STRUCTURAL AND FUNCTIONAL APPROACHES TO EVALUATE NOVEL MISSENSE GERMLINE MUTATIONS IN BRCA1

By

Lumbini Ramraj Yadav
(LIFE09200804002)

TATA MEMORIAL CENTRE
MUMBAI

A thesis submitted to the
Board of Studies in Life Sciences
In partial fulfilment of the requirements
for the Degree of

DOCTOR OF PHILOSOPHY

OF

HOMI BHABHA NATIONAL INSTITUTE

January, 2016

Homi Bhabha National Institute

Recommendations of the Viva Voce Committee

As members of the Viva Voce Committee, we certify that we have read the dissertation prepared by Lumbini Ramraj Yadav entitled "**STRUCTURAL AND FUNCTIONAL APPROACHES TO EVALUATE NOVEL MISSENSE GERMLINE MUTATIONS IN BRCA1**" and recommend that it may be accepted as fulfilling the thesis requirement for the award of Degree of Doctor of Philosophy.

Chairman - Dr. Vinay Kumar — **Date:** 29/01/2016

Guide / Convener - Dr. Ashok K. Varma — **Date:** 29/01/2016

Co-guide - Dr. Rajiv Sarin — **Date:** 29.1.16

Member - Dr. Sanjay Gupta — **Date:** 29/1/16

Member - Dr. Pradnya Kowtal — **Date:** 29/1/16.

External Examiner- Dr. Punit Kaur — **Date:** 29/1/2016.

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

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

Date: 29.01.2016

Place: Kharghar

Dr. Rajiv Sarin
Co-guide

Dr. Ashok K. Varma
Guide

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfilment 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.

Lumbini 5/1/2016

Navi Mumbai

January 2016

Lumbini R.Yadav

DECLARATION

I, hereby declare that the investigation presented in the thesis has been carried out by me. The 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.

Lumbini 5/1/2016

Navi Mumbai

January 2016

Lumbini R.Yadav

List of Publications arising from the thesis

Journal

1. "Tetrameric ZBRK1 DNA binding domain has affinity towards cognate DNA in absence of zinc ions." **Lumbini R. Yadav,** Mahamaya N. Biswal, Vikrant, M.V. Hosur, Ashok K. Varma. Biochemical and Biophysical Research Communications, 2014, 450, 283-288.
2. "Functional assessment of intrinsic disorder central domains of BRCA1." **Lumbini R.Yadav**, Sharad Rai, M.V.Hosur, & Ashok K. Varma. Journal of Biomolecular Structure and Dynamics, 2015, DOI:10.1080/07391102.2014.1000973.
3. "Structural basis to characterise transactivation domain of BRCA1." **Lumbini R. Yadav**, Mahamaya N. Biswal, M.V. Hosur , Nachimuthu Senthil Kumar and Ashok K. Varma. Journal of Biomolecular Structure and Dynamics, 2016.
4. "Structural and functional characterization of transactivation Domain of BRCA1." **Lumbini R. Yadav**, Ashok K. Varma. Acta Crystallographica Section A, 2011, A67, C221-C222.

Other Publications:

1. Optimized hydrophobic interactions and hydrogen bonding at the target-ligand interface leads the pathways of drug-designing. Patil R, Das S, Stanley A, **Yadav L**, Sudhakar A, Varma AK. PLoS One. 2010, 5. doi:10.1371/journal.pone.0012029
2. Preliminary crystallographic studies of BRCA1 BRCT-ABRAXAS complex. Badgujar DC, Sawant U; Vikrant, **Yadav L,** Hosur MV, Varma AK. Acta Crystallographica. Section F, Structural Biology Communications, 2013, F69, 1401–1404.
3. Structural and functional implication of RAP80 ΔGlu81 mutation Vikrant, Kumar R, **Yadav LR**, Nakhwa P, Waghmare SK, Goyal P,Varma AK. PLOS One. 2013.

Conferences

1. Indo-US symposium /Workshop Modern, Trends in Macromolecular structure, Indian Institute of Technology, February, Mumbai, INDIA 2011. Poster Presentation.
2. XXII Congress and General Assembly, International Union of Crystallography (IUCr), Madrid, Spain, in August 2011. Poster Presentation.
3. International Cancer Genome Consortium, Workshop/ conference 2014, October TMC-ACTREC, Mumbai, INDIA. Poster and Oral Presentation.
4. Recent Advances in Crystallography, Indian Institute of Technology, November Mumbai, INDIA 2011. Poster Presentation.
5. National Research Scholar Meet, TMC-ACTREC, December 2014, INDIA. Oral Presentation.

Lumbini 5/1/2016

Lumbini R.Yadav

I DEDICATE THIS THESIS TO ALL THOSE WHO HAVE SACRIFICED THEIR LIVES FOR EMANCIPATION AND EMPOWERMENT OF WOMEN

ACKNOWLEDGEMENTS

I wish to express my heartfelt gratitude to my mentor Dr. Ashok K. Varma for giving me an opportunity to be a part of his research group and introducing me to the field of structural biology. His guidance and help in experiments, data management and documentation has helped me in completing the assigned project. He has also helped me in improving my professional skills. His ever enthusiastic, confident, constantly optimistic and his damn care attitude have helped me to sail through all thicks and thin during my Ph.D. days. I am thankful to my Co-guide Dr. Rajiv Sarin a very inquisitive and a person full of energy all throughout since, the years I have known him. I am glad to get his guidance in analysing, managing and documenting the data. His attention to detail, patience, motivation and trust in me has helped me evolve in my research endeavours. I express my gratitude to him for being by my side and listening to me patiently whenever needed.

I am thankful to Dr. Shubhada Chiplunkar (Director, ACTREC), Dr. Rajiv Sarin (Ex-Director, ACTREC) and Dr. Surekha Zingde (Ex-Deputy Director, ACTREC) for their help and support directly or indirectly. I'm also thankful to Department of Biotechnology (DBT) and ICMR, Homi Bhabha National Institute (HBNI), Patel Kantaben Kantilal trust, SIA-TMC and International Union of Crystallography (IUCr) for funding.

I was privileged to have critical, supportive and inspiring Doctoral Committee members. I am indebted and highly obliged to Prof. Hosur (Ex-chairperson) for all his support and guidance in my research work and in critically evaluating my thesis. I feel blessed to have an opportunity to interact with him to such a large extent and also thank him for all the help, support and guidance he has extended towards me. I am thankful to Dr. Vinay Kumar (Chairperson) for critical evaluation of data, critical evaluation of scientific understanding, his guidance and motivation that he provided me during this tenure and hope to get so in future too. I would like to thank Dr. Sanjay Gupta for his invaluable suggestions and guidance that had an impact in shaping up my scientific aptitude. I would like to thank Dr. Pradnya Kowtal not only as my DC member, but also for her support in providing the patient data and chemicals very promptly whenever I needed.

I would like to express my deepest gratitude to Mr. Uday Dandekar and the team of common instrument facility. I would like to acknowledge Dr. Kakoli Bose incharge of maintaining CD and fluorescence instrument. I am thankful to Mrs Maya from Program office, Mr. Shashi Dolas and all members of Mass Spectrometry facility, Mrs Sharada and Mr. Naresh from sequencing facility, Mahadev and Dinesh (Hostel Caretaker) for all their help. I am also thankful to the electrical department, Security department for the safety and watchfulness.

I am gratified with all the members of Cancer Genetics Clinic and the HBOC family patients for their active participation and help in providing the information.

I want to express my warmest thank to all Varma lab members. Ulka Ma'am a pillar of Varma lab for always being there for her support. Sopan and Sachin for their help in autoclaving and cleaning glass wares, Dr. Dilip, Dr.Vikrant, Quadir, Bhanu, Rajan, Pankaj, Lipi, Sucheta and Nikhil Gadewal (BTIS) for their help, support and co-operation. I am extremely thankful and indebted to all my trainees (Mahamaya, Sharad, Rashmi, Elavarasi, Megha, Poornima and all others who have helped me) for their presence, honest support and affection during their tenure in my Ph.D.

I am thankful to my batchmates of 2008 for their support and all the good times we had together. I am glad and thankful to all the seniors and specifically all our seniors from 2006 batch for their support and affection. I am thankful to all the students of Actrec for providing timely support whenever needed. I would specifically like to acknowledge Kalraiya lab, Prasanna Lab, Vaidya lab, Robin lab, Dibyendu lab and Sorab lab.

I am specially thankful to Anamika Pandey, Amit Ranjan, Praveen Patil, Ira, Poornima, Dilip, Mahamaya Biswal, Melissa, Neha behare, Dr. Sarins Mom (Dadiji) and his wife, Shalini Ma'am, My graduation teachers (Pradhan Ma'am, Shinde Ma'am, Karni Ma'am, Kulkarni Sir, and all teachers of Zoology, Botany and Chemistry Department who made me realise the world is a very beautiful place to live in. Special thanks to rishi for his help and guidance during my initial Ph.D. days. I owe my peace, happiness, success and progress to all such people in my life.

I am extremely thankfull to my parents: an image of god for me, for their sacrifice, trust, support, and confidence in me and for always standing by me whenever needed. I am thankful to Keerti, Tejaswini and Yesashwini the little angels in my life for their unconditional love and affection. I thank all my cute sisters for the strength, courage, love and affection they provided.

I would like to thank each and every one who interacted with me and taught me lessons of life that will help me grow in future. I am thankful to the supreme power of the universe for making me one of the luckiest being and for conspiring with nature to bring me at this stage and helping me in all odd times

Thank you all

Lumbini Ramraj Yadav

Contents

Chapter 2: MATERIALS AND METHODS

Homi Bhabha National Institute

SYNOPSIS OF Ph. D. THESIS

1. Name of the Student	**:**	Lumbini R. Yadav
2. Name of the Constituent Institution	**:**	Tata Memorial Centre, Advanced Centre for Treatment, Research and Education in Cancer
3. Enrolment No.	**:**	LIFE09200804002
4. Title of the Thesis	**:**	"STRUCTURAL AND FUNCTIONAL APPROACHES TO EVALUATE NOVEL MISSENSE GERMLINE MUTATIONS IN BRCA1".
5. Board of Studies	**:**	Life Sciences

Synopsis

Cancer is a major public health challenge worldwide, including developing countries. The annual incidence of cancer is one million in India, of which, approximately 100,000 are breast cancers. Cancer is characterized as uncontrolled division of cells with underlying mutations in tumor suppressor genes or proto-oncogenes. Breast cancer ranks second as a cause of cancer death in women. Germline mutations in *BRCA1, BRCA2 and a few other genes* are responsible for hereditary breast and ovarian cancer (HBOC) syndrome which accounts for 5-10% of all breast and ovarian cancers [1]. Carriers of deleterious germ-line *BRCA1* mutations have a 50–80% lifetime risk of developing breast cancer, and 30–50% risk of developing ovarian cancer [2]. Gain of function mutations in oncogenes (e.g. RET or RAS) occur at specific loci and their functional and phenotypic effect are well characterized. In contrast, mutations are scattered across the tumour suppressor genes, majority being missense changes with variable biological effect. While nonsense and frame shift mutations which truncate the protein are always pathogenic, missense mutations that have weak or no biological consequence (polymorphisms) or severe biological effect (deleterious mutations), confer high risk for disease. Missense germline variants for which the cancer risk has not been determined or cannot be predicted with great confidence are termed as variants of uncertain significance (VUS), accounting for >90% of missense BRCA1 variants. VUS pose a major challenge in risk estimation and counseling of individuals inheriting these VUS and in interpreting research results. There is a need for comprehensive characterization of known and novel missense germline variants of uncertain significance using *in silico,* clinical, biophysical, structural and functional approaches for making recommendation for genetic screening and surveillance of carriers of these variants. Several novel missense variants of *BRCA1* have been discovered in Tata Memorial Centre (TMC) cohort of Indian HBOC families. The objectives of the present study are

- Structural and functional analysis of novel BRCA1 variants identified in Indian HBOC families
- Studies on structure of central regions of BRCA1 and their interaction with DNA.
- Studies on ZBRK1- a BRCA1 central region binding protein.

