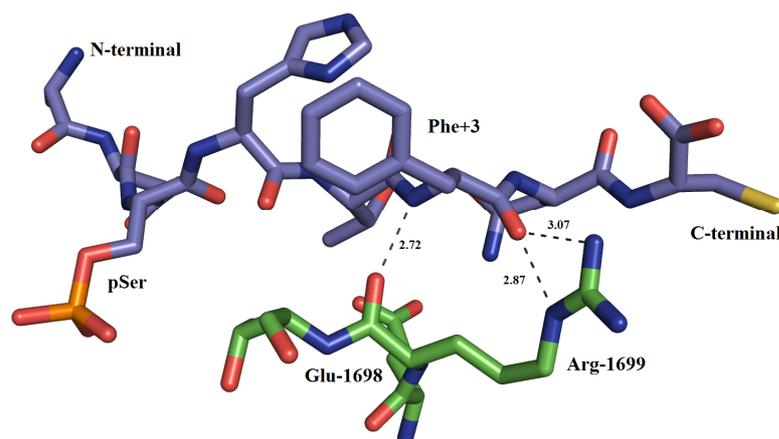
The pSer<sub>(0)</sub> residue forms hydrogen bonds with Ser-1655, Gly-1656 and Lys-1702 residues as shown in **Figure-3.12**. It may be noted that the two hydrogen bonds from the phosphate oxygen to 1702 (Lys) [NZ] and to 1655 (Ser) [OH] are short (2.35 and 2.49 Å respectively), and hence very strong. The shortness suggests that partial electrostatic charges are residing on the hydrogen bond donor and/or hydrogen bond acceptor atoms in the crystals, which were grown at pH 6.5. Interestingly, in the case of BACH1 complex, which also was prepared at pH 6.5, the Lys NZ – OPO<sub>3</sub> hydrogen bond is a normal hydrogen bond of length 2.9 Å.



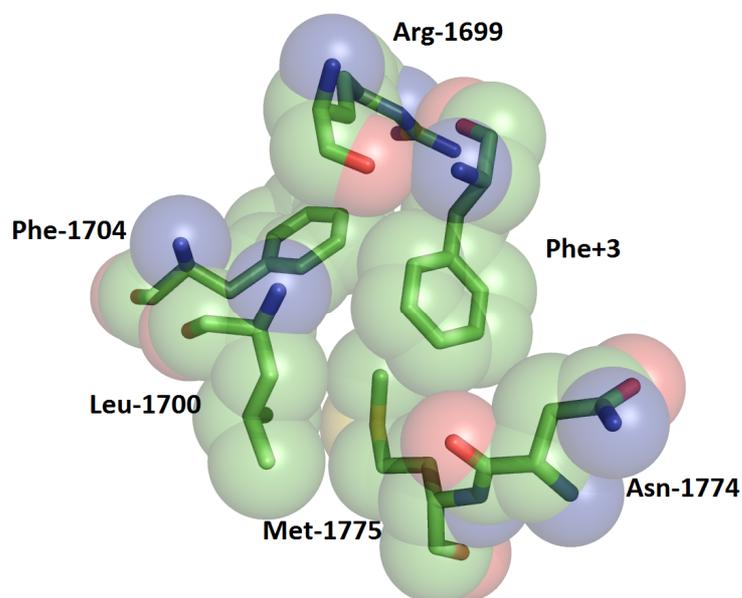
**Figure-3.12: The phosphate group of peptides from NCoA2 (A) and BACH1 (B) forms three hydrogen bonds with BRCA BRCT domain,**

### 3.3.5.2. Molecular geometry around Phe<sub>(+3)</sub> residues

The carbonyl oxygen and the N-H group of Phe (+3) residue form three hydrogen bonds with Arg-1699 as shown in (**Figure-3.13**). The aromatic side chain of Phe (+3) residue stacks against Phe-1704 residue in the hydrophobic pocket formed by residues from both the N and C-terminal BRCTs as shown in **Figure-3.14**.

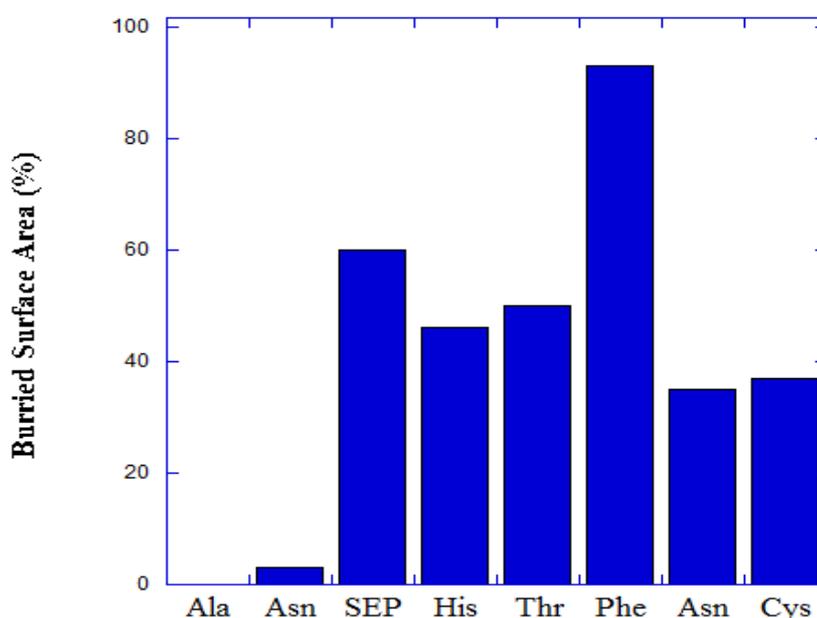


**Figure-3.13:** The hydrogen bond network formed by Arg-1699 with NcoA2 peptide.



**Figure-3.14:** Space filling model representing environment around Phe+3 of NCoA2 peptide.

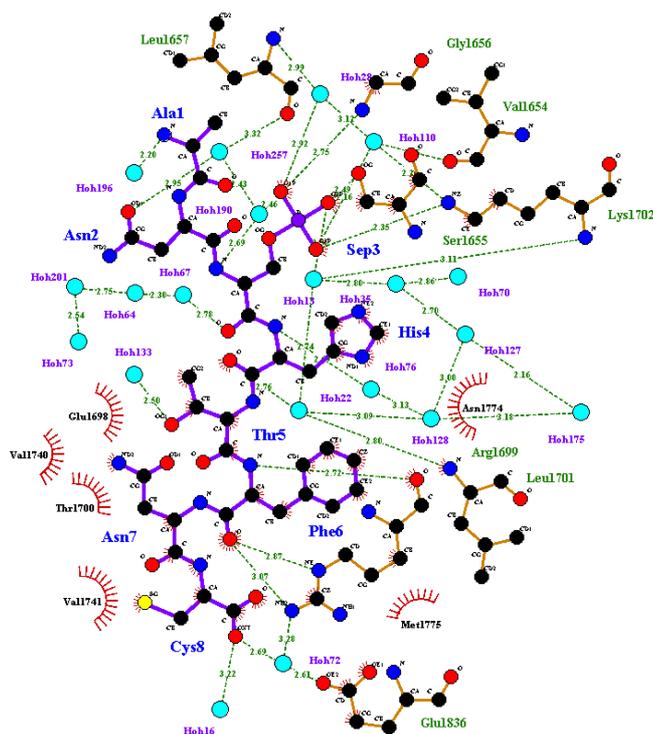
The Phe (+3) residue is buried in the hydrophobic pocket surrounded by Arg-1699, Leu-1700, Phe-1704, Asn-1774 and Met-1775, and is totally inaccessible as may be seen from **Figure 3.15**. The residues of BRCA1 BRCT contributing in interactions with pSer (0) and Phe (+3) of NCoA2 oligopeptide, are among those found in the earlier complex structures of BRCA1 BRCT with BACH1, CtIP1 and ACC1 [21] [27] [28].



**Figure-3.15:** Contribution of individual residues in the phospho-peptide to the surface area burial in the BRCT-NCoA2 complexes.

### 3.3.5.3. Residues other than pSer<sub>(0)</sub> and Phe<sub>(+3)</sub>

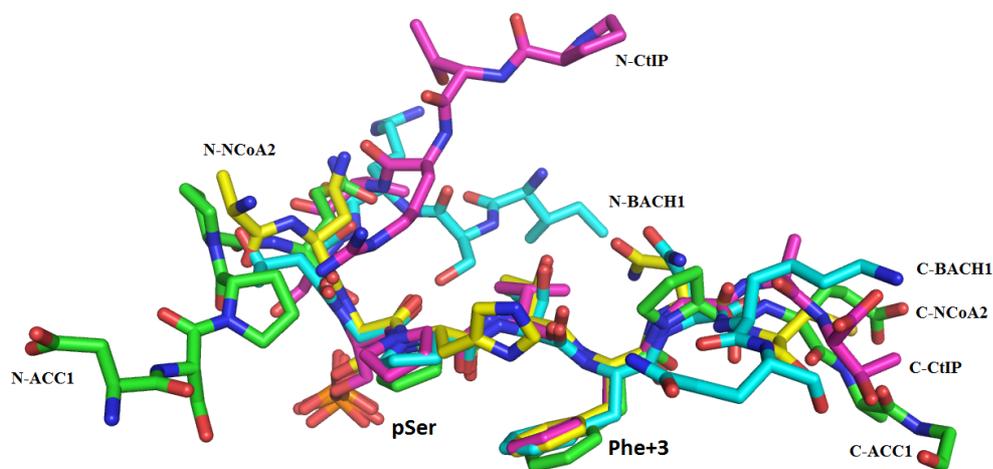
Good electron density is observed for eight residues of the ligand from amino acids 193-200. The residues Asn-194 and His-196 form water mediated hydrogen bonds with carbonyl oxygen of Leu-1657 and Leu-1701 residues respectively. The Cysteine at +5 position is forming two hydrogen bonds with water molecule-16 and Glu-1836 (**Figure-3.16**).



**Figure-3.16:** Two-dimensional representation of the interactions between BRCA1 BRCT (purple bonds) and NCoA2 residues (orange bonds). Water molecules (W) are shown as cyan spheres, hydrogen bonds as dashed lines, and hydrophobic interactions as arcs with radial spokes. This figure was made using LIGPLOT [246]

### 3.3.6. Comparison of oligopeptides conformation

**Figure-3.17** shows the relative positions of oligopeptide ligand when BRCT is structurally superposed in the four complexes.



**Figure-3.17:** Structural superposition of oligopeptides represented in stick model, where BACH1 (Cyan) CtIP (red), ACC1 (green) and NCoA2 (yellow),

It may be seen that the peptide residues from 0 to +3 are almost identically positioned (with the average rmsd of 0.36 Å), however, there are differences in the positions of residues at N and C-terminal ends of the phosphorylated Serine residue. Consequently their interactions with BRCT domain could also be different, thereby contributing to differences in the binding affinities.

### **3.4. Conclusion**

The oligopeptide corresponding to the residues 190-200 from the N-terminal region of NCoA2 protein, and selected for study here contains the signature sequence pSXXF. The oligopeptide is found to have 8.7 μM binding affinity for the BRCA1 BRCT domain. The NCoA2 oligopeptide has been co-crystallized with BRCA1 BRCT domain, and the crystal structure was determined/refined at 1.7 Å resolution. Of the 11 residues in the oligopeptide, electron density is observed for eight internal residues including the central tetrapeptide.

The complex structure revealed a two knob mode of binding for the NCoA2 peptide with BRCA1 BRCT domain. The pSer<sub>(0)</sub> and Phe<sub>(+3)</sub> form main interaction centres: pSer interacts with N-terminal of BRCT through a number of hydrogen bonds, while the Phe<sub>(+3)</sub> interacts with interface region and the C-terminal BRCT through hydrophobic interactions. The crystal structure of BRCA1 BRCT complexed with NCoA2 provides new insight into specifics of BRCA1 binding to various cellular partners. Further, this crystal structure could be exploited in the design of small molecules that can activate androgen receptor (AR) regulated signalling pathways for cell cycle arrest.

*Characterization of Nup153 and RBP-12 as  
interacting partner of BRCA1 BRCT domain*

---

## 4.1. Introduction

DNA repair, like other DNA dependent processes, is highly compartmentalized in the nucleus, and BRCA1 plays a very important role in the DNA damage repair pathway (described in Chapter1). [3]. There is, however, conflicting data on the sub-cellular location of BRCA1 protein. One group proposes that, it is a nuclear protein in most of the cell lines, and is cytoplasmic only in the breast and ovarian cell lines [247]. The other group claims that BRCA1 is found exclusively in the nucleus [248].

Since BRCA1 BRCT domain recognizes proteins with pS<sub>0</sub>-X-X-F<sub>+3</sub> sequence, to identify new interacting partners of the BRCA1 BRCT domain, we have searched for [S-X-X-F] motif in the UniProtKB protein database (<http://www.uniprot.org/>). This search resulted in identification of two proteins, Nup153 and RNA binding domain, as potential partners, and one of this had been reported by Rodriguez *et al* [108]. In the Nup153 protein the consensus sequence is at the C-terminal region (as shown in **Figure-4.1**), while in the RNA binding protein 12 (RBP-12), it is present in the N-terminal region (as shown in **Figure-4.2**). Nup153 is one of the mobile nucleoporins, and is known to be an important part of the nuclear pore complex [249]. The nuclear pore complex is comprised of more than 30 different types of nucleoporins, and these help in the transport of proteins from cytoplasm to nucleus through nuclear translocation signals [250]. Recently, it has been shown that Nup153 promotes 53BP1 to the nuclear foci, and facilitates repair of the DNA double strand breaks [16]. NUP153 is essential for the proper activation of the DNA damage checkpoints, and is also known to regulate the choice between NHEJ and HR [16].

RNA binding protein 12 (RBP-12), associated with pre mRNA, is part of a heterogenous ribonucleoprotein complex, which helps in gene splicing and polyadenylation [251]. We

have obtained oligophosphopeptides corresponding to Nup153 and RBP-12 sequences to study the interactions between BRCA1 BRCT and phosphopeptides. The location of these oligopeptides in Nup153 and RBP-12 is shown in **Figures-4.1** and **4.2**.



**Figure-4.1: Nup153 domain organisation.** *The N-terminal region of Nup153 contains Nuclear Pore Associated Region (NPAR) and central Zn finger region and C-terminal FG rich region. The Nup153 peptide is mapped in the FG rich region.*



**Figure-4.2: Domain organisation of RBP-12 protein.** *The N-terminal region RBP-12 contains proline rich region and the signature motif for BRCA1 BRCT binding mapped from this region. The RNA Recognition Motif (RRM) present at central and C-terminal region of protein.*

## 4.2. Materials and methods

### 4.2.1. Expression and purification of BRCA1 BRCT domain.

The BRCA1 BRCT domain was purified using affinity chromatography as discussed in chapter 3.

### 4.2.2. Binding affinities

The two oligopeptides, of Nup153 and RBP-12 sequences were commercially obtained, and used without further purification. The interaction studies between BRCA1 BRCT domain with these oligopeptides were carried out at 25°C, using ITC-200 (GE, Healthcare). The concentration of BRCA1 BRCT protein was 0.02 mM, while the

concentrations of Nup153 and RBP-12 peptides were both 0.2 mM. The peptide was mixed in the same FPLC buffer containing 50mM Tris, pH-7.5 and 300mM NaCl, which was used

**Table-4.1:** ITC experimental and injection parameters

<b>Experimental parameters</b>	
Total injections	16
Cell Temperature (°C)	25
Ref power (µcal/sec.)	5
Initial Delay (sec.)	250
Syringe conc. (mM)	0.2
Cell conc. (mM)	0.02
Stirring speed (RPM)	1000
<b>Injection parameters</b>	
Injection volume (µl)	2
Duration (sec.)	40
Spacing (sec.)	210
Pipette volume (µl)	~38
Filter period (sec.)	5

for protein preparation. In a typical ITC experiment, BRCA1 BRCT protein was kept in the sample cell, and peptide solution was injected as 2 µl aliquots under constant stirring of 1000 rpm. The ITC experiment consisted of a total 16 injections with a 210 seconds gap between two successive injections. The detailed experimental and injection parameters are reported in **Table-4.1**. The base line was corrected by subtracting the heat of dilution from sample heat isotherm. The ITC experimental data were analysed using Origin software (Version 7.2) to calculate the affinity and stoichiometry of the binding.

#### 4.2.3. Crystallization

The FPLC purified BRCA1 BRCT protein was concentrated up to 25mg/ml as described in chapter 3. The concentrated protein was mixed with appropriate volume of the peptide, at a concentration of 10 mg/ml, to obtain 1:1.5 molar ratio, and the mixture was incubated

overnight at 4°C. The crystallization experiment was carried out at 295 K. Initial crystallization experiments were performed using Hampton crystallization screening kits (Hampton Research, Inc). We attempted crystallization by both sitting-drop and hanging-drop methods. The 4 µl crystallization drop consisted of 2 µl protein-peptide complex solutions and 2 µl of precipitant, and was equilibrated through vapour phase against 0.5 ml of precipitant solution. Single crystals of both the complexes were obtained in “0.2M Ammonium Sulphate, 0.1M MES monohydrate pH6.5, and 30% Poly Ethylene Glycol mono methyl ether” condition. The crystals were confirmed as protein crystals by staining with IZIT dye (Hampton Research). Attempts were made to improve the diffraction quality of the crystals by systematically varying the pH and concentration of precipitant and salt during crystallisation trials.

#### **4.2.4. Crystal screening and data processing**

The BRCA1 BRCT-Nup153 and BRCA1 BRCT-RBP-12 complex crystals were screened for diffraction analysis using the rotating anode x-ray generator available at ACTREC. The radiation damage to the protein crystals are minimized by maintaining the crystals at 100 K during exposure to X-rays. Ice formation during rapid freezing is avoided by using 30 % glycerol as the cryoprotectant solution. However, crystals of BRCA1 BRCT-Nup 153 complex, and BRCA1 BRCT-RBP-12 complex diffracted poorly to around 5.5 Å resolution. The post crystallization treatments such as dehydration and cryo -annealing were tried to improve the diffraction quality [252]. But these treatments could not improve the diffraction resolution.

#### **4.2.5. Molecular dynamics simulation**

The molecular models of RBP-12 peptide and Nup-153 peptide were generated from the coordinates of BRCA1 BRCT-CtIP complex structure (PDB ID-1Y98) [102], using build

mutant protocol of Discovery Studio 2.5 [253] software. These modelled structures were energy minimised using CHARMM force field [254]. The minimum-energy structures were subjected to Molecular Dynamics Simulations (MDS) for 2ns, using the GROMACS 4.5.1 software package [255]. Atomic coordinates were saved after every 1.2 picoseconds, and were analyzed using the analytical tools of the Desmond package [256].

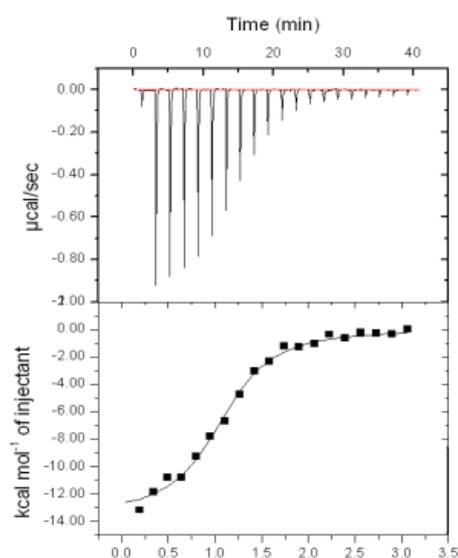
### 4.3. Results and discussion

#### 4.3.1. Interaction analysis

In order to understand functional basis of protein-protein interactions for BRCA1 BRCT, we have performed binding with FPLC purified BRCT and purified synthetic peptides of Nup153 and RNA binding domain

##### 4.3.1.1. Nup153 peptide interacts with BRCA1 BRCT domain

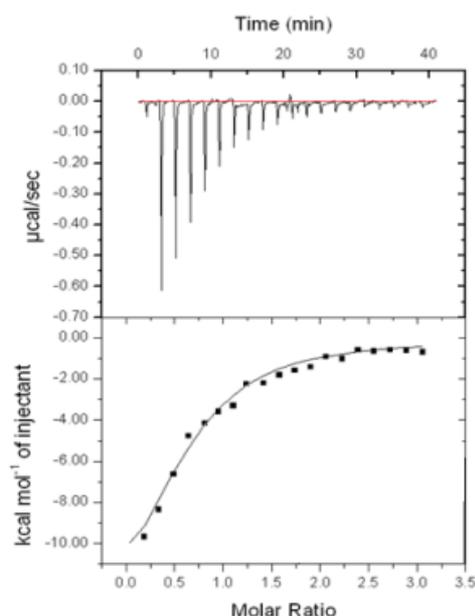
**Figure-4.3** shows raw heat change observed when Nup 153 peptide was titrated with protein solution as a function of time. The isothermal titration analysis shows that BRCT domain of BRCA1 interacts with Nup153 peptide.



**Figure-4.3: Representation of ITC result obtained from interaction of BRCA1 BRCT domain with Nup153 peptide**

#### 4.3.1.2. RBP-12 interacts with BRCA1 BRCT domain

**Figure-4.4** shows raw heat change observed when RBP-12 peptide solution was titrated with the protein solution as a function of time. The isothermal titration analysis shows that BRCT domain of BRCA1 interacts with RBP-12 peptide. The binding affinity of interaction was calculated to be 5.1  $\mu\text{M}$  and the stoichiometry of the interaction was found to be 1:1. RBP-12 peptide is binding to BRCA1 BRCT with lower affinity compared to NUP153 and the earlier reported peptides: BACH1 [18], CtIP [13] and ACC1 [19].

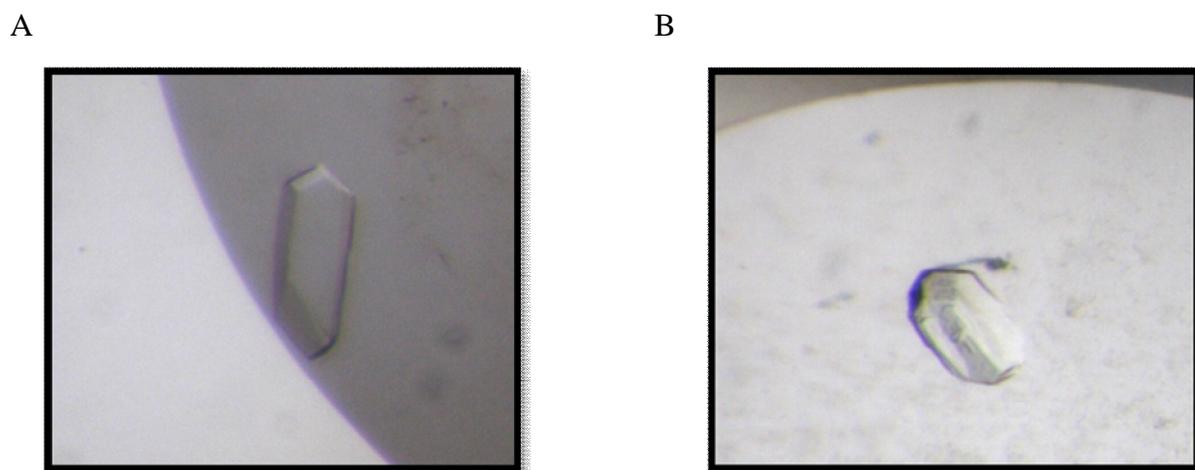


**Figure-4.4: Representation of ITC result obtained from interaction of BRCA1 BRCT domain with RBP-12 peptide**

#### 4.3.2. Crystallization of BRCA1 BRCT domain complexed with Nup153 and with RBP-12

**Figure-4.5** shows single crystals of BRCA1 BRCT domain complexed with Nup 153 (A) and with RBP-12 (B) peptides. Both the crystals look hexagonal with one long arm. The crystals grow to a maximum size of 0.15x 0.07x 0.07 mm<sup>3</sup> within two to three days,

However, there is a drastic change in the morphology after a week. We have tested freshly grown, two to three days old and a week old crystals for their ability to diffract x-rays. They all diffracted poorly.

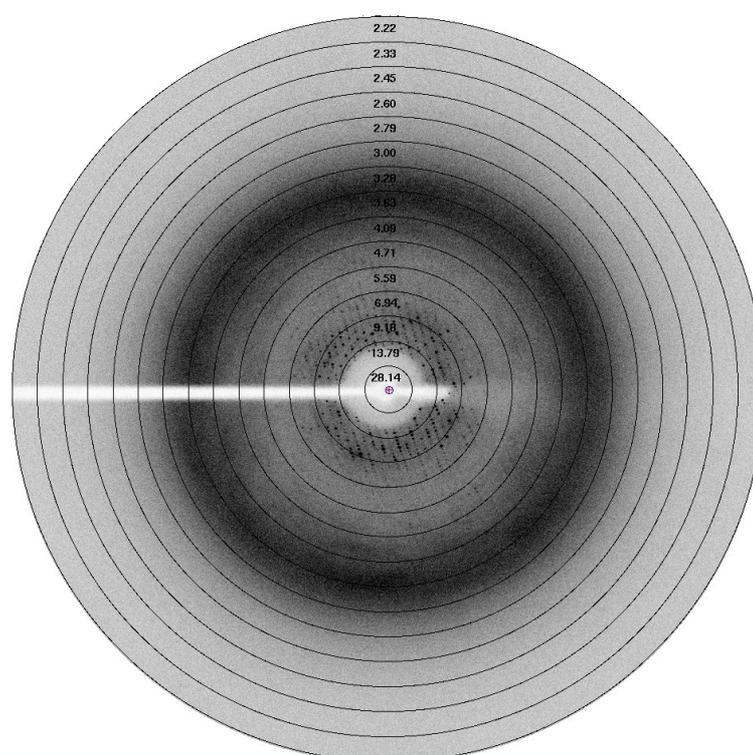


**Figure-4.5:** A) Single crystal of BRCA1 BRCT: Nup153 complex in crystallization drop of 4 $\mu$ l volume. B) Single crystal of BRCA1 BRCT: RBP-12 complex in crystallization drop of 4 $\mu$ l volume.

### 4.3.3. Data collection and data processing

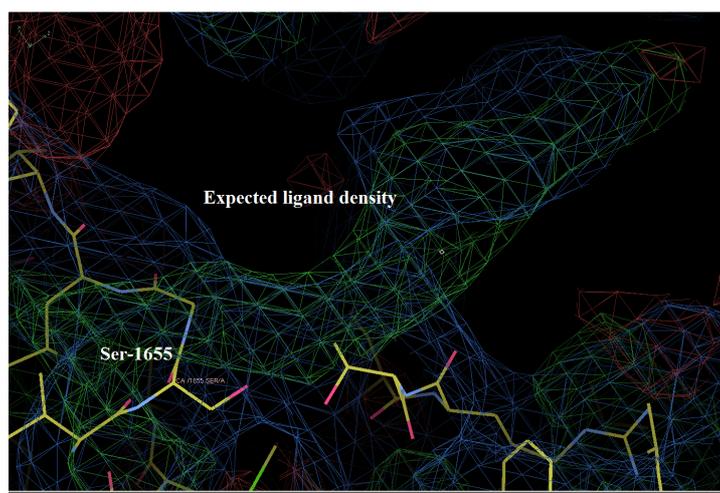
The BRCA1 BRCT-Nup153 complex crystals were screened for diffraction analysis on the rotating anode x-ray source operated at 50 kV and 100 mA (XRD facility, ACTREC). Initially, most of the good looking crystals did not diffract beyond 8 $\text{\AA}$  resolution. After some post crystallization treatment and soaking with metals, the BRCA1 BRCT-Nup153 complex crystals diffraction improved only marginally. One of these diffractions photographs collected by oscillation of crystal for 1 $^\circ$  is shown in **Figure-4.5**. The BRCA1 BRCT-Nup153 complex crystals diffracted up to around  $\sim$ 6 $\text{\AA}$  resolution. Few diffraction images of BRCT-Nup153 complex were processed in iMosflm software [189]. This processing shows that the crystals are orthorhombic, and the unit cell parameters are  $a= 85.0$ ,  $b=185.72$ ,  $c= 196.5$   $\text{\AA}$ . Molecular replacement (MR) [198] method suggested a solution characterised by a TFZ score of 5.8 and log likelihood gain (LLG) score of 1033.

The complex crystal belongs to P222 space group and initial solution from MR indicates two molecules in the asymmetric unit. Although the given structure is of low resolution the difference density corresponding to peptide was observed near Ser-1655 residue as shown in **Figure-4.7**. This indicates, the crystal indeed is a complex crystal of BRCA1 BRCT and Nup153 peptide. Further attempts are being made to improve the crystal quality to get better diffraction.



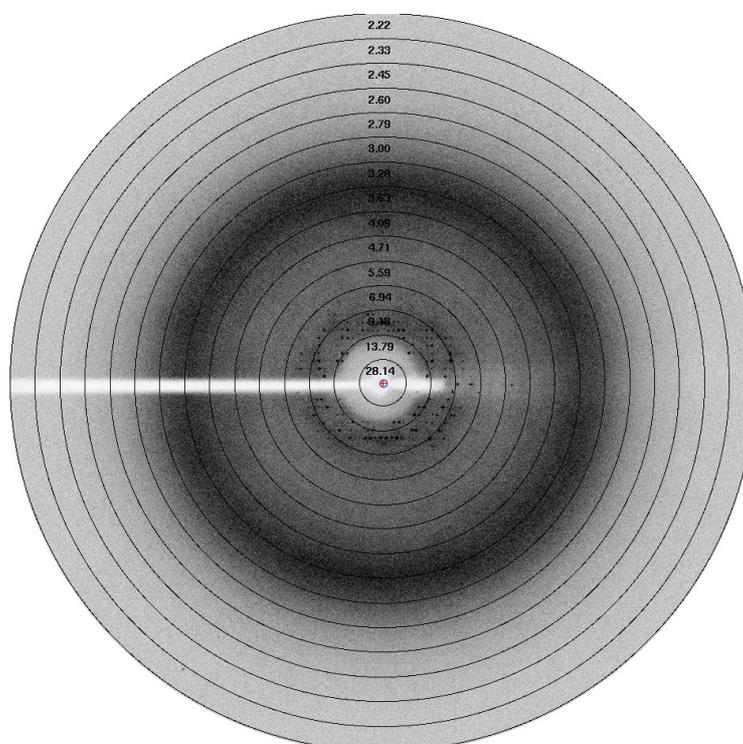
**Figure-4.6: Diffraction image of BRCA1 BRCT-Nup153 complex**

The BRCA1 BRCT-RBP-12 complex crystals also diffracted to around  $\sim 6\text{\AA}$  resolution as may be seen in **Figure-4.8**. The processing of this single image indicates the cell parameters to be:  $a=b=185.72$ ,  $c=87\text{\AA}$  and  $\alpha=\beta=\gamma=90^\circ$ . Post crystallization treatment



**Figure-4.7:** BRCA1 BRCT-Nup153 structure solved by molecular replacement method shows the difference density map ( $2\sigma$ ) which appeared at expected binding pocket of BRCA1 BRCT domain. Figure was prepared in coot [257].

such as annealing to the crystal does not yield any better diffraction. Further attempts are being made to improve the crystal quality to get better diffraction.