Chapter 1 provides brief introduction about the structure and function of *BRCA1* gene (17q21) with 24 exons [3,4]. Developmental defects and embryonic lethality in *BRCA1* null mice and chromosomal abnormalities in *BR*CA1 deficient cell lines clearly indicate its role in maintaining genomic stability. *BRCA1* is a pleiotropic tumor suppressor gene with autosomal dominant inheritance. BRCA1 is a 220 kDa protein, which has two recognizable domains, an N-terminal ring finger (1-109) domain and a C- terminal (1646-1859) BRCT domain which is established as a phospho-peptide recognition domain. The central region, reported to be unstructured, has two nuclear localization signals in exon 11, and a 'SQ' cluster between amino acids 1280–1524 [5]. BRCA1 plays a role in transcription regulation, by acting both as a transcriptional activator and transcriptional repressor. BRCA1 has role in regulating G1/S, S-phase, and G2/M checkpoints during cell proliferation. BRCA1 participates in both types of double strand DNA break repair, non-homologous end joining and homologous repair [6,7,8,9,10,11].

Chapter 2 provides a brief description of the various methodologies and techniques used in this study. Sticky-end-generating, restriction enzyme based cloning technique was used to ensure correct orientation of the insert in the vectors. Site-directed mutagenesis was performed on wild type template to incorporate the mutations identified in TMC cohort. Bacterial strains Rosetta2 (DE3) and BL21 (DE3) were used for protein expression. Protein purification methods employed were ammonium sulphate precipitation, affinity chromatography (GST and Ni-NTA), and size exclusion chromatography. The purity of protein was confirmed by SDS-PAGE. Further, peptide mass fingerprinting was used to

reveal the identity of purified proteins. Conformational and oligomeric characterization was performed using gel permeation chromatography and native gel electrophoresis. The secondary and tertiary structural characterization was explored using Circular dichroism and Fluorescence spectroscopy, respectively. Molecular interaction studies were performed using Isothermal Titration Calorimetry (ITC), electrophoretic mobility shift and pull down assays. *In silico* analysis for predicting the biological effect of BRCA1 variants was performed using Polyphen, Align GVGD, PhD SNP, and Sift servers. Anchor, Metadisorder, MFDp and Cspritz servers were used for disorder prediction. Composition Profiler and PONDR were used for amino acid composition analysis. Molecular modeling studies were performed using software Desmond to understand the alterations in hydrogen bond and hydrophobic interactions of BRCA1 variants.

Crystallization attempts were made at 22°C, by the hanging drop and sitting drop vapor diffusion methods. Crystallization screening was performed using Hampton's Crystal Screen 1 and 2, SaltRx 1, PEG/Ion and additive screen solutions. X-ray diffraction data were collected on the mar dtb image plate system mounted on a rotating microStar X-ray generator available in- house. Diffraction data frames were processed using iMosflm and were scaled using SCALEPACK. Finally, structure was solved by the molecular replacement method and subsequent structure refinement was done using Refmac-5. Electron density maps were examined using Coot software.

Chapter 3 describes the systematic *in silico,* genetic and clinico-pathological evaluation of the germline missense *BRCA1* variants identified in 48 HBOC families through Sanger sequencing of the cancer affected cases. Of the 19 missense variants identified, 10 were already reported in the BIC/ClinVar database, 5 were synonymous variants and the remaining 4 (C1697Y, S1722P F1124I and Q210H) were novel unreported variants. The likelihood ratio for these variants being deleterious were calculated by examining 1) co-occurrence of variant

with deleterious mutation in *trans*, have very low likelihood of being deleterious, as homozygosity for BRCA1 pathogenic mutations are embryologically lethal, 2) co-segregation of variant with the disease in family more than expected by chance, 3) proband and family history of cancer which considers age at cancer diagnosis, number of members affected, and 4) pathological characteristics like tumor grade and ER/PR status [12]. The four likelihood ratios obtained were then multiplied to get the joint likelihood value. These final likelihood values was used to calculate the posterior probability, considering the prior probability value obtained from Align GVGD. The value of the posterior probability was used to place the variants in different (International Agency for Research on Cancer) IARC class and also to define the protocol for genetic testing and surveillance recommendations [13]. *In silico* studies were carried out to predict the pathogenicity of variants, based on, either sequence or structure or both sequence and structure. The variants S1722P and C1697Y belong to the definitely pathogenic IARC class 5. Expectedly, these residues are at positions which are conserved across the species. F1124I and Q210H are not conserved and belong to IARC class 2, which indicates class of little clinical significance. Co-occurrence of Q210H variant with a deleterious mutation suggests that this variant is very unlikely to be a deleterious mutation. Even though E809 is not a conserved residue, its high posterior probability value classifies it into IARC class 4, a likely pathogenic variant. All other identified variants classified belong to class 2/1, which is a likely not pathogenic class. Therefore, S1722P, C1697Y and E809G mutants were further pursued in this research study.

Chapter 4 deals with functional and biophysical characterization of S1722P and C1697Y variants. BRCA1 transactivation domain (1560-1859) (TAD) was cloned in pGBKT-7 and pGEX-KT vector for expression in yeast and bacterial system, respectively. Site directed mutagenesis was carried out to generate mutant clones S1722P and C1697Y for use, both in yeast and bacteria. These variants occurred in the TAD of BRCA1 and therefore the influence

of these variants on transactivation was assessed. Transcription activation assay in yeast resulted in the loss of the β-galactosidase activity for both variants and M1775R, which was used as a negative control. However, wild type protein showed β-galactosidase activity suggesting that the variants perturb optimal protein-protein interactions. In bacterial system, all the constructs were expressed as observed in SDS-gel electrophoresis. However, C1697Y could not be purified in soluble form, as it was prone to aggregation and precipitation. S1722P and TAD were successfully purified. ITC was carried out to explore the binding of S1722P and TAD with (pS-X-X-F) phosphopeptide motif. However, while TAD binds to phosphopeptide motif, S1722P did not. The observation that S1722P elutes in Superdex-200 column in void volume, suggests protein aggregation, providing a further possible explanation for nonbinding of the phosphopeptide. Intrinsic tryptophan fluorescence also suggests unfolding and aggregation of S1722P. Far UV CD analysis suggests loss in alpha helical content in S1722P. In summary, multimodal *in vitro* and *in silico* studies, led to the conclusion that these substitutions destabilize the proteins' three dimensional folding thereby affecting their functionality. Thus comprehensive analysis suggests that C1697Y and S1722P are high risk variants.

Chapter 5 Crystallization and biophysical characterization of N terminal deletion constructs of BRCA1 are reported in this chapter. Although, the carboxyl terminal region 1760–1863 is the minimal region required for BRCA1 to activate transcription, it is also known that extension of this region in the N terminal direction enhances the transcriptional activity in a context dependent manner [4]. The following N-terminal deletion constructs (1314-1863, 1396-1863, 1560-1859 and 1646-1859) of BRCA1 were generated by cloning in pGEX-KT vector at the BamH1/ECoR1 site. The different BRCA1 fragments were expressed in bacterial system and were purified using ammonium sulphate precipitation and affinity chromatography methods. Attempts were made to crystallize these constructs by hanging and

sitting drop vapour diffusion method. However, crystals were obtained only for the BRCA1 (1560-1859) fragment. These crystals were screened for diffraction analysis using laboratory X-ray source. Crystals screened, diffracted to a low resolution of around 4 Å. These crystals belong to space group P 6_1 2 2 with unit cell parameters a=b= 114.2 Å and c=120.9 Å. Cell content analysis calculates the Matthews coefficient (V_M) value of 3.40 $Å^3$Da-1 which corresponds to a solvent content of 63%, and suggests presence of one protein molecule of 33 kDa per asymmetric unit. Structure solution was performed by PHASER using the structure with PDB ID 1T29 as a search model. Interestingly, good electron density was observed only for residues forming the BRCT domain (1646-1859). Crystal packing shows that the polypeptide is either disordered or totally absent as there is enough space to accommodate the extra 86 amino acid upstream of BRCT. As the cell parameters matched with those of BRCA1 BRCT crystals (PDB ID: 1JNX), SDS-gel electrophoresis was employed to investigate the actual protein contents of the crystals and the mother liquor. A band corresponding to 33 kDa was observed in some crystal samples although majority of samples showed band at 24 kDa suggesting sample degradation to BRCT domain. The reason for degradation was further explored by incubation of BRCA1 (1560-1859) alone and in the presence of protease inhibitor cocktail. It was found that total degradation of protein resulted within 31 hours in either case. Limited proteolysis study of other N-terminal deletion constructs show that while BRCT was resistant, all others were susceptible to quick tryptic digestion yielding 24 kDa band corresponding to BRCT. Interestingly, far UV CD (250-195 nm) signal due to BRCT domain progressively decreases as the number of N-terminal residues added onto BRCT increases. *In silico* study, in concordance with other biophysical studies, using PSIPRED and Composition Profiler showed that region upstream of BRCT (1646-1859) has no defined structure, and is enriched with disorder promoting residues.

Chapter 6 describes: 1. The characterization of central regions of BRCA1 which are involved in binding to DNA and proteins like p53, ZBRK1, STAT1 and RAD 51, 2. Structural alterations in BRCA1 (502-802) in presence of DNA, 3: Biophysical characterization of a novel missense variant E809G. Cloning of BRCA1 fragments (260-553) in pGEX-4T, (341-748) in pET 41a+, (502-802) and (758-1064) in pGEX-KT vectors were carried out using BamH1/ECoR1 restriction site. The proteins were expressed in bacterial expression system and purified using affinity chromatography. *In silico* study using PONDR and composition profiler demonstrated that the central region spanning these fragments are rich in residues conferring disorder, and has high net negative charge [14,15,16]. Far UV CD spectra showed negative ellipticity at 200 nm, confirming that the central region of BRCA1 is unstructured [17]. Limited proteolysis with trypsin also demonstrated degradation of protein within 10 min, in sharp contrast to BSA, which under identical conditions, did not show any degradation even after 2 h. Thermal denaturation study indicated incorporation of structure with increase in temperature, which is termed as 'turned out' response, providing further support to intrinsically disordered nature of the central fragments. BRCA1 (502-802) interaction with duplex and cruciform DNA were studied. EMSA study showed two and three different type of complex formation for (502-802) with duplex and cruciform DNA, respectively. Increase in retention volume was observed for DNA/BRCA1 (502-802) complex as compared to BRCA1 (502-802) alone in gel permeation chromatography, indicating compaction in the structure. Increase in negative ellipticity at λ_{222} nm in far UV CD scan, and increase in fluorescence intensity, on DNA binding suggests increase in ordered component in BRCA1 (502-802). Although all the above studies suggest increase in structural component and compaction of protein, no change in susceptibility to protease was observed.

Biophysical study explored the effects of variant E809G classified under likely pathogenic IARC class 4. Elution profile in gel permeation chromatography indicates less compact structure for E809G variant compared to BRCA1 (758-1064) wild type. Far UV CD scan suggest that incorporation of E809G variant leads to increase in disorder. Red shift of 5 nm of E809G variant in fluorescence spectroscopy study also suggests loss of tertiary structure. In conclusion, central region of BRCA1 is intrinsically disordered. This unstructured pattern even in the presence of its binding partner is possibly responsible for providing BRCA1, the ability to interact with multiple binding partners and act as a scaffold protein in the DNA repair process.

Chapter 7 describes purification and interaction study of ZBRK1 zinc finger domain. ZBRK1 is a zinc finger transcription regulatory protein that plays a crucial role in cell-cycle regulation, DNA damage response and tumorogenesis [18]. ZBRK1 interacts with BRCA1 central region to repress various cell-cycle regulatory proteins. In this study, we have cloned ZBRK1 zinc finger domain (ZFD) (206-424) in pRSET-A vector. The protein was expressed in Rosetta 2 (DE3) and purified using Ni-NTA resin. ZFD binding with zinc ions and DNA were explored using biophysical techniques. Dissociation and pull down assays suggest that ZFD forms a higher order oligomer. The elution profile of the purified protein suggests that this oligomer is a homo tetramer in solution [19]. Thermal denaturation study indicated ZFD acquires higher stability in the presence of zinc ions and DNA. Near UV CD spectra indicate structure of the DNA in the complex with the protein is significantly altered from the standard B-DNA conformation. ITC study of ZFD with zinc and DNA showed exothermic reaction with stoichiometry of 1:42 and 1:2, respectively.

Chapter 8 summarizes the findings of the study. In this study, we could establish the clinical significance of BRCA1 missense variants identified in the Indian HBOC families using comprehensive *in silico*, clinico-pathological and structure–function analysis. Of the 14

variants identified, 3 variants (E809G, C1697Y and S1722P), which were classified as pathogenic were characterized further with detailed functional and structural studies. The C1697Y and S1722P variants were shown to impair the transcriptional activity of BRCA1 in yeast. C1697Y was expressed but was not soluble. S1722P variant aggregates and abrogates interaction with signature phosphopeptide. Biophysical analysis of S1722P revealed loss in secondary structure and distortion in the tertiary structure of the protein. While *In silico* study predicted E809G to be neutral variant, biophysical studies of this variant show increased hydrodynamic radius due to altered secondary and tertiary structure. Comprehensive characterization of 3 novel *BRCA1* variants (E809G, C1697Y and S1722P) confirmed the pathogenic nature of these variants. To reveal the three-dimensional structural details of these variants crystallization attempts were made. Structural characterization of the N-terminal deletion constructs of full length BRCA1 suggested the constructs to be enriched in disorder promoting residues with defined secondary structure only in BRCA1 BRCT region.

Study on the central region revealed the region to be intrinsically disordered. The central region has random coil like structure and is extremely sensitive to trypsin digestion. To explore structural changes on ligand binding, interaction study of BRCA1 (502-802) with DNA was carried out, and these studies revealed decrease in the hydrodynamic radius of the protein on binding to DNA. Biophysical study demonstrated incorporation of secondary and tertiary structure on DNA binding. The central region of BRCA1 binds to zinc finger domain and acts as a co repressor of ZBRK1. Study on ZBRK1 zinc finger domain indicated that it forms a tetramer. ZFD binds exclusively to double stranded DNA with stoichiometry of 1:2. It was found that the zinc finger of ZBRK1 binds DNA even in the absence of zinc ions. In the complex with ZFD, the DNA is distorted from standard B-geometry. This study confirms the utility of a multi-modality approach for comprehensive characterization of missense variants in clinically important tumor suppressor genes like

BRCA1. Establishing the pathogenicity of variants of unknown significance in such genes greatly helps in genetic counseling and risk management of mutation carriers.

References:

1. Robson, M.E., Boyd, J., Borgen, P.I. and Cody, H.S., 3rd. (2001) Hereditary breast cancer. *Curr Probl Surg*, **38**, 387-480.
2. Rahman, N. and Stratton, M.R. (1998) The genetics of breast cancer susceptibility. *Annual review of genetics*, **32**, 95-121.
3. Easton, D.F., Bishop, D.T., Ford, D. and Crockford, G.P. (1993) Genetic linkage analysis in familial breast and ovarian cancer: results from 214 families. The Breast Cancer Linkage Consortium. *American journal of human genetics*, **52**, 678-701.
4. Hu, Y.F., Miyake, T., Ye, Q. and Li, R. (2000) Characterization of a novel trans-activation domain of BRCA1 that functions in concert with the BRCA1 C-terminal (BRCT) domain. *The Journal of biological chemistry*, **275**, 40910-40915.
5. Paterson, J.W. (1998) BRCA1: a review of structure and putative functions. *Disease markers*, **13**, 261-274.
6. Larner, J.M., Lee, H. and Hamlin, J.L. (1997) S phase damage sensing checkpoints in mammalian cells. *Cancer surveys*, **29**, 25-45.
7. Zhao, G.Y., Sonoda, E., Barber, L.J., Oka, H., Murakawa, Y., Yamada, K., Ikura, T., Wang, X., Kobayashi, M., Yamamoto, K. *et al.* (2007) A critical role for the ubiquitin-conjugating enzyme Ubc13 in initiating homologous recombination. *Molecular cell*, **25**, 663-675.
8. Lee, J.-H. and Paull, T.T. (2005) ATM activation by DNA double-strand breaks through the Mre11-Rad50-Nbs1 complex. *Science*, **308**, 551-554.
9. Lee, J.H. and Paull, T.T. (2004) Direct activation of the ATM protein kinase by the Mre11/Rad50/Nbs1 complex. *Science*, **304**, 93-96.
10. Yu, X. and Chen, J. (2004) DNA damage-induced cell cycle checkpoint control requires CtIP, a phosphorylation-dependent binding partner of BRCA1 C-terminal domains. *Molecular and cellular biology*, **24**, 9478-9486.
11. Xu, B., Kim, S. and Kastan, M.B. (2001) Involvement of Brca1 in S-phase and G(2)-phase checkpoints after ionizing irradiation. *Molecular and cellular biology*, **21**, 3445-3450.

12. Goldgar, D.E., Easton, D.F., Deffenbaugh, A.M., Monteiro, A.N., Tavtigian, S.V. and Couch, F.J. (2004) Integrated evaluation of DNA sequence variants of unknown clinical significance: application to BRCA1 and BRCA2. *American journal of human genetics*, **75**, 535-544.
13. Plon, S.E., Eccles, D.M., Easton, D., Foulkes, W.D., Genuardi, M., Greenblatt, M.S., Hogervorst, F.B., Hoogerbrugge, N., Spurdle, A.B. and Tavtigian, S.V. (2008) Sequence variant classification and reporting: recommendations for improving the interpretation of cancer susceptibility genetic test results. *Human mutation*, **29**, 1282-1291.
14. Dunker, A.K., Lawson, J.D., Brown, C.J., Williams, R.M., Romero, P., Oh, J.S., Oldfield, C.J., Campen, A.M., Ratliff, C.M., Hipps, K.W. *et al.* (2001) Intrinsically disordered protein. *Journal of molecular graphics & modelling*, **19**, 26-59.
15. Dunker, A.K., Brown, C.J. and Obradovic, Z. (2002) Identification and functions of usefully disordered proteins. *Advances in protein chemistry*, **62**, 25-49.
16. Uversky, V.N., Gillespie, J.R. and Fink, A.L. (2000) Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins: Structure, Function, and Bioinformatics*, **41**, 415-427.
17. Mark, W.Y., Liao, J.C., Lu, Y., Ayed, A., Laister, R., Szymczyna, B., Chakrabartty, A. and Arrowsmith, C.H. (2005) 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*, **345**, 275-287.
18. Zheng, L., Pan, H., Li, S., Flesken-Nikitin, A., Chen, P.L., Boyer, T.G. and Lee, W.H. (2000) Sequence-specific transcriptional corepressor function for BRCA1 through a novel zinc finger protein, ZBRK1. *Mol Cell*, **6**, 757-768.
19. Tan, W., Kim, S. and Boyer, T.G. (2004) Tetrameric oligomerization mediates transcriptional repression by the BRCA1-dependent Kruppel-associated box-zinc finger protein ZBRK1. *The Journal of biological chemistry*, **279**, 55153-55160.

Publications in Refereed Journal:

1. Functional assessment of intrinsic disorder central domains of BRCA1.
   Lumbini R. Yadav, S Rai, MV Hosur,AK Varma.
   J Biomol Struct Dyn. 2015 January 23: 1-10.

2. Tetrameric ZBRK1 DNA binding domain has affinity towards cognate DNA in absence of zinc ions.

Lumbini R. Yadav, MN Biswal, Vikrant, MV Hosur, AK Varma.
Biochem Biophys Res Commun. 2014 Jul 18; 450(1):283-8.

Other Publications:

1. Structural and functional implication of RAP80 ΔGlu81 mutation.
   Vikrant, R Kumar, **Lumbini R. Yadav**, P Nakhwa, SK Waghmare, P Goyal,AK Varma.
   PLoS One 2013 Sep 9; 8(9):e72707.

2. Preliminary crystallographic studies of BRCA1 BRCT-ABRAXAS complex.
   DC Badgujar,U Sawant U, Vikrant, **Lumbini R. Yadav**, MV Hosur, Varma AK.
   Acta Crystallogr Sect F Struct Biol Cryst Commun. 2013 Dec;69(Pt 12):1401-4.

3. Optimized hydrophobic interactions and hydrogen bonding at the target-ligand interface leads the pathways of drug-designing.
   R Patil, S Das, A Stanley, **Lumbini R. Yadav**, A Sudhakar, A.K. Varma.
   PLoS One. 2010 Aug 16; 5(8):e12029.

Date: 27-3-15

Signature of Student:

Doctoral committee members:

S. No.	Name	Designation	Signature	Date
1.	Dr. Vinay Kumar	Chairman		27/3/15
2.	Dr. Ashok K. Varma	Guide & Convener		26/03/15
3.	Dr. Rajiv Sarin	Co-Guide		26/3/15
4.	Dr. Sanjay Gupta	Member		26/03/15
5.	Dr. Pradnya Kowtal	Member		26/3/15

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

Dr. S. V. Chiplunkar
Director
Advanced Centre for Treatment, Research & Education in Cancer (ACTREC)
Tata Memorial Centre
Kharghar, Navi Mumbai 410210.