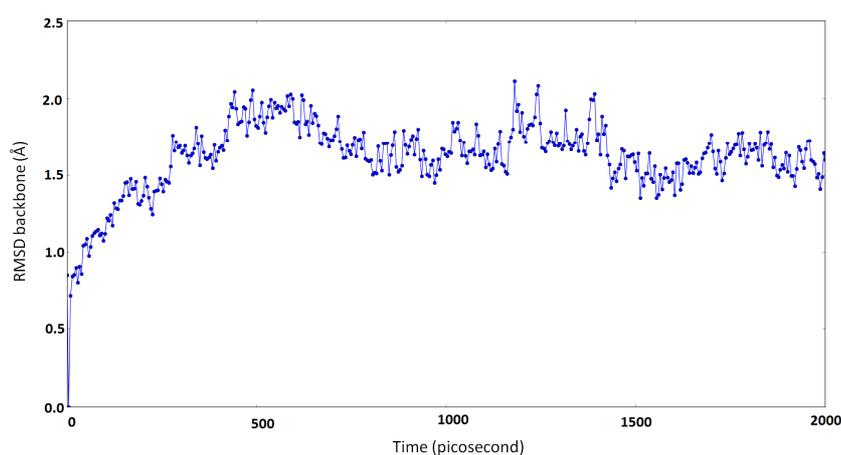


**Figure-4.8:** Diffraction image of BRCA1 BRCT-RBP-12 complex

### 4.3.4. Model building and molecular dynamics simulations

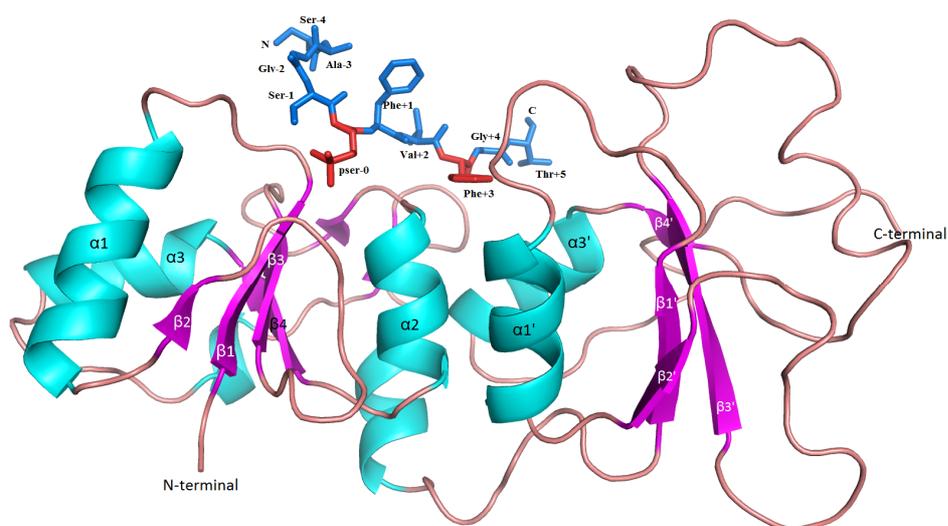
#### 4.3.4.1. BRCA1 BRCT-Nup153 complex

Noting the failure of crystallization attempts, it was decided to explore in-silico model building approach to visualise the binding mode. BRCA1 BRCT-Nup 153 complex model were subjected to MD simulations for 2ns to get an idea about the changes in the interactions between Nup-153 modelled peptide and BRCA1 BRCT domain. **Figure-4.9** shows the RMSD of the changes in positions of the backbone atoms of the model, as a function of time. There are no significant changes in the overall structure. There are only minor changes after 500 ps, and the model seems to be stabilised to within an RMSD of 1.5Å.



**Figure-4.9: RMSD of BRCA1 BRCT-Nup modelled structure over 2000 pico second.**

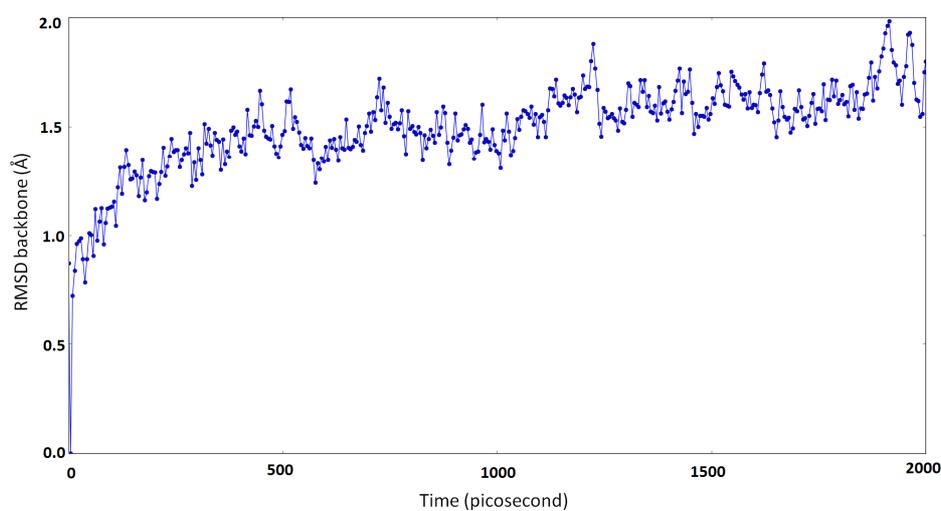
In the BRCA1 BRCT-Nup model, phosphorylated serine forms three hydrogen bonds with Ser-1655, Gly-1656 and Lys-1702. And Phe +3 is found to interact in the hydrophobic pocket formed by Phe-1704, GLu-1698 and Leu-1701. Most of these interactions are consistent with the earlier reported complex structures. But the N-terminal terminal residues upstream to pSer (0), and residues C-terminal to Phe (+3) show some variation in binding to BRCA1 BRCT.



**Figure-4.10: Ribbon representation of BRCA1 BRCT complex with Nup153 peptide**

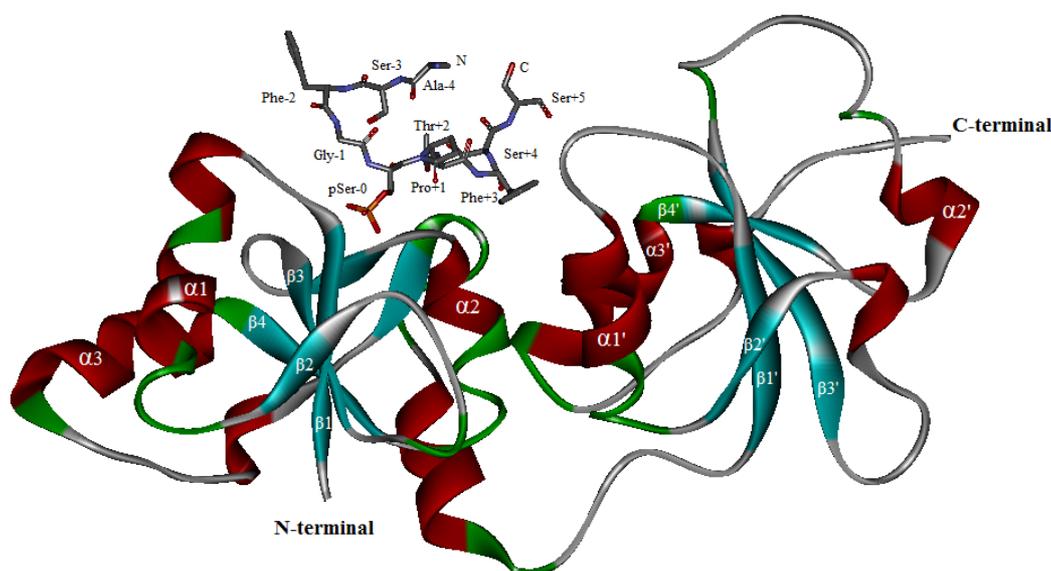
#### 4.3.4.2. BRCA1 BRCT-RBP-12 complex

Similarly, the RBP-12 peptide was modelled in BRCA1 BRCT complex structure with CtIP peptide used as a template. **Figure-4.12** shows the changes in the BRCA1 BRCT-RBP-12 complex model during 2 ns molecular dynamics simulations. Model does not show drastic changes in RMSD and model is found to be stabilised after 500ps.



**Figure-4.11: RMSD of BRCA1 BRCT-RBP modelled structure over 2 nano seconds.**

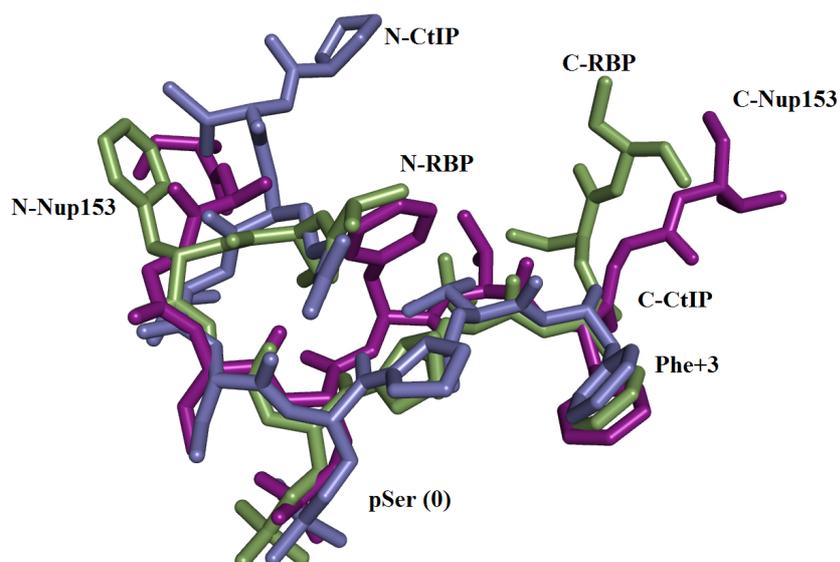
In the BRCA1 BRCT-RBP model also phosphorylated serine forms three hydrogen bonds with Ser-1655, Gly-1656 and Lys-1702. And Phe +3 is found to interact in the hydrophobic pocket formed by Phe-1704, Asn-1774, Glu-1698, Val-1741, Met-1775 and Leu-1701. Thr-1710 is observed to form hydrogen bond with Glu-1698, and this interaction is unique to this structure, and has not been reported in earlier complex structures.



**Figure-4.12: Ribbon representation of BRCA1 BRCT complexed with RBP peptide.** Figure was prepared in discovery studio [253].

#### 4.3.4.3. Structural comparison

**Figure-4.15** shows the structural superposition of Nup153, RBP and CtIP peptides. The Phe+3 of the modelled structure of Nup153 and RBP shows similar orientation but the pSer shows slightly different orientation with respect to CtIP peptide. There are also some changes at N and C-terminal ends of modelled peptides. The Nup153 peptide binds with BRCA1 BRCT with extended “S” shaped conformation as reported in the earlier complex structures, while the RBP peptide bound in compact “C” shaped conformation.



**Figure-4.13: Structural comparison of Nup153, RBP and CtIP peptides bound to BRCA1 BRCT domain.** *The Nup153, RBP and CtIP peptides are shown in red, green, and blue colours respectively.*

Both the model structures were compared with the crystallographically determined BRCT-CtIP complex (PDB ID-1Y98). The rmsd values for pSer(0) and Phe(+3) residues is found to be 1.14 Å and 1.2 Å for Nup153 and RBP respectively.

#### 4.4. Conclusion

The Nup153 and RBP-12 are potential interacting partners of BRCA1 BRCT domain, identified by motif search analysis. The Nup153 is one of the important mobile nucleoporins which helps in transport of proteins from cytoplasm to nucleus. Recently, Nup153 is found to be associated with 53BP1 protein, and takes part in DNA damage repair. Since 53BP1 contains BRCT domain, it would be interesting to study the interactions between Nup153 and BRCA1 BRCT. The RBP-12 is transcription activator and is suggested to be binding partner of BRCA1 BRCT.

The ITC experiment shows that Nup153 interacts strongly with BRCT, whereas RBP-12 is interacting weakly. Single crystals of BRCA1 BRCT complex with Nup153 and RBP-12 were obtained by co-crystallization. Both the complex crystals diffracted poorly. Data processing revealed that BRCA1 BRCT-Nup153 complex crystals belong to space group  $P2_12_12_1$  with unit cell parameters:  $a=85.0, b=185.72$  and  $c=196.5$  Å. Structure solution was attempted using molecular replacement method and it suggested presence of two molecules in the asymmetric unit. There is difference density near the expected peptide binding region of BRCT. Further progress could not be made because of low resolution diffraction. In-situ annealing and dehydration did not improve the diffraction. We have modelled both the Nup153 and RBP-12 peptides using earlier reported complex structure of BRCA1 BRCT-CtIP (PDB ID-1Y98). The lowest energy structures from both the complexes were used to measure the interactions between BRCA1 BRCT domain and respective peptides. Both the peptides bind to BRCA1 BRCT domain in standard ‘two knob manner’ and the interactions between the BRCA1 BRCT and peptide is consistent with earlier reported structure.

*Structural studies of BRCA1 BRCT-Abraxas complex*

---

## 5.1. Introduction

As already described, BRCA1 is one of the key players which performs different functions to provide genomic stability to every cell [49]. BRCA1 is required for homologous recombination (HR) DNA repair [52], which is an error free repair of double stranded breaks. BRCA1 is also required for resolving of stalled replication fork through HR [52, 258]. BRCA1 acts as a mediator protein which interacts with multiple other regulator proteins in different DNA repair and checkpoint pathways [71]. BRCA1 performs diverse functions by interacting with different proteins through its N-terminal RING domain and the C-terminal BRCT domain.

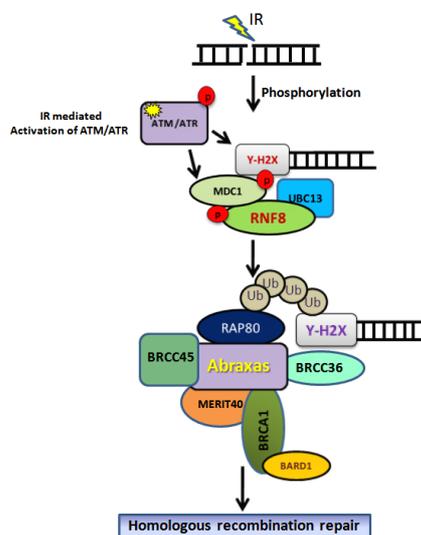
Some of the identified BRCA1 interacting partners are: Abraxas, BACH1 and CtIP proteins [73, 101, 102]. BRCA1 BRCT domain, which interacts with these three proteins through the pS-X-X-F motif, forms three different types of complexes (A, B, & C) in a mutually exclusive manner [73]. These different complexes have redundant roles in DNA repair processes and participate in different protein–protein interactions. The BRCA1-A complex is formed by interaction with Abraxas and is required for DNA double strand break repair and checkpoint activation [73, 259]. BRCA1 associates with BACH1 and forms BRCA1-B complex, which is required for replication stress induced checkpoint activation as well as for DNA inter strand crosslink repair [260, 261]. The BRCA1-C complex is formed due to binding with the CtIP protein, and is required for HR mediated DNA double strand break DNA repair [262].

In these three complexes, only an oligopeptide segment of about ten residues containing the signature motif pS-X-X-F, is involved in the direct binding to BRCA1 BRCT domain. Crystal structures of BRCA1 BRCT complexed to cognate oligopeptides from BACH1 and CtIP have been determined [101, 102]. However, the crystal structure of the A

complex is not reported so far. Recent genetic analysis of cancer patient families has identified *Abraxas* as a new breast cancer susceptibility gene, and also as a novel target for therapeutic intervention [263]. Cells depleted of *Abraxas* or *RAP80*, exhibit hypersensitivity toward the ionising radiation and show defects not only in G2-M checkpoint activation but also in HR repair [73]. The protein *Abraxas* is one of the key adaptor proteins, which not only directly binds *BRCA1*, but also recruits other proteins to the ionisation radiation induced foci. Lieu *et al* have shown that *BRCA1* directly interacts with *Abraxas* and that this binding is independent of DNA damage [259]. Wang *et al* have shown that *Abraxas* specifically bound to wild type *BRCA1* BRCT domain and not to BRCT domain mutant (M1775R). As shown in **Figure-5.1**, *BRCA1-A* complex consists of six proteins: *BRCA1*, *Abraxas*, *RAP80*, *MERIT40*, *BRCC36*, and *BRCC45*, which associate with each other directly or indirectly [264-266] in a sequential manner. Upon DNA damage, central kinase *ATM* and *ATR* get activated and phosphorylate histone *H2AX* at the damage site [267]. The activated histone *H2AX* recruits Mediator DNA damage Checkpoint1 protein (*MDC1*) to the damage site. *MDC1* is also a substrate of *ATM* and *ATR* kinases and the subsequent phosphorylation of *MDC1* results into recruitment of ubiquitin conjugase (*RNF 8*) and ligase (*UBC 13*) to the DNA damage site [268-271]. These enzymes further form Lys-63 linked poly ubiquitin chains on histone *H2AX*, which then binds at the ubiquitin interacting motifs of *RAP80* [270, 272]. This complex then recruits other repair proteins such as *BRCA1* and *Abraxas* to the damage site [73, 273, 274]. The *RAP80* protein interacts with *Abraxas* through the *Abraxas* interacting region.

The mediator role of *Abraxas* is found to be dependent upon its phosphorylation state. *Abraxas* actually contains two potential *ATM*-phosphorylation serines, Ser-404 and Ser-

406, at the C-terminus [73]. It is observed that Abraxas phosphorylated at both Ser-404 and 406 residues is selectively enriched on incidence of ionising radiation [73]. It is also



**Figure-5.1: Homologous recombination repair by BRCA1-A DNA repair complex.**

found that Abraxas phosphorylated at Ser-406 alone binds with BRCA1 BRCT, while that phosphorylated at Ser-404 alone does not bind BRCA1 BRCT. To gain insight into formation of BRCA1-A complex and also into BRCT-phosphopeptide binding, we have undertaken biophysical and crystallographic studies of the interaction between BRCA1 BRCT domain and differently phosphorylated C-terminal oligopeptides from Abraxas, and the results are described in this chapter.

## 5.2. Material and Methods

### 5.2.1. Expression and purification of BRCA1 BRCT domain

The BRCA1 BRCT domain was purified using affinity chromatography as discussed in chapter 3.

### 5.2.2. Interaction analysis using ITC

The following differently phosphorylated Abraxas synthetic peptides were commercially obtained.

**A1-** NH<sub>2</sub>-G-F-G-E-Y-S-R-pS<sub>406</sub>-P-T-F-COOH

**A2-** NH<sub>2</sub>-G-F-G-E-Y-pS<sub>404</sub>-R-pS<sub>406</sub>-P-T-F-COOH

**A3-** NH<sub>2</sub>-G-F-G-E-Y-pS<sub>404</sub>-R-S-P-T-F-COOH

All above Abraxas peptides were dissolved in autoclaved distilled water before use for interaction analysis and for co-crystallization with BRCA1 BRCT domain.

**Table-5.1:** ITC experimental and injection parameters.

<b>Experimental parameters</b>	
Total injections	13/19
Cell Temperature (°C)	25
Ref power (μcal/sec.)	5
Initial Delay (sec.)	250
Syringe conc. (mM)	0.2
Cell conc. (mM)	0.02
Stirring speed (RPM)	1000
<b>Injection parameters</b>	
Injection volume (μl)	2
Duration (sec.)	40
Spacing (sec.)	210
Pipette volume (μl)	~38
Filter period (sec.)	5

The interaction analysis was carried out at 25°C by using ITC-200 (GE, Healthcare). The concentration of BRCA1 BRCT protein was 0.02 mM, while the concentration of the three Abraxas peptides was 0.2 mM. In a typical ITC experiment, BRCA1 BRCT protein was kept in the sample cell, and Abraxas peptide solution was injected as 2 μl aliquots

under constant stirring at 1000 rpm. The ITC experiment consisted of a total of 19 injections with a time gap of 210 seconds between successive injections. The detailed experimental and injection parameters are recorded in **Table-5.1**. The base line correction was applied by subtracting the heat of dilution from the sample heat. The ITC experimental data were analysed using Origin software (Version 7.2) to calculate the affinity and the stoichiometry of the binding.

### **5.2.3. Protein structure determination**

#### **5.2.3.1. Crystallization**

The purified BRCA1 BRCT protein was concentrated to 25mg/ml as described in chapter 2. The concentrated protein was mixed in 1:1.5 molar ratio with A1 and A2 peptides at a concentration of 10 mg/ml, and the mixture was incubated overnight at 4°C. The crystallization experiments carried out at 22°C were performed using Hampton crystallization kits (Hampton Research, Inc). The crystallization attempts used both sitting drop and hanging drop methods, where a 4 µl drop consisting of 2 µl complex solution and 2 µl precipitant was concentrated through vapour diffusion against 0.5 ml of precipitant solution. This crystallization condition was further optimised by systematically varying the concentration of precipitant and salt solutions.

#### **5.2.3.2. Diffraction data collection and processing**

Crystals of both the complexes were screened for diffraction using the rotating anode x-ray source as well as the synchrotron source of x-rays. The radiation damage to the protein crystals were minimized by maintaining the crystals at 100K during exposure, and by using 30 % glycerol as the cryoprotectant solution. BRCA1 BRCT-A1 complex crystals diffracted to around 5Å resolution at the home source however, at the synchrotron, the this complex crystals diffracted up to around 3.8 Å resolution.

Diffraction data were collected as a number of contiguous frames, each frame being for a crystal oscillation angle of 1°. Similarly, the BRCA1 BRCT-A2 complex crystals diffracted up to around 4 Å resolution at home source and complete data set were collected by the oscillation method [188] with an oscillation angle of 1°. The diffraction data were processed using iMOSFLM software [240] and were scaled by SCALA program in CCP4 software suite [241].

### 5.2.3.3. Structure determination and refinement

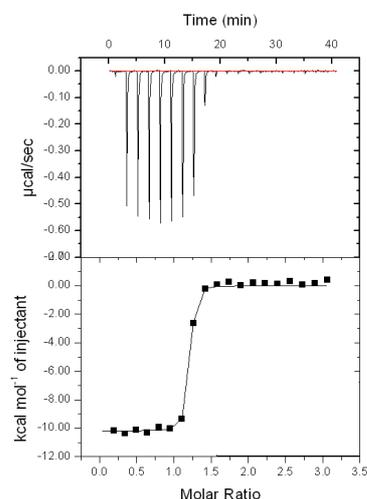
Both the complex structures were solved by the Molecular Replacement method [198], as implemented in the PHASER program from CCP4 software suite [241]. The molecular model with PDB ID-1T15 [101] was used as the search model to perform the rotational and translational search. The ligand and water molecules were removed from the search model prior to Molecular Replacement. The crystallographic refinement for both the complex structures was done by using REFMAC5 [218] software from CCP4 suite, and the PHENIX - REFINE software [243]. TLS refinement and jelly body refinement along with NCS restraints were used for the A1 complex crystal refinement [218, 275]. The building of the atomic model into the electron density was carried out using COOT [257].

## 5.3. Result and Discussion

### 5.3.1. Isothermal Titration Calorimetry (ITC)

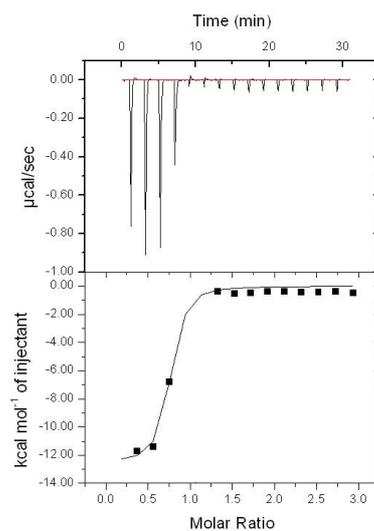
**Figure-5.2** shows the enthalpic change on interaction of BRCA1 BRCT domain with the A1 peptide. The titration profile shows that BRCT domain of BRCA1 interacts with the A1 peptide, and that the interaction is exothermic in nature. The binding affinity of interaction was calculated to be 1.2 µM and the stoichiometry of the interaction was found to be 1:1. The binding affinity of A1 peptide with BRCA1 BRCT is stronger than those observed for oligopeptides CtIP (3.7 µM) [102] and ACC1 (3.4 µM) [113], but is

comparable with that of BACH1 (0.9  $\mu\text{M}$ ) [101]. These differences may be reflecting the influence of amino acid residues other than the phosphoserine and phenyl alanine residue at the positions 0 and +3 respectively.



**Figure 5.2: Representation of ITC results obtained from interaction between BRCA1 BRCT domain and A1 peptide.**

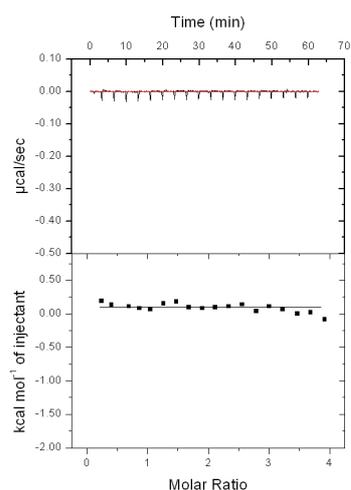
Similarly, **Figure-5.3** shows enthalpic change observed when A2 peptide was titrated with BRCA1 BRCT domain. The exothermic titration profile shows that BRCT domain of BRCA1 interacts also with the A2 peptide.



**Figure-5.3: Representation of ITC results obtained from interaction between BRCA1 BRCT domain and A2 peptide.**

The binding affinity of interaction was calculated to be 0.2  $\mu\text{M}$ , and the stoichiometry of the interaction was found to be 1:1. The A2 peptide not only has stronger binding affinity compared to A1 peptide (Ser-406), but also compared to those observed for other oligopeptides: BACH1 (0.9  $\mu\text{M}$ ) [101], CtIP (3.7  $\mu\text{M}$ ) [102] and ACC1 (3.4  $\mu\text{M}$ ) [113]. The stronger binding affinity of A2 peptide may be due to presence of two phosphorylated Serine residues (Ser-404 and Ser-406).

**Figure-5.4** shows the enthalpic change observed when A3 peptide was mixed with BRCA1 BRCT domain. There is no heat change observed in the isothermal titration profile which indicates that the A3 peptide does not interact with BRCA1 BRCT domain.



**Figure-5.4:** The ITC results from interaction between BRCA1 BRCT with A3 Abraxas peptide.

Though it is tempting to suggest that the A2 peptide has 6 times higher binding affinity because of additional phosphorylated serine (Ser -404) compared to A1 peptide, it is intriguing that A3 peptide which is singly phosphorylated at Ser-404 does not bind at all to BRCA1 BRCT domain. To investigate this phenomenon we have co-crystallized A1 and A2 peptides with BRCA1 BRCT domain.

### 5.3.2. Structure determination

#### 5.3.2.1. Crystallization and structure determination

Single crystals of BRCA1 BRCT complexed with A1 and A2 oligopeptides were obtained under following conditions.

**Table-5.2:** Initial crystallization condition for BRCA1:A1 and BRCA1:A2 complex

<b>BRCA1: A1 complex</b>	<b>BRCA1: A2 complex</b>
0.01 M Cobalt Chloride (Co-Cl <sub>2</sub> ) 0.1 M(MES), pH-6.5 30% PEG Mono Methyl Ether (MME) 5000	0.2M Ammonium Acetate (A/A) 0.1M Sodium Acetate (S/A), pH-4.6 30% Poly Ethylene Glycol (PEG) Mono Methyl Ether (MME) 4000
0.2M Calcium Acetate (C/A) 0.1 M sodium Cacodylate, pH-6.5 (SC) 30% PEG 8000	0.01 M Ammonium sulphate 0.1 M(MES), pH-6.5 30% PEG Mono Methyl Ether

The crystals were confirmed as protein crystals by staining with IZIT dye from Hampton Research. A1 and A2 complex crystals belong to tetragonal and hexagonal system respectively. The space groups were established as P4<sub>1</sub>2<sub>1</sub>2 and P6<sub>1</sub>22 respectively, based on intensity equivalents, systematic absences and translation function calculations. MR suggested a single solution for A1 complex characterised by a TFZ score of 19.8, LLG score of 684.9, and also indicated the presence of three molecules in the asymmetric unit. The three molecules in the asymmetric unit indicate a solvent content of 75.68%. Crystals containing large amount of solvent are generally not well ordered, and the low diffraction resolution observed here might indeed be due to the large solvent content. MR calculations also yielded a single solution for the A2 complex characterised by a TFZ score of 17.5 and LLG score of 240.9, and indicated the presence of a single molecule in the asymmetric unit. Subsequently, both the structures were refined using REFMAC-5 [218], and the refinement statistics are given in **Table-5.3**.

**Table-5.3:** Data collection and refinement statistics

<b>Complex</b>	<b>A1 complex</b>	<b>A2 complex</b>
No. of crystals used	1	1
X-ray source	BM-14, ESRF	Home source (ACTREC)
Wavelength (Å)	0.9000	1.54179
Crystal to detector distance (mm)	156.04	250
Space group	P4 <sub>1</sub> 2 <sub>1</sub> 2/ P4 <sub>3</sub> 2 <sub>1</sub> 2	P6 <sub>1</sub> 22
Unit cell parameters (Å)	a=b= 187.18,c=85.31	a=b= 113.6,c=121.61 $\alpha=\beta= 90^\circ, \gamma=120^\circ$
V <sub>m</sub> value	5.06	2.3
Mosaicity (°)	1.12	0.8
Resolution limit (Å)	28.47 – 3.8 (4.01-3.8) <sup>a</sup>	28.4-3.8 (4.01-3.8)
Total No. Of reflections	106886	562011
Unique reflections	15491	6221
Redundancy	6.9	21.1
I/ $\sigma$ I	9.5 (2.0)	12.3 (4.4)
Completeness (%)	100 (100)	100 (100)
R <sub>merge</sub> (%) <sup>b</sup>	9.1 (87.2)	37.0 (83)
Wilson B factor	135.7	20.4
<b>Refinement statistics</b>		
R <sub>work</sub> (%) <sup>c</sup>	31.47	32.00
R <sub>free</sub> (%) <sup>d</sup>	35.79	35.7
Total no. of residues	613	213
Total no. of water molecules	-	-
RMSD bond length (Å)	0.015	0.0115
RMSD bond angle (°)	1.171	0.9890
Overall B factor	123.0	42.9
Ligand B factor	131.0	85.29
<b>Ramachandran plot analysis</b>		
Most favoured (%)	90.28	88.00
Additionally allowed (%)	7.72	11.5
Disallowed region (%)	2.0	0.5

<sup>a</sup> Values in parentheses are for the highest-resolution shell (4.01-3.8Å).

<sup>b</sup>  $R_{merge} = \frac{\sum |I - \langle I \rangle|}{\sum (I)}$  where  $I$  is the observed integrated intensity,  $\langle I \rangle$  is the average integrated intensity obtained from multiple measurements, and the summation is over all observed reflections.

<sup>c</sup>  $R_{work} = \frac{\sum_{hkl} \|F_{obs}(hkl) - K F_{calc}(hkl)\|}{\sum_{hkl} |F_{obs}(hkl)|}$  Where  $F_{obs}$  and  $F_{calc}$  are observed and calculated structure factors respectively.

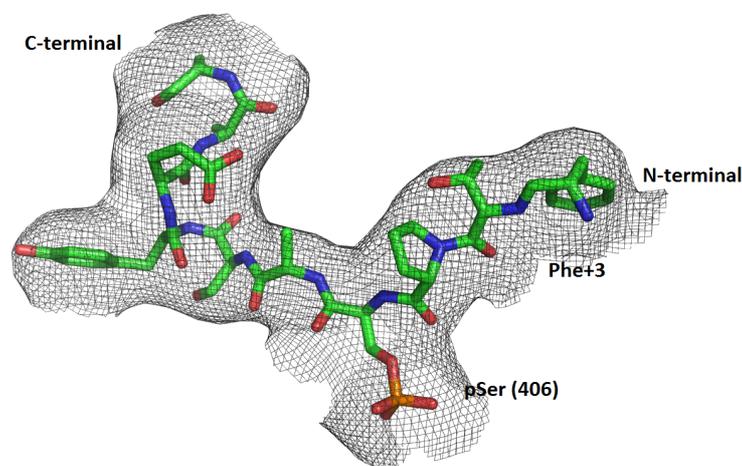
<sup>d</sup>  $R_{free}$  was calculated as for  $R_{work}$  but only 5% data left out of refinement procedure has been used in the calculations

### 5.3.2. BRCA1 BRCT-Abraxas peptide structures

We have determined crystal structures of BRCA1 BRCT in complex with Abraxas, A1 and A2 peptides.