Prof. K. Sharma
Director, Academics
Tata Memorial Centre

PROF. K. S. SHARMA
DIRECTOR (ACADEMICS)
TATA MEMORIAL CENTRE
PAREL, MUMBAI

ABBREVIATIONS

µM	micro Molar
53BP1	53 Binding Protein
Align GVGD	Align Grantham Variation Grantham Deviation
AR	Androgen Receptor
ATM	Ataxia Telangiectasia Mutated
ATR	ATM and RAD3 related
ATRIP	ATR Interacting Protein
BACH1	BTB and CNC homology 1
BAP1	BRCA1 Associated Protein 1
BARD1	BRCA1 Associated Ring Domain -1
BIC	Breast cancer Information Core
BRCA1	BReast CAncer susceptibility gene1
BRCT	BRCA1 C-Terminal domain
CCP4	Collaborative Computational Project No.4
CD	Circular Dichroism
Chk1	Checkpoint kinase 1
COOT	Crystallographic Object-Oriented Toolkit
CtIP	CtBP Interacting Protein
DNA	Deoxyribonucleic Acid
DNA Pol λ	DNA polymerase λ
E. coli	*Escherichia coli*
EMSA	El□ctrophoretic Mobility Shift assay
ENIGMA	Evidence-based Network for the Interpretation of Germline□Mutant Alleles
ER	Estrogen Receptor
Etbr	Ethidium bromide
FPLC	Fast Protein Liquid C□ro□atography
GST	Glutathione S-Transferase
HBOC	Hereditary Breast and Ovarian Cancer Syndrome
HDAC	Histone DeACetylases
HIC	Hydrophobic Interaction Chromatography
HR	Homologous Recombination
IARC	International Agency for Research on Cancer
IPTG	Isopropyl β-D-1 thiogalactopyranoside
ITC	Isothermal Titration Calorimetry
IARC	International Agency for Research on Cancer
LB	Luria Bertani
LLG	Log Likelihood Gain
LOVD	Leiden Open Variation Database
LR	Likelihood Ratio
MALDI-TOF	Matrix-Assisted Laser Desorption and Ionisation-Time Of Flight

MDC1	Mediator of DNA Damage Checkpoint -1
MR	Molecular Replacement
MRE11	Meiotic Recombination 11 homolog
MRN	Mre11-Rad50-Nbs1 complex
MS	Mass Spectrometry
NHEJ	Non-Homologous End Joining
NMR	Nuclear Magnetic Resonance
OPLS	Optimized Potentials for Liquid Simulations
PARP1	Poly (ADP Ribose) Polymerase 1
PCR	Polymerase Chain Reaction
PolyPhen2	Polymorphism Phenotyping v2
PR	Progesterone Receptor
RAP80	Receptor Associated Protein-80
RB	Retinoblastoma protein
RING	Really Interesting New Gene
RMSD	Coot Mean Square Deviation
RNA	Ribonucleic acid
RNAse E	Endoribonuclease E
RNF	RING finger containing protein
SDM	Site directed mutagenesis
SDS	Sodium Dodecyl Sulfate
SDS PAGE	Sodium Dodecyl Sulfate Poly-Acrylamide Gel electrophoresis
SEC	Size exclusion Chromatography
SIFT	Sorting intolerant from tolerant
SNAP	Screening for Non Acceptable Polymorphisms
ssDNA	Single-stranded DNA
TEV	Tobacco Etch Virus protease
TGF-β	Transforming growth factor beta
TNF	Tumor necrosis factor
TOF	Time of Flight Mass Spectrometry
To□ BP1	Topoisomerase (DNA) II Binding Protein 1
UBC	Ubiquitin Conjugating Enzyme
UV	Ultra Violet
VEGF	Vascular endothelial growth factor
ZBRK1	Zinc Finger And BRCA1-Interacting Protein with a KRAB Domain

LIST OF FIGURES

LIST OF TABLES

Chapter 1

Review of Literature

1.1 Cancer

Cancer is the second most deadly disease that occurs due to uncontrolled division of cells, and this uncontrolled division of cells leads to formation of tumour, which may be either benign or malignant. A benign tumour is a mass of cells that lacks the ability to invade the neighbouring tissue, a process that is described as metastasis. Although benign tumours do not metastasize they may cause nerve damage, ischemia, necrosis and organ damage. However, malignant tumours multiply aggressively, degrade the extracellular matrix components, migrate and colonise at distant organs. The reasons for uncontrolled division of cells are attributed to mutations and deregulation in genes that play a crucial role in maintaining the genetic integrity and homeostasis. The genes involved in multiplication and maintenance of genomic integrity include: (1) Tumour suppressor genes such as RB **(R**etino**b**lastoma), TP53 (**t**umor **p**rotein p53), APC (**A**denomatous **p**olyposis **c**oli**)**, (2) Proto-oncogenes such as RAS (**Ra**t **s**arcoma), Myc (avian **My**elo**c**ytomatosis viral oncogene homolog), RAF (**R**apidly **A**ccelerated **F**ibrosarcoma), and (3) Stability/ caretaker genes, which include BRCA1/2 (breast cancer type 1/2 susceptibility protein), ATM (**A**taxia **t**elangiectasia **m**utated), and MLH1(MutL homolog1). The mutations in the tumour suppressor genes are recessive in nature, and hence it requires mutations in both the copies of the genes to have an effect. A proto-oncogene can become an oncogene due to mutations or increased expression, and is responsible for sensing and repairing the DNA damage. Stability genes control the rate of mutation as they are responsible for sensing and repairing the DNA damage [20]. In the process of tumorigenesis, cells must acquire at least six different properties to become abnormal or malignant by overcoming all the regulatory mechanisms [21]. These six basic properties, known as ‘hallmarks of cancer’, for malignant transformation are as follows:

1.1.1 Self-sufficiency in growth signal

Normal cells enter into the proliferative state via paracrine and systemic endocrine signalling. Normal cells can replicate only in the presence of mitogen signals and proper substratum whereas the tumour cells show a reduced dependence on the external mitogen signal. The common mechanisms of tumour cells for this reduced dependence on mitogen signal are as under: (1) Alterations in extracellular growth circuit; overexpression of cell surface receptor enables normal cells to become over responsive to growth stimulatory signals, for example HER2/ neu overexpression in mammary and stomach carcinomas [22]. (2) Modification of intracellular growth circuits. Ligand-independent signalling is caused due to overexpression and structural alterations, e.g. Ras protein in human colon cancer. (3) Autocrine signalling which creates a positive feedback loop. Cancer cells acquire the competence to synthesise the growth factor to which they are responsive, e.g. platelet-derived growth factor in glioblastomas [23,24].

1.1.2 Insensitivity to antigrowth signals

In a normal situation, cells maintain tissue homeostasis by responding to the anti-proliferative signals or by entering into quiescent phase. The anti-proliferative signals include both soluble growth inhibitors and immobilised inhibitors of an extracellular matrix. Growth inhibitory signals are transmitted through proteins like retinoblastoma (pRb) (**Figure1.1),** p107 and p130 [25]. To constitutively multiply, cancer cells turn off expression of integrin and other cell adhesion molecules. Tumour cells also use various strategies to avoid terminal differentiation, for example, Myc in association with Max elicits growth stimulatory signal while Max-Mad complex elicits differentiation inducing signal. In tumour cells, Myc is overexpressed shifting the equilibrium towards Myc-Max complex promoting growth [26]. Similar examples are seen in colon carcinogenesis and avian erythroblastosis.

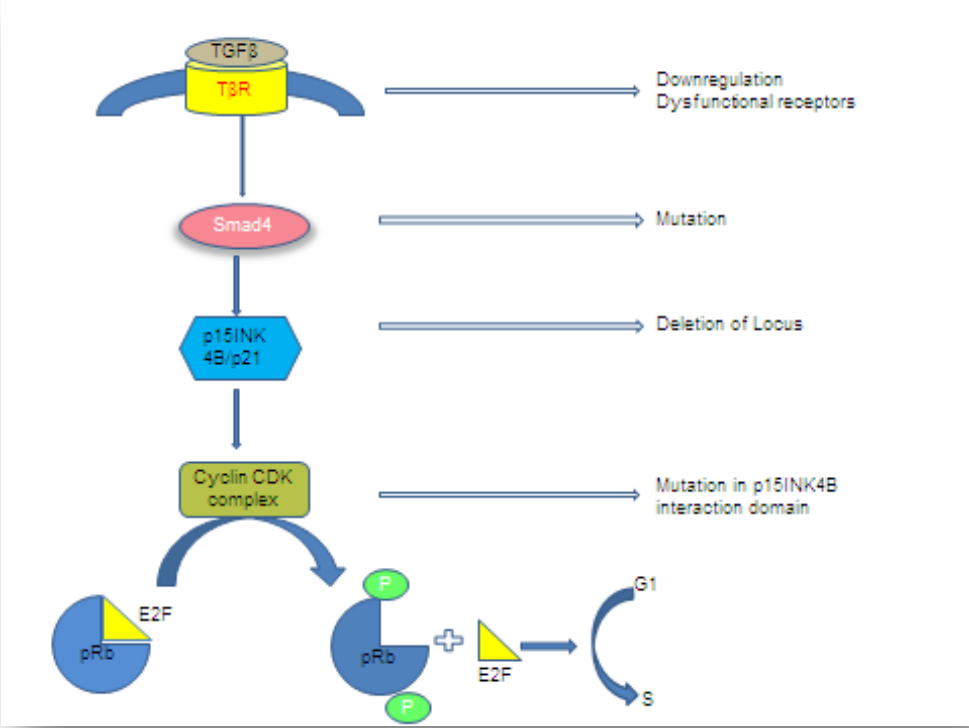


***Figure 1.1: Signal transduction of transforming growth factor beta (TGF-β) signalling**: TGF-β ligand binds to TβR receptor and activates its serine-threonine kinase activity. Activated TβR further activates Smad4and transduces the signal from receptor to p15INK4b/p21 which in turn activates cyclin-CDK kinase activity. The cyclin-CDK kinase then phosphorylates pRb which liberates E2F. This liberated E2F now controls the expression of genes essential for G1-S progression. Arrows indicate different ways in which the pathway may be hindered.*

1.1.3 Evading apoptosis

Apoptosis is a process of programmed cell death where cells commit suicide in a defined pattern either to help for development in the case of survival factor deficiency or in irreparable DNA damage. Apoptosis in cells is initiated by sensor molecules which show their effect through effector molecules via mitochondrial or direct signal transduction pathways. In mitochondrial pathway, Bcl-2 family proteins that include pro-apoptotic (Bax,Bak, Bid, Bim) and anti-apoptotic (Bcl-2, Bcl-XL, Bcl-W) proteins, act by governing

mitochondrial death signal through cytochrome C. Direct signal transduction occurs via TNF pathway or Fas ligand pathway. Mutations and deregulation of proteins of the apoptotic pathway lead to cancer. Up-regulation of Bcl-2 is shown to be responsible for follicular lymphoma [27], overexpression of survivin has been reported in several cancers [28], and reduced expression of CD95 has been reported in neuroblastoma and lymphoma [28].

1.1.4 Limitless replicative potential

Mammalian cells carry an intrinsic, cell-autonomous programme that decides the limits of their multiplication described as Hay flick limit. The Hay flick limit/ senescence is defined as the number of times a normal human cell divides until cell division stops [29]. This Hay flick limit is determined by the stretch of nucleotide sequence present at the end of chromosomes known as telomeres. Telomeres are a 6 base pair sequence repeated thousands of times at the end of chromosomes. During replication, the DNA polymerase is not able to copy a short stretch of DNA present at the end of chromosome leading to telomere loss in each replicative cycle. This shortened telomere is not able to protect the chromosomal ends leading to apoptosis. However, this process of telomere erosion does not take place in malignant cells [30].The malignant cells are able to maintain their telomere length in successive cell-cycle by two different ways: 1) Up-regulation of enzyme telomerase which adds hexanucleotide to telomere ends. 80 to 90% of tumours show up-regulation of telomerase activity. 2) Through recombination-based interchromosomal exchanges of sequence information.

1.1.5 Sustained angiogenesis

Angiogenesis is a process of formation of new blood vessels from the pre-existing blood vessels. In normal tissues, oxygen and nutrients are supplied by vasculature for cell function. Positive and negative signals, extracellular proteases and integrin’s exist for angiogenesis

regulation. Angiogenesis initiating signals are exemplified by vascular endothelial growth factor (VEGF) and fibroblast growth factors (FGF1/2) through transmembrane tyrosine kinase receptors displayed by endothelial cells [31]. Increased expression of VEGF and/or FGFs and downregulation of angiogenesis inhibitor thrombospondin-1 or β-interferon is demonstrated in many tumours.

1.1.6 Tissue invasion and metastasis

Metastasis is the spread of cancer cells from one organ to another. Cancer cells migrate to other secondary sites of the body through lymphatic or blood stream, and this phenomenon is known as haematogenous /lymphatic spread [32].The most common sites for metastasis include lung, liver, bone and brain. For the cells to metastasise from the site of the tumour to secondary sites they need the ability to separate them from neighbouring cells and pass through the extracellular matrix component. Therefore, cell-cell adhesion molecules like, E-cadherin and integrin are altered in metastatic cells.

To meet the increased nutritional requirement of rapidly dividing tumour cells reprogramming of energy metabolism and evasion of the immune response are now considered as additional hallmarks to the above list. To fulfill their energy requirement, cancer cells exhibit increased expression of glucose transporter genes (Glut1 and Glut3) to facilitate the import of glucose [33]. Although information on evasion of the immune response by cancer cells is still rudimentary, certain reports suggest that cancer cells may paralyse infiltrating cytotoxic T lymphocytes and natural killer cells by secreting immunosuppressive factors [34]. Transplantation experiments in mice suggested that only weakly immunogenic cells are able to grow and form the solid tumour.

1.2 Types of cancer

Cancers are grouped into broader categories based on the site of origin (**Table1.1)**.

Table 1.1: *Nomenclature of cancer based on site of origin*

Sr.no	Cancer	Site of origin
1	Carcinoma	skin or in tissues
2	Sarcoma	connective/supportive tissue
3	Leukemia	Bone marrow
4	Lymphoma/myeloma	Immune system.
5	Central nervous system	Brain and spinal cord.

1.3 Cancer Statistics

Cancer is the second most common disease in India responsible for about 0.3 million deaths per year [35]. Rates of cancer deaths in India are about 40% lower in adult men and 30% lower in women than in USA or UK. In India, in the year 2010, more than 556,000 cancer deaths were estimated for people of all age. 71% of cancer deaths occurred in the age group of 30–69 years. The three most common fatal cancers reported in this age group are oral-22·9%, stomach-12·6%, and lung-11·4% in men, and cervical-17·1%, stomach-14·1%, and breast-10·2% in women. [36].

Table 1.2: *5-year relative survival rate at different stages of breast cancer*

Sr. No.	Stages	5-year relative survival rate
1	localized stage (no spread to lymph nodes or other locations	99%
2	If cancer has spread to tissues or lymph nodes under the arm (regional stage)	84%.
3	If the spread is to lymph nodes around the collarbone or to distant lymph nodes or organs (distant stage)	24%

Breast cancer is one of the leading causes of cancer death in women. 232,670 new cases of invasive breast cancer are expected to be diagnosed among women and about 2,360 new cases are expected in men in the US during 2014. The 5-year survival rate by race for breast cancer is 90% for white women and 79% for African American women [37] (**Table 1.2**).

As per records from National Cancer Registry, breast cancer is common cancer in urban Indian females and the second most common in the rural Indian women. A case-control study in Mumbai indicated that nulliparous women had a 2.2-fold higher risk than parous women for developing breast cancer in the age group of above 40 years [38,39]. Premenopausal breast cancer patients in India constitute ~ 50% of all patients (SGPGIMS Lucknow data). The proportion of Indian breast cancer patients, younger than 35 years varies between 11% (Tata Memorial Hospital (TMH) Mumbai) to 26% (SGPGIMS Lucknow) [40]. A study from hospitals in Mumbai, Chennai and Thiruvananthapuram shows that breast cancer rates were reported to be highest among Christians and Parsis [41,42]. Incidence of breast cancer is predicted to increase by two-thirds in 2025 among older women [43].

1.4 Breast cancer

1.4.1 Anatomy of breast

The breast is an organ that has the function of producing milk during the period of lactation. It is located anteriorly to the pectoralis major muscle of the thoracic wall. It extends anteriorly up to second rib, posteriorly up to sixth or seventh ribs. The longitudinal section of the breast is divided into 5 parts 1) Lobes and Lobules, 2) Ducts, 3) Areola, 4) Nipple and 5) Fatty, connective tissue and lymph nodes (**Figure 1.2).** The different lymph nodes present in the breast, and which help in understanding disease progression are supraclavicular, infraclavicular, axillary and internal mammary nodes. An abnormal mammogram includes

lump in the breast, swelling in the armpit, abnormal nipple discharge and change in the size, contour, texture, or temperature of the breast.

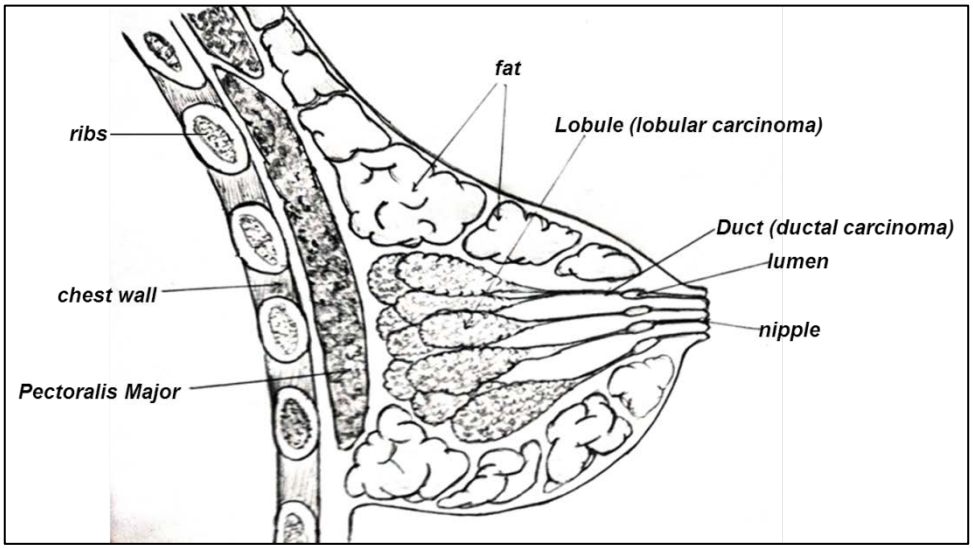


Figure 1.2: *Longitudinal section demonstrating the anatomy of breast*

1.4.2 Types of breast cancer

Two different types of breast cancers are described: 1) sporadic or non- inherited cancer, and 2) inherited cancer. Sporadic breast cancer occurs at the late stage of life and is due to environmental factors while familial breast cancer occurs at a relatively early age, and is due to inheritance of defective cancer susceptibility genes i.e. BRCA1/2. Mutations in a number of genes other than BRCA1/2 also cause inherited breast cancer (**Table 1.3).** Inheritance of mutation in other different genes has 20% lifetime risk of developing cancer [44].

a. Sporadic breast cancer

Parity, hormones, environment and diet are the major factors that play an important role in the development of sporadic cancer [26]. 90 % of breast cancer cases are sporadic. Sporadic breast cancer occurs late in life, is mostly unilateral cancer, and increases by 1.2 fold the susceptibility to cancer of first-degree relatives [27]. Cytoplasmic localisation of Δ11b splice variant with reduced production of full-length BRCA1 has been demonstrated in sporadic

tumour [28]. In sporadic breast and ovarian cancer cases, BRCA1 was found to be down regulated by epigenetic silencing [29] and transcriptional repression. Over expression of HMGA1 that represses transcription of the BRCA1 gene leading to downregulation of BRCA1 is observed in aggressive mammary carcinoma [30]. Hyper methylation of BRCA1 promoter leads to reduced BRCA1 expression interfering with the functions of BRCA1 protein. Promoter hyper methylation in BRCA1 gene account for 11 to 33% of sporadic breast cancer cases [31]. All these observations suggest that BRCA1 has a key role in the development of sporadic cancers also.

Table 1.3: *High penetrance breast cancer susceptibility genes*

Sr No	**Gene**	**Syndrome associated with gene**
1	BRCA1/ 2	Hereditary breast and ovarian cancer
2	TP53	Li-Fraumeni syndrome
3	PTEN	Cowden syndrome
4	STK11	Peutz-Jeghers syndrome
5	CDH1	Familial diffuse gastric cancer
6	ATM	Ataxia-telangiectasia
7	NBS1	Nijmegen Breakage syndrome

b. Inherited breast cancer

Hereditary breast and ovarian cancer syndrome (HBOC) is mostly attributed to mutations present in BRCA1/2 genes. About 3-5% cases of breast cancer and 10% of ovarian cancer are estimated to be due to mutations in BRCA1/2 genes [1]. Carriers of even a single missense mutation in BRCA1/2 may have a 50-80% lifetime risk of developing breast cancer, and 30-50% risk of developing ovarian cancer [2]. The mutations reported in BRCA1 include frame shift, missense, insertion and deletion of bases. The prevalence of the deleterious mutation in BRCA1/2 genes in the general population is estimated to be 1:140 to 1:800 [45,46]. Founder

mutations are also reported in certain populations. Patients with mutations in BRCA1 gene generally show triple negative phenotype (triple negative breast cancer cells do not express estrogen receptor (ER), progesterone receptor (PR) or Her2/neu; hence cannot use drugs that target these receptors). BRCA1 mutations show an excess of medullary/atypical medullary carcinoma, high mitotic count pushing tumour margins and a lymphocytic infiltrate [47,48]. The inherited breast cancer, therefore, is of higher grade than sporadic cancer.

1.5 BRCA1; an Overview

1.5.1 Domain organisation of BRCA1

BRCA1is encoded by the gene which is present on the chromosome 17 q21, and contains 24 exons. It is a 220 kDa protein that comprises of 1863 amino acid residues [3,4]. BRCA1, a tumour suppressor protein has two reported domains, an N-terminal RING finger domain and the C-terminal BRCT domain [5], and an intrinsically unstructured central region.

The RING (Really Interesting New Gene) finger domain at the N-terminus is encoded by exons 2-7 encompassing (1-109) amino acids [49]. This RING finger domain interacts with BARD1 and forms a BRCA1/BARD1 heterodimer, which possesses E3 ubiquitin ligase activity. This heterodimer is known to specifically ubiquitylate lysines at 127 and 129 of histone H2A in nucleosomal context upon DNA damage [50]. Pathogenic mutations reported in the RING finger domain abrogate the interactions with BARD1, and further impair the ubiquitin activity [51,52].

The C-terminal region of BRCA1 has two tandem repeats of BRCT (BRCT1 and BRCT2) motifs, which are also found in many DNA repair proteins. The BRCA1 BRCT is a Class-I BRCT domain which recognises phospho-serine residue. The BRCA1 BRCT domain recognizes the consensus sequence containing pSer-X-X-Phe (pSer-phosphoserine, X-any

amino acids) in its binding partners. The BRCA1 BRCT domain interacts with a number of proteins including CtIP [10], BACH1, and CCDC98/Abraxas [53,54].