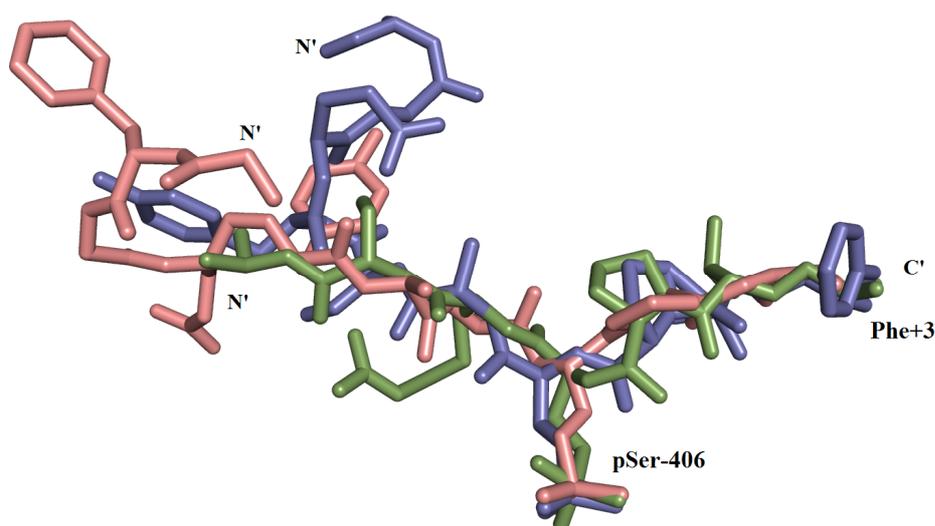
### 5.3.3. Crystals structure of BRCA1 BRCT-A1 complex:

In the A1 complex, the asymmetric unit comprises of three molecules A, B and C. Of these three molecules, molecules A and B are better defined with electron density compared to molecule C. The conformation of all the three molecules was acceptable. Each BRCT repeat comprised of three  $\alpha$  helices and one  $\beta$  sheet with a topology  $\beta 1-\alpha 1-\beta 2-\beta 3-\alpha 2-\beta 4-\alpha 3$ . The RMSD values for A-B, B-C and A-C superpositions are: 0.1 Å, 0.1 Å and 0.08 Å respectively. Positive 2Fo-Fc electron density is observed near Ser-1655 residue of each molecule



**Figure-5.5: The atomic model and 2Fo-Fc electron density map for the A1 peptide at 3.8 Å resolution.** The electron density map is contoured at  $1\sigma$  level.

Therefore, the present structure is that of a molecular complex between BRCA1 BRCT and the A1 peptide. A1 peptide was built into the electron density in all the three subunits. Eight out of eleven residues of the A1 peptide could be built in Subunit A, while in subunit B and C there is clear density for ten and eight residues respectively (**Figure-5.5**). All the three peptides superpose well indicating that the A1 peptide binds with BRCA1 BRCT in similar conformation (**Figure-5.6**). Because of lower resolution water molecules have not been identified.

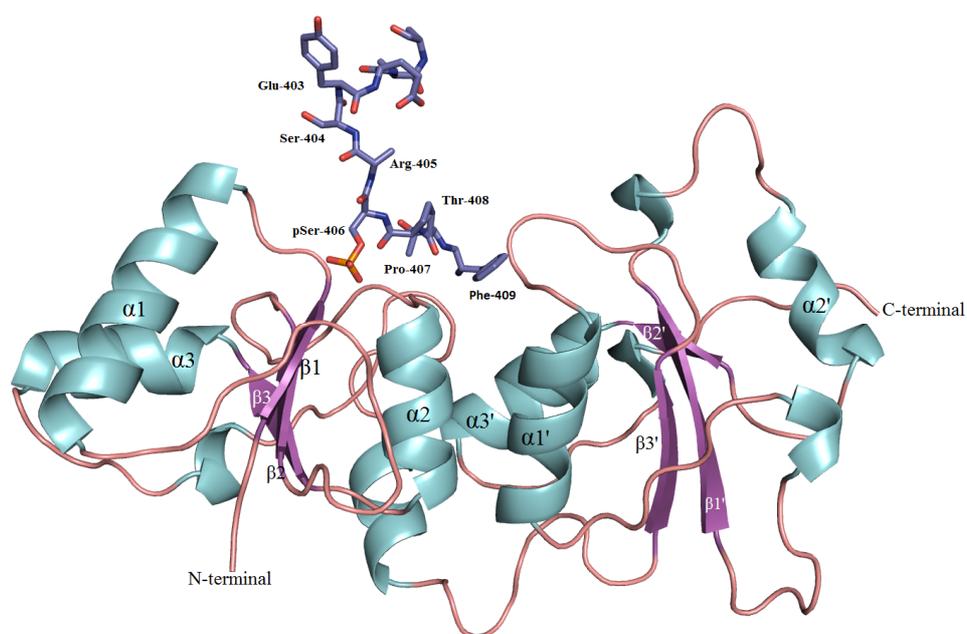


**Figure-5.6: Superposition of three BRCA1 BRCT-A1 complex peptides, where A1 peptide from A, B and C subunit shown in green, blue and brown colour respectively.**

Structural superposition with unliganded protein [10] yields RMSD of  $0.8\text{\AA}$  over 213 C $\alpha$  atom pairs showing that the protein conformation is not altered much upon ligand binding. This is in agreement with earlier reported complex structures.

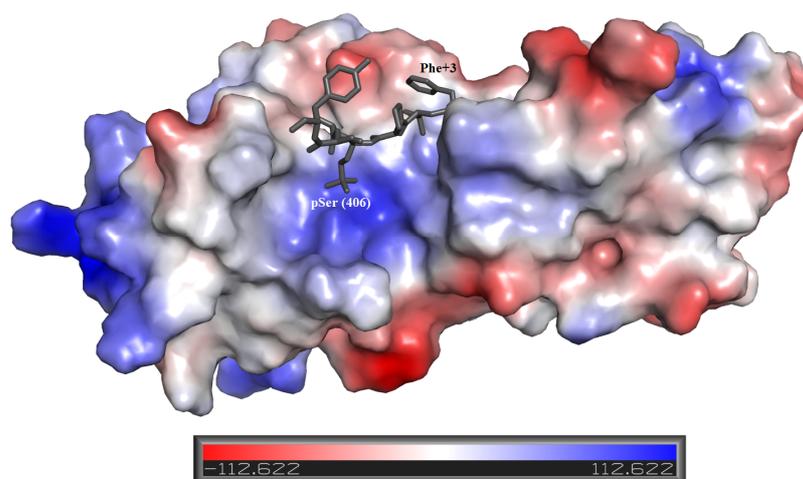
#### 5.3.3.3.1. Interaction analysis

Crystal structure of A1 complex shows that the A1 peptide binds to BRCA1 BRCT domain in “two knob manner” (**Figure-5.7**).



**Figure-5.7: A1 peptide recognition by BRCA1 BRCT domain.** Ribbon representation of BRCA1 BRCT bound to A1 peptide (blue colour stick model). The  $\alpha$ -helices and  $\beta$ -sheets coloured cyan and pink respectively. The critical residues of A1 peptide denoted as pSer-406 and Phe-409. The figure was prepared using PyMOL [245].

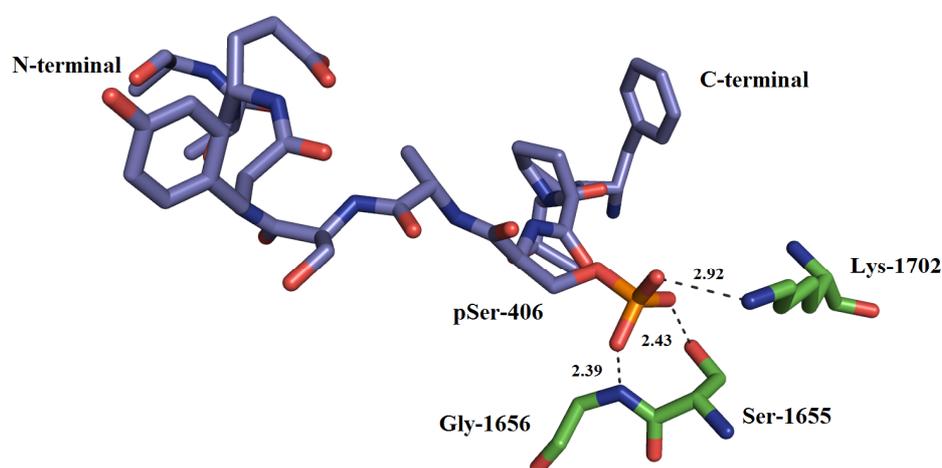
The pSer-406 interacts with the N-terminal BRCT and Phe-409 interacts at the interface region between the two tandem BRCT domains. The pSer is negatively charged, and interacts with positively charged region in N-terminal BRCT while the Phe+3 is found to interact with the hydrophobic area from the interface region between the two BRCT domains (**Figure-5.8**).



**Figure-5.8: Electrostatic surface representation of BRCA1 BRCT-A1 complex structure.** The surface is colored by the electrostatic potential (red-negative, blue-positive).

### A) pSer-406 residue

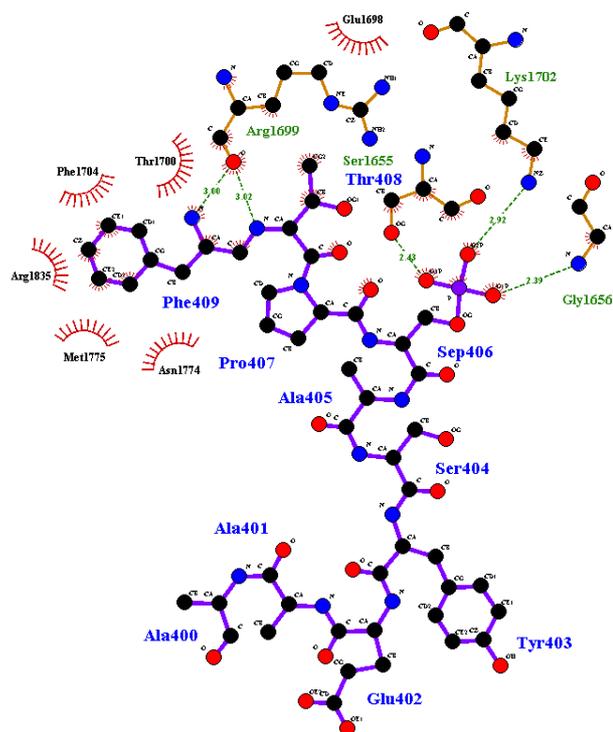
**Figure-5.9** shows the pSer-406 residue of A1 peptide forms three hydrogen bonds with Ser-1655, Gly-1656 and Lys-1702 residues. These hydrogen bonding interactions are conserved and are observed also in earlier reported structures. It has been observed that the residues responsible for interaction with phosphopeptide are found to be conserved in BRCT domain, indicating that the phosphopeptide binding is an evolutionarily conserved feature.



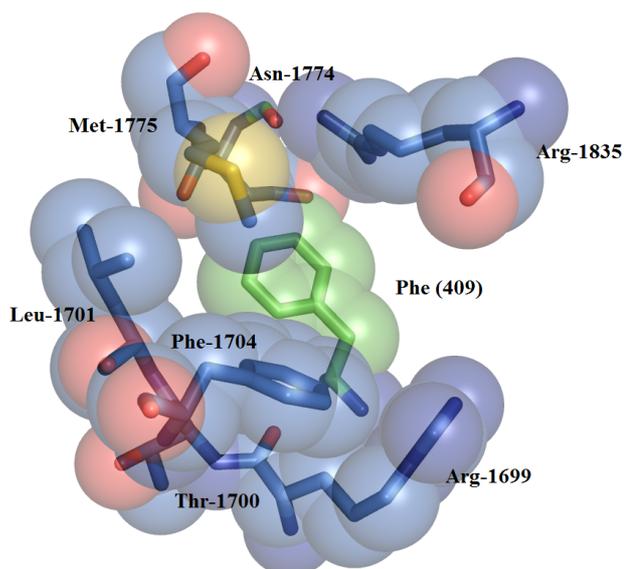
**Figure-5.9:** The phosphorylated serine present at 406 position in A1 peptide form three hydrogen bonds BRCA1 BRCT domain.

### B) Phe-409 binding

The amide nitrogens of Thr-408 and Phe-409 form two hydrogen bonds with carbonyl oxygen of Arg-1699 (**Figure-5.10**). In earlier reported structures of complexes with other oligopeptides, the side chain of Arg-1699 forms one extra hydrogen bond with carbonyl oxygen of Phe-409. The Phe-409 forms hydrophobic interactions in the hydrophobic pocket formed by Arg-1699, Thr-1700, Leu-1701, Phe-1704 from N-terminal BRCT and Asn-1774, Met-1775 and Arg-1835 from C-terminal BRCT (**Figure 5.11**).



**Figure-5.10:** Schematic representation of weak intermolecular interaction between AbraxasA1 peptide (purple bonds) and BRCA1 BRCT domain (orange bonds), the green dotted line indicates the hydrogen bond while the wheel spoke indicates the hydrophobic interaction.(Figure is prepared by LigPlus [246])

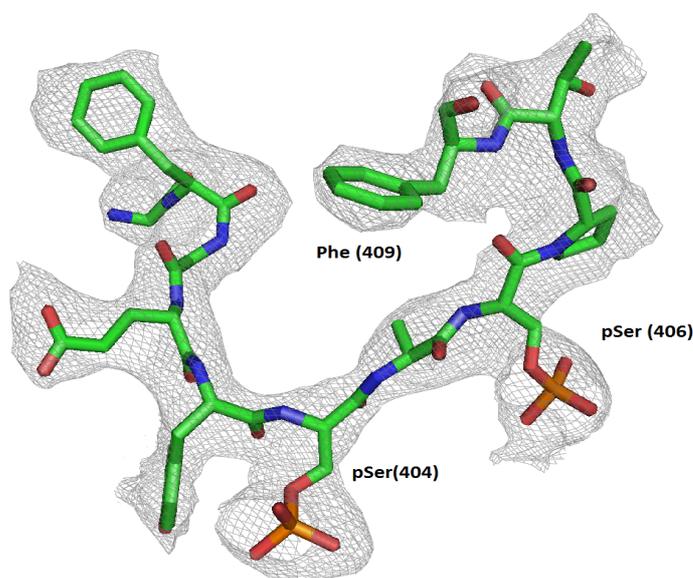


**Figure-5.11:** Space filling model representing environment around Phe (409) of A1 peptide

### 5.3.4. Crystal structure of BRCA1 BRCT-A2 complex

The complex structure is solved by Molecular Replacement method, and refined using REFMAC-5 program [218] (**Table-5.3**). The electron density of A2 peptide was clearly visible after few cycles of refinement. The complete A2 peptide modelled in the 2Fo-Fc electron density map countered at  $1\sigma$  is given below.

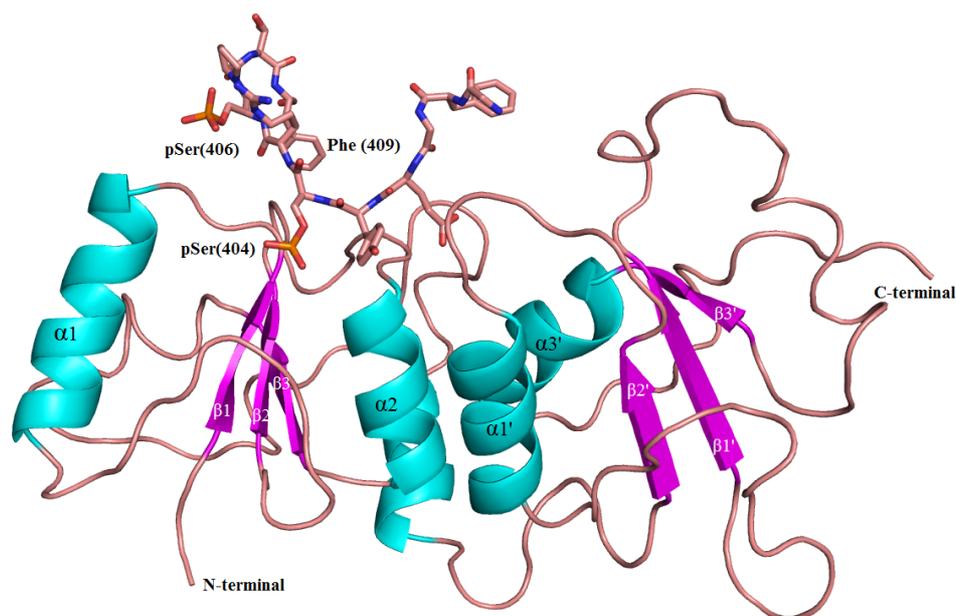
All the eleven residues of the A2 peptide could be modelled in the electron density map, except the side chain for Arg-405. The complex structure has cylindrical shape and consists of N-terminal BRCT and C-terminal BRCT repeat arranged in a head to tail manner. Each BRCT repeat comprised of three  $\alpha$  helices and four  $\beta$  strands. The overall conformation of the protein is similar to unliganded BRCA1 BRCT domain (r.m.s. deviation  $\sim 0.6$  for all  $C\alpha$  atoms).



**Figure-5.12:** Electron density map (2fo-fc) covering the A2 Abraxas peptide.

#### 5.3.4.3.1. Interaction analysis

The A2 Abraxas peptide contains two phosphate groups (pSer-404 and pSer-406) and their location in the complex is shown in **Figure-5.13**.

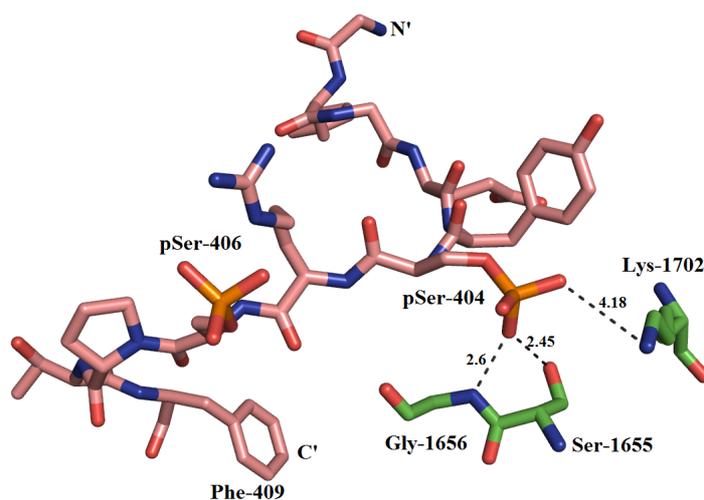


**Figure-5.13: A2 peptide recognition by BRCA1 BRCT domain.** Ribbon representation of BRCA1 BRCT bound to A2 peptide (Brown colour stick model). The  $\alpha$ -helices and  $\beta$ -sheets coloured cyan and pink respectively. The critical residues of A2 peptide denoted as pSer-404. The figure was prepared using PyMOL [245].

These phosphorylated serines are the main interacting centers in the A2 peptide complex. It is very interesting that all the interactions between the A2 peptide and the protein are with the N-terminal BRCT domain. This is a very novel mode of phosphopeptide binding by BRCA1 BRCT domain, and has not been reported so far.

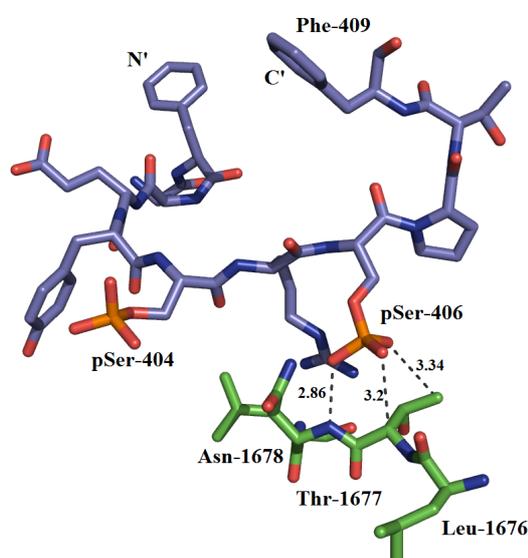
#### A) pSer-404 and pSer-406 binding

BRCA1 BRCT domain possesses a phosphate-binding pocket in the N-terminal BRCT domain. In the present structure, pSer-404 rather than the expected pSer-406 binds in this pocket and forms two hydrogen bonds with Ser-1655 and Gly-1656 (**Figure-5.14**). It has been observed from earlier reported structures, that pSer of phosphopeptide forms three hydrogen bonds. But here the pSer is oriented slightly differently and does not form the third hydrogen bond with Lys-1702.



**Figure-5.14:** Phosphate group of Serine residue present at 404 position forms two hydrogen bonds with BRCA1 BRCT domain.

The phosphate group from Ser-406 does not interact with the BRCT bound by Ser – 404. But it forms three hydrogen bonds with Leu-1676, Thr-1677 and Asn-1678 residues from a single neighboring symmetry related molecule in the crystal (**Figure-5.15**).

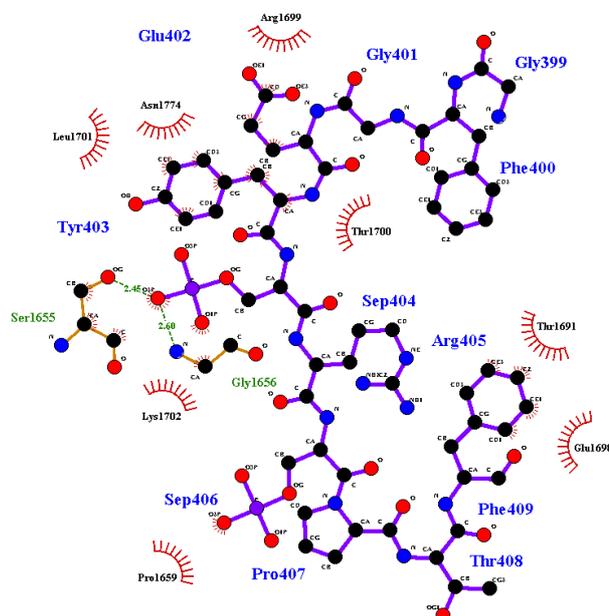


**Figure-5.15:** The phosphoserine at 406 positions is stabilized by three hydrogen bonds with the symmetry related molecule Thr-1677.

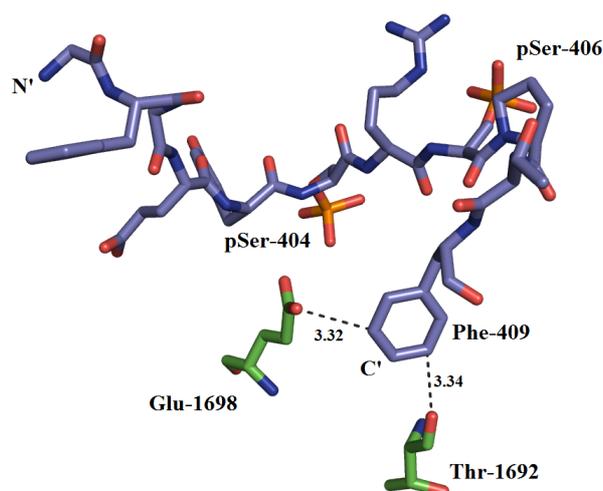
### B) Phe+3 binding

In the present structure, Phe-409 is not surrounded by hydrophobic residues. Instead it is stabilized by electrostatic interactions with Glu-1698 and Thr-1692 residues from the N-

terminal BRCT (**Figure-5.17**). This is the unique way of phosphopeptide binding with the BRCT domain.



**Figure-5.16:** Schematic representation of weak intermolecular interaction between Abraxas A2 peptide (purple bonds) and BRCA1 BRCT domain (orange bonds), the green dotted line indicates the hydrogen bond while the wheel spoke indicates the hydrophobic interaction.(Figure is prepared by LigPlus [246])

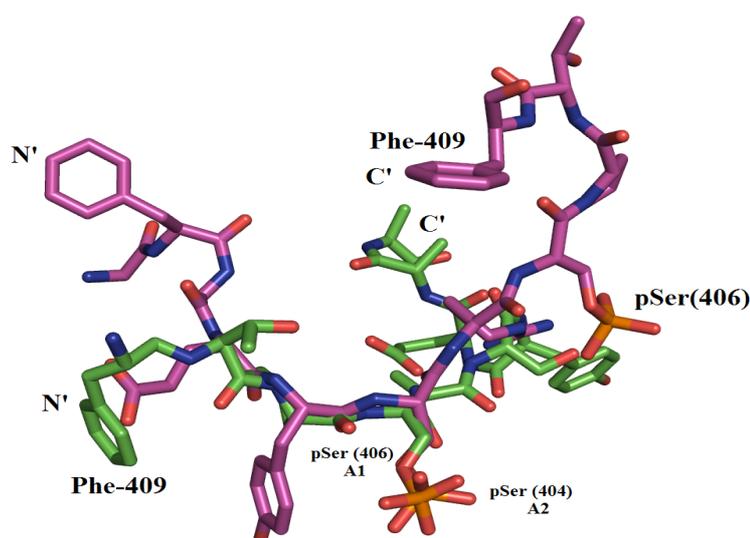


**Figure-5.17:** The phenyl ring is stabilized by hydrogen bonds from Glu-1698 and Thr-1692.

In all the earlier reported complex structures [101, 102, 113], Phe+3 residue forms hydrophobic interactions with residues present in the interface region between the two BRCT repeats.

#### 5.3.4.4. Comparison of A1 and A2 complex structures

**Figure 5.19** shows the superposition of A1 and A2 complex structures. The two structures superpose well with an rmsd of 0.93 Å over 213 Ca atom pairs, indicating that the protein conformation is not altered much. A1 and A2 peptides acquire totally different conformations compared to each other (as shown **Table-5.4**). In A2 peptide, pSer-404 superposes well on pSer-406, and both bind in the phosphate-binding pocket of BRCA1 BRCT domain. However, the Phe-409 residue in the two peptides does not superpose, and also bind to distinct sites on the BRCT domains. The Phe-409 of A1 peptide binds to the hydrophobic pocket region at the interface between two BRCT repeats, which is consistent with the earlier complex structure reports [101, 102, 113]. But the Phe-409 of A2 peptide binds to the Thr-1692 and Glu-1698 residues from N-terminal BRCT.



**Figure-5.18:** Superposition of stick model of A1 peptide (green) with A2 peptide (magenta).

**Table-5.4** The phi and psi angle for the residues of A1 and A2 peptide.

A.A. No.	Residues name	A1		A2	
		phi	psi	Phi	Psi
400	Ala	40.11	118.96	-98.07	-29.83
401	Ala	178.82	-69.88	-93.39	-111.0
402	Glu	-62.62	-106.6	-162.99	158.41
403	Tyr	41.29	87.28	67.56	96.27
404	pSer/Ser	-174.71	-96.5	95.96	91.76
405	Ala	-157.94	-112.73	-161.09	-160.8
406	pSer	159.98	137	-133.77	-124.36
407	Pro	-131.86	-85.41	-38.08	-73.58
408	Thr	-166.95	-177.95	-93.51	51.57
409	Phe	117.10	128.16	-65.01	164.77

## 5.4. Conclusion

Abraxas is a component of BRCA1 A complex and the amino acid sequence of Abraxas involved in the binding to BRCA1 BRCT has following special features: 1) the pS-X-X-F signature motif recognised by BRCT domain forms the carboxy terminus, and 2) in addition to the signature motif serine residue, there is one more serine residue present upstream to the signature motif. Both the serine residues are found to be phosphorylated in-vivo upon incidence of ionizing radiation. In order to understand the significance of phosphorylation we have undertaken the structural studies on mono (A1) and di phosphorylated (A2) Abraxas peptides complexed with BRCA1 BRCT.

The ITC experiments show that Abraxas A2 peptide binds strongly (0.2  $\mu$ M affinity) to BRCA1 BRCT domain compared to Abraxas A1 peptide (1.2  $\mu$ M affinity). The stoichiometry of binding to BRCT domain of both the peptides was found to be 1:1. Single crystals of BRCT: A1 and BRCT: A2 complexes were obtained by co-crystallization. Crystals of the two complexes diffracted to about 3.8 Å resolution. Data

processing revealed that the crystals of A1 complex belong to the space group  $P4_12_12/P4_32_12$ , while crystals of A2 complex belong to the space group  $P6_122$ . While there are three molecules in the asymmetric unit of A1 complex, there is a single molecule in the asymmetric unit of A2 complex. Both crystal structures have been solved by MR methods and refined using REMAC-5. The structure of BRCA1 BRCT-A1 complex was solved in  $P4_12_12$  space group. Clear difference density enabled building and refinement of A1 and A2 peptides. The overall structure of protein in both the complexes is similar to the unliganded structure, with a RMSD of about 0.8 Å over 213 superposed  $C\alpha$  atom pairs. The A1 peptide binds to BRCA1 BRCT in standard “two knob manner”, where pSer-406 binds to N-terminal BRCT while the Phe-409 interacts at the interface region between the two tandem BRCT domains. On the other hand, the A2 peptide is found to be binding in a totally unexpected and novel way. The A2 peptide binds entirely to the N-terminal BRCT domain rather than spanning both BRCT repeats. The two phosphate groups in A2 bind to two different BRCT molecules in the crystal. The phosphate group of pSer-404 rather than the expected pSer-406, occupies the phosphate binding pocket in the N-terminal BRCT domain. The pSer-404 forms only two hydrogen bonds with Ser-1655 and Gly-1656, and it does not form a hydrogen bond with the Lys-1702 which was present in earlier reported BRCA1 BRCT domain complex structures. The phenyl ring of Phe-409 is not surrounded by any hydrophobic moieties. Instead, the charged carboxyl oxygen of Glu-1698 is positioned in the phenyl ring plane at a distance of 3.32 Å suggesting an electrostatic interaction. The hydroxyl group of Thr-1692 residue is also similarly positioned. This is very unique interaction between Phe-409 and BRCA1 BRCT domain which was not seen in earlier reported complex structures.

*Structural basis of pathogenicity of mis-sense  
mutations discovered in BRCA1 BRCT domain*

---

## 6.1. Introduction

The breast cancer susceptibility gene-1 (*BRCA1*) is found to be mutated in most of the Breast and Ovarian cancers [37, 276]. *BRCA1* mutations account for 5-10% of all breast cancers, and 40-45% of hereditary breast cancers. There are different kinds of mutations reported in *BRCA1*, such as deletions, insertions or mis-sense variants. A nonsense mutation in *BRCA1*, which removes last 11 amino acids (Tyr-1853) has been found to be deleterious and responsible for early onset of cancer [277]. The deletion of last eight amino acids from *BRCA1* is found to affect the function [278]. Most of the frame shift and nonsense mutations resulting into truncated *BRCA1* protein are found to be functionally deleterious [279]. All these *BRCA1* mutations are reported in Breast cancer Information Core (BIC) database [280]. BIC database serves as a repository for large number of *BRCA1* and *BRCA2* mis-sense variants however, pedigree information for most of them is unavailable. Hence, they are classified as variants with unknown clinical significance (UCV). These UCVs remain unclassified until their segregation with disease is shown. Although the mutations are present in complete *BRCA1*, they are specifically concentrated in the RING finger and BRCT domains [139]. However, the molecular mechanistic details about how *BRCA1* mutations cause the cancer progression remain unclear.

The BRCT domain is known to act as a phosphopeptide interaction module and perform its tumor suppressor functions by interacting with multiple phosphorylated proteins [108, 109, 281, 282]. The BRCT domain, when fused to GAL4 DNA binding domain is also known to activate transcription [5]. The phosphopeptide recognition is through major interactions of two residues: the phosphorylated serine at the 0 position and the phenylalanine at +3 position. The residues which are involved in direct contact with phosphorylated serine are: Ser-1655, Gly-1656, Arg-1699, Thr-1700, Leu-1701, Met-

1775, and Glu-1836, and mutations in these residues definitely affect the function of BRCA1 BRCT domain. The residues interacting with phenylalanine are Lys-1702, Asn-1774, Met-1775, Val-1740 and Glu-1698. Mis-sense mutations found in the hydrophobic pocket residues are responsible for disruption of hydrophobic environment, and ultimately lead to loss of interaction with the phosphopeptide. It has been reported that BRCA1 mis-sense mutations M1775R and M1775K are responsible for abrogation of peptide binding [117, 283]. The structural stability of the BRCT domain is also essential, and a few mutations that are known to disrupt the protein folding cause loss of function of BRCT domain [283]. Structural analysis of following four BRCA1 BRCT domain variants could enable us to predict the clinical significance of these mutations.