The central region is not well ordered structurally and contains two nuclear localization signals located in the exon 11, and a 'SQ' cluster between amino acids 1280–1524. The central region of BRCA1 interacts with different DNA repair and cell cycle regulatory proteins. It interacts with retinoblastoma protein (RB) [55], c-Myc [56], Rad50 and Rad51 [57,58]. The SQ/TQ cluster domain or (SCD-serine cluster domain) contains multiple sites phosphorylated by ATM (ataxia-telangiectasia mutated). The phosphorylation of BRCA1 by ATM after DNA damage is essential for cell cycle checkpoint activation [59,60] **(Figure 1.3).**

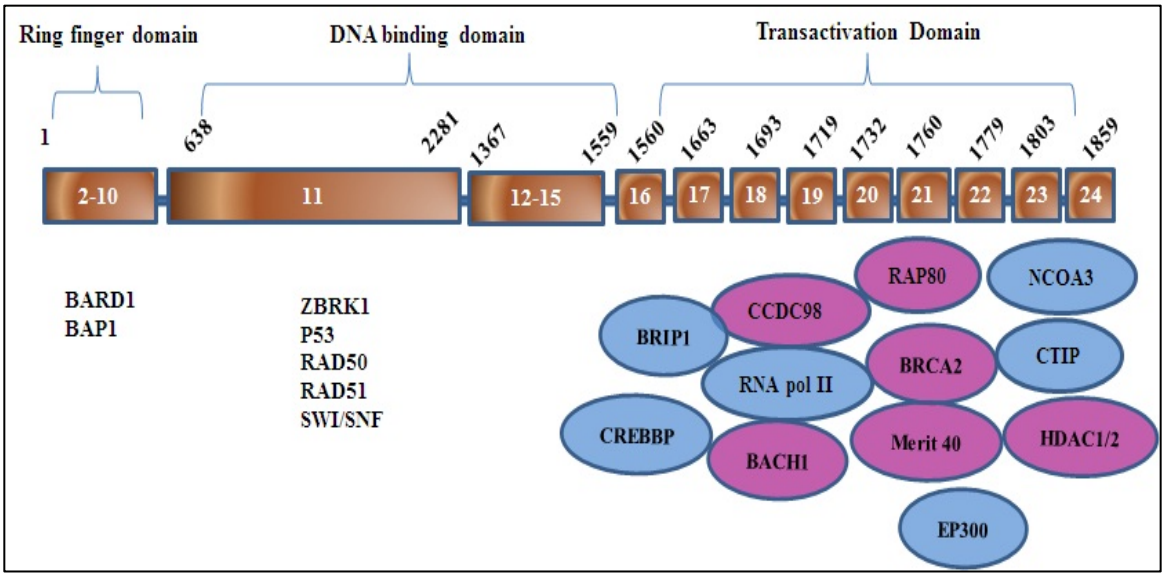


Figure 1.3: Domain organisation of BRCA1 and its interacting partners

1.6 Interacting partners and functions of BRCA1

BRCA1 protein performs various functions like transcription activation, transcription repression and DNA damage repair by interacting with multiple cellular binding partners (**Table 1.4).** Following are few of the reported interacting partners, the region of BRCA1 with which they interact, and function the complex performs post- interaction.

Table 1.4: *Interacting partners of BRCA1*

BRCA1 interacting region	**Interacting protein partner**	**Function of the complex**	**Reference**
C-terminus (BRCT)	CtIP	Transcription repression	[61]
AD1	Jun	Transactivation	[62]
N-terminus (RING)	ATF1	Transactivation	[63]
N-terminus (RING)	BARD1	Ubiquitin ligase activity	[64]
Middle	Rad51	Homologous recombination	[58]
C-terminal	RNA Pol II	Transactivation	[65,66]
C-terminal	RNA helicase A	Transactivation	[67]
C-terminus	p53	Act as a scaffold protein for phosphorylation of p53, act as p53 co-activator	[68,69]
Central	STAT1	Transactivation	[70]
Central	c-myc	Transcription repressor of c-myc	[71]
Central	ZBRK1	Transcription co-Repressor	[72]
C-terminal	BACH1	DNA Repair	[73]
C-terminal	FHL2	Enhances transactivation function of FHL2	[74]
C-terminal	AKT	Regulates AKT activity through ubiquitination	[75]
C-terminal	DBC1	Represses the transcriptional activation function of BRCT	[76]
N-terminus	BAP1	Enhances BRCA1-mediated cell growth suppression	[77]
N-terminus	E2F	Regulates cell cycle	[78]

1.6.1 Transactivation functions of BRCA1

The C-terminal region, comprising exons 16-24 (amino acids 1560-1863) of BRCA1 can activate transcription, both in yeast and mammalian cells. Deletion mutations of exons 16-24 of BRCA1 demonstrated that the region between exon 21-24 (amino acids 1760–1863) is the minimal transactivation domain that activates transcription, although less efficiently

compared to larger fragment encoded by exons 16-24. Germline mutations (A1708E, R1699Q, M1775R, Y1853Stop) found in patients with breast or ovarian cancer, abolish the transcription activity, strongly supporting the idea that BRCA1 protein is involved in transcription. Further studies in yeast and human cells demonstrated a new (AD1) activation domain (amino acids 1293–1558) in BRCA1 that can function cooperatively with the (AD2) activation domain 2 (amino acids 1560-1863). Further, the study established that AD1 activity is more dependent on cellular environment, whereas AD2 activates transcription in a ubiquitous manner [4]. BRCA1 via a coiled-coil motif in AD1 interacts with the basic leucine zipper (bZIP) region of the Jun proteins. The Jun-interacting domain in BRCA1 is critical for AD1-mediated transcriptional activation. In particular, the strength of AD1 in transcriptional activation is limited by the JunB level as explained by the observed correlation between ectopic expression of Jun B and the transcriptional activity of AD1. Furthermore, it has been observed that Jun B mRNA expression is down-regulated in many ovarian tumour tissues examined. Thus, the BRCA1 and Jun B interactions may facilitate the tumour suppression function of BRCA1 in the tissue-specific manner [62]. BRCA1 is also known to stimulate transcription machinery directly without the requirement for DNA-tethering partners. It could stimulate transcription of the p53-responsive promoter, without p53, when overexpressed. BRCA1 does not require functional TATA box sequence to stimulate transcription. The reported results, indicated the role of BRCA1 in the stabilisation of transcription machinery [79]. Results on direct transactivation by BRCA1 have been demonstrated, in the expression of $p27^{Kip1}$, a member of the universal cyclin-dependent kinase inhibitor family. The transactivation domain of BRCA1 binds to the BRCA1-responsive element located at position -615 to -511 of the $p27^{Kip1}$ promoter. These results suggest a mechanism for BRCA1-induced growth inhibition through transcriptional regulation of $p27^{Kip1}$ [80]. BRCA1 transactivates the expression of the cyclin-dependent kinase inhibitor p21 (WAF1/CIP1).

Trans-activation deficient mutants of BRCA1 are defective in both transactivation of p21 and cell-cycle inhibition [81]. The BRCT 2 of BRCA1 is reported to interact with LIM domains of FHL2 (four and a half LIM-only protein 2) and enhance the transactivation function of FHL2. FHL2 mRNA levels are down regulated in many breast cancer cell lines indicating its growth regulatory function [74]. In conclusion, BRCA1 acts both as a direct as well as an indirect transactivator protein, which performs its function in cell-specific and ubiquitous manner.

1.6.2 Transcription repression functions of BRCA1

BRCA1 regulates AREG transcription directly by binding to the AREG promoter. It binds to BRCA1-response elements located at the positions −202/−182 and +19/+122 in the AREG promoter. BRCA1 is responsible for the repression of EGR1, indirectly by associating with ZBRK1 [82]. BRCA1 BRCT interacts with CtIP, which associates with the CtBP transcriptional co-repressor. Tumour-associated mutations discovered in BRCA1 BRCT motifs abrogates interaction with CtIP, suggesting BRCA1-CtIP interaction is necessary for tumour suppression [61]. Study of the role of BRCA1 in the regulation of estrogen receptor-α has revealed the underlying basis for the tissue-restricted tumor-suppressive properties of BRCA1. In BRCA1-deficient human ovarian cancer cells estrogen receptor- α exhibited ligand-independent transcriptional activity that was not observed in BRCA1-proficient cells [83]. BRCA1 is also able to perform repression function in ubiquitin-dependent manner. In a study by Horwitz et al, a unique BRCA1 ubiquitin-dependent mechanism of repression is demonstrated. BRCA1 is reported to ubiquitinate the transcriptional pre-initiation complex. The ubiquitin moiety interferes with the assembly of basal transcription factors at the promoter, and thus blocks the initiation of mRNA synthesis [84]. The BRCA1-BRCT binds to phosphorylated Akt leading to Akt ubiquitination by BRCA1 N-terminal E3 ligase activity

and ultimately to degradation. Mutant cells lacking the BRCA1 BRCT repeats accumulate nuclear pAkt and consequently inactivate the transcription functions of FOXO3a. Phosphorylated FOXO3a is translocated out of the nucleus. FOXO3a is responsible for upregulation of proapoptotic genes Bim and PUMA and downregulation of anti-apoptotic proteins such as FLIP [75]. HP1γ regulates BRCA1 mediated transcription via modulation of promoter occupancy and histone modification. Time-lapse studies on promoter association and histone methylation revealed that HP1γ occupies the promoter, in the quiescent unstressed state to repress transcription. In case of DNA damage, BRCA1 is recruited to the promoter site while HP1 γ is disassembled. Once the damaged DNA is repaired, HP1γ/SUV39H1 is restored at the promoter leading to BRCA1 disassembly and histone methylation, after which transcription repression resumes [85].

1.6.3 BRCA1 in DNA repair

The developmental defects, the embryonic lethality of BRCA1 null mice and chromosomal abnormalities in cell lines deficient in BRCA1 clearly indicated the role of BRCA1 in maintaining genomic stability during cell proliferation [86,87]. Translocation of BRCA1 to DNA damage sites, co-localization with RAD51 and hypersensitivity of BRCA1-deficient cells to IR, UV and DNA alkylating agents confirm the role of BRCA1 in DNA repair.

a. BRCA1 is recruited to the DNA damage sites

ATM a double-strand breaks sensor protein, autophosphorylates on residue Ser1981 on DNA damage and releases active ATM monomers from inactive dimer. Further, it performs the function by interacting with protein like MRN complex, NBS1, MDC1 and MRE11 and phosphorylating downstream target [88]. ATR interacts with ATRIP and recognizes RPA coated single-stranded DNA. Activated ATR phosphorylates Chk1 ultimately leading to cell cycle arrest [89].

Upon DNA damage, H2AX is phosphorylated at serine 139 after ataxia telangiectasia mutated (ATM), ataxia telangiectasia and RAD3-related (ATR) proteins are activated [90,91]. Phosphorylated γH2AX mark the sites of DNA breaks which leads to recruitment of an E3 ubiquitin ligases-RNF8 and E2 ubiquitin conjugating enzyme 13 (UBC13/UBE2N) by MDC1. This complex exclusively catalyzes lys63-linked ubiquitination of histone and other substrate ubiquitylation reactions close to double-strand breaks (DSBs) [92]. These ubiquitin chains are recognized by the ubiquitin interacting motif (UIM) of RAP80, which recruits BRCA1 to double-stranded break site [93,94]. Phosphopeptide affinity proteomics analysis could identify the coiled-coil domain containing protein ABRAXAS (CCDC98), one of the binding partners of the BRCA1 BRCT domain that forms a part of the Rap80-ABRAXAS-BRCA1 complex. Abraxas phosphorylated at Ser 406 position interacts specifically with the BRCA1 BRCT [95]. Hence, Abraxas bridges the interactions between RAP80 and the BRCA1 BRCT domain in RAP80-ABRAXAS-BRCA1 complex and this mediates the localization of BRCA1 to the sites of DNA damage. Mutations affecting the BRCT domains of BRCA1 disrupt the recruitment of BRCA1 to DNA (DSBs) site.