**Table-6.1.** List of BRCA1 BRCT domain mis-sense mutations selected for study

Sr.No.	Mis-sense Mutation	No. recorded in BIC Database	Time and Structural Functional reports	References
1	H1686Q	Unreported	Not Available (N.A.)	[284]
2	S1715R	4	N.A.	[103]
3	P1749R	1	N.A.	[103]
4	C1697R	1	N.A.	[103]

## 6.2. Material and Methods

### 6.2.1. Gene cloning and site -directed- mutagenesis

The gene encoding BRCA1 BRCT domain (1646-1859) was PCR amplified using pGEX-BRCA1 BRCT construct. NdeI site was added in the forward primer (5'-ATCATATG GCCATGGTCAACAAAAGAATGTCC-3') and BamHI site was added in the reverse primer (5'-TAGGATC CTCACTAGGGGATCTGGGGTATCAGG-3'). The amplicon of

~670 base pair was first cloned in pJET blunt end ligation vector and later sub-cloned into a pET3a Vector using Quick T4 DNA ligase (NEB, USA). Site-directed-mutants H1686Q, P1749R, S1715R, and C1697R were produced using primers listed in **Table-6.2**.

**Table-6.2:** List of site- directed- mutagenesis (SDM) primers for BRCA1 mis-sense mutants.

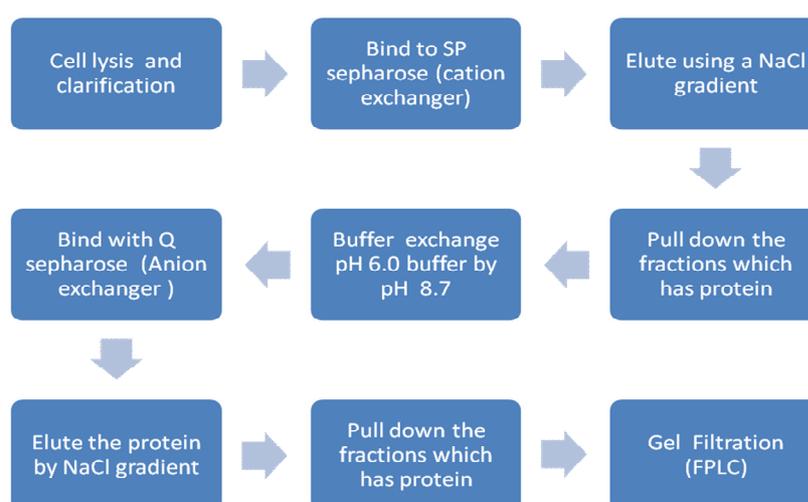
H1686Q	5'-CTGAAGAGACTACTCAGGTTGTTATGAAAACAG-3' 5'-CTGTTTTTCATA ACAACCTCAGTAGTCTCTTCAG-3',
P1749R	5'-GAAACCACCAAGGTCGAAAGCGAGCAAGAG-3' 5'-CTCTTGCTCGCTTTCGACCTTGGTGGTTTC-3'
S1715R	5'-GAAAATGGGTAGTTAGGTATTTCTGGGTGACC-3' 5'-GGTCACCCAGAAATACCTAACTACCCATTTTC-3
C1697R	5'-GATGCTGAGTTGTGCGTGAACGGACACTG-3' 5'-CAGTGTCCGTTACGCACAAACTCAGCATC-3'

The SDM primers were first used to amplify the respective mutant strand using pfu DNA polymerase [285]. The amplified product is then subjected to DpnI digestion to cleave the parent methylated strand. The respective mutations were confirmed by DNA sequencing.

### 6.2.2. Protein expression and purification

The BRCA1 mis-sense mutants expression and purification was first attempted using pGEX-kt vector and affinity chromatography. The GST fused BRCA1 mutant proteins were expressed as described in chapter3. The BRCA1 BRCT variants (H1686Q, P1749R, C1697R and S1715R) in pET3a vector were incorporated in BL21 (pLysS). Cells were grown at 37°C in LB medium containing 100 µg/ml Ampicillin and 34 µg/ml chloramphenicol till A600 OD reached a value between 0.6-0.8. Then the culture was induced at 18°C by adding IPTG to a concentration of 0.4 mM under constant agitation. All the further protein purification steps were carried out at 4°C. The induced culture were harvested in Sorvall SLC-3000 rotor at 6000 rpm for 10 min. Harvested induced

cells were then resuspended in lysis buffer 20 mM Sodium phosphate pH- 6.0, 50 mM NaCl, 1 mM EDTA, 1 mM  $\beta$ ME, 1 mM PMSF and 0.5% Triton X-100, and further sonicated for 5-8 times. Cell lysate was centrifuged at 18000 rpm for 30 min and loaded onto SDS-PAGE gel. It was observed that BRCA1 BRCT mutants C1697R, S1715R and P1749R were not soluble. Therefore, we proceeded with further purification of only the BRCA1 BRCT variant, H1686Q. Cleared bacterial cell lysate was bound to SP sepharose resins, preequilibrated by buffer B, 20 mM Sodium phosphate pH- 6.0, 50 mM NaCl, 1 mM EDTA, 1 mM  $\beta$ ME, 1 mM PMSF. Bound protein was eluted with salt gradient ranging from 100 mM – 800 mM. The fractions which showed proteins were pooled and concentrated to a final volume of 2 ml. The buffer was exchanged to sodium borate buffer of pH 8.7, using G-50 desalting column, and then the sample was applied on Q sepharose column preequilibrated with 10 mM Sodium borate buffer. Bound protein was eluted using NaCl gradient ranging in concentration from 100 mM to 500 mM. Protein expression and purification were monitored by loading the fractions on SDS PAGE. The whole mis-sense mutant purification process is schematically represented in **Figure-6.1**.



**Figure-6.1: Schematic representation of purification of BRCA1 H1686Q mis-sense mutant.**

Gel filtration chromatography was done using superdex-75 Hiload column (GE Healthcare). The elution fractions containing protein from Q sepharose were pooled down and concentrated to a volume of 2 ml, and applied on superdex-75 column pre-calibrated with 10 mM Sodium borate pH 8.7, 100 mM NaCl buffer. Fractions were collected by monitoring the absorbance at  $\lambda=280$  nm.

### **6.2.3. Protein characterization**

#### **6.2.3.1. Circular Dichroism**

Circular dichroism (CD) spectra were collected using JASCO J-715 spectropolarimeter (Jasco, Easton, MD), in the far-UV region (180-260 nm). The JASCO J-715 is well equipped with a Jasco PTC 348 WI temperature controller and sealed quartz cuvettes (Jasco, USA). In the far-UV region (180–260 nm) BRCA1 BRCT and BRCA1 H1686Q (concentration 10  $\mu$ M in borate buffer, pH-9.0, 300 mM NaCl) was loaded onto a 0.1 cm path-length quartz cuvette (Hellma, Germany). Seven spectra were accumulated and averaged for each experiment, with a resolution of 1 nm at a scan speed of 50 nm/min with a response time of 1 s. Buffer blank spectra, obtained at identical conditions, have been subtracted from the raw data. The results in all experiments have been expressed in molar ellipticity  $[\theta]$  ( $^{\circ}\text{cm}^2 \text{dmol}^{-1}$ ).

#### **6.2.3.2. Fluorescence spectroscopy**

Emission spectra were collected on HORIBA FL3-21 spectrofluorometer. The variant and wild type BRCA1 BRCT of 10  $\mu$ M concentration were taken in a cuvette of 1 mm path length, and emission spectrum was recorded with excitation at 280 nm.

### **6.2.4. Prediction of pathogenicity using online servers**

Align-GVGD [21] analysis was carried out to predict the pathogenicity of mutant protein. This server uses multiple sequence alignment available for BRCA1 from human to frog and grades the pathogenicity in different classes beginning from C0, C15, up to C65. Any

mutation with category C0 is less likely to be pathogenic and C65 is most likely to be pathogenic.

### 6.2.5. In silico modelling

To study the structural effect of mutations, atomic models for the four mutants H1668Q, S1715R, C1697R and P1749R were built based on the reported crystal structure of BRCA1 BRCT (PDB ID:1Y98) [102], and these molecular models were subjected to Molecular Dynamics Simulation (MDS) [286] using Desmond 2010 software package [287]. Optimized Potentials for Liquid Simulations (OPLS) all-atom force field was used to analyze model stability. These pre-equilibrated systems were subsequently used in the 1 ns production MDS with a time-step of 2 femto seconds. Structural coordinates were saved every 1.2 picoseconds and analyzed using the analytical tools present in the Desmond package. The lowest energy structures were obtained for each mutant and molecular interactions of wild-type and mutant were plotted using Ligplot [246].

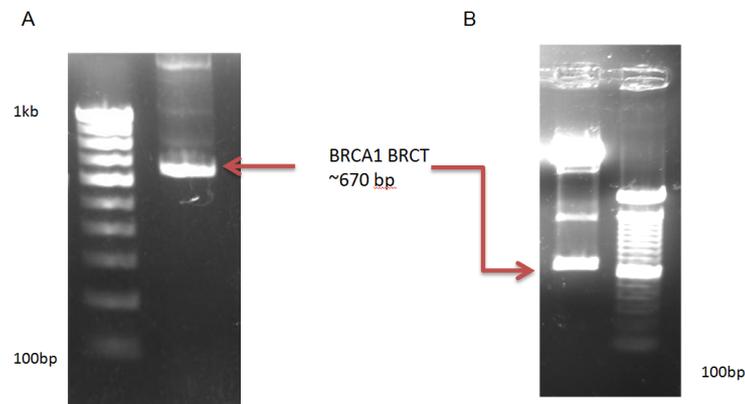
## 6.3. Results and discussion

We have attempted to study the effect of these mis-sense mutations on structural integrity of BRCT domain using biophysical and *in-silico* approach. Three of the four mutants BRCA1 P1749R, C1697R and S1715R are discovered to be highly pathogenic with a grade of C65 class by A-GVGD analysis. On the other hand BRCA1 H1686Q mutation was calculated to be moderately pathogenic with a grade of C15 [21], even though it has been found in some cancer patients.

### 6.3.1. Cloning of BRCA1 BRCT domain in pET3a and SDM

**Figure-6.2A** shows the amplification of the BRCA1 BRCT domain (1646-1859) using the forward and reverse primers as described before. PCR product of ~650 base pairs and empty vector pET3a were digested by NdeI and BamHI restriction enzymes. Both the digested products were mixed in 3:1 molar ratio, and ligated using Quick DNA ligase.

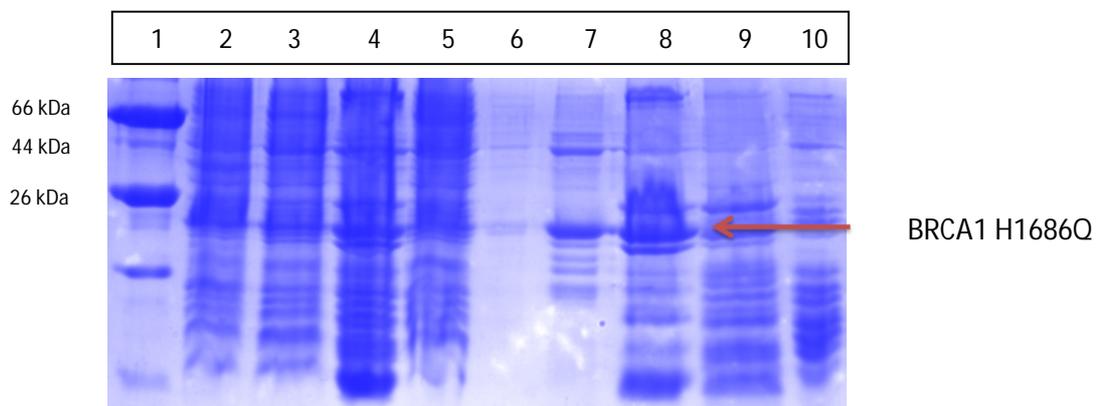
**Figure-6.2B** shows the insert release of BRCA1 BRCT domain (670 bps) which was further confirmed by DNA sequencing. The mutations, generated using site-directed mutagenesis, were also confirmed by DNA sequencing.



**Figure-6.2:** PCR amplification of BRCA1 BRCT domain (~650bp) A), The restriction digestion of potential clones B).

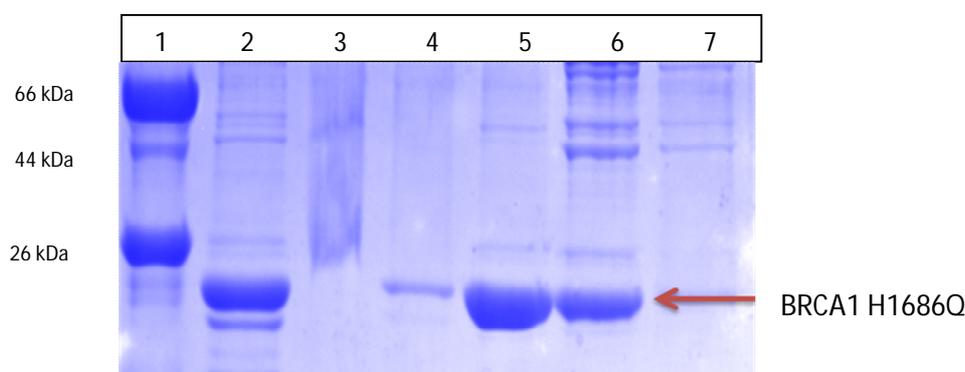
### 6.3.2. BRCA1 BRCT domain mis-sense mutant expression and purification

We have attempted to express BRCA1 mutant proteins as GST fusion proteins in pGEX-kt vector. Though the fusion protein expressed in BL21 (DE3) cells, the mutant protein could not be cleaved off from the GST sepharose beads, perhaps because of the inaccessibility of the cleavage site. Further, we sub-cloned the mutants in pET3a expression vector, without any affinity tag.



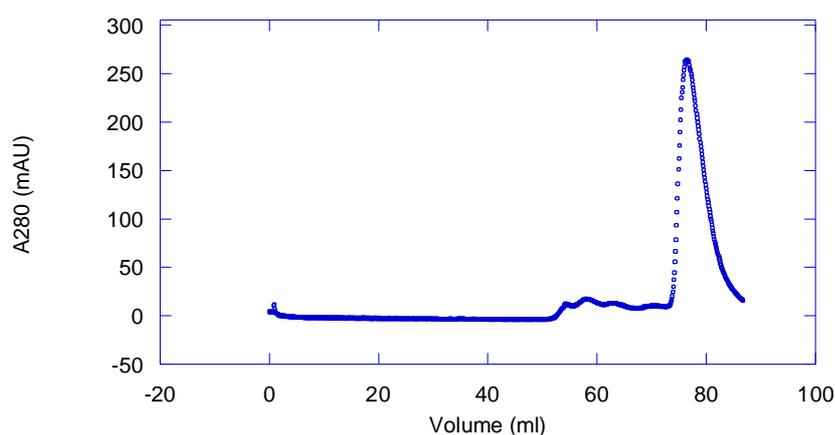
**Figure-6.3: Cation exchange chromatographic profile of BRCA1 H1696Q mutant**, Lane 1 is marker, lane 2-Induced whole cell, lane3- Induced soluble fraction, lane4- protein bound on SP sepharose resin, lane5- unbound protein, lane 6 to 9- Different NaCl elution fraction (mM), lane10- Beads after elution.

Three of them are insoluble and are difficult to purify, which indicates deleterious effect of mis-sense mutations on protein structure and conformation. We were, however, able to purify the BRCA1 BRCT H1686Q mutant by the combination of cation (**Figure-6.3**) and anion exchange chromatography (**Figure-6.4**).



**Figure-6.4: Anion exchange chromatographic profile BRCA1 H1696Q mutant**, lane1- molecular weight marker, lane2 total loaded protein, lane3- protein bound on Q sepharose, lane 4 to7- Gradient NaCl elations

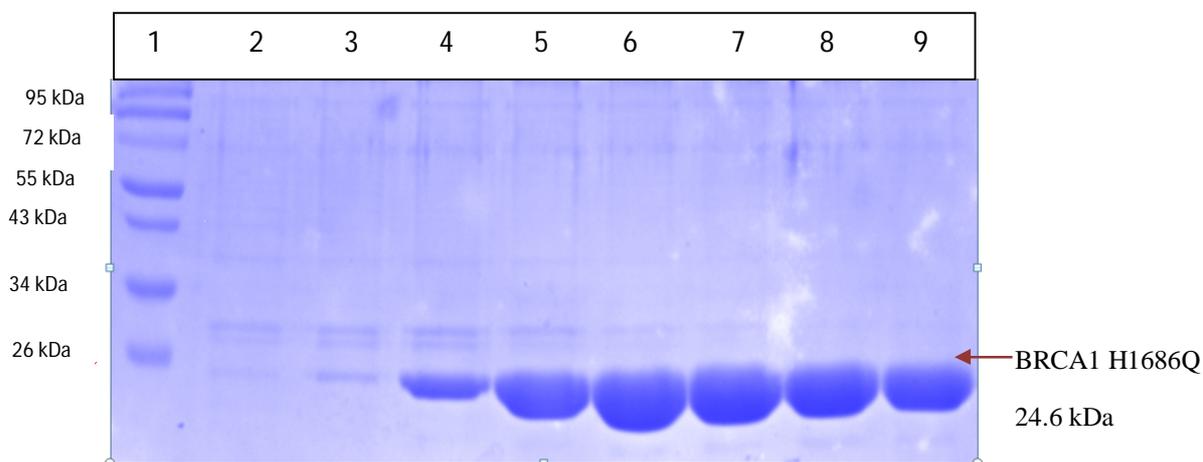
The fractions which showed the presence of protein were pooled together and passed through superdex-75 gel filtration column. The BRCA1 H1686Q mutant protein eluted in same elution volume as that of wild -type BRCA1 BRCT domain, which indicates the homogeneous form of protein (**Figure-6.5**).



**Figure-6.5:** Gel filtration chromatogram of BRCA1 H1686Q mutant protein.

The purity of purified BRCA1 H 1686 Q mutant is validated by running SDS-PAGE

(**Figure-6.6**).



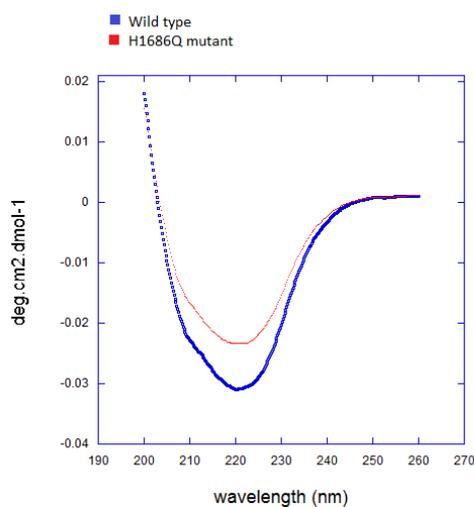
**Figure-6.6:** The SDS-PAGE profile of FPLC purified BRCA 1 H 1686 Q protein, lane 1 is molecular weight ladder while lanes 2-9 are different FPLC fractions.

### 6.3.3. Characterization of BRCA1 H1686Q variant

The BRCA1 H1686Q was biophysically characterised using Circular Dichroism and fluorescence spectroscopy to explore changes in the secondary and tertiary structures of the mutant.

#### 6.3.3.1. Circular Dichroism analysis

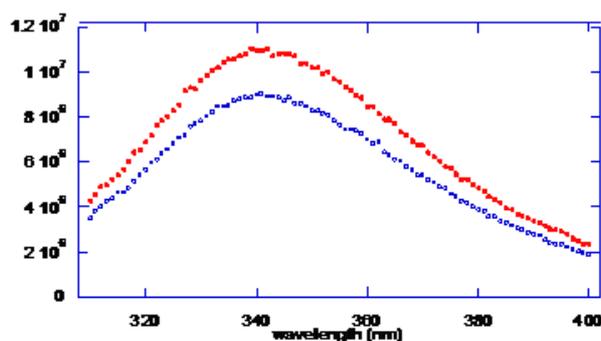
The CD spectra taken at near UV range (200- 260 nm) for both BRCA1 H1686Q and wild -type BRCA1 BRCT are overlaid in (**Figure-6.7**). The spectra indicate there is slight change due to mutation from histidine to glutamine at BRCA1 1686 position.



**Figure-6.7:** CD spectra for wild- type BRCA1 BRCT (blue) and BRCA1 H 1686 Q mutant (red) in near UV range.

### 6.3.3.2. Fluorescence spectrometry

The BRCA1 BRCT contains eight tryptophan residues which can be used to investigate if the overall fold of BRCT domain is altered by the H1686Q mutation. The wild -type and BRCA1 H1686Q mutant has been excited at 280 nm and their emission maxima were recorded (**Figure-6.8**). No alterations in the position of the fluorescence maximum were observed, however there is slight change in intensity, unravelling the fact that the tertiary structure is moderately changed.



**Figure-6.8:** Fluorescence emission spectra of BRCA1 BRCT (blue) and BRCA1 His 1686 Gln mutant (red) excited at 280nm.

### 6.3.4. Prediction of pathogenicity of BRCA1 BRCT mutants

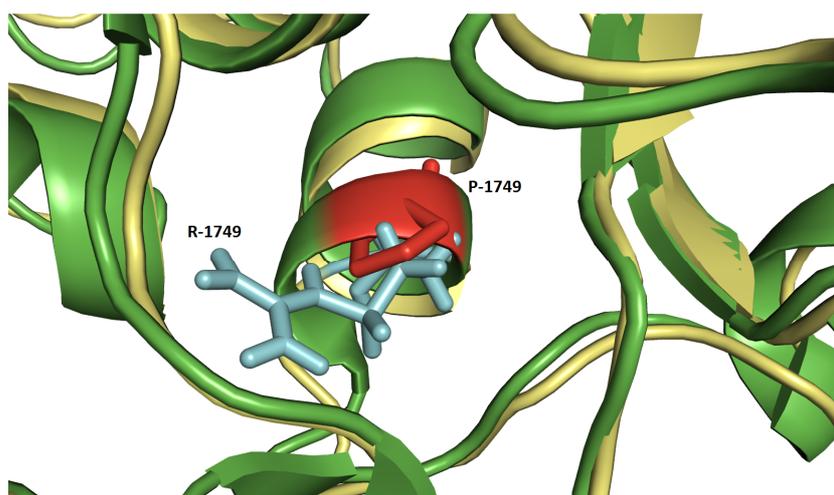
Online softwares such as the *Align-GVGD* [21], *Polyphen* [22], *Mutpred* [23] are also useful to predict the pathogenicity of a particular mutation. *Align-GVGD* software predicts three mutants (P1749R, C1697R and S1715R) to be highly pathogenic with a grade of C65 class. On the other hand, H1686Q mutation was calculated to be moderately pathogenic, with a grade of C15 [21].

### 6.3.5. Intramolecular interactions involving pathogenic mutations

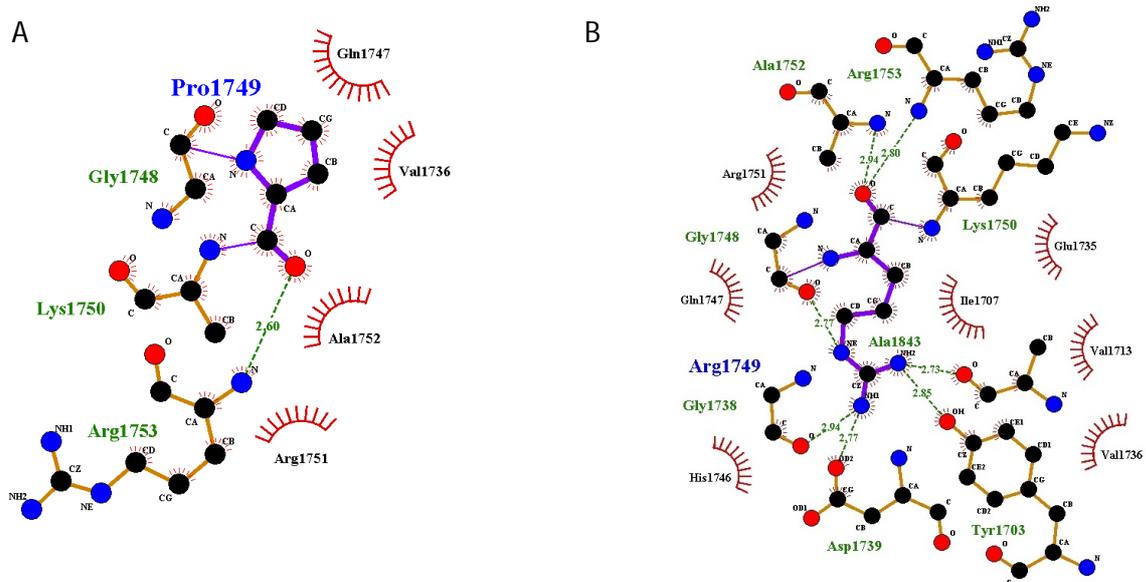
To gain structural insights into why these mutations are pathogenic, we have carried out molecular model building of these mutant proteins.

**BRCA1 Proline 1749 to Arginine:**

The BRCA1 P1749R has been reported to be associated with ovarian cancer [22]. Proline (P) is a medium-sized, rigid, and hydrophobic amino acid, while arginine (R) is a large-sized and positively charged basic amino acid. Therefore, this mutation is expected to have significant structural effect. **Figure-6.9** show the superposition of wild- type and mutant protein structures highlighting the effect due to change in the conformations of residues. Though the overall rmsd is 1.1 Å, which is not significant, the local changes are substantial. The extent and type of intramolecular interactions in wild- type and P1749R mutant are shown in **Figures-6.10A** and **6.10B** respectively. Arginine 1749 is forming different hydrogen bonds with Gly-1738, Asp-1739, Tyr-1707, Ala-1843, and Gly-1748. However, arginine has also re-oriented the hydrophobic pocket by forming hydrophobic interactions with Lys-1750, His-1746, Val-1713, and Ile-1707. This alteration could affect the ability of BRCA1 BRCT to interact with other protein partners, thereby affecting the functionality.



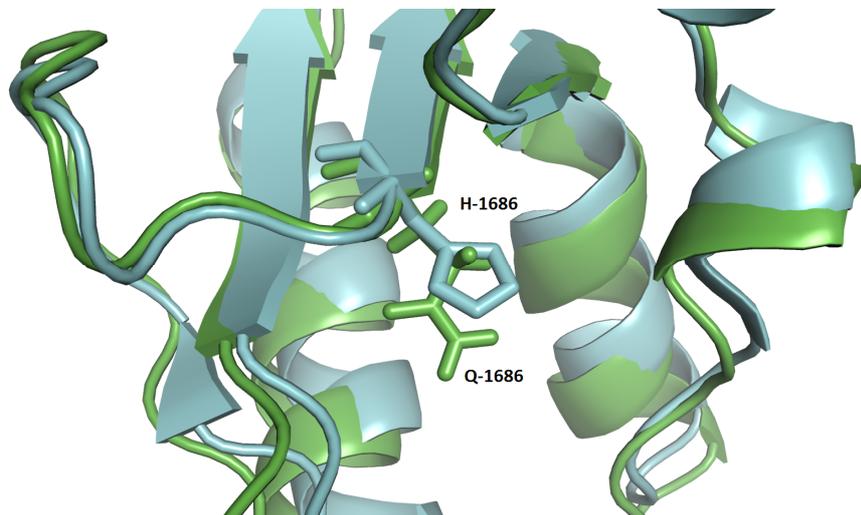
**Figure-6.9:** The superposition of wild type BRCA1 BRCT (green) with modelled mutant BRCA1 P1749R (yellow), where the proline 1749 (red) of wild type and arginine 1749 (cyan) of mutant protein represented by stick model.



**Figure-6.10:** Schematic representations of weak intramolecular interactions between BRCA1 BRCT wt (A) and pathogenic mutations P1749R (B), figure prepared in Ligplot [246].

#### BRCA1 Histidine 1686 to Glutamine:

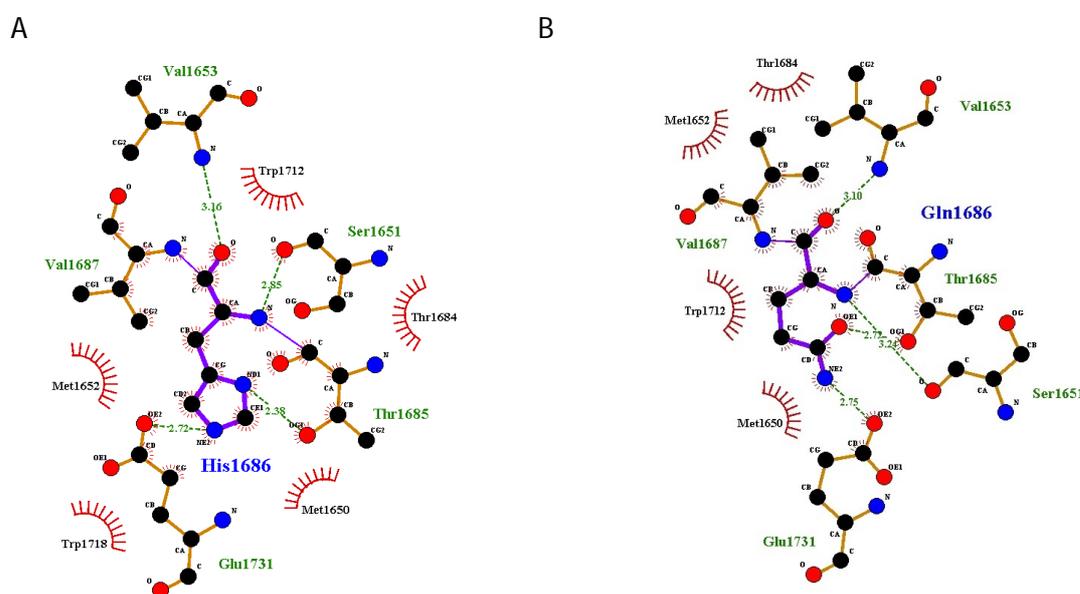
This mutation has been discovered in Italian breast and ovarian cancer families [284]. It is located on exon 17 of BRCA1 and is characterized as pathogenic mutation.



**Figure-6.11:** The superposition of wild type BRCA1 BRCT (cyan) with modelled mutant BRCA1 H1686Q (green), where the histidine 1686 (cyan) of wild type and glutamine 1686 (green) of mutant protein represented by stick model.

Structurally, BRCA1 His-1686 is stabilized by hydrophobic interactions with Met-1650, Met-1652, Val-1687 and Trp-1712. It also forms hydrogen bonds with residues Thr-1685,

Val-1653, and Glu-1731 as shown in **Figure-6.12**. However, the modelled mutant BRCA1-H1686Q is stabilized by a single hydrophobic interaction with Val-1687 and four hydrogen bonds with Thr-1685, Val-1653, Ser-1651, and Glu-1731. Though there are not many differences in overall structure, rmsd (1.2 Å), there (**Figure-6.11**) are significant differences around the hydrophobic core, pointing to the reason for pathogenicity of the mutation

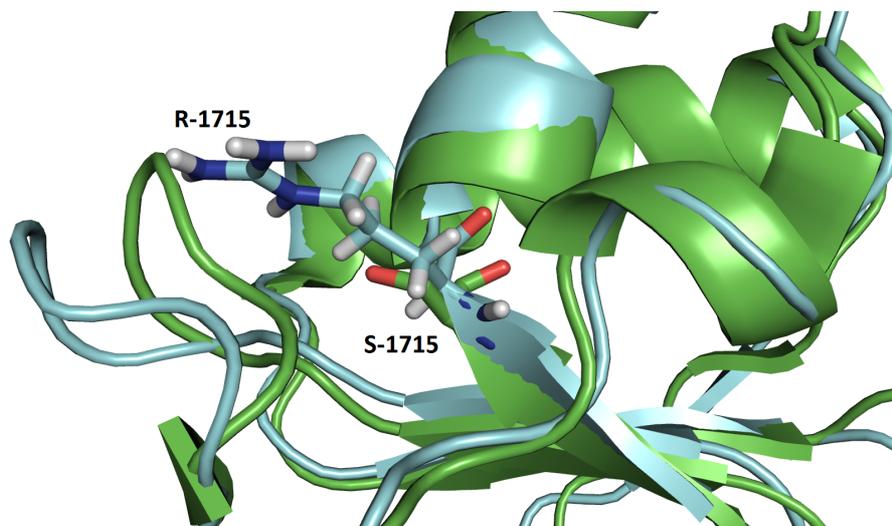


**Figure-6.12:** Schematic representations of weak intramolecular interactions between BRCA1 BRCT wild-type (A) and pathogenic mutation H 1686 Q (B), figure prepared using Ligplot [246].

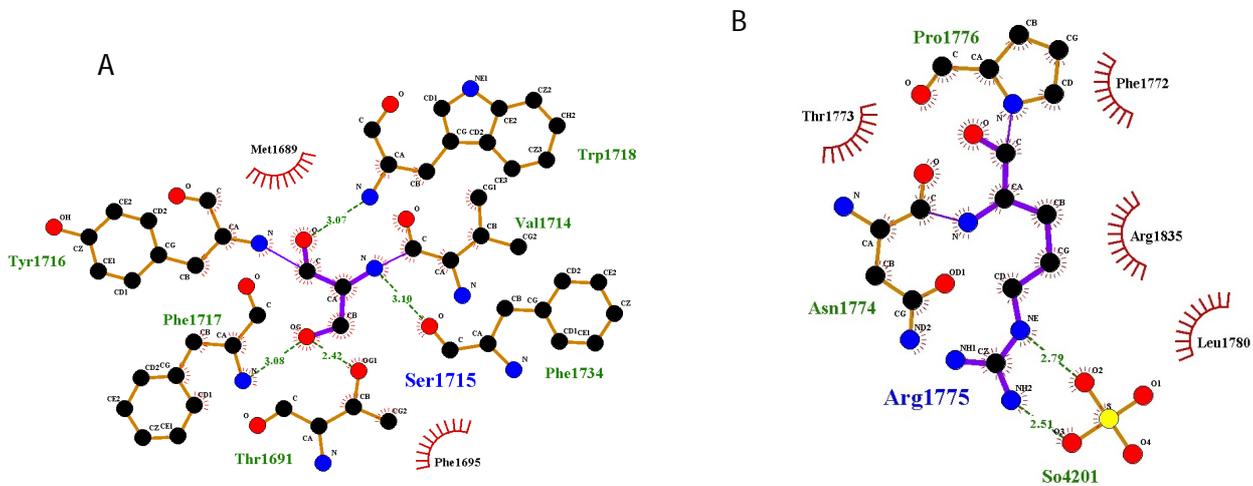
### BRCA1 Serine 1715 to Arginine:

Ser-1715 is located on exon 18 of *BRCA1* and its mutation to Arg has been reported on one occasion in a Danish patient [24]. Ser-1715 is present in the hydrophobic core and when it is mutated to Arg, its impact on structure and function is unknown [25]. Mutation from a small size and polar Ser (S) to a large size and basic Arg (R) can affect the structure significantly. Model building, however, shows that there is not much conformational change due to the mutation (RMSD-1.05 Å) (**Figure-6.13**). The Ser-1715 residue is close to the surface shielded by three phenyl side chains (1695, 1717 and 1734).

These phe side chains take up different conformation and make room for Arg-1715 side chain in the mutant structure. In the wild type, conformation of Ser-1715 is stabilized by three hydrophobic interactions, with Val-1714, Tyr-1716, and Phe-1695, and three hydrogen bonds with Phe-1734, Trp-1718, and Thr-1691. In the mutant BRCA1- there are fewer hydrogen bonds and more hydrophobic interactions as seen in **Figures-6.14 A** and **6.14 B**.



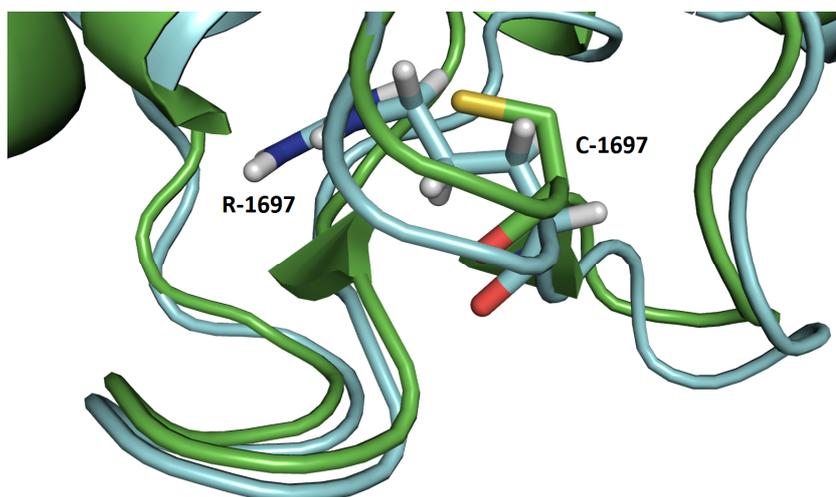
**Figure-6.13:** The superposition of wild type BRCA1 BRCT (green) with modelled mutant BRCA1 S1715 (cyan), where the serine 1715 of wild- type and arginine 1715 of mutant protein represented by stick model.



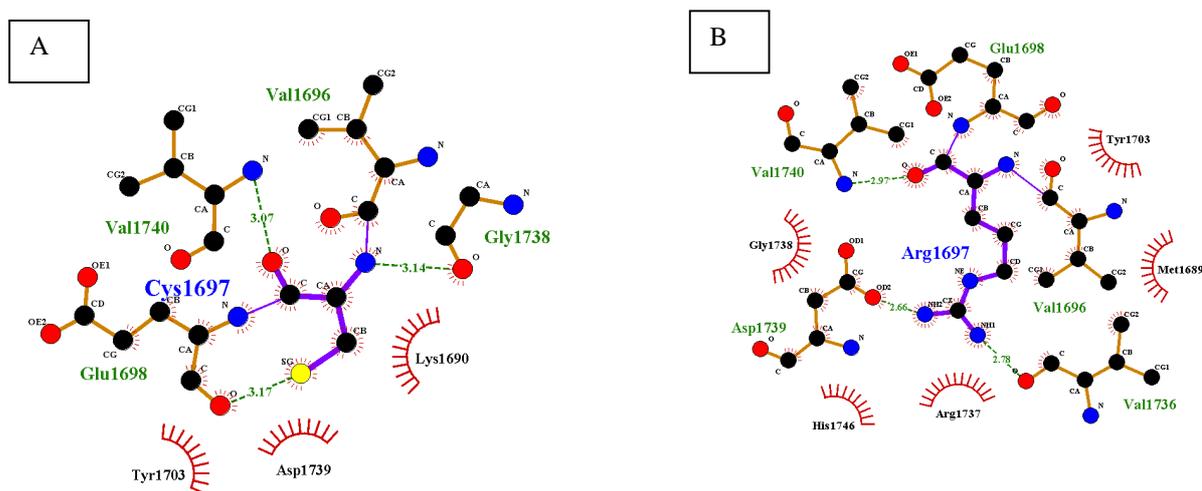
**Figure-6.14:** Schematic representations of weak intramolecular interactions between BRCA1 BRCT wild-type (A) and pathogenic mutations S1715R (B), figure prepared in Ligplot [246].

### BRCA1 Cystine 1697 to Arginine

Cys-1697 is located on exon 18 of *BRCA1* and its mutation to Arg has been reported in BIC database [6]. Hence it is very crucial to characterise pathogenicity of this mutation based on protein-protein interactions.



**Figure-6.15:** The superposition of wild type BRCA1 BRCT (green) with modelled mutant BRCA1 C1697 (cyan), where the cystine 1697 of wild type and arginine 1697 of mutant protein represented by stick model.



**Figure-6.16:** Schematic representations of weak intramolecular interactions between BRCA1 BRCT wild-type (A) and pathogenic mutations BRCA1 C1697R (B), figure prepared in Ligplot [246].

Model building has shown that there is not much change occurring in the overall structure as the rmsd is only 1.09 Å (**Figure-6.15**). In the wild-type structure, BRCA1 Cys-1697

forms hydrogen bonds with Val-1740 and Gly-1738 while in case of mutant model the hydrogen bonding residues to Arg-1697 are altered to Val-1736 and Asp-1739 (**Figure-6.16**). Due to mutation there are alterations in the hydrophobic interactions compared to wild- type. This indicates that there are local conformational changes which may be significant.

#### **6.4. Conclusion**

BRCA1 is found to be mutated in breast and ovarian cancers. As per the BIC database BRCA1 mutations account for 5-10% of all breast cancers and 40-45% of hereditary breast cancers. Most of these mutations are concentrated in RING finger and BRCT domains of BRCA1 protein. The tools to definitely establish the correlation between a given mutation and the disease are lacking. However sequence alignment and three dimensional structure of mutant protein have been advanced as tools to establish this linkage. If a mutation is to be associated with a disease, then it should be found in highly conserved region of sequence. Similarly a disease associated mutation would lead to an altered 3D structure resulting in abrogation of phosphopeptide binding. The mutations C1697R, P1749R, S1775R and H1686Q are in the conserved region and hence are likely to be pathogenic. To understand the exact mechanism of pathogenicity, we have undertaken the structural studies on these mutants.

These four mutations were first generated in pGEX-BRCA1 BRCT construct using site directed mutagenesis. The mutant protein was over expressed as fusion proteins in BL21 (DE3) cells. The mutant proteins could not be cleaved off from the GST sepharose beads. Subsequently, we sub-cloned the mutants in pET3a expression vector, without any affinity tag. Three of the four mutants are insoluble, and difficult to purify, suggesting deleterious influence of these mis-sense mutations. We were however able to purify the BRCA1 BRCT H1686Q mutant by ion exchange chromatography. CD and fluorescence

studies indicate that BRCA1 H1686Q mutant shows some changes in secondary and tertiary structure when compared with wild type protein. In-silico analysis revealed that there is alteration in the intramolecular hydrophobic interactions upon mis-sense mutations.

*Analysis of BRCT domain containing proteins*

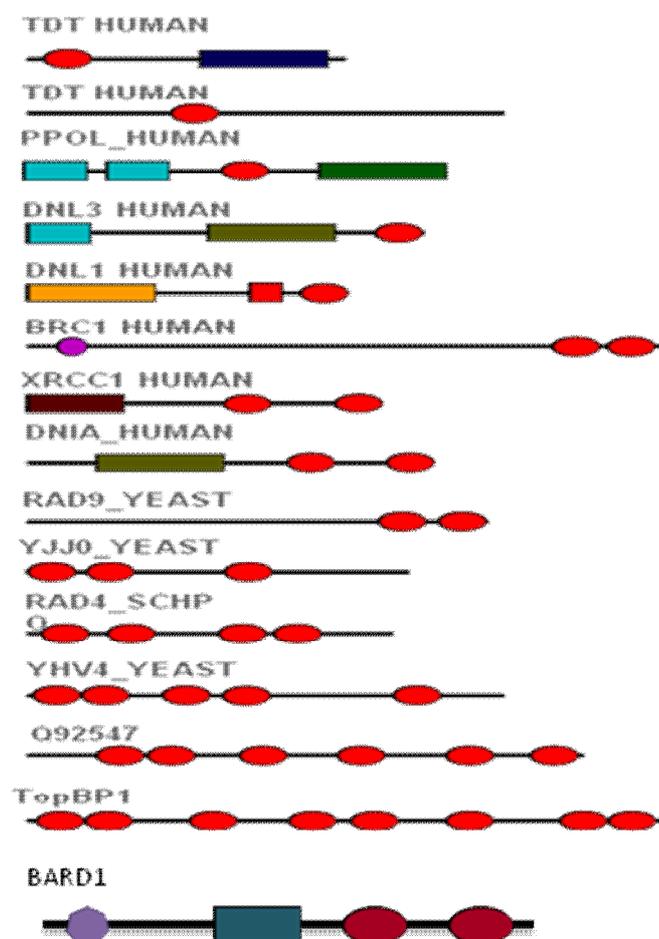
---

## 7.1. Introduction

Every aspect of cellular function including proliferation, gene expression, metabolism and survival, in a multi-cellular organism, depends upon the signal transduction [288]. The basic prerequisite of the signalling process is the activation of specific molecules by unique post translational modification such as phosphorylation, acetylation, ubiquitination, methylation etc. In the cell, upon extrinsic signals, the writer molecules, like the ATM and ATR kinases are responsible for the phosphorylation and activation of specific proteins [289]. For example the proteins which are phosphorylated at the Ser and Thr residues are recognized by mediator proteins, like BRCA1 and MDC1, which are known to possess the BRCT domain [87]. ATM and ATR is responsible for phosphorylation of number of different proteins in the cell [290]. However, it is the interacting mediator proteins which transfer the signals to the effector molecules for the specific functional response. The BRCT-domain-containing-proteins belong to an important class of mediator proteins which are involved in DNA damage repair and checkpoint regulation [87].

The BRCT domain has been initially identified in BRCA1, and is found subsequently in many different prokaryotic and eukaryotic proteins [87]. The BRCT domain, shown as a space filled ellipse in **Figure-7.1**, is comprised of 85-95 amino acids, with conserved hydrophobic amino acids forming the core. The BRCT domain is made up of a central 4-stranded  $\beta$  sheet surrounded on either side by  $\alpha$ -helices, two on one side and one on the opposite side. The topology of the BRCT domain is  $\beta 1-\alpha 1-\beta 2-\beta 3-\alpha 2-\beta 4-\alpha 3$  [10]. The BRCT domain is found either as a single unit or as tandem multiple units within an individual protein (**Figure-7.1**). In the case of multiple BRCT domain containing proteins, the linker length varies from 20-60 amino acids, providing different arrangements for BRCT domains in the individual proteins [94]. The two BRCT domains

are separated by a linker of ~25 amino acids in BRCA1 [87]. Although the BRCT domain does not have enzymatic ability, it plays an important role in ensuring proper DNA repair and checkpoint regulation through the phosphomediated protein-protein interactions[94].



**Figure-7.1: Schematic representation of different BRCT domain family members**

There are 134 redundant BRCT domain structures available in the protein data bank. Similar structural reports of each BRCT domain are considered as redundant structural entry. The non-redundant distribution of these structures into different categories such as single BRCT domain, multiple BRCT domains, BRCT domain complexed with phosphopeptide and BRCT domain mis-sense mutants is given in **Table7.1**.

**Table-7.1:** List of non-redundant BRCT domain containing protein structures from PDB

Domain architecture	Examples of BRCT repeat containing proteins	PDB ID
Single BRCT domain	XRCC1, Lig3, PARP-1, TDT, ECT2, DNA pol $\mu$ , REV1, RFC, PES1, DNA pol $\lambda$ , MCPH1, FCP1, TopBP1,	1CDZ, 1IN1, 2COK, 2COE, 2HTF, 2EBW, 2EBU, 2EP8, 2JW5, 2WT8, 3EF0, 3JVE
Tandem BRCT domain	BRCA1, 53BP1, MDC1, BARD1, <i>S.pombe</i> Crb2, <i>S.pombe</i> Brc1, TopBP1	1JNX,1KZY,2ADO,2NTE, 2VXB, 3L40, 3AL2
BRCT domain complexed with phosphopeptide	BACH1, CtIP, ACC1, ATRIP, BAAT,	1T15, 1Y98, 3COJ, 4IGK, 4IFI
BRCA1 mis-sense mutations	M1775R, M1775K, D1840T, G1656D, T1700A, R1699Q, R1835P, E1836K, V1809F	1N5O, 2ING, 3K15, 3PXA, 3PXB,3PXC, 3PXD, 3PXE, 1T2U

The single unit BRCT domain is mostly present in prokaryotes while the tandem BRCT domain containing proteins are predominantly found in multi-cellular organisms [291]. This correlation may be indicative of the evolution in the functionality of the BRCT-domain-containing-protein in maintaining genomic integrity through protein-protein interactions. BRCA1 Associated Ring Domain (BARD1) protein contains two BRCT domains and is therefore expected to form functional complexes with phosphorylated proteins. However, so far, no such complex has been reported.

## 7.2. Materials and methods

### 7.2.1. Amino acid sequence alignment

The three dimensional structures of BRCT-domain-containing-proteins and their amino acid sequence in FASTA format were obtained from protein data bank (PDB). Multiple

sequence alignment was carried out using DALI [292], and MULTALIN softwares [140]. The final multiple sequence alignment figure was prepared by ESPript program (<http://esprict.ibcp.fr/ESPript/ESPript/>). The RMSD values are as reported by the DALI software [292]. The superposition figures were prepared in pymol [245].

### **7.2.2. Cloning of BARD1 ARD-BRCT domain in pGEX-Kt vector**

The BARD1 ARD-BRCT (425-777) were PCR amplified from full length BARD1 cDNA ( Kind gift from Prof. Richard Bair US) using forward 5'- GTCGGATCCCA TATG GAGAACCTGTACTTTCAGG GTAACCATCGTGGTGAGACTTTGCTCCT-3' and reverse 5'- GTCGGATC CGAATCCCTATTAGCTCTCAAGAGGAAGCAA CTC-3' primers. The PCR product and pGEX-kT vector were digested by BamHI and EcoRI restriction enzymes. The digested products were purified and ligated using Quick DNA ligase. The potential clones were screened using restriction digestion method, and were further confirmed by DNA sequencing.

### **7.2.3. Expression and purification of BARD1 ARD-BRCT**

The pGEX-BARD1-ARD-BRCT construct is incorporated into BL21 (DE3) bacterial strain by the process of transformation. The starting culture was prepared by inoculating a single colony of cells containing the pGEX-BARD1ARD-BRCT construct into 100 ml LB broth with 100 µg/ml ampicillin, then allowed the culture to grow overnight at 37°C. The protein expression was scaled up by seeding 1% starting culture into a larger volume (8 Litre) of growth medium, incubated at 37°C until OD at  $\lambda = 600$  nm reaches between 0.6-0.8. The IPTG was added in the culture flask to the final concentration of 0.3 mM, and the culture was incubated at 24°C for 16 hours. The bacterial culture is pelleted at 4°C by centrifugation at 6000 rpm for 10 min. All the protein expression and purification steps were carried out at 4°C. The bacterial cells were resuspended in lysis buffer

containing 50mM Tris, pH-7.5 and 300mM NaCl and further sonicated. The cell lysate was cleared by centrifugation at 18000 rpm for 30 min. Cleared cell lysate was applied on GST sepharose 4B column which was pre-equilibrated with lysis buffer. Further, after allowing the protein to bind to the column for 2 hours, the unbound protein was washed by lysis buffer. The BARD1 ARD-BRCT protein was cleaved from beads by incubation with TEV protease. Cleaved protein was then passed through Ni-NTA beads to remove the His-tagged-TEV protease. This partially purified protein fraction was concentrated to a final volume of 2ml, and loaded on superdex-200 gel filtration column for further purification.

### **7.3. Results and discussion**

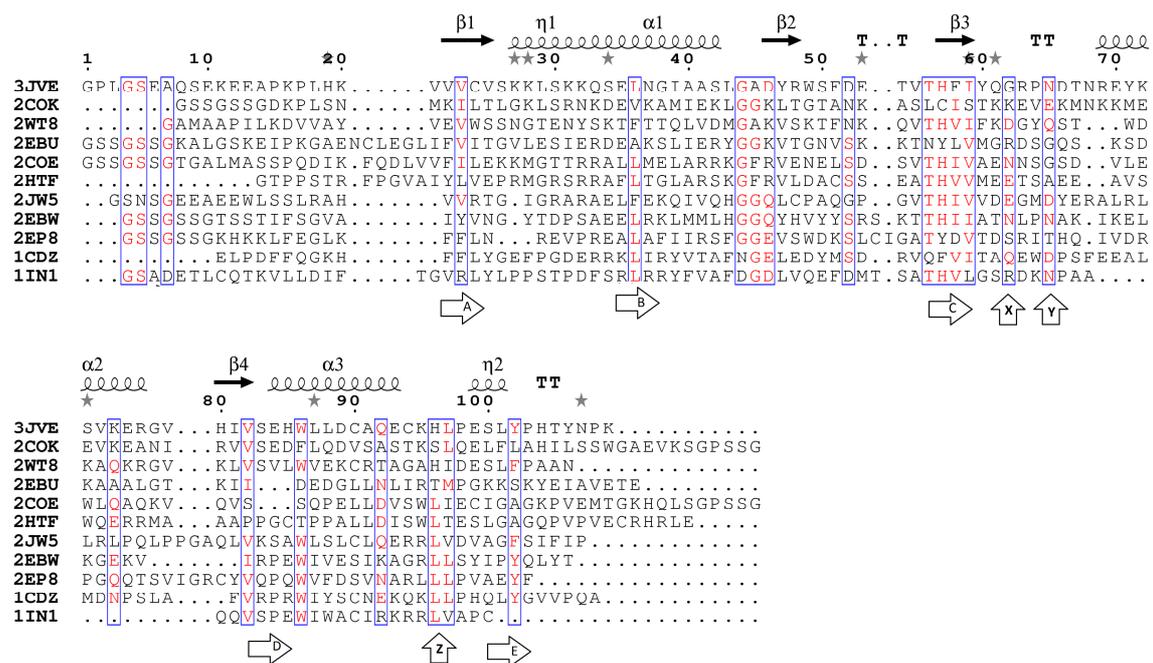
#### **7.3.1. In-silico analysis**

The structure and amino acid sequence of single and double BRCT domain containing proteins were analysed for functional significance.

##### **7.3.1.1. Structural alignment of single BRCT domain**

The list of protein structures with single BRCT domain has been reported in **Table-7.1**. Among these reported 11 structures, 5 were solved by crystallography and the remaining by NMR. We have separately superposed the structures determined by X-ray and NMR methods. The rmsd values for these superpositions are given in **Table-7.2** and **Table-7.3**. Interestingly, even though the sequence similarity is very less, the three dimensional fold is similar. **Figure-7.2** shows the sequence alignment derived from structural superposition of the single BRCT-domain-containing-structures. There are regions in the sequence, where amino acids residues are of similar chemical character and these residues are boxed in **Figure-7.2**. It is interesting to note that majority of boxed residues occur within secondary structure. Alignment of BRCT domain sequences has been reported earlier [87]. These investigators identified five regions, labelled A-E and ranging from 3-7

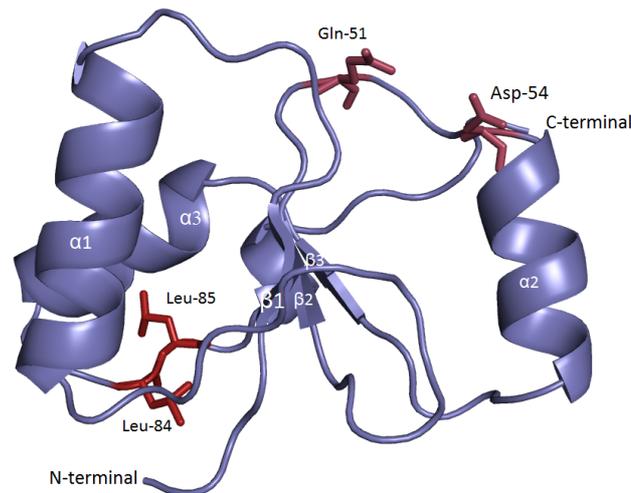
residues, as conserved region in BRCT domain [87]. The boxed region obtained in the present alignment matches with the five conserved regions A-E. However, the number of residues in each conserved element is much smaller than what has been reported earlier [87]. For example in element A, the earlier report has six contiguous residues, whereas in the present alignment based on the structure, a single residue overlaps the A element. Additionally, the present alignment reveals new regions of conservations labelled as X, Y and Z in **Figure-7.2**. It is interesting that these residues map on to the loop regions in the BRCT domain (**Figure-7.3**). In XRCC1 protein the residues Gln-51 and Asp-54 are not part of either a  $\alpha$ -helix or a  $\beta$ -sheet, but are found to contribute significantly for intramolecular interactions through hydrogen bonding.



**Figure-7.2:** The multiple sequence alignment of single BRCT repeat containing proteins, PDB ID: 1CDZ, 1IN1, 2COK, 2COE, 2HTF, 2EBW, 2EBU, 2EP8, 2JW5, 2WT8, 3JVE correspond to XRCC1, Lig3, PARP-1, TDT, ECT2, DNA pol $\mu$ , REV1, RFC, PES1, DNA pol $\lambda$ , MCPHI, TopBP1 protein respectively.

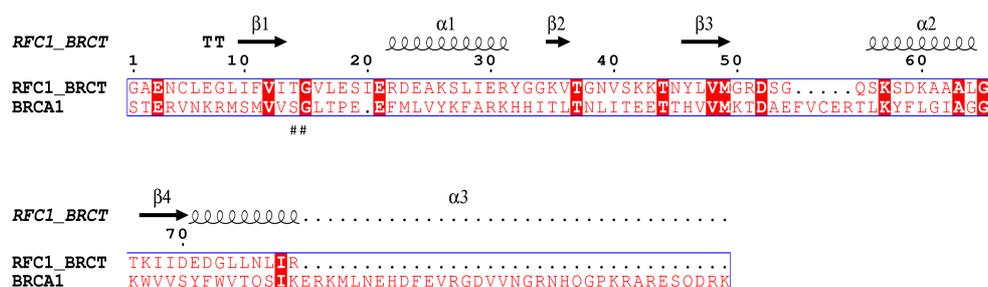
The Gln-51 of XRCC1 forms two hydrogen bonds with Arg-71 and Thr-49. While the Asp-54 forms three hydrogen bonds with Glu-52, Ser-56 and Phe-57. This indicates that these residues may be important for providing structural stability to BRCT domain and

especially to the loop region. The single BRCT domain is postulated to have the primary function of phosphate binding [293]. Interestingly, sequence alignment **Figure-7.2** does not reveal presence of standard phosphate binding motifs such as, Walker A, Walker B or P peptide motif [294, 295].



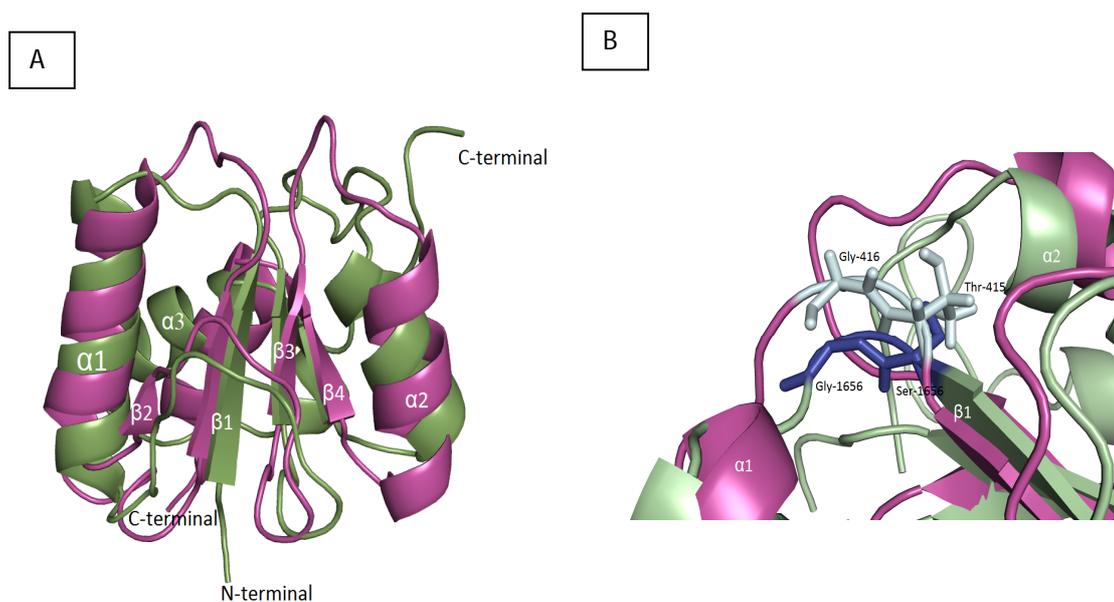
**Figure-7.3: Mapping of conserved residues on XRCC1 BRCT domain (PDB ID:1CDZ).** The residues Gln-51, Asp-54 and Leu84-85 are shown as stick model which representing X, Y, and Z conserved motifs.

The three dimensional solution structure of single BRCT domain containing protein RFC1 has been determined as a complex with DNA fragment (29 bp). In this structure the residues Thr-415 and Gly-416 are observed to form hydrogen bonds with the phosphate group. Secondary structure based sequence analysis reveals that phosphate group interacting residues of RFC1 protein overlap with phosphate interacting residues of N-terminal BRCA1 BRCT as shown in **Figure-7.4**.

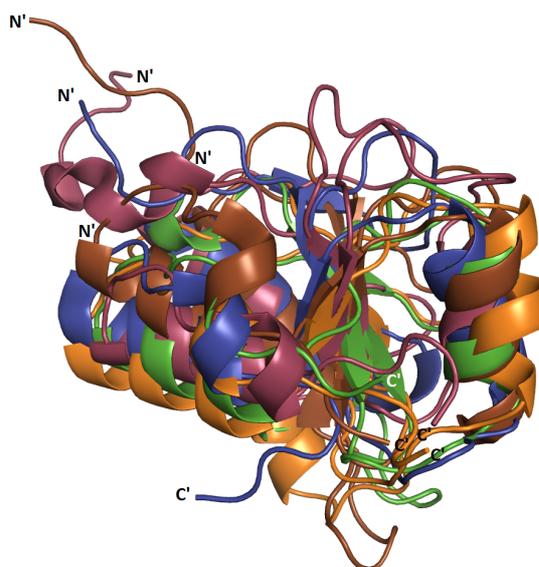


**Figure-7.4:** Pair wise secondary structure alignment of RFC1 BRCT with N-terminal BRCT, where phosphate interacting residues are marked as # and invariant residues are shaded.

**Figure-7.5B** shows the zoomed image of phosphate binding pocket, it is comprised of Thr-415, Gly-416 and Lys-458 which is matching with Ser-1655, Gly-1656 and Lys-1702.



**Figure-7.5:** A-Structural superposition of RFC1 BRCT (PDB ID:-2EBU) (magenta) with N-terminal BRCT (PDB ID:1JNX) (green).B-Zoomed image of superposition showing the phosphate binding residues.

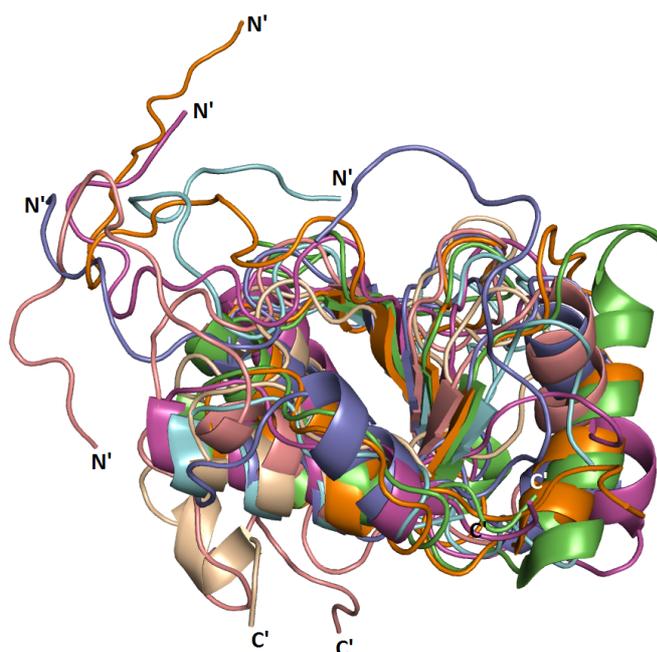


**Figure-7.6:** Superposition of different crystal structures of single BRCT repeat containing proteins, such as Top BPI (PDBID-3JVE) shown blue, MCPH1 (PDBID-2WT8) shown in brown, DNA pol $\lambda$  (PDBID- 2JW5) shown in red, XRCC1 (PDBID- 1CDZ) shown in orange, while the FCP1 BRCT repeat (PDBID-3EF0) is shown in green.

**Table-7.2:** Comparison of BRCT domains contained protein in different crystal structures, rmsd values are given in Å.

Five NMR structures were superposed and rmsd varies from 1.6-3.0 Å.

	Top BP1	MCPH1	DNA pol $\lambda$	XRCC1
FCP1	1.3	1.9	2.4	1.9
XRCC1	1.8	1.9	2.3	
DNA pol $\lambda$	2.2	2.2		
MCPH1	1.7			



**Figure-7.7:** Superposition of different crystal structures of single BRCT repeat containing proteins *Lig3* (PDBID-1I1N1) shown wheat color, *TDT* (PDBID-2COE) shown in orange, *PARP1* (PDBID- 2COK) shown in blue, *RFC* (PDBID- 2EBU) shown in brown, *REV1* (PDBID-2EBW) shown in red, *PES1* (PDBID-2EP8) shown in cyan and *DNA pol* (PDBID- 2HTF) shown in green.