BRCA1 also plays a role in all the three checkpoints G1/S, S-phase and G2/M checkpoint to maintain homeostasis.

b. G1/S checkpoint

The G1/S checkpoint arrests cells at G1/S boundary, following DNA damage. BRCA1 acts as a scaffold protein facilitating the phosphorylation of bound p53 by ATM and this further leads to p53-mediated induction of p21 and induced G1/S arrest. BRCA1 is required for both IR and UV-induced p53 phosphorylation by ATM/ATR. Downregulation of BRCA1 using si-RNA showed that BRCA1 depleted cells had defective G1/S checkpoint in response to DNA damage. BRCA1 was only required for IR-induced G1/S checkpoint, but not for UV-induced G1/S arrest [96].

c. S phase checkpoint

S-phase checkpoint stops DNA replication following DNA damage [6]. ATM and ATR are activated on DNA damage and these subsequently activate Chk1 and Chk2. Activated Chk1 and Chk2 regulate Cdc25 A/B/C, which governs the activity of cyclins and cyclin-dependent kinases during S phase progression [7]. Upon IR irradiation, ATM is known to phosphorylate BRCA1 at Ser 1387 which is required to recruit other binding partners to regulate Chk1 kinase activity. BACH1 is a part of BRCA1/BRCA2 complex that is required for timely progression through the synthetic phase of cell-cycle. BRCA1-deficient HCC1937 cells are defective in S-phase checkpoint. Cells with defective BACH1 helicase activity display enhanced genomic instability [97]. BRCA1 is also reported to interact and co-localise with Mre11/Rad50/NBS1 (MRN) complex, which is a sensor for DSBs, and also to directly activate ATM [8,9].

d. G2/M checkpoint

G2/M checkpoint repairs and prevents DNA lesions from being passed on to daughter cells. Phosphorylated and activated Chk1 and Chk2 phosphorylate mitotic kinase weel and Cdc25A/B/C proteins, which suppress the activity of cyclin B and Cdc2 and block the cells from entering into mitosis. BRCA1 regulates Chk1 kinase activity during G2/M checkpoint activation [10]. Here, ATM phosphorylates serine1423 of BRCA1 which is required for G2/M checkpoint activation. BRCA1 defective cells demonstrate a defective G2/M checkpoint action [11].

1.6.4 BRCA1 in Non-homologous end joining (NHEJ) and Homologous Recombination (HR) repair

NHEJ and HR are repair pathways for DSBs in DNA, and BRCA1 participates in both types of repair [98]. NHEJ is a relatively error-prone repair, unlike HR, which faithfully repairs the damage. NHEJ mostly occurs during the G1 phase of the cell cycle while HR mediated repair

works during the S and G2 phases of the cell cycle. BRCA1 binds to MRN complex that plays an important role in NHEJ and HR repair. BRCA1, through its BRCT domain, promotes HR by blocking the ability of DNA-PKcs to auto phosphorylation at the Serine 2056 position. Blocking auto phosphorylation of DNA-PKcs promotes HR-required DNA end processing and loading of HR factors to DSBs [99].

HR-mediated DNA repair requires the recombinase RAD51, a single strand DNA (ss DNA) binding protein. BRCA2 is responsible for recruitment and assimilation of RAD51 onto the ssDNA forming a nucleoprotein filament. PAlB2 is required to assist in chromatin loading of BRCA2 and RAD51 [100] and it also act as a bridging factor required for the BRCA1-BRCA2 association. The loss of BRCA1 interferes with HR- mediated repair, and the mechanism by which BRCA1 promotes HR through the PAlB2-BRCA2-RAD51 complex still remains unclear [101,102]. However, recent studies on β-HPV infections in skin cancer demonstrated the role of BRCA1 in homology dependent repair. Viral proteins β-HPV and E6 proteins destabilise histone acetyltransferase and p300, which interfere with transcription of BRCA1/2 that curtails the ability of these proteins to form repair foci at DSBs [103].

In NHEJ pathway, damage DNA ends are directly ligated without using an additional template which may lead to removal or addition of bases. BRCA1 is a part of the Mre11/Rad50/Nbs1 complex, which plays a role in double-strand break repair. BRCA1 is reported to suppress the nuclease activity of MRE11 and is required for phosphorylation of NBS1 following DNA damage [104,105]. Studies in BRCA1-deficient cell lines MEFs and HCC1937 have provided evidence for defective NHEJ [106,107]. However, previous conflicting results and indirect assays still leave a room to understand the exact mechanism of action of NHEJ pathway and a role of BRCA1 in the NHEJ pathway.

1.7 Regulation of BRCA1

BRCA1 being a multifunctional protein, a stringent control on the regulation of BRCA1 expression is essential for the cell to maintain genomic stability. Chromatin immunoprecipitation assay established the interactions of Myc with BRCA1 promoter which acts as a transcriptional activator of BRCA1. C-Myc activates BRCA1 via the E boxes present in the distal BRCA1 promoter region [108]. A repressive multi-component transcriptional complex containing E2F-1 RB and BRCA1 inhibits BRCA1 promoter transcription. Chromatin immune-precipitation studies demonstrated that this repressive complex is disrupted by DNA damaging agents. This disassembly releases the repression on BRCA1 promoter and leads to transcription of BRCA1 in response to genotoxic insult [109]. The aberrant methylation of the BRCA1 promoter in patient sample leads to decrease in expression of BRCA1 suggesting its role in sporadic mammary carcinogenesis [110] and other techniques. The different regions of BRCA1 have been expressed and purified to homogeneity. Structural details of each one of these were elucidated using circular dichroism spectroscopy and NMR. NMR spectra and CD for different central regions of BRCA1 displayed features of disordered proteins [17,111,112]. No crystal structure has been reported for the central domains of BRCA1 perhaps due to its disordered nature. BRCA1 DNA binding was visualised using electron microscopy which revealed the formation of nonspecific complexes involving multiple DNA strands and multiple BRCA1 molecules at a given site [104]. NMR solution structure of the complex of the N-terminal region of BRCA1 with the N-terminal region of BARD1 and X-ray crystal structure of C-terminal BRCA1 BRCT are reported till date. Co-crystal structures of BRCA1 BRCT with the interacting phosphopeptides are also reported.

1.8 Structure of BRCA1 protein

1.8.1 X-ray structure of BRCA1 BRCT

BRCA1 protein at its carboxyl region has tandem repeats of two BRCTs. The crystal structure of BRCA1 BRCT has been determined to a resolution of 2.5Å (PDB ID: 1JNX). The study revealed that the molecule is shaped like a cylinder of length ~70 Å and ~30-35 Å diameter. Furthermore, the two BRCTs are arranged in the head to tail fashion **(Figure 1.4).** The BRCT fold has central parallel four-stranded β-sheet, with a pair of α-helices (α1 and α3) packed against one face of this sheet and a single α-helix (α2) packed against the opposite face of the sheet. The core of this interface between the BRCT repeat is formed by α1′ and α3′ from the C-terminal repeat, and α2 of the N-terminal repeat. The two tandem BRCTs are linked by a 23-amino acid linker which is poorly defined in the electron density, possibly indicating flexibility [113]. Crystal structures of BRCA1 BRCT in complex with oligopeptide derived from binding partners have been determined. These structures PDB IDs are listed in **Table 1.5.**

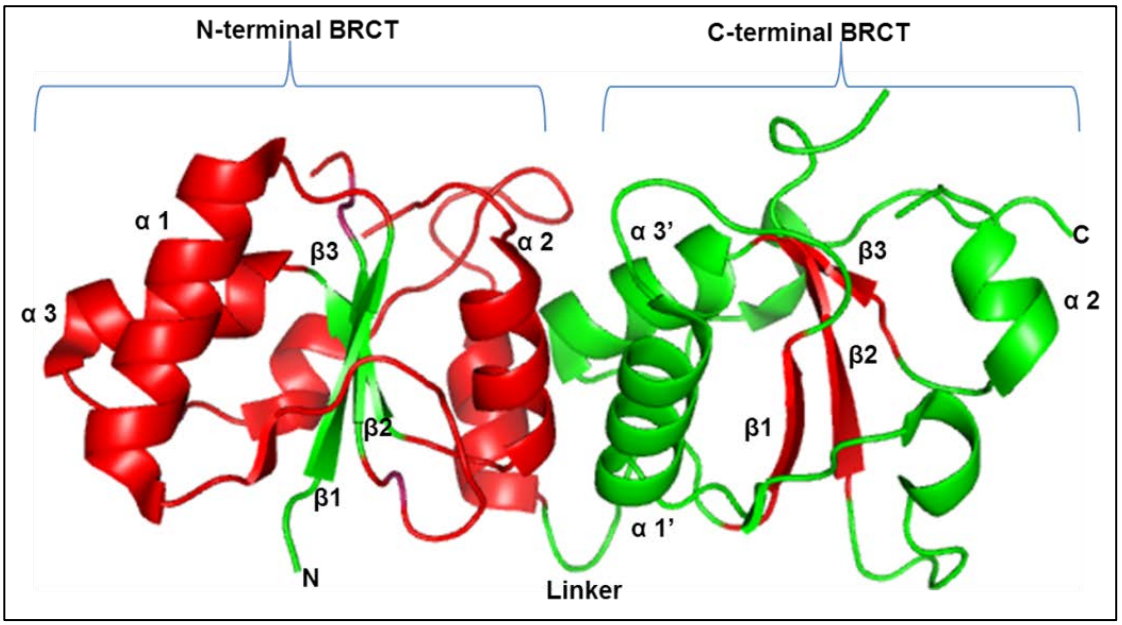


Figure 1.4: BRCA1 BRCT crystal structure

The BRCA1 BRCT has been co-crystallised with phosphorylated oligopeptides from binding partners containing the signature sequence pS_0-X_1-X_2-F_{+3} (where phosphorylated serine at 0

position, X can be any amino acid and phenylalanine residue at +3 position) [114]. The crystal structures give a detailed picture of the interaction between the signature motif and the protein. It has been observed that phosphopeptide interacts with BRCA1-BRCT via the pS_0 and F_{+3} residues.

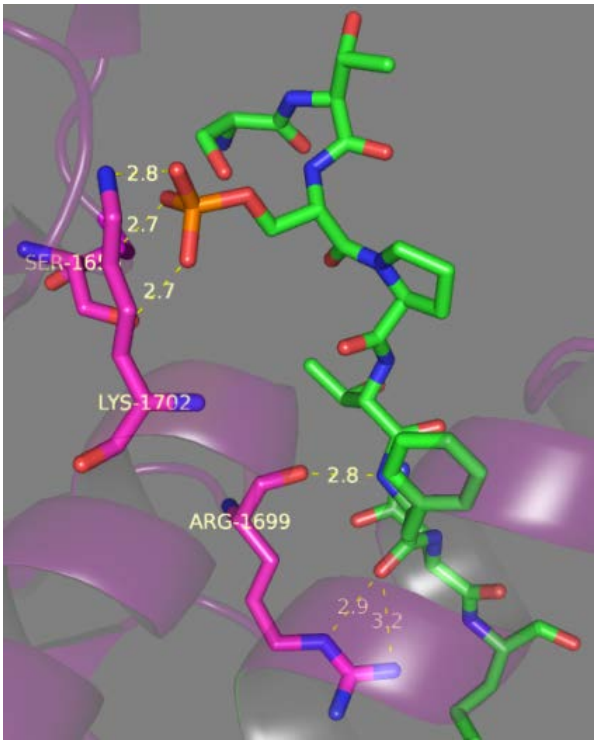


Figure 1.5: Peptide binding site in BRCA1 BRCT- BACH1 complex *(PDB ID: 1T15). Interaction between pSer and Phe residues of peptide forms network of hydrogen bonds*

The phosphate group of the phosphorylated serine in phosphopeptide interacts with residues from the N-terminal BRCT via the salt bridge and hydrogen bonding interactions. The backbone of Phe (+3) residue from the phosphopeptide is recognized by conserved Arg-1699 residue. Arg 1699 forms a strong hydrogen bond with the carbonyl oxygen of phenylalanine residue in the +3 position of the signature sequence. This interaction ensures proper positioning of the hydrophobic side chain of Phe in the hydrophobic pocket from BRCT domain [115]. Out of the reported crystal structures of the complexes, it has been observed that the association between pSer (0)-X-X Phe (+3) and BRCA1-BRCT is very much conserved [116].