However, the phosphate group interacting residues (Thr-415 and Gly-416) are not conserved across all the members of single BRCT domain containing proteins (**Figure-**

7.2). This observation may be suggesting that single BRCT-domain-containing proteins have not converged to a single mode of phosphate binding.

**Table-7.3:** RMSD in Å upon comparison of solution structures of different single BRCT domain containing

	1IN1	2COE	2COK	2EBU	2EBW	2EP8
2HTF	2.4	2.0	3.0	2.4	2.7	2.2
2EP8	2.0	2.2	2.3	1.9	1.6	
2EBW	2.4	2.3	2.3	2.6		
EBU	2.0	2.5	1.8			
2COK	2.1	2.6				
2COE	2.4					

### 7.3.1.2. Tandem BRCT domain

The RMSD values obtained on the superposition of the structures are listed in **Tables-7.4-**

**7.6.** While in Table-7.4, the C- $\alpha$  atom from both the BRCT domains are superposed on

**Table-7.4:** RMSD values expressed in Å upon comparison of two different BRCT domain containing proteins. In parenthesis, the percentage identities of superposed amino acids is given

	Top BRCT(7/8)	Brc2 BRCT	BARD1 BRCT	MDC1 BRCT
BRCA1 BRCT	3.4 (19)	3.1 (14)	2.2 (18)	2.9 (18)
MDC1 BRCT	2.9 (14)	2.7 (28)	2.4 (10)	
BARD1 BRCT	2.8 (20)	3.2 (15)		
Brc2 BRCT	3.8 (13)			

one another, the superposition for **Table-7.5** and **Table-7.6** are respectively for either the N-terminal and C-terminal BRCT domains. The identity of residues derived from structure based sequence alignment is also given in **Tables-7.4-7.6**.

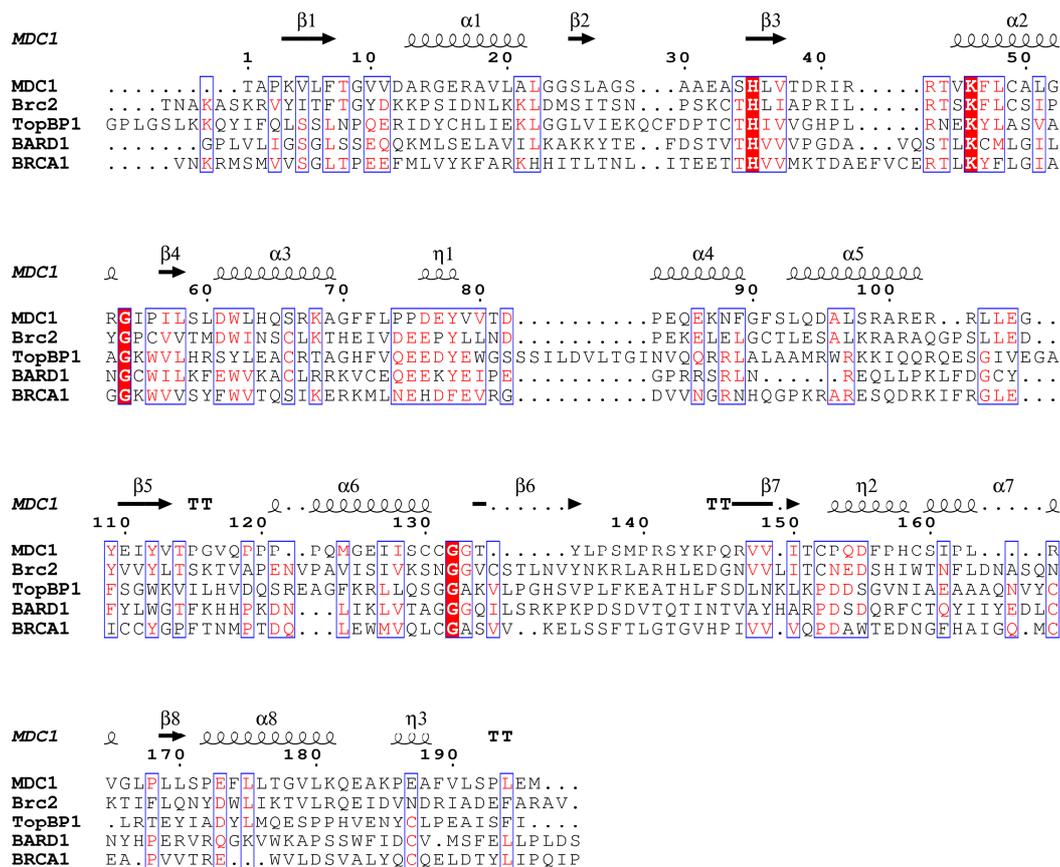
**Table-7.5:** The structural superposition of N-terminal BRCT domains from BRCA1, MDC1, BARD1, TopBp1 and Brc2 protein. RMSD values in Å and percent identity given in parenthesis.

	BRCA1-N	MDC1-N	BARD1-N	TopBP1-N
Brc2-N	2.5 (17)	1.7 (31)	1.7 (20)	2.6 (19)
TopBp1-N	2.9 (22)	2.1 (18)	1.5 (26)	
BARD1-N	1.1 (25)	1.7 (14)		
MDC1-N	2.6 (18)			

**Table-7.6:** The structural superposition of C-terminal BRCT domains from BRCA1, MDC1, BARD1, TopBp1 and Brc2, protein. RMSD values in Å and percent identity given in parenthesis.

	BRCA1-C	MDC1-C	BARD1-C	TopBP1-C
Brc2-C	3.2 (7)	2.3 (25)	3.5 (18)	2.8 (9)
TopBp1-C	2.4 (17)	2.3 (10)	2.3 (15)	
BARD1-C	2.7 (17)	2.5 (15)		
MDC1-C	2.5 (18)			

It is noteworthy that the N-terminal BRCT domain can structurally superpose better compared to C-terminal BRCT domain. Interestingly, the sequence identity averaged over the N-terminal domains is 21% compared to C-terminal domains (15%). On the other hand the average rmsd values, when single N and C terminal BRCT domain were included in the superposition, are significantly different at 2.0 Å and 2.6 Å respectively.

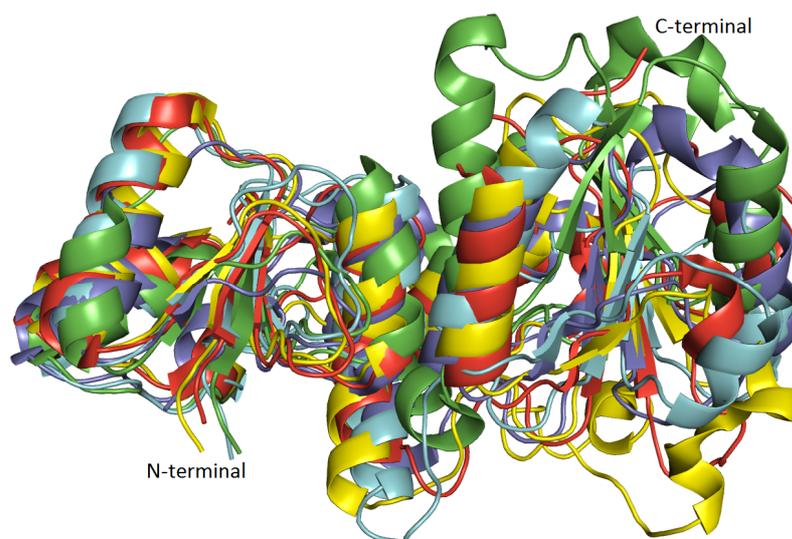


**Figure-7.8: Multiple sequence alignment of tandem BRCT domain containing proteins using Multalin software.**

**Figure-7.8** shows the sequence alignment derived from the structural superposition of double BRCT domain containing proteins. Interestingly, there are four residues which are invariant (His-1686, Lys-1702, Gly-1710 and Gly-1788). One of these is Lys-1702, which is found to form a strong hydrogen bond with the phosphate group in the crystal structures of BRCT-domain/ phosphopeptide complexes [102, 110, 112, 113]. On the other hand the residues Ser-1655 and Gly-1656, which also form hydrogen bonds with phosphate group, are not conserved. These observations may imply that majority of free energy of interaction is derived from hydrogen bonding to Lys-1702 residue. It may be recalled that such a conserved Lys residue was not observed in the case of single BRCT domain

structures. This may be of significance to the evolution of BRCT domain for phosphopeptide binding [293].

Although there is very weak sequence similarity between different BRCT domain containing proteins their overall fold is very similar (**Figure-7.9**).



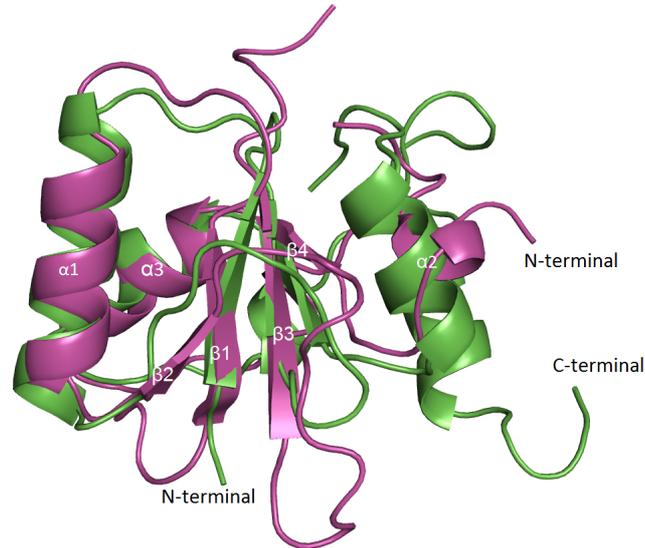
**Figure-7.9: Structural superposition of BRCT domain containing proteins, BRCA1 BRCT (1JNX) shown in red, BARD1 BRCT (2NTE) shown in yellow, MDC1 BRCT (2ADO) shown in blue, TopBP1 BRCT (3AL2) shown in cyan, Brc2 BRCT (3LA0) shown in green.**

The structural superposition of these structures yields rmsd values ranging from 2.3-3.0Å, which is higher than that obtained for single BRCT domain. When only the N-terminal BRCT domains were superposed the RMSD obtained varied from 1.1-2.9 Å. Similarly, when C-terminal BRCT domains were superposed the RMSD obtained varies from 2.3-3.2Å.

#### 7.3.1.2.1. BRCA1 BRCT

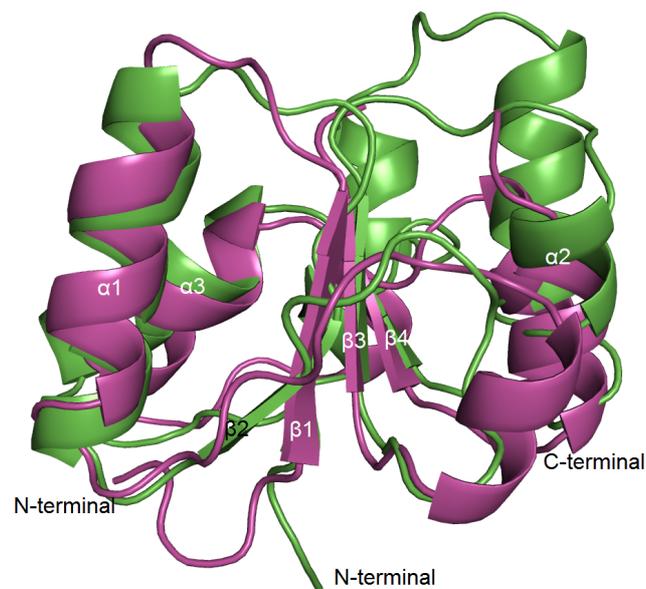
The two BRCT domains are often described as BRCT repeats, and we have therefore examined their similarity both in terms of sequence and structure. The residues 1646-1759 are considered to be part of N-terminal BRCT, and the residues 1760-1859 were taken to belong to C-terminal BRCT domain. The RMSD values obtained on

superposition of these two domains are shown in **Table-7.7**. **Figures-7.10.A-E** show the overlay of the two BRCT domains from five BRCT domain containing proteins.

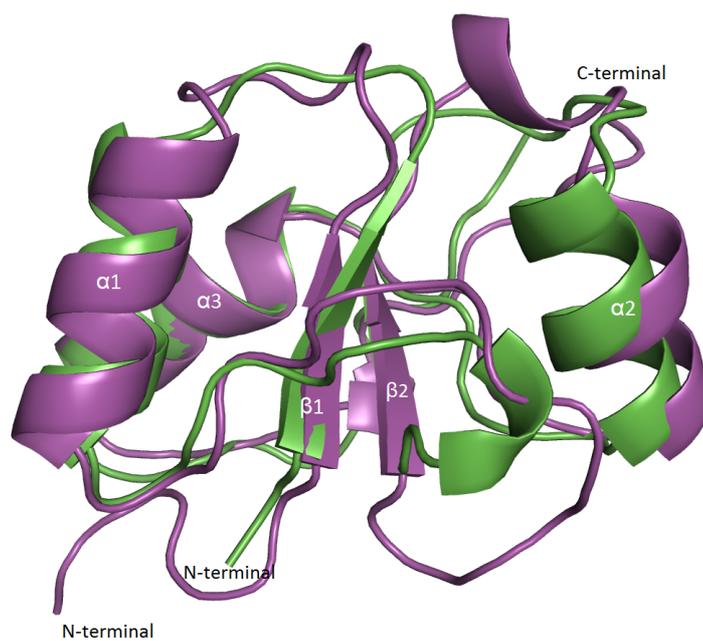


**Figure-7.10A: Structural superposition of BRCT domains of BRCA1**, where N-terminal BRCT domain shown in green while C-terminal BRCT shown in red of BRCA1 protein (PDB ID: 1JNX).

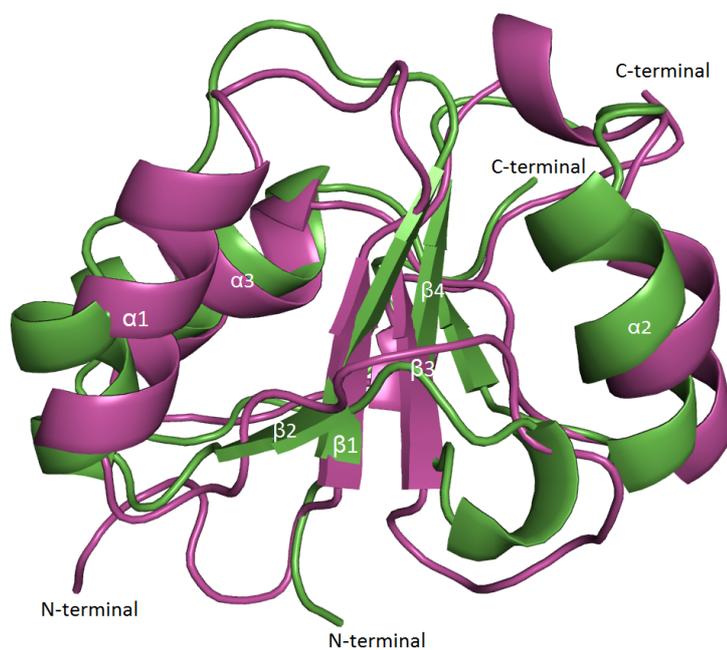
It is interesting that the central  $\beta$  sheet does not superpose well, while  $\alpha$ -helices excepting  $\alpha_2$  superpose better.



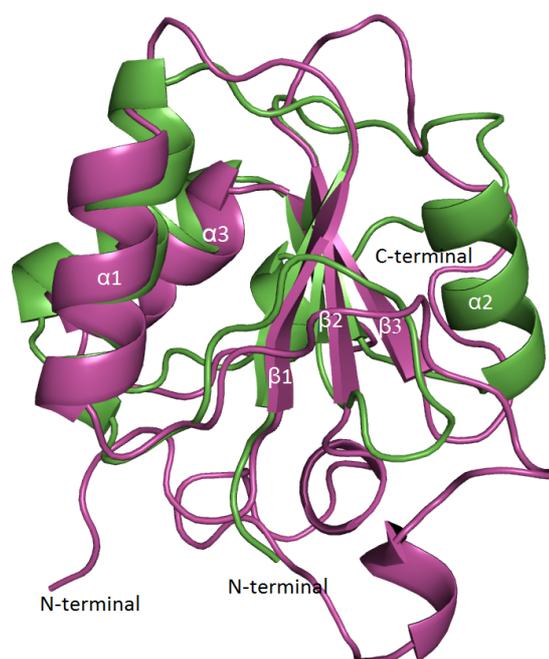
**Figure-7.10B: Structural superposition of BRCT domains of TopBp1 protein**, where, 7 BRCT domain shown in brown while 8 BRCT domain shown in blue (PDB ID: 3AL2).



**Figure-7.10C:** The structural superposition of MDC1 N-terminal BRCT shown in green with C-terminal BRCT shown in magenta (PDB ID: 2ADO).



**Figure-7.10D:** The structural superposition of N-terminal BRCT of Brc1 shown in green colour with C-terminal BRCT shown in red (PDB ID: 3L40).



**Figure-7.10E:** The structural superposition N-terminal BRCT of BARD1 shown in red with C-terminal BRCT shown in cyan (PDB ID: 2NTE).

**Table-7.7:** The sequence similarity and rmsd between N-terminal and C-terminal BRCT repeat for respective proteins.

Sr. No	BRCT domain example	Sequence similarity between N and C terminal BRCT repeat (%)	RMSD between after superposition of N terminal and C-terminal BRCT repeat (Å)
1	BRCA1	13	2.7
2	MDC1	21	2.9
3	Brc1	14	2.3
4	TopBP1	14	2.2
5	BARD1	8	2.0

It may be pointed out that the average rmsd of 2.5Å is higher than the average rmsd obtained when single domain of BRCT's were superposed. This may be additional evidence to support evolution of the BRCT domain via gene duplication and divergence to acquire ability to bind with different ligands.

### BRCT domain and ligand binding

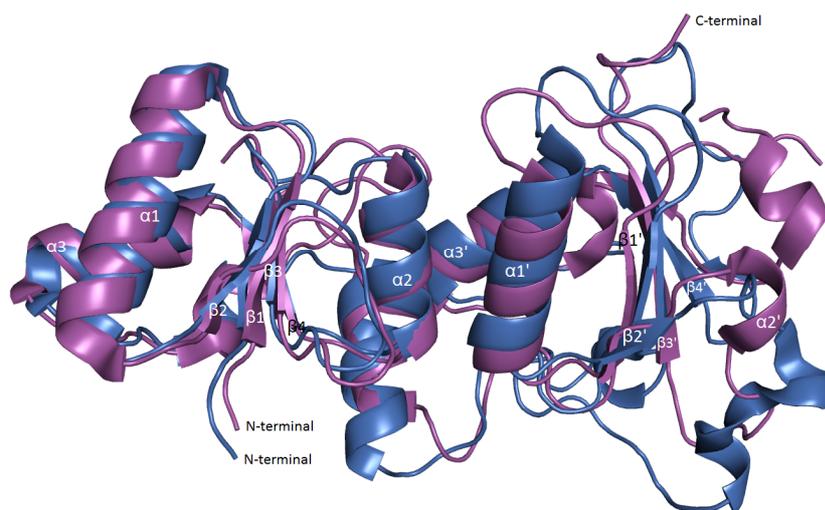
The double BRCT domain is recognised as a module for phosphopeptide recognition.

**Table-7.8** shows the list of BRCT domain containing proteins with their preferred phosphopeptide consensus motif, where X denotes any amino acid residue.

**Table-7.8:** List of BRCT domain containing proteins with preferred consensus sequence of interaction.\*Indicate the lack of structural report, just the biophysical analysis report the binding.

Sr. No.	BRCT domain containing proteins	Consensus motif	PDBID
1	BRCA1 BRCT (1646-777)	pS-X-X-F	1T29, 1T15, 1Y98, 3COJ,
2	MDC1 BRCT (1891-2086)	pS-X-X-Y	2AZM
3	Top bp1 7/8 (1259-1493)	pT-X-X-X	3AL3
4	BARD1 BRCT (568-777)	pS-X-X-F* (pS-D-D-E)	-
5	<i>S.pombe</i> Brcl	pS-X-X-X	3L41
6	<i>S.pombe</i> Crb2	pS-X-X-X	2VXC

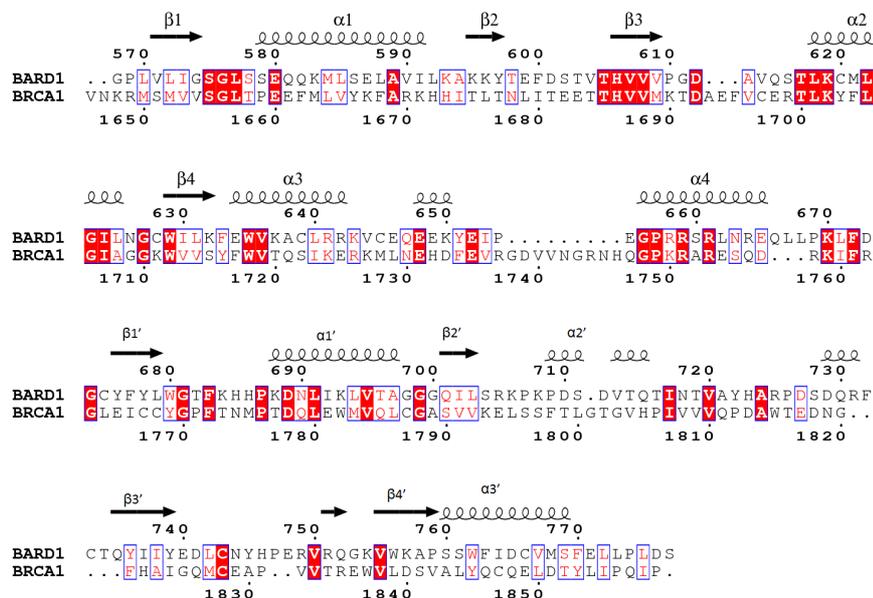
Two different classes of phosphate binding pockets have been identified as of now. They are either pSer binding pocket present in the BRCA1 BRCT and MDC1 BRCT or pThr binding pocket present in the Top BP1 BRCT.



**Figure-7.11:** The structural superposition of BARD1 BRCT (PDBID-2NTE) shown in blue color with BRCA1 BRCT domain (PDBID-1JNX) shown in green color.

With the position of the phosphorylated residue labelled as “0”, the residue at the +3 position towards the carboxyl end of the phosphopeptide is found to determine the specificity of binding. While in most phosphopeptides recognized by the double BRCT domain, this residue is hydrophobic (F/Y): for some it is predicted to be a polar residue. For example BARD1 BRCT domain is predicted to bind phosphopeptide having glutamic acid residue at the +3 position. **Figure-7.11** shows the three dimensional superposition of BRCA1 BRCT with BARD1 BRCT domain. These two structures are very similar in the overall fold.

Sequence alignment (**Figure-7.12**) shows that while the phosphate binding residues Ser-1655, Gly-1656 and Lys-1702 are conserved, there are substantial differences at Phe (+3) binding site of BARD1 BRCT. The Arg-1699 hydrogen bonding with the carbonyl oxygen of Phe (+3) is changed to a much smaller residue, Ser. Similarly the large hydrophobic residue Leu at 1839 position is changed to Ala.



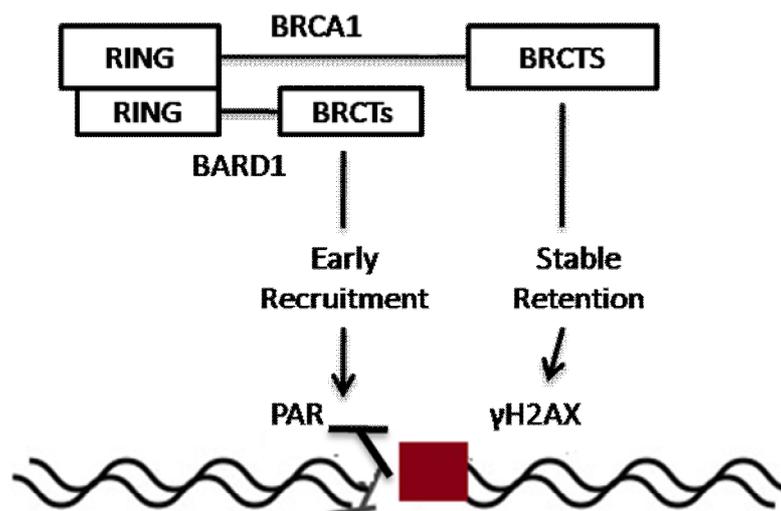
**Figure-7.12: Pair wise alignment of BARD1 BRCT and BRCA1 BRCT domain.**

These changes could rationalise the absence of binding preference to Phe (+3) containing proteins to BARD1. More recently the binding partner of BARD1 BRCT and BRCA1 BRCT are shown to be different [296]. BARD1 BRCT binds to poly ADP polymerase (PAR) very tightly, but BRCA1 BRCT does not bind to PAR. To further investigate binding specificity of BARD1 BRCT, we have attempted to express the BARD1 protein in bacterial system. BARD1 comprises 777 amino acids, and the domain organization is shown in the **Figure-7.13**.



**Figure-7.13 Functional domains BARD1 protein.** *BARD1* comprised of three distinct domains i.e. RING finger domain at the N-terminal, central ankyrin repeat domain and at c-terminal two BRCT domains.

BARD1 recruits BRCA1 to the DNA damage site as per the model shown in the **Figure-7.14**.

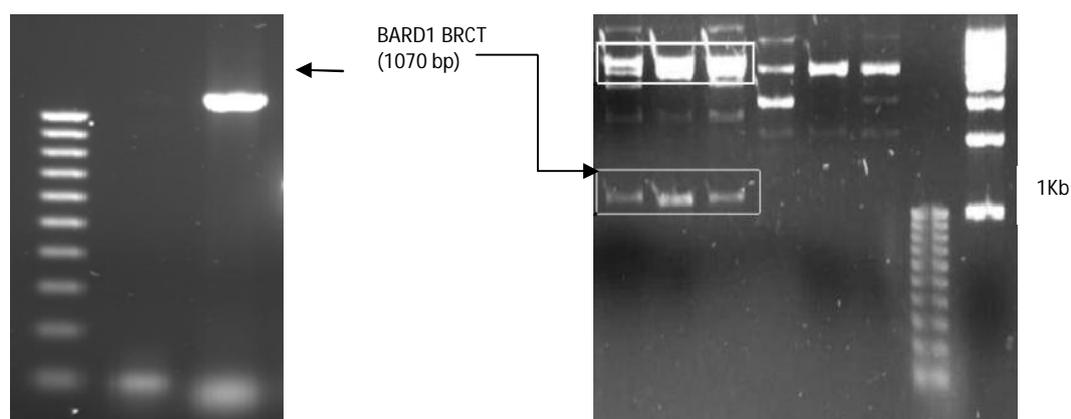


**Figure-7.14: Schematic representation of DNA repair function of BRCA1 and BARD1** BRCA1 and BARD1 forms heterodimer, and BRCT domain of BARD1 interacts with PAR at double stranded DNA breaks and recruits BRCA1- BARD1 complex rapidly to damage site [296].

We have initiated a program to gain insight into BARD1 interaction with its partners. As a start in this direction we have cloned BARD1-ARD-BRCT domain in pGEX-KT vector as described below.

### 7.3.2. Cloning of BARD1 ARD-BRCT in pGEX-KT

The *BARD1* BRCT domain (425-777) with Ankyrin Repeat domain (ARD)[297] was PCR amplified from full length cDNA of human-*BARD1*. **Figure-7.15A** shows the BARD1 ARD-BRCT domain PCR product of ~1070 bp. The potential clones were screened by restriction digestion method. **Figure-7.15B** shows the insert release corresponding to BARD1 ARD-BRCT domain. The clone was further confirmed by DNA sequencing.

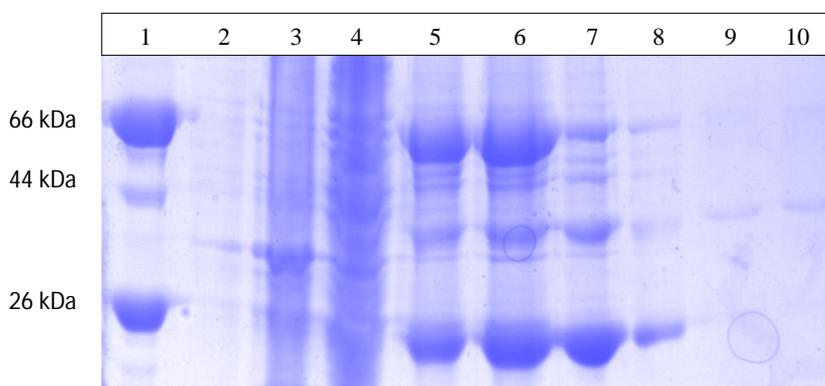


**Figure-7.15: A) PCR amplification of BARD1 ARD-BRCT domain, B) Restriction digestion of potential clone of BARD1 BRCT.**

### 7.3.3. Expression and purification of BARD1 BRCT-ARD

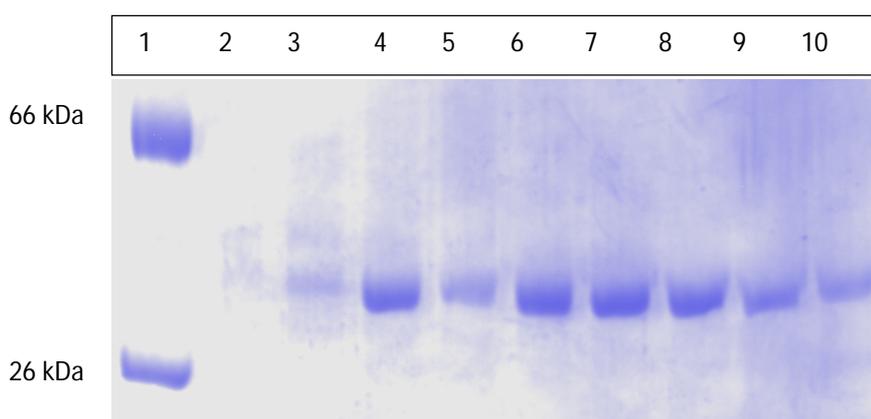
The BARD1 BRCT-ARD domain was expressed and purified using affinity chromatography. **Figure-7.16** shows the expression profile of BARD1 BRCT-ARD. The BARD1 ARD-BRCT domain was cleaved from GST using TEV protease. FPLC was

used for final purification; the gel filtration pattern indicates that BARD1 ARD-BRCT exists as a monomer in solution.



**Figure-7.16: Protein expression profile of BARD1 ARD-BRCT Protein samples used to monitor the protein purification process is listed below, (Lane 1: indicate the molecular weight ladder, Lane 2: Uninduced Whole Cell (UWC)- The cells which does not induced by IPTG, Lane 3: Induced Whole Cell (IWC) - The BRCT over expressing bacterial cells induced with 0.3 mM IPTG and grown at 24°C, Lane 4: Induced Soluble Fraction (ISF) - The protein cell lysate sample taken after sonication and centrifugation of BRCT over expressing bacterial cells, Lane 5, 6: Beads bound (B/B) – The sample of 10µl GST beads, taken after passing ISF and washing with lysis buffer, Lane 7, 8: Beads after TEV cleavage, Lane 9,10: Cleavage sample (CL) – The protein sample got after TEV on beads cleavage).**

**Figure-7.17** indicates more than 95% homogeneity of the FPLC purified BARD1 ARD-BRCT protein which is further concentrated till 25mg/ml and used for crystallization trial.



**Figure-7.17: SDS-PAGE profile of BARD1 ARD-BRCT domain FPLC fractions. Lane 1: Molecular weight ladder, Lane 2-10: FPLC fractions of BARD1 ARD-BRCT**

#### **7.3.4. Crystallization of BARD1 ARD-BRCT**

We have attempted to crystallize BARD1 ARD-BRCT using Hampton research crystal screen 1 and crystal screen 2, by the sitting drop vapour diffusion method at 22°C and also at 4°C. The 2 µl crystallization drop consists of 1 µl protein solution and 1 µl reservoir solution. Single crystals however, have not been obtained so far. Attempts are being made to alter the construct and to co-crystallize with different ligands.

#### **7.4. Conclusion**

BRCT domain is approximately 85-95 amino acids long and is made up of four stranded parallel  $\beta$  sheet and three  $\alpha$  helices located on either side of the  $\beta$  sheet. BRCT domain is present either as a single unit or as units in tandem within a single protein. The sequence alignment of single BRCT domains derived from structural superposition yields three new conserved regions. These regions contain amino acids that make hydrogen bonding and hydrophobic interactions to confer structural stability to BRCT domain. Single BRCT domain is known to bind with phosphate group, but the phosphate binding residues are not conserved among single BRCT domains. This may indicate that the single BRCT-domain-containing proteins have not evolutionarily converged to a single mode of phosphate binding. Double BRCT domain is known to act as phosphopeptide recognition module. Sequence alignment obtained from structural superposition of these proteins shows conservation of Lys-1702 and partial conservation of Ser-1655 and Gly-1656 as phosphate interacting residues. When N-terminal domain was structurally compared with C-terminal BRCT domain from the tandem BRCT repeat containing proteins, the average rmsd is higher than that for single BRCT domain. This observation may be supportive of evolution of the tandem repeats of BRCT domain via gene duplication and divergence to acquire ability to bind with different ligands. The BARD1 ARD-BRCT domain was expressed in BL21-DE3 cells, and was purified using affinity chromatography. Initial

crystallization attempts have not yet yielded single crystals. Attempts are being made to alter the construct and to co-crystallize with different ligands.

*Summary and future directions*

---

Macromolecular crystallography is the only technique that enables scientists to determine the atomic level structures of biomolecules of any size. The high resolution atomic structure can elucidate structure function relationships in biologically important proteins. The major stumbling block, however, is the requirement of single crystals of the sample. Recent developments of robots to conduct crystallization experiments has enabled thorough and quick screening for crystallization conditions, using limited amount of protein sample. The advancement in recombinant DNA technology, which can produce large amounts of any protein of interest in a short period of time, has been of major help to the field of protein crystallography. This branch of science has provided a very good momentum in determining the relationships between protein structures and their clinical applications. Easy access to the synchrotron source of x-rays can allow scientists to obtain the diffraction data from small and even weakly diffracting crystals. Synchrotron facility has radically transformed the post-genomic era for protein target identification. Three-dimensional structures of proteins have long been recognised as important for structure guided drug design. Many drugs have been designed using structure based inhibitor design approach [298-300]. High throughput biophysical studies of binding affinity between the target and drug have contributed in drug discovery. Circular dichroism and fluorescence spectroscopy have assisted in the analysis of the protein structure. Isothermal Titration Calorimetry (ITC) [301] and Surface Plasmon Resonance (SPR) [302] have provided alternative tools for the study of protein-protein interactions (PPIs) for drug discovery. Hence, structural biology, biophysics and bioinformatics based interdisciplinary approaches have applications in translational research. Considering the importance of the subject and noting the growing number of breast cancer cases in India, I decided to explore the structure of the functional domains of the breast cancer gene product BRCA1, specifically the BRCT domain. BRCT domains have long been

documented as protein-protein interaction modules. The BRCA1 BRCT domain is essential for BRCA1 to perform tumor suppressor functions, as it is found to be mutated in breast and ovarian cancers. The protein-protein interactions are mediated through recognition, by BRCT domain, of a stretch of amino acid residues phosphorylated at one or two residues. The singly phosphorylated binding partners are NCoA2, Nup153 and the RNA binding domain, whereas Abraxas is doubly phosphorylated; each binding domain has different functions in association with BRCT. The biophysical characterization and interaction analysis for BRCA1 BRCT phosphopeptide complexes were carried out using ITC analysis and the x-ray diffraction method. It has been determined that NCoA2, Nup153, RNA binding domain, and Abraxas (singly phospho-A1 as well as doubly phosphorylated-A2) peptides have different binding affinities to BRCA1 BRCT. The crystal structures of BRCA1 BRCT complexed with NCoA2, Abraxas A1 peptide and the Abraxas A2 peptide have been determined. This is the first crystal structure of BRCT domain complexed with a doubly phosphorylated peptide. We have analysed some of the mis-sense mutations discovered in the BRCT domain of BRCA1. We have also structurally aligned different single and tandem BRCT domain containing proteins to shed some light on their function and evolution.

The interaction of singly phosphorylated oligopeptide (NH<sub>2</sub>-<sub>190</sub>-PPRRNSHTFNC-<sub>200</sub>-COOH) from NCoA2 with BRCA1 BRCT has been studied. The binding analysis using ITC predicts stronger binding compared to ATRIP peptide [110] and weaker binding compared to earlier reported complexes [101, 102, 110, 113]. The crystal structure of the complex has been determined to 1.7 Å resolution. In the BRCA1 BRCT-NCoA2 complex crystal, the oligopeptide binds to the BRCT domain in a “two knob” manner, spanning both the BRCT repeats. The pSer (0) residue from the peptide interacts with the N-terminal BRCT and the Phe (+3) residue interacts with the C-terminal BRCT. The pSer

residue is stabilized by three hydrogen bonds formed by Ser-1655, Gly-1656 and Lys-1702 residues, whereas Phe +3 residue is buried in the hydrophobic core formed by the N-terminal BRCT residues Asn-1774, Met- 1775, Val -1740, Glu -1698, and Thr-1700 residue of the C-terminal BRCT. The residues Asn and His, at -2 and +1 positions, form water mediated hydrogen bonds with N-atoms of Leu-1657 and Leu-1701 respectively. The Cysteine at +5 position is forming two hydrogen bonds with water molecule-16 and Glu-1836. Here we report for the first time the structure of BRCA1 BRCT complexed with phosphopeptide having His rather than Pro residue at +1 position.

An attempt has been made to determine crystal structures of BRCA1 BRCT complexed with oligopeptides from the Nup153 (NH<sub>2</sub><sup>-(1299)</sup>-SAGSSFVFGT<sup>-(1308)</sup>-COOH) and RNA binding domain, RBP-12 (NH<sub>2</sub><sup>-(164)</sup>-ASFGSPTFSS<sup>-(173)</sup>-COOH). Nup 153 is the nuclear core complex protein and consensus sequence for BRCT domain binding is present in the C-terminal FG-rich region. RNA binding domain (RBP-12) is required for the transcription activation process and consensus sequence for BRCT domain binding is present in the C-terminal proline rich region. Nup153 and RBP-12 bind to BRCA1 BRCT domain with affinities of 0.1 $\mu$ M and 5.1 $\mu$ M respectively. Single crystals of the two complexes have been obtained under similar condition; however the diffraction from the crystals was very poor, extending to about 6Å resolution. In an aggressive attempt to obtain better diffraction, we have screened the series of crystals on home source as well as on the synchrotron using the remote data collection facility. However, we could not, so far obtain good quality/high resolution data.

BRCA1 A complex consists of RAP80, BRCC35, BRCC36, RNF8 and Abraxas proteins, and Abraxas acts as a scaffold protein playing a key role in the architecture of this A complex [73, 264, 274]. Upon DNA damage Abraxas is known to be phosphorylated at

Ser-404 and Ser-406 positions by ATM and ATR kinase, and the doubly phosphorylated abraxas is selectively enriched in the BRCA1 A complex recruited to the damage site [73]. To unravel the molecular complexity associated with the BRCA1-A complex, we have characterised the interactions of differently phosphorylated peptides NH<sub>2</sub>-(<sup>399</sup>)-G-F-G-E-Y-S-R-pS<sub>406</sub>-P-T-F-(<sup>409</sup>)-COOH (A1) and NH<sub>2</sub>-(<sup>399</sup>)-G-F-G-E-Y-pS<sub>404</sub>-R-pS<sub>406</sub>-P-T-F-(<sup>409</sup>)-COOH (A2) with BRCA1 BRCT. In our experiments, it has been observed that the doubly phosphorylated abraxas peptide binds more strongly to the BRCA1 BRCT domain (0.2 μM affinity) compared to singly phosphorylated peptide (1.2 μM affinity). We have co-crystallized both the complexes and determined their crystal structures. The crystals of BRCA1 BRCT complexes with A1 (single phosphorylated) and A2 (doubly phosphorylated) Abraxas peptides diffracted to around 3.8 Å resolution, and the structures were solved by molecular replacement method. The A1 complex structure is similar to earlier reported complexes, where pSer-406 binds to N-terminal BRCT and Phe-409 interacts with hydrophobic residues from the interface region. In A2 complex structure the pSer-404 rather than the expected pSer-406, occupies the P1-phosphate binding pocket of BRCT domain. The pSer-404 forms two hydrogen bonds with Ser1655 and Gly1656. However, it does not form hydrogen bond with Lys-1702. Further, Phe (+3) residue does not bind at the interface region of N and C-terminal BRCT domains. Thus, the doubly phosphorylated phosphopeptide binds entirely to the N-terminal BRCT domain.

### **BRCA1 BRCT domain mis-sense variants**

The *BRCA1* is a breast cancer tumor suppressor gene and is found to be frequently mutated in breast and ovarian cancer. Most of the breast cancer mutations have been reported in the BIC database and a few BRCA1 BRCT mis-sense mutations BRCA1 H1686Q, P1749R, S1715R and C1697R have been selected for study here. These mutant

proteins have been cloned and expressed for structural studies. Unfortunately, they are insoluble and difficult to purify, except for the BRCA1 H1686Q mutant. BRCA1 H1686Q variant has been successfully expressed and purified by ion exchange chromatography. The CD spectrum of the mutant suggests that there slight alterations in the secondary structures of BRCA1 H1686Q variant, when compared to wild type, even though H1686Q mutation has been reported as a pathogenic mutation.

### **Other BRCT domain containing proteins**

The BRCT domain is present as a single copy in some proteins (XRCC1, DNA ligase III, PARP1, MCPH1 etc.), and as tandem repeats in others (BRCA1, BARD1, MDC1, 53BP1 etc). The single copy proteins bind to phosphate groups, while the tandem repeat protein binds to phosphopeptide. We have structurally aligned BRCT domain present in a protein as a single copy to understand the phosphate binding determinants. We have found that hydrogen bonding with the Lys-1702 is not a conserved interaction in these proteins. We have further compared the N and C-terminal BRCT domains from tandem proteins. We find that the structural variation among these two domains is larger than that between BRCT domain in single copy containing proteins. This indicates that, the two domains of BRCA1 have evolved by gene duplication and divergence. The C-terminal domain has diverged to acquire peptide binding capability. The Lysine residue conserved in proteins possessing single BRCT domain is absent in C-terminal BRCT domain. Although tandem BRCT domain of BARD1 is structurally very similar to BRCT domain of BRCA1, the efforts to find out the binding partner of BARD1 BRCT have been futile so far. We have cloned the ARD and BRCT domain (425-777) of BARD1, and have purified it using affinity chromatography. We have attempted to crystallize BARD1 ARD-BRCT domain, but these attempts were unsuccessful. The probable reason may be the flexibility between the ARD and BRCT domains.

**Future Directions**

To my best abilities, I have compiled several BRCA1 BRCT oligopeptide complexes and have determined their crystal structures. The future goals will be:

- (1) to obtain better diffraction quality crystals for all the singly phosphorylated complexes to unravel the exact mode of binding with BRCT domain,
- (2) to study doubly phosphorylated Abraxas at higher resolution, and
- (3) to investigate the role of reported pathogenic mutations by crystallizing the mutant protein.

I am sure the above stated results will help in designing small molecule inhibitors for cancer treatment, and also in the development of a multi-model based approach for mutational analysis and clinical management.



1. Hall, J.M., et al., *Linkage of early-onset familial breast cancer to chromosome 17q21*. Science, 1990. **250**(4988): p. 1684-9.
2. Futreal, P.A., et al., *BRCA1 mutations in primary breast and ovarian carcinomas*. Science, 1994. **266**(5182): p. 120-2.
3. Hu, Y.-F., Z.L. Hao, and R. Li, *Chromatin remodeling and activation of chromosomal DNA replication by an acidic transcriptional activation domain from BRCA1*. Genes & development, 1999. **13**(6): p. 637-642.
4. Deng, C.-X., *Roles of BRCA1 in centrosome duplication*. Oncogene, 2002. **21**(40): p. 6222-6227.
5. Monteiro, A.N., A. August, and H. Hanafusa, *Evidence for a transcriptional activation function of BRCA1 C-terminal region*. Proceedings of the National Academy of Sciences, 1996. **93**(24): p. 13595-13599.
6. Lokesh, G., et al., *Thermodynamics of phosphopeptide tethering to BRCT: the structural minima for inhibitor design*. Journal of the American Chemical Society, 2007. **129**(35): p. 10658-10659.
7. Brzovic, P.S., et al., *Binding and recognition in the assembly of an active BRCA1/BARD1 ubiquitin-ligase complex*. Proceedings of the National Academy of Sciences, 2003. **100**(10): p. 5646-5651.
8. Roehm, P.C. and J.M. Berg, *Sequential metal binding by the RING finger domain of BRCA1*. Biochemistry, 1997. **36**(33): p. 10240-10245.
9. Manke, I.A., et al., *BRCT repeats as phosphopeptide-binding modules involved in protein targeting*. Science Signaling, 2003. **302**(5645): p. 636.
10. Williams, R.S., R. Green, and J.M. Glover, *Crystal structure of the BRCT repeat region from the breast cancer-associated protein BRCA1*. Nature Structural & Molecular Biology, 2001. **8**(10): p. 838-842.
11. Yu, X., et al., *The BRCT domain is a phospho-protein binding domain*. Science Signaling, 2003. **302**(5645): p. 639.
12. Bailey, S., *The CCP4 suite: programs for protein crystallography*. 1993: Daresbury Laboratory.
13. Emsley, P., et al., *Features and development of Coot*. Acta Crystallographica Section D: Biological Crystallography, 2010. **66**(4): p. 486-501.
14. Thompson, J.D., D.G. Higgins, and T.J. Gibson, *CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice*. Nucleic acids research, 1994. **22**(22): p. 4673-4680.
15. Krissinel, E. and K. Henrick, *Inference of macromolecular assemblies from crystalline state*. Journal of molecular biology, 2007. **372**(3): p. 774-797.
16. Lemaitre, C., et al., *The nucleoporin 153, a novel factor in double-strand break repair and DNA damage response*. Oncogene, 2012. **31**(45): p. 4803-4809.
17. Anderson, S.F., et al., *BRCA1 protein is linked to the RNA polymerase II holoenzyme complex via RNA helicase A*. Nature genetics, 1998. **19**(3): p. 254-256.
18. Davis, I.W., et al., *MolProbity: all-atom contacts and structure validation for proteins and nucleic acids*. Nucleic acids research, 2007. **35**(suppl 2): p. W375-W383.
19. Bic, L. and J.P. Gilbert, *Learning from AL: new trends in database technology*. Computer;(United States), 1986. **19**(3).
20. Ashok, K., *Pathogenicity of Mutations Discovered in BRCA1 BRCT Domains is Characterized by Destabilizing the Hydrophobic Interactions*. Journal of Cancer Science & Therapy, 2012.
21. Tavtigian, S.V., et al., *Comprehensive statistical study of 452 BRCA1 missense substitutions with classification of eight recurrent substitutions as neutral*. Journal of medical genetics, 2006. **43**(4): p. 295-305.

22. Adzhubei, I.A., et al., *A method and server for predicting damaging missense mutations*. Nature methods, 2010. **7**(4): p. 248-249.
23. Li, B., et al., *Automated inference of molecular mechanisms of disease from amino acid substitutions*. Bioinformatics, 2009. **25**(21): p. 2744-2750.
24. Birrane, G., et al., *Crystal structure of the BARD1 BRCT domains*. Biochemistry, 2007. **46**(26): p. 7706-7712.
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*. Journal of Biological Chemistry, 2005. **280**(37): p. 32053-32056.
26. Joo, W.S., et al., *Structure of the 53BP1 BRCT region bound to p53 and its comparison to the Brca1 BRCT structure*. Genes & development, 2002. **16**(5): p. 583-593.
27. Schultz, M., et al., *Oldest known case of metastasizing prostate carcinoma diagnosed in the skeleton of a 2,700-year-old Scythian king from Arzhan (Siberia, Russia)*. International Journal of Cancer, 2007. **121**(12): p. 2591-2595.
28. Schlumberger, H.G. and B. Lucké, *Tumors of fishes, amphibians, and reptiles*. Cancer Research, 1948. **8**(12): p. 657-753.
29. Dawe, C.J., *Phyletic Approaches to Cancer: Proceedings of the Symposium, Tokyo, 1980*. 1981: Japan Scientific Societies Press.
30. Hanahan, D. and R.A. Weinberg, *The hallmarks of cancer*. cell, 2000. **100**(1): p. 57-70.
31. Vogelstein, B. and K.W. Kinzler, *The genetic basis of human cancer*. Vol. 821. 2002: McGraw-Hill, Medical Pub. Division New York.
32. Kinzler, K.W. and B. Vogelstein, *Gatekeepers and caretakers*. Nature, 1997. **386**(6627).
33. Weinberg, R.A., *Tumor suppressor genes*. Science, 1991. **254**(5035): p. 1138-1146.
34. Sieber, O.M., K. Heinimann, and I.P. Tomlinson, *Genomic instability—the engine of tumorigenesis?* Nature reviews cancer, 2003. **3**(9): p. 701-708.
35. Jemal, A., et al., *Global cancer statistics*. CA: a cancer journal for clinicians, 2011. **61**(2): p. 69-90.
36. Futreal, P.A., et al., *BRCA1 mutations in primary breast and ovarian carcinomas*. Science, 1994. **266**(5182): p. 120-122.
37. Miki, Y., et al., *A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1*. Science, 1994. **266**(5182): p. 66-71.
38. Wooster, R., et al., *Localization of a breast cancer susceptibility gene, BRCA2, to chromosome 13q12-13*. Science, 1994. **265**(5181): p. 2088-2090.
39. Datta, S.R., A. Brunet, and M.E. Greenberg, *Cellular survival: a play in three Acts*. Genes & development, 1999. **13**(22): p. 2905-2927.
40. Plo, I., et al., *AKT1 inhibits homologous recombination by inducing cytoplasmic retention of BRCA1 and RAD51*. Cancer research, 2008. **68**(22): p. 9404-9412.
41. Raičević-Maravić, L. and S. Radulović, *Breast cancer susceptibility gene 1 (BRCA 1)*. Archive of Oncology, 2000. **8**(1): p. 21-23.
42. Brody, L.C. and B.B. Biesecker, *Breast cancer susceptibility genes: BRCA1 and BRCA2*. Medicine, 1998. **77**(3): p. 208-226.
43. Solomon, E. and D.H. Ledbetter, *Report of the committee on the genetic constitution of chromosome 17*. Cytogenet Cell Genet, 1990. **55**(1-4): p. 198-215.
44. Rohlfs, E.M., et al., *An Alu-mediated 7.1 kb deletion of BRCA1 exons 8 and 9 in breast and ovarian cancer families that results in alternative splicing of exon 10*. Genes, Chromosomes and Cancer, 2000. **28**(3): p. 300-307.
45. BENNETT, M.L., et al., *Isolation of the mouse homologue of BRCA1 and genetic mapping to mouse chromosome 11*. Genomics, 1995. **29**(3): p. 576-581.
46. Gowen, L.C., et al., *Brca1 deficiency results in early embryonic lethality characterized by neuroepithelial abnormalities*. Nature genetics, 1996. **12**(2): p. 191-194.

47. Hakem, R., et al., *The Tumor Suppressor Gene *Brca1* Is Required for Embryonic Cellular Proliferation in the Mouse*. Cell, 1996. **85**(7): p. 1009-1023.
48. Chapman, M.S. and I.M. Verma, *Transcriptional activation by BRCA1*. Nature, 1996. **382**(6593): p. 678-679.
49. Venkitaraman, A.R., *Breast cancer genes and DNA repair*. Science, 1999. **286**(5442): p. 1100-1102.
50. Deng, C.-X., *BRCA1: cell cycle checkpoint, genetic instability, DNA damage response and cancer evolution*. Nucleic acids research, 2006. **34**(5): p. 1416-1426.
51. Andrews, H.N., et al., *BRCA1 regulates the interferon gamma-mediated apoptotic response*. J Biol Chem, 2002. **277**(29): p. 26225-32.
52. Powell, S.N. and L.A. Kachnic, *Roles of BRCA1 and BRCA2 in homologous recombination, DNA replication fidelity and the cellular response to ionizing radiation*. Oncogene, 2003. **22**(37): p. 5784-5791.
53. Zhang, J. and S.N. Powell, *The role of the BRCA1 tumor suppressor in DNA double-strand break repair*. Molecular cancer research, 2005. **3**(10): p. 531-539.
54. Sims, R.J., 3rd, R. Belotserkovskaya, and D. Reinberg, *Elongation by RNA polymerase II: the short and long of it*. Genes Dev, 2004. **18**(20): p. 2437-68.
55. Chapman, M.S. and I.M. Verma, *Transcriptional activation by BRCA1*. Nature, 1996. **382**(6593): p. 678-9.
56. Gowen, L.C., et al., *BRCA1 required for transcription-coupled repair of oxidative DNA damage*. Science, 1998. **281**(5379): p. 1009-12.
57. Anderson, S.F., et al., *BRCA1 protein is linked to the RNA polymerase II holoenzyme complex via RNA helicase A*. Nat Genet, 1998. **19**(3): p. 254-6.
58. Zhang, H., et al., *BRCA1 physically associates with p53 and stimulates its transcriptional activity*. Oncogene, 1998. **16**(13): p. 1713-21.
59. MacLachlan, T.K., R. Takimoto, and W.S. El-Deiry, *BRCA1 directs a selective p53-dependent transcriptional response towards growth arrest and DNA repair targets*. Molecular and cellular biology, 2002. **22**(12): p. 4280-4292.
60. Meloni, A.R., E.J. Smith, and J.R. Nevins, *A mechanism for Rb/p130-mediated transcription repression involving recruitment of the CtBP corepressor*. Proceedings of the National Academy of Sciences, 1999. **96**(17): p. 9574-9579.
61. Yu, X., et al., *The C-terminal (BRCT) domains of BRCA1 interact in vivo with CtIP, a protein implicated in the CtBP pathway of transcriptional repression*. J Biol Chem, 1998. **273**(39): p. 25388-92.
62. Scully, R., et al., *BRCA1 is a component of the RNA polymerase II holoenzyme*. Proc Natl Acad Sci U S A, 1997. **94**(11): p. 5605-10.
63. Krum, S.A., et al., *BRCA1 associates with processive RNA polymerase II*. J Biol Chem, 2003. **278**(52): p. 52012-20.
64. Halle, D.T. and J.D. Parvin, *Activation of transcription in vitro by the BRCA1 carboxyl-terminal domain*. J Biol Chem, 1999. **274**(4): p. 2113-7.
65. Parvin, J.D., *BRCA1 at a branch point*. Proc Natl Acad Sci U S A, 2001. **98**(11): p. 5952-4.
66. Hanawalt, P.C., *Transcription-coupled repair and human disease*. Science, 1994. **266**(5193): p. 1957-8.
67. Scully, R., et al., *Association of BRCA1 with Rad51 in mitotic and meiotic cells*. Cell, 1997. **88**(2): p. 265-275.
68. Sung, P., *Catalysis of ATP-dependent homologous DNA pairing and strand exchange by yeast RAD51 protein*. Science, 1994. **265**(5176): p. 1241-3.
69. Baumann, P., F.E. Benson, and S.C. West, *Human Rad51 protein promotes ATP-dependent homologous pairing and strand transfer reactions in vitro*. Cell, 1996. **87**(4): p. 757-66.

70. Tauchi, H., et al., *Nbs1 is essential for DNA repair by homologous recombination in higher vertebrate cells*. Nature, 2002. **420**(6911): p. 93-98.
71. Huen, M.S., S.M. Sy, and J. Chen, *BRCA1 and its toolbox for the maintenance of genome integrity*. Nature reviews Molecular cell biology, 2009. **11**(2): p. 138-148.
72. 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.
73. Wang, B., et al., *Abraxas and RAP80 form a BRCA1 protein complex required for the DNA damage response*. Science, 2007. **316**(5828): p. 1194-1198.
74. 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.
75. Luo, R.X. and D.C. Dean, *Chromatin remodeling and transcriptional regulation*. Journal of the National Cancer Institute, 1999. **91**(15): p. 1288-1294.
76. Neish, A.S., et al., *Factors associated with the mammalian RNA polymerase II holoenzyme*. Nucleic Acids Res, 1998. **26**(3): p. 847-53.
77. Narod, S.A. and W.D. Foulkes, *BRCA1 and BRCA2: 1994 and beyond*. Nature Reviews Cancer, 2004. **4**(9): p. 665-676.
78. Bochar, D.A., et al., *BRCA1 is associated with a human SWI/SNF-related complex: linking chromatin remodeling to breast cancer*. Cell, 2000. **102**(2): p. 257-265.
79. Yarden, R.I. and L.C. Brody, *BRCA1 interacts with components of the histone deacetylase complex*. Proceedings of the National Academy of Sciences, 1999. **96**(9): p. 4983-4988.
80. Kastan, M.B. and J. Bartek, *Cell-cycle checkpoints and cancer*. Nature, 2004. **432**(7015): p. 316-323.
81. Yarden, R.I., et al., *BRCA1 regulates the G2/M checkpoint by activating Chk1 kinase upon DNA damage*. Nature genetics, 2002. **30**(3): p. 285-289.
82. Xu, X., et al., *Centrosome Amplification and a Defective G<sub>2</sub>-M Cell Cycle Checkpoint Induce Genetic Instability in BRCA1 Exon 11 Isoform-Deficient Cells*. Molecular cell, 1999. **3**(3): p. 389-395.
83. Aprelikova, O.N., et al., *BRCA1-associated growth arrest is RB-dependent*. Proceedings of the National Academy of Sciences, 1999. **96**(21): p. 11866-11871.
84. Saurin, A.J., et al., *Does this have a familiar RING?* Trends in biochemical sciences, 1996. **21**(6): p. 208-214.
85. Mark, W.-Y., et al., *Characterization of segments from the central region of BRCA1: an intrinsically disordered scaffold for multiple protein-protein and protein-DNA interactions?* Journal of molecular biology, 2005. **345**(2): p. 275-287.
86. Koonin, E.V., S.F. Altschul, and P. Bork, *BRCA1 protein products ... Functional motifs*. Nat Genet, 1996. **13**(3): p. 266-8.
87. Callebaut, I. and J.-P. Mornon, *From BRCA1 to RAP1: a widespread BRCT module closely associated with DNA repair*. Febs Letters, 1997. **400**(1): p. 25-30.
88. Castilla, L.H., et al., *Mutations in the BRCA1 gene in families with early-onset breast and ovarian cancer*. Nature genetics, 1994. **8**(4): p. 387-391.
89. Wu, L.C., et al., *Identification of a RING protein that can interact in vivo with the BRCA1 gene product*. Nature genetics, 1996. **14**(4): p. 430-440.
90. Jensen, D.E., et al., *BAP1: a novel ubiquitin hydrolase which binds to the BRCA1 RING finger and enhances BRCA1-mediated cell growth suppression*. Oncogene, 1998. **16**(9): p. 1097-1112.
91. Paull, T.T., et al., *Direct DNA binding by Brca1*. Proceedings of the National Academy of Sciences, 2001. **98**(11): p. 6086-6091.
92. Miki, Y., et al., *A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1*. Science, 1994. **266**(5182): p. 66-71.
93. Garcia, V., K. Furuya, and A.M. Carr, *Identification and functional analysis of TopBP1 and its homologs*. DNA repair, 2005. **4**(11): p. 1227-1239.

94. Leung, C.C.Y. and J.M. Glover, *BRCT domains: easy as one, two, three*. Cell Cycle, 2011. **10**(15): p. 2461-2470.
95. Stewart, G.S., et al., *MDC1 is a mediator of the mammalian DNA damage checkpoint*. Nature, 2003. **421**(6926): p. 961-966.
96. Brzovic, P.S., et al., *Structure of a BRCA1–BARD1 heterodimeric RING–RING complex*. Nature Structural & Molecular Biology, 2001. **8**(10): p. 833-837.
97. Hashizume, R., et al., *The RING heterodimer BRCA1-BARD1 is a ubiquitin ligase inactivated by a breast cancer-derived mutation*. Journal of Biological Chemistry, 2001. **276**(18): p. 14537-14540.
98. Bellon, S.F., et al., *Crystal structure of the RAG1 dimerization domain reveals multiple zinc-binding motifs including a novel zinc binuclear cluster*. Nature Structural & Molecular Biology, 1997. **4**(7): p. 586-591.
99. Zheng, N., et al., *Structure of the Cul1–Rbx1–Skp1–F boxSkp2 SCF ubiquitin ligase complex*. Nature, 2002. **416**(6882): p. 703-709.
100. Zhang, X., et al., *Structure of an XRCC1 BRCT domain: a new protein–protein interaction module*. The EMBO journal, 1998. **17**(21): p. 6404-6411.
101. Clapperton, J.A., et al., *Structure and mechanism of BRCA1 BRCT domain recognition of phosphorylated BACH1 with implications for cancer*. Nature structural & molecular biology, 2004. **11**(6): p. 512-518.
102. Varma, A.K., et al., *Structural basis for cell cycle checkpoint control by the BRCA1-CtIP complex*. Biochemistry, 2005. **44**(33): p. 10941-10946.
103. Shen, D. and J.V. Vadgama, *BRCA1 and BRCA2 gene mutation analysis: visit to the Breast Cancer Information Core (BIC)*. Oncol Res, 1999. **11**(2): p. 63-9.
104. Lee, M.S., et al., *Comprehensive analysis of missense variations in the BRCT domain of BRCA1 by structural and functional assays*. Cancer research, 2010. **70**(12): p. 4880-4890.
105. Williams, R.S. and J.M. Glover, *Structural consequences of a cancer-causing BRCA1-BRCT missense mutation*. Journal of Biological Chemistry, 2003. **278**(4): p. 2630-2635.
106. Manke, I.A., et al., *BRCT repeats as phosphopeptide-binding modules involved in protein targeting*. Science, 2003. **302**(5645): p. 636-9.
107. Yu, X., et al., *The BRCT domain is a phospho-protein binding domain*. Science, 2003. **302**(5645): p. 639-42.
108. Rodriguez, M., et al., *Phosphopeptide binding specificities of BRCA1 COOH-terminal (BRCT) domains*. Journal of Biological Chemistry, 2003. **278**(52): p. 52914-52918.
109. Rodriguez, M.C. and Z. Songyang, *BRCT domains: phosphopeptide binding and signaling modules*. Front Biosci, 2008. **13**: p. 5905-15.
110. Liu, X. and J.A. Ladas, *Structural Basis for the BRCA1 BRCT Interaction with the Proteins ATRIP and BAAT1*. Biochemistry, 2013. **52**(43): p. 7618-7627.
111. 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.
112. 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.
113. Shen, Y. and L. Tong, *Structural Evidence for Direct Interactions between the BRCT Domains of Human BRCA1 and a Phospho-peptide from Human ACC1†*. Biochemistry, 2008. **47**(21): p. 5767-5773.
114. 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. 149-160.
115. 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.
116. Litman, R., et al., *BACH1 is critical for homologous recombination and appears to be the Fanconi anemia gene product FANCI*. Cancer cell, 2005. **8**(3): p. 255-265.

117. Tischkowitz, M., et al., *Pathogenicity of the BRCA1 missense variant M1775K is determined by the disruption of the BRCT phosphopeptide-binding pocket: a multi-modal approach*. European Journal of Human Genetics, 2008. **16**(7): p. 820-832.
118. Schaeper, U., et al., *Interaction between a cellular protein that binds to the C-terminal region of adenovirus E1A (CtBP) and a novel cellular protein is disrupted by E1A through a conserved PLDLS motif*. J Biol Chem, 1998. **273**(15): p. 8549-52.
119. Fusco, C., A. Reymond, and A.S. Zervos, *Molecular cloning and characterization of a novel retinoblastoma-binding protein*. Genomics, 1998. **51**(3): p. 351-8.
120. Wong, A., et al., *Characterization of a carboxy-terminal BRCA1 interacting protein*. Oncogene, 1998. **17**(18): p. 2279-2285.
121. Wong, A.K., et al., *Characterization of a carboxy-terminal BRCA1 interacting protein*. Oncogene, 1998. **17**(18): p. 2279-85.
122. 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.
123. Varma, A.K., et al., *Structural basis for cell cycle checkpoint control by the BRCA1-CtIP complex*. Biochemistry, 2005. **44**(33): p. 10941-6.
124. Tong, L., *Acetyl-coenzyme A carboxylase: crucial metabolic enzyme and attractive target for drug discovery*. Cell Mol Life Sci, 2005. **62**(16): p. 1784-803.
125. Wakil, S.J., J.K. Stoops, and V.C. Joshi, *Fatty acid synthesis and its regulation*. Annu Rev Biochem, 1983. **52**: p. 537-79.
126. Magnard, C., et al., *BRCA1 interacts with acetyl-CoA carboxylase through its tandem of BRCT domains*. Oncogene, 2002. **21**(44): p. 6729-39.
127. Ray, H., et al., *ACCA phosphopeptide recognition by the BRCT repeats of BRCA1*. J Mol Biol, 2006. **359**(4): p. 973-82.
128. Brunet, J., et al., *BRCA1 and acetyl-CoA carboxylase: the metabolic syndrome of breast cancer*. Mol Carcinog, 2008. **47**(2): p. 157-63.
129. Bandyopadhyay, S., et al., *Mechanism of apoptosis induced by the inhibition of fatty acid synthase in breast cancer cells*. Cancer Res, 2006. **66**(11): p. 5934-40.
130. Alo, P.L., et al., *Expression of fatty acid synthase (FAS) as a predictor of recurrence in stage I breast carcinoma patients*. Cancer, 1996. **77**(3): p. 474-82.
131. Shen, Y. and L. Tong, *Structural evidence for direct interactions between the BRCT domains of human BRCA1 and a phospho-peptide from human ACC1*. Biochemistry, 2008. **47**(21): p. 5767-73.
132. Venere, M., et al., *Phosphorylation of ATR-interacting protein on Ser239 mediates an interaction with breast-ovarian cancer susceptibility 1 and checkpoint function*. Cancer research, 2007. **67**(13): p. 6100-6105.
133. Aglipay, J.A., et al., *ATM activation by ionizing radiation requires BRCA1-associated BAAT1*. Journal of Biological Chemistry, 2006. **281**(14): p. 9710-9718.
134. Claus, E.B., N. Risch, and W.D. Thompson, *Genetic analysis of breast cancer in the cancer and steroid hormone study*. Am J Hum Genet, 1991. **48**(2): p. 232-42.
135. Ralhan, R., et al., *Links between DNA double strand break repair and breast cancer: accumulating evidence from both familial and nonfamilial cases*. Cancer Lett, 2007. **248**(1): p. 1-17.
136. Szabo, C., et al., *The breast cancer information core: database design, structure, and scope*. Hum Mutat, 2000. **16**(2): p. 123-31.
137. Shattuck-Eidens, D., et al., *A collaborative survey of 80 mutations in the BRCA1 breast and ovarian cancer susceptibility gene. Implications for presymptomatic testing and screening*. JAMA, 1995. **273**(7): p. 535-41.
138. Couch, F.J. and B.L. Weber, *Mutations and polymorphisms in the familial early-onset breast cancer (BRCA1) gene*. Breast Cancer Information Core. Hum Mutat, 1996. **8**(1): p. 8-18.

139. Clark, S.L., et al., *Structure-function Of the tumor suppressor BRCA1*. Computational and structural biotechnology journal, 2012. **1**(1).
140. Corpet, F., *Multiple sequence alignment with hierarchical clustering*. Nucleic acids research, 1988. **16**(22): p. 10881-10890.
141. Ekblad, C.M., et al., *Characterisation of the BRCT domains of the breast cancer susceptibility gene product BRCA1*. J Mol Biol, 2002. **320**(3): p. 431-42.
142. 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.
143. Williams, R.S., et al., *Structural basis of phosphopeptide recognition by the BRCT domain of BRCA1*. Nature structural & molecular biology, 2004. **11**(6): p. 519-525.
144. Campbell, S.J., R.A. Edwards, and J. Glover, *Comparison of the structures and peptide binding specificities of the BRCT domains of MDC1 and BRCA1*. Structure, 2010. **18**(2): p. 167-176.
145. Coquelle, N., R. Green, and J.M. Glover, *Impact of BRCA1 BRCT domain missense substitutions on phosphopeptide recognition*. Biochemistry, 2011. **50**(21): p. 4579-4589.
146. Makrides, S.C., *Strategies for achieving high-level expression of genes in Escherichia coli*. Microbiological reviews, 1996. **60**(3): p. 512-538.
147. Gräslund, S., et al., *The use of systematic N-and C-terminal deletions to promote production and structural studies of recombinant proteins*. Protein expression and purification, 2008. **58**(2): p. 210-221.
148. McGuffin, L.J., K. Bryson, and D.T. Jones, *The PSIPRED protein structure prediction server*. Bioinformatics, 2000. **16**(4): p. 404-405.
149. Ward, J.J., et al., *The DISOPRED server for the prediction of protein disorder*. Bioinformatics, 2004. **20**(13): p. 2138-2139.
150. Zhang, Y., et al., *DNA cloning by homologous recombination in Escherichia coli*. Nature biotechnology, 2000. **18**(12): p. 1314-1317.
151. Aslanidis, C. and P.J. de Jong, *Ligation-independent cloning of PCR products (LIC-PCR)*. Nucleic acids research, 1990. **18**(20): p. 6069-6074.
152. Sambrook, J., E.F. Fritsch, and T. Maniatis, *Molecular cloning*. Vol. 2. 1989: Cold spring harbor laboratory press New York.
153. Nilsson, J., et al., *Affinity fusion strategies for detection, purification, and immobilization of recombinant proteins*. Protein expression and purification, 1997. **11**(1): p. 1-16.
154. Waugh, D.S., *Making the most of affinity tags*. TRENDS in Biotechnology, 2005. **23**(6): p. 316-320.
155. Mandel, M. and A. Higa, *Calcium-dependent bacteriophage DNA infection*. Journal of molecular biology, 1970. **53**(1): p. 159-162.
156. Meyers, J.A., et al., *Simple agarose gel electrophoretic method for the identification and characterization of plasmid deoxyribonucleic acid*. Journal of Bacteriology, 1976. **127**(3): p. 1529-1537.
157. Saiki, R.K., et al., *Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase*. Science, 1988. **239**(4839): p. 487-491.
158. Ho, S.N., et al., *Site-directed mutagenesis by overlap extension using the polymerase chain reaction*. Gene, 1989. **77**(1): p. 51-59.
159. Fisher, C.L. and G.K. Pei, *Modification of a PCR-based site-directed mutagenesis method*. Biotechniques, 1997. **23**(4): p. 570-1, 574.
160. Janknecht, R., et al., *Rapid and efficient purification of native histidine-tagged protein expressed by recombinant vaccinia virus*. Proceedings of the National Academy of Sciences, 1991. **88**(20): p. 8972-8976.
161. Kellermann, O.K. and T. Ferenci, *[75] Maltose-binding protein from Escherichia coli*. Methods in enzymology, 1982. **90**: p. 459-463.

162. Chong, S., et al., *Single-column purification of free recombinant proteins using a self-cleavable affinity tag derived from a protein splicing element*. *Gene*, 1997. **192**(2): p. 271-281.
163. Gopalakrishna, R. and W.B. Anderson, *Ca<sup>2+</sup>-induced hydrophobic site on calmodulin: Application for purification of calmodulin by phenyl-Sepharose affinity chromatography*. *Biochemical and biophysical research communications*, 1982. **104**(2): p. 830-836.
164. Peterson, E.A. and H.A. Sober, *Chromatography of proteins. I. Cellulose ion-exchange adsorbents*. *Journal of the American Chemical Society*, 1956. **78**(4): p. 751-755.
165. Müller, W., *New ion exchangers for the chromatography of biopolymers*. *Journal of Chromatography A*, 1990. **510**: p. 133-140.
166. Caspers, M.L., Y. Posey, and R.K. Brown, *Separator isoelectric focusing: an improved method of protein analysis and purification*. *Analytical biochemistry*, 1977. **79**(1): p. 166-180.
167. Queiroz, J., C. Tomaz, and J. Cabral, *Hydrophobic interaction chromatography of proteins*. *Journal of Biotechnology*, 2001. **87**(2): p. 143-159.
168. Greenfield, N.J. and G.D. Fasman, *Computed circular dichroism spectra for the evaluation of protein conformation*. *Biochemistry*, 1969. **8**(10): p. 4108-4116.
169. Venyaminov, S.Y., et al., *Circular dichroic analysis of denatured proteins: inclusion of denatured proteins in the reference set*. *Analytical biochemistry*, 1993. **214**(1): p. 17-24.
170. Sreerama, N. and R.W. Woody, *A self-consistent method for the analysis of protein secondary structure from circular dichroism*. *Analytical biochemistry*, 1993. **209**(1): p. 32-44.
171. Perez-Iratxeta, C. and M.A. Andrade-Navarro, *K2D2: estimation of protein secondary structure from circular dichroism spectra*. *BMC Struct Biol*, 2008. **8**: p. 25.
172. Provencher, S.W. and J. Gloeckner, *Estimation of globular protein secondary structure from circular dichroism*. *Biochemistry*, 1981. **20**(1): p. 33-37.
173. Greenfield, N.J., *Using circular dichroism spectra to estimate protein secondary structure*. *Nature protocols*, 2007. **1**(6): p. 2876-2890.
174. Lakowicz, J.R., *Principles of fluorescence spectroscopy*. 2009: Springer.
175. Ladokhin, A.S., *Fluorescence spectroscopy in peptide and protein analysis*. *Encyclopedia of analytical chemistry*, 2000.
176. Zhuang, X., et al., *Fluorescence quenching: A tool for single-molecule protein-folding study*. *Proceedings of the National Academy of Sciences*, 2000. **97**(26): p. 14241-14244.
177. Weber, G. and D.J. Laurence, *Fluorescent indicators of adsorption in aqueous solution and on the solid phase*. *Biochem J*, 1954. **56**(325th Meeting): p. xxxi.
178. Rosen, C.G. and G. Weber, *Dimer formation from 1-amino-8-naphthalenesulfonate catalyzed by bovine serum albumin. A new fluorescent molecule with exceptional binding properties*. *Biochemistry*, 1969. **8**(10): p. 3915-20.
179. Tanaka, K., et al., *Protein and polymer analyses up to m/z 100 000 by laser ionization time-of-flight mass spectrometry*. *Rapid Communications in Mass Spectrometry*, 1988. **2**(8): p. 151-153.
180. Fenn, J.B., et al., *Electrospray ionization for mass spectrometry of large biomolecules*. *Science*, 1989. **246**(4926): p. 64-71.
181. O'Brien, R., B. Chowdhry, and J. Ladbury, *Protein-Ligand Interactions: Hydrodynamics and Calorimetry*, edited by SE Harding & BZ Chowdry. 2001, Oxford University Press.
182. D'arcy, A., et al., *A novel approach to crystallising proteins under oil*. *Journal of crystal growth*, 1996. **168**(1): p. 175-180.
183. Chayen, N.E., et al., *An automated system for micro-batch protein crystallization and screening*. *Journal of applied crystallography*, 1990. **23**(4): p. 297-302.

184. Zeppezauer, M., H. Eklund, and E.S. Zeppezauer, *Micro diffusion cells for the growth of single protein crystals by means of equilibrium dialysis*. Archives of biochemistry and biophysics, 1968. **126**(2): p. 564-573.
185. Van Der Woerd, M., D. Ferree, and M. Pusey, *The promise of macromolecular crystallization in microfluidic chips*. Journal of Structural Biology, 2003. **142**(1): p. 180-187.
186. Christopher, G., A. Phipps, and R. Gray, *Temperature-dependent solubility of selected proteins*. Journal of crystal growth, 1998. **191**(4): p. 820-826.
187. Rhodes, G., *Crystallography made crystal clear: a guide for users of macromolecular models*. 2010: Academic press.
188. Arndt, U.t. and A. Wonacott, *The rotation method in crystallography*. Amsterdam: North-Holland. AUBURY, A., VITUOX, B., BOUSSARD, G. & MARRAUD, M.(1981). Int. J. Pept. Protein Res, 1977. **8**: p. 195-202.
189. Battye, T.G., et al., *iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM*. Acta Crystallogr D Biol Crystallogr, 2011. **67**(Pt 4): p. 271-81.
190. Steller, I., R. Bolotovskiy, and M. Rossmann, *An algorithm for automatic indexing of oscillation images using Fourier analysis*. Journal of applied crystallography, 1997. **30**(6): p. 1036-1040.
191. Pflugrath, J., *The finer things in X-ray diffraction data collection*. Acta Crystallographica Section D: Biological Crystallography, 1999. **55**(10): p. 1718-1725.
192. Kabsch, W., *Xds*. Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 2): p. 125-32.
193. Otwinowski, Z. and W. Minor, *Processing of X-ray diffraction data*. Methods in enzymology, 1997. **276**: p. 307-326.
194. Evans, P.R., *An introduction to data reduction: space-group determination, scaling and intensity statistics*. Acta Crystallographica Section D: Biological Crystallography, 2011. **67**(4): p. 282-292.
195. Weiss, M.S., *Global indicators of X-ray data quality*. Journal of applied crystallography, 2001. **34**(2): p. 130-135.
196. Leslie, A.G., *The integration of macromolecular diffraction data*. Acta Crystallogr D Biol Crystallogr, 2006. **62**(Pt 1): p. 48-57.
197. Rossmann, M., *The position of anomalous scatterers in protein crystals*. Acta Crystallographica, 1961. **14**(4): p. 383-388.
198. Rossmann, M.G., *The molecular replacement method*. Acta Crystallographica Section A: Foundations of Crystallography, 1990. **46**(2): p. 73-82.
199. Rossmann, M.G. and D.M. Blow, *The detection of sub-units within the crystallographic asymmetric unit*. Acta Crystallographica, 1962. **15**(1): p. 24-31.
200. de La Fortelle, E. and G. Bricogne, *[27] Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods*. Methods in enzymology, 1997. **276**: p. 472-494.
201. Hendrickson, W.A., J.R. Horton, and D.M. LeMaster, *Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure*. The EMBO journal, 1990. **9**(5): p. 1665.
202. Vagin, A. and A. Teplyakov, *MOLREP: an automated program for molecular replacement*. Journal of Applied Crystallography, 1997. **30**(6): p. 1022-1025.
203. McCoy, A.J., et al., *Phaser crystallographic software*. Journal of applied crystallography, 2007. **40**(4): p. 658-674.
204. Navaza, J., *AMoRe: an automated package for molecular replacement*. Acta Crystallographica Section A: Foundations of Crystallography, 1994. **50**(2): p. 157-163.
205. Brünger, A.T., et al., *Crystallography & NMR system: A new software suite for macromolecular structure determination*. Acta Crystallographica Section D Biological Crystallography, 1998. **54**(5): p. 905-921.

206. Green, D., V. Ingram, and M. Perutz, *The structure of haemoglobin. IV. Sign determination by the isomorphous replacement method*. Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences, 1954. **225**(1162): p. 287-307.
207. Blow, D. and M. Rossmann, *The single isomorphous replacement method*. Acta Crystallographica, 1961. **14**(11): p. 1195-1202.
208. Harker, D., *The determination of the phases of the structure factors of non-centrosymmetric crystals by the method of double isomorphous replacement*. Acta Crystallographica, 1956. **9**(1): p. 1-9.
209. Ramaseshan, S., K. Venkatesan, and N. Mani. *The use of anomalous scattering for the determination of crystal structures-KMnO<sub>4</sub>*. in *Proceedings of the Indian Academy of Sciences-Section A*. 1957. Springer.
210. Hendrickson, W.A., *Determination of macromolecular structures from anomalous diffraction of synchrotron radiation*. Science, 1991. **254**(5028): p. 51-58.
211. Karle, J., *Some developments in anomalous dispersion for the structural investigation of macromolecular systems in biology*. International Journal of Quantum Chemistry, 1980. **18**(S7): p. 357-367.
212. McCoy, A.J., L.C. Storoni, and R.J. Read, *Simple algorithm for a maximum-likelihood SAD function*. Acta Crystallographica Section D: Biological Crystallography, 2004. **60**(7): p. 1220-1228.
213. Goulet, A., et al., *Getting the best out of long-wavelength X-rays: de novo chlorine/sulfur SAD phasing of a structural protein from ATV*. Acta Crystallographica Section D: Biological Crystallography, 2010. **66**(3): p. 304-308.
214. Sarma, G.N. and P.A. Karplus, *In-house sulfur SAD phasing: a case study of the effects of data quality and resolution cutoffs*. Acta Crystallographica Section D: Biological Crystallography, 2006. **62**(7): p. 707-716.
215. Brunger, A.T., J. Kuriyan, and M. Karplus, *Crystallographic R factor refinement by molecular dynamics*. Science, 1987. **235**(4787): p. 458-460.
216. Pannu, N.S. and R.J. Read, *Improved structure refinement through maximum likelihood*. Acta Crystallographica Section A: Foundations of Crystallography, 1996. **52**(5): p. 659-668.
217. Briinger, A.T., *Free R value: a novel statistical quantity for assessing the accuracy of crystal structures*. Nature, 1992. **355**: p. 472-475.
218. Murshudov, G.N., et al., *REFMAC5 for the refinement of macromolecular crystal structures*. Acta Crystallographica Section D: Biological Crystallography, 2011. **67**(4): p. 355-367.
219. Blanc, E., et al., *Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT*. Acta Crystallographica Section D: Biological Crystallography, 2004. **60**(12): p. 2210-2221.
220. Sheldrick, G.M. and T.R. Schneider, *[16] SHELXL: High-resolution refinement*. Methods in enzymology, 1997. **277**: p. 319-343.
221. Laskowski, R.A., et al., *PROCHECK: a program to check the stereochemical quality of protein structures*. Journal of applied crystallography, 1993. **26**(2): p. 283-291.
222. Berman, H.M., et al., *The protein data bank*. Nucleic acids research, 2000. **28**(1): p. 235-242.
223. RAMACHANDRAN, G.t., *Conformation of polypeptides and proteins*. Adv. Protein Chem., 1968. **23**: p. 283-437.
224. Boyd, S., *Remarks on Oöphorectomy in the Treatment of Cancer of the Breast*. British medical journal, 1899. **1**(1988): p. 257.

225. Moulder, S.L., et al., *Epidermal growth factor receptor (HER1) tyrosine kinase inhibitor ZD1839 (Iressa) inhibits HER2/neu (erbB2)-overexpressing breast cancer cells in vitro and in vivo*. *Cancer research*, 2001. **61**(24): p. 8887-8895.
226. Young, S., et al., *The prevalence of BRCA1 mutations among young women with triple-negative breast cancer*. *BMC cancer*, 2009. **9**(1): p. 86.
227. Henderson, B.E. and H.S. Feigelson, *Hormonal carcinogenesis*. *Carcinogenesis*, 2000. **21**(3): p. 427-433.
228. Park, J.J., et al., *Breast cancer susceptibility gene 1 (BRCA1) is a coactivator of the androgen receptor*. *Cancer research*, 2000. **60**(21): p. 5946-5949.
229. Lupulescu, A., *Estrogen use and cancer risk: a review*. *Experimental and Clinical Endocrinology & Diabetes*, 1993. **101**(04): p. 204-214.
230. Paech, K., et al., *Differential ligand activation of estrogen receptors ER $\alpha$  and ER $\beta$  at AP1 sites*. *Science*, 1997. **277**(5331): p. 1508-1510.
231. Fan, S., et al., *BRCA1 as a potential human prostate tumor suppressor: modulation of proliferation, damage responses and expression of cell regulatory proteins*. *Oncogene*, 1998. **16**(23): p. 3069.
232. Fan, S., et al., *BRCA1 inhibition of estrogen receptor signaling in transfected cells*. *Science*, 1999. **284**(5418): p. 1354-1356.
233. Weihua, Z., et al., *Estrogen receptor (ER)  $\beta$ , a modulator of ER $\alpha$  in the uterus*. *Proceedings of the National Academy of Sciences*, 2000. **97**(11): p. 5936-5941.
234. Shiau, A.K., et al., *The structural basis of estrogen receptor/coactivator recognition and the antagonism of this interaction by tamoxifen*. *Cell*, 1998. **95**(7): p. 927-937.
235. Poulin, R., D. Baker, and F. Labrie, *Androgens inhibit basal and estrogen-induced cell proliferation in the ZR-75-1 human breast cancer cell line*. *Breast cancer research and treatment*, 1988. **12**(2): p. 213-225.
236. Garay, J.P. and B.H. Park, *Androgen receptor as a targeted therapy for breast cancer*. *American Journal of Cancer Research*, 2012. **2**(4): p. 434.
237. Kamei, Y., et al., *A CBP integrator complex mediates transcriptional activation and AP-1 inhibition by nuclear receptors*. *Cell*, 1996. **85**(3): p. 403-414.
238. Yuan, C.-X., et al., *The TRAP220 component of a thyroid hormone receptor-associated protein (TRAP) coactivator complex interacts directly with nuclear receptors in a ligand-dependent fashion*. *Proceedings of the National Academy of Sciences*, 1998. **95**(14): p. 7939-7944.
239. Schmidt, S., et al., *Multiple receptor interaction domains of GRIP1 function in synergy*. *Nucleic acids research*, 1998. **26**(5): p. 1191-1197.
240. Leslie, A., *Recent changes to the MOSFLM package for processing film and image plate data*. 1992.
241. Winn, M.D., et al., *Overview of the CCP4 suite and current developments*. *Acta Crystallographica Section D: Biological Crystallography*, 2011. **67**(4): p. 235-242.
242. Matthews, B.W., *Solvent content of protein crystals*. *Journal of molecular biology*, 1968. **33**(2): p. 491-497.
243. Adams, P.D., et al., *PHENIX: a comprehensive Python-based system for macromolecular structure solution*. *Acta Crystallographica Section D: Biological Crystallography*, 2010. **66**(2): p. 213-221.
244. Schuttelkopf, A.W. and D.M. Van Aalten, *PRODRG: a tool for high-throughput crystallography of protein-ligand complexes*. *Acta Crystallographica Section D: Biological Crystallography*, 2004. **60**(8): p. 1355-1363.
245. DeLano, W.L., *The PyMOL molecular graphics system*. 2002.
246. Wallace, A.C., R.A. Laskowski, and J.M. Thornton, *LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions*. *Protein engineering*, 1995. **8**(2): p. 127-134.

247. Chen, Y., et al., *Aberrant subcellular localization of BRCA1 in breast cancer*. SCIENCE-NEW YORK THEN WASHINGTON-, 1995: p. 789-789.
248. Chen, C.F., et al., *The nuclear localization sequences of the BRCA1 protein interact with the importin-alpha subunit of the nuclear transport signal receptor*. J Biol Chem, 1996. **271**(51): p. 32863-8.
249. Nakielny, S., et al., *Nup153 is an M9-containing mobile nucleoporin with a novel Ran-binding domain*. The EMBO journal, 1999. **18**(7): p. 1982-1995.
250. Cronshaw, J.M., et al., *Proteomic analysis of the mammalian nuclear pore complex*. The Journal of cell biology, 2002. **158**(5): p. 915-927.
251. Pederson, T., *Nuclear RNA-protein interactions and messenger RNA processing*. The Journal of cell biology, 1983. **97**(5): p. 1321-1326.
252. Heras, B. and J.L. Martin, *Post-crystallization treatments for improving diffraction quality of protein crystals*. Acta Crystallographica Section D: Biological Crystallography, 2005. **61**(9): p. 1173-1180.
253. Studio, D., *version 2.5*. Accelrys Inc.: San Diego, CA, USA, 2009.
254. Brooks, B.R., et al., *CHARMM: the biomolecular simulation program*. Journal of computational chemistry, 2009. **30**(10): p. 1545-1614.
255. Hess, B., et al., *GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation*. Journal of chemical theory and computation, 2008. **4**(3): p. 435-447.
256. Bowers, K.J., et al. *Scalable algorithms for molecular dynamics simulations on commodity clusters*. in *SC 2006 Conference, Proceedings of the ACM/IEEE*. 2006. IEEE.
257. Emsley, P. and K. Cowtan, *Coot: model-building tools for molecular graphics*. Acta Crystallographica Section D: Biological Crystallography, 2004. **60**(12): p. 2126-2132.
258. Moynahan, M.E., et al., *Brca1 controls homology-directed DNA repair*. Molecular cell, 1999. **4**(4): p. 511-518.
259. Liu, Z., J. Wu, and X. Yu, *CCDC98 targets BRCA1 to DNA damage sites*. Nature structural & molecular biology, 2007. **14**(8): p. 716-720.
260. Gong, Z., et al., *BACH1/FANCI acts with TopBP1 and participates early in DNA replication checkpoint control*. Molecular cell, 2010. **37**(3): p. 438-446.
261. Gupta, R., et al., *Analysis of the DNA substrate specificity of the human BACH1 helicase associated with breast cancer*. Journal of Biological Chemistry, 2005. **280**(27): p. 25450-25460.
262. Chen, L., et al., *Cell cycle-dependent complex formation of BRCA1· CtIP· MRN is important for DNA double-strand break repair*. Journal of Biological Chemistry, 2008. **283**(12): p. 7713-7720.
263. Solyom, S., et al., *Breast cancer-associated Abraxas mutation disrupts nuclear localization and DNA damage response functions*. Sci Transl Med, 2012. **4**(122): p. 122-123.
264. Shao, G., et al., *MERIT40 controls BRCA1–Rap80 complex integrity and recruitment to DNA double-strand breaks*. Genes & development, 2009. **23**(6): p. 740-754.
265. Shao, G., et al., *MERIT40 controls BRCA1-Rap80 complex integrity and recruitment to DNA double-strand breaks*. Genes Dev, 2009. **23**(6): p. 740-54.
266. Wang, B., et al., *NBA1, a new player in the Brca1 A complex, is required for DNA damage resistance and checkpoint control*. Genes Dev, 2009. **23**(6): p. 729-39.
267. Chen, J., et al., *Stable interaction between the products of the BRCA1 and BRCA2 tumor suppressor genes in mitotic and meiotic cells*. Mol Cell, 1998. **2**(3): p. 317-28.
268. 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.
269. Kolas, N.K., et al., *Orchestration of the DNA-damage response by the RNF8 ubiquitin ligase*. Science, 2007. **318**(5856): p. 1637-40.

270. 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.
271. 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.
272. Ciccia, A. and S.J. Elledge, *The DNA damage response: making it safe to play with knives*. Molecular cell, 2010. **40**(2): p. 179-204.
273. Kim, H., J. Huang, and J. Chen, *CCDC98 is a BRCA1-BRCT domain-binding protein involved in the DNA damage response*. Nature structural & molecular biology, 2007. **14**(8): p. 710-715.
274. Sobhian, B., et al., *RAP80 targets BRCA1 to specific ubiquitin structures at DNA damage sites*. Science, 2007. **316**(5828): p. 1198-1202.
275. Winn, M.D., G.N. Murshudov, and M.Z. Papiz, *Macromolecular TLS refinement in REFMAC at moderate resolutions*. Methods in enzymology, 2003. **374**: p. 300-321.
276. Hall, J.M., et al., *Linkage of early-onset familial breast cancer to chromosome 17q21*. Science, 1990. **250**(4988): p. 1684-1689.
277. Friedman, L.S., et al., *Confirmation of BRCA1 by analysis of germline mutations linked to breast and ovarian cancer in ten families*. Nature genetics, 1994. **8**(4): p. 399-404.
278. Hayes, F., et al., *Functional assay for BRCA1: mutagenesis of the COOH-terminal region reveals critical residues for transcription activation*. Cancer research, 2000. **60**(9): p. 2411-2418.
279. Couch, F.J. and B.L. Weber, *Mutations and Polymorphisms in the familial early-onset breast cancer (BRCA1) gene*. Human mutation, 1996. **8**(1): p. 8-18.
280. Szabo, C., et al., *The breast cancer information core: database design, structure, and scope*. Human mutation, 2000. **16**(2): p. 123-131.
281. Manke, I.A., et al., *BRCT repeats as phosphopeptide-binding modules involved in protein targeting*. Science, 2003. **302**(5645): p. 636-639.
282. Yu, X., et al., *The BRCT domain is a phospho-protein binding domain*. Science, 2003. **302**(5645): p. 639-642.
283. Williams, R.S., et al., *Detection of protein folding defects caused by BRCA1-BRCT truncation and missense mutations*. Journal of Biological Chemistry, 2003. **278**(52): p. 53007-53016.
284. Giannini, G., et al., *Clinical classification of BRCA1 DNA missense variants: H1686Q is a novel pathogenic mutation occurring in the ontogenetically invariant THV motif of the N-terminal BRCT domain*. Journal of Clinical Oncology, 2008. **26**(25): p. 4212-4214.
285. Lundberg, K.S., et al., *High-fidelity amplification using a thermostable DNA polymerase isolated from *Pyrococcus furiosus**. Gene, 1991. **108**(1): p. 1-6.
286. Hansson, T., C. Oostenbrink, and W. van Gunsteren, *Molecular dynamics simulations*. Current opinion in structural biology, 2002. **12**(2): p. 190-196.
287. Bowers, K.J., R.O. Dror, and D.E. Shaw, *The midpoint method for parallelization of particle simulations*. The Journal of chemical physics, 2006. **124**: p. 184109.
288. Schwartz, D.W. and C.P. Barry, *Cellular communication through signal transduction: the background*. J Cardiovasc Nurs, 1994. **8**(3): p. 1-27.
289. Shiloh, Y., *ATM and ATR: networking cellular responses to DNA damage*. Current opinion in genetics & development, 2001. **11**(1): p. 71-77.
290. Abraham, R.T., *Cell cycle checkpoint signaling through the ATM and ATR kinases*. Genes & development, 2001. **15**(17): p. 2177-2196.
291. Mesquita, R.D., et al., *Tandem BRCT Domains DNA's Praetorian Guard*. Genes & cancer, 2010. **1**(11): p. 1140-1146.
292. Hasegawa, H. and L. Holm, *Advances and pitfalls of protein structural alignment*. Current opinion in structural biology, 2009. **19**(3): p. 341-348.

293. Sheng, Z.-Z., Y.-Q. Zhao, and J.-F. Huang, *Functional evolution of BRcT Domains from Binding DnA to protein*. Evolutionary bioinformatics online, 2011. **7**: p. 87.
294. Saraste, M., P.R. Sibbald, and A. Wittinghofer, *The P-loop—a common motif in ATP- and GTP-binding proteins*. Trends in biochemical sciences, 1990. **15**(11): p. 430-434.
295. Walker, J.E., et al., *Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold*. The EMBO journal, 1982. **1**(8): p. 945.
296. Li, M. and X. Yu, *Function of BRCA1 in the DNA Damage Response Is Mediated by ADP-Ribosylation*. Cancer cell, 2013. **23**(5): p. 693-704.
297. Fox, D., et al., *Crystal structure of the BARD1 ankyrin repeat domain and its functional consequences*. Journal of Biological Chemistry, 2008. **283**(30): p. 21179-21186.
298. Congreve, M., et al., *Progress in structure based drug design for G protein-coupled receptors*. Journal of medicinal chemistry, 2011. **54**(13): p. 4283-4311.
299. Scott, D.E., et al., *Using a Fragment-Based Approach To Target Protein-Protein Interactions*. Chembiochem, 2013. **14**(3): p. 332-342.
300. Whittle, P.J. and T.L. Blundell, *Protein structure-based drug design*. Annual review of biophysics and biomolecular structure, 1994. **23**(1): p. 349-375.
301. Leavitt, S. and E. Freire, *Direct measurement of protein binding energetics by isothermal titration calorimetry*. Current opinion in structural biology, 2001. **11**(5): p. 560-566.
302. Homola, J., S.S. Yee, and G. Gauglitz, *Surface plasmon resonance sensors: review*. Sensors and Actuators B: Chemical, 1999. **54**(1): p. 3-15.



## List of Publications arising from the thesis

### Journal

1. **Badgujar DC**, Sawant U, Mahadik H, Gadewal N, Varma AK (2012) Pathogenicity of Mutations Discovered in BRCA1 BRCT Domains is Characterized by Destabilizing the Hydrophobic Interactions. *Journal of Cancer Science & Therapy* **4**: 386-393.
2. **Badgujar DC**, Sawant U, Yadav L, Hosur M, Varma AK (2013). Preliminary crystallographic studies of BRCA1 BRCT-ABRAXAS complex. *Acta Crystallographica Section F: Structural Biology and Crystallization Communications* **69**: 1401-1404.