Table 1.5: *List of PDB ID of complexes of BRCA1 BRCT and different phosphopeptide partners*

Sr.No.	PDB ID	Interacting peptide	Resolution
1	4JLU	Doubly phosphorylated Abraxas	3.5
2	4IFI	BAAT	2.2
3	4IGK	ATRIP	1.74
4	3COJ	ACC1	3.21
5	1Y98	CtIP	2.5
6	1T15	BACH1 helicase	1.85
7	1T29	BACH1	2.3

1.8.2 Structure of RING finger domain of BRCA1 BARD1

BRCA1 through its N-terminal region interacts with BARD1 and performs the ubiquitin ligase activity. Interaction between BRCA1 with BARD1 is also necessary to retain BRCA1 in the nucleus. Structurally the RING finger domain of BRCA1 is characterized by a central α-helix, two large Zn^{2+} binding loops and a short three-stranded antiparallel β-sheet (**Figure 1.6).** The RING motif has a conserved pattern of eight Cys and His residues arranged in pairs in the primary sequence that form zinc binding sites I and II [117]. The solvent exposed regions surrounding the second Zn^{2+} binding loop is the most variable part in the RING motifs of both BRCA1 and BARD1. The central RING motif of BRCA1 (residues 23-76) is flanked by antiparallel α-helices encoded by residues (8-22) and (81-96) of the BRCA1. In BRCA1, the side chains of Gln 19, Leu 22, Val 83, Glu 84 and Leu 87 in the helices are in contact with Leu 30, Ile 31 and Lys 32. Similarly, in BARD1, central RING motif (residues49-100) is surrounded by helices formed by residues (36-48) and (101-116) in the BARD1. Likewise, in BARD1 the side chains of helix residues Glu 45, Leu 48 and Ile 105

interact with Ile 56 and Leu 57 [118]. These helices combine as a four-helix bundle to form a BRCA1-BARD1 heterodimer complex interface.

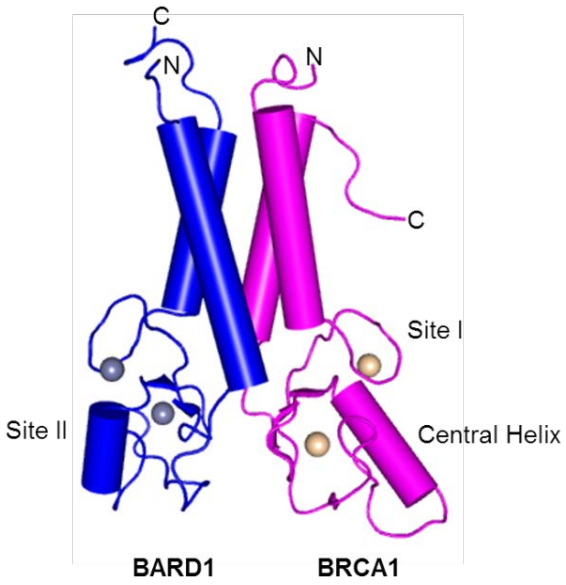


Figure 1.6: Structure of BRCA1-BARD1 heterodimer complex (PDB ID 1JM7)

1.9 Missense mutations in BRCA1 accountable for HBOC syndrome

Germline mutations reported throughout 1863 amino acids of BRCA1 are responsible for hereditary breast and ovarian cancer (HBOC) syndrome. Missense, indels, truncating and frame shift mutations are reported in BRCA1. While all the mutations alter the primary sequence of the protein, missense mutations identified alter a single residue, thus posing a challenge to classify the variant as a deleterious or just polymorphic variant (**Table 1.6)**. Such unclassified variants are known as variants of uncertain significance (VUS). Breast Cancer Information Core (BIC) Database http://research.nhgri.nih.gov/bic/) has all the information regarding different kinds of mutations reported in BRCA1/BRCA2 and their classifications. Of the thousands of mutations reported in BIC database, 90% are variants of uncertain significance. Classification of the variants as deleterious or neutral is necessary for genetic testing and surveillance recommendations. Classification of VUS as either deleterious or polymorphic is

done on the basis of frequency of occurrence of variants, linkage analysis, clinical, functional and structural studies.

Table 1.6: Mutations reported in different regions of BRCA1

Different region of BRCA1 *	**No. of mutations reported**	**Missense mutations**
1-109	56	M18K, C24R, C64R, D96N, G98R, A102G
110-200	23	E111A, I124V, S157P, R170W, Y179C, D201N
200-340	38	R213G, H239R, L246V, S264R, G275S, T276R
341-758	109	K355R, Q356R, S377N, R504C, D522N, N609S
758-1064	85	R979H, M1008I, S1027N, R1028H, S1040N, E1060A
1065-1169	30	Q1069R, G1087A, V1088D, S1101N, N1102S, S1140G
1170-1270	39	R1203Q, P1238R, S1241Y, T1242A, K1254E, N1259T
1271-1366	22	A1293D, S1301N, D1337E, E1346K, R1347G, R1347K
1367-1467	24	S1377R, S1389N, Q1395R, M1400T, L1404P, H1421R
1468-1641	58	P1544L, D1546N, S1613C, M1628I, M1628T, S1631N
1642-1736	54	M1689R,V1696L, C1697R, R1699Q, S1722F, S1722P
1737-1755	12	G1738E, V1741G, H1746N, P1749R, R1751Q, A1752P
1756-1863	50	L1764P, M1775R, E1836K, W1837C, P1859R, H1862L

* indicates different functional regions selected for analysing the different mutations reported.

To calculate the likelihood ratio for these variants being deleterious each of the following independent observations and criteria were considered: 1) co-occurrence of variants with deleterious mutations in trans, have a very low likelihood of being deleterious, as homozygosity for BRCA1 pathogenic mutations are embryonically lethal, 2) co-segregation of variants with the disease in family more often than expected by chance, 3) proband and family history of cancer which considers age at cancer diagnosis, number of members affected, and 4) pathological characteristics like tumour grade and ER/PR status [12]. The four likelihood ratios obtained are then multiplied to get the joint likelihood value. These final likelihood values are used to calculate the posterior probability, considering the prior probability value obtained from Align GVGD. The value of the posterior probability was used to place the variants in different IARC (International Agency for Research on Cancer) classes and also to define the protocol for genetic testing and surveillance recommendations [13].

Each of the above clinical classification methods is dealt in detail in chapter 3 of the thesis. This IARC class further provides guidelines for genetic testing. **L**eiden **O**pen **V**ariation **D**atabase (LOVD) database includes the entries of all the missense mutations along with IARC classification and available information of functional data (http://brca.iarc.fr/LOVD/variants.php?action=view_unique&select_db=BRCA1). This IARC classified variant is further validated with in-vitro experiments for genetic testing and surveillance recommendations [13]. Since the variants reported in BRCA1 are very rare and are ethnic specific, getting information that is statistically significant poses major challenges in classifying the variants by focusing only on the clinical data. Therefore, a study on the functional impact of the variants should be carried out by transactivation assay, small colony assay, rescue of radiation resistance and ubiquitin ligase assay. Structural changes at the atomic level have been revealed by using the techniques like X-ray crystallography and

NMR. The biophysical analyses of the variants are also being exploited to understand the effect of variants on their biochemical properties and on the nature of interactions with interacting partners. In-silico approach considers sequence conservation, charge, bulkiness, solvent accessibility and structural alterations. Using any of the above mentioned approaches in isolation has its own advantages and disadvantages [12].

1.10 Conclusion

Breast cancer ranks second as a cause of cancer death in women worldwide. The mutations can be either hereditary or sporadic. ~ 90 % of breast cancer cases are sporadic in nature [119]. About 3-5% of familial breast cancer and 10% of ovarian cancer are estimated to be due to mutations discovered in BRCA1/2 [1]. Carriers of an inherited missense mutation in BRCA1, have a 50-80% lifetime risk of developing breast cancer, and 30-50% risk of developing ovarian cancer [2]. Classification of a given missense variant as either deleterious or polymorphic determines the nature of recommendation for genetic screening and surveillance of the patients and their relatives. The present study aims at the clinical classification of the novel missense variants of BRCA1 discovered in TMC cohort of the Indian population. The study further deals with the structural, functional and biophysical characterization of the different functional domains of wild-type/ variant BRCA1 and its binding partners.

Chapter 2

Materials and Methods

2.1 Materials

i. Vectors used

Pet41a+, pGEX-KT, pRSET-A, pGEX-4Tand pGBKT7 are the vectors used for protein expression. Complementary DNA (cDNA) of BRCA1 is a generous gift from Prof. Richard Baer, Columbia University. pGEX-KT vector and pET-TEV construct a generous gift from John. Ladias, BIDMC, Eric Sundberg, BBRI, US. pGBKT7 is a kind gift from Dr. Alvaro Monterio of H. Lee Moffitt Cancer Centre & Research Institute, USA.

ii. Molecular biology reagents

Plasmid DNA extraction, DNA gel extraction kits, Bradford reagent, Ethidium Bromide are obtained either from Sigma-Aldrich (USA) or Qiagen (USA), Restriction enzymes, Pfu polymerase, DNA ligase, DpnI are procured either from Fermentas or New England Biolabs inc (NEB). Agarose is obtained from Himedia (India).

iii. Media and Antibiotics: Ampicillin, Kanamycin, Chloramphenicol, Agar powder, Luria Bertani (LB) Broth, Yeast Nitrogen Base, Isopropyl β-D-1-thiogalactopyranoside (IPTG) are obtained from Himedia, India.

Antibiotic stock solutions:

a) Ampicillin (100mg/ml): 1 gm of ampicillin is dissolved in 10 ml of sterile Milli-Q water and filtered using 0.22 µm filters
b) Kanamycin (50mg/ml): 0.5 gm of kanamycin is dissolved in 10 ml of sterile Milli-Q water and was filter sterilised.
c) Chloramphenicol (34µg/ml): 0.340 mg is dissolved in 10 ml of 100% isopropanol.

iv. Buffer components and stock solution preparation:

NaCl, Tris, EDTA, Glycerol, β-mercaptoethanol, TEMED, Glycine, Sodium Dodecyl Sulphate (SDS), Ammonium sulphate, Sodium phosphate dibasic, Sodium phosphate

monobasic, Potassium chloride, Magnesium sulphate, CTAB, sodium deoxycholate, Triton X-100 other salts, buffers, precipitant, detergents and organic solvents are obtained from Hi Media (India), Sigma (USA), Merck (Germany), Fluka (Germany), SRL diagnostics, Qualigens (India).

a) 5M Sodium chloride: 292 gm of sodium chloride is dissolved in Milli-Q water and volume is adjusted up to 1 litre. This stock solution is then filtered using Whatman filter paper and autoclaved. It is stored at room temperature for further use.
b) 1M Tris: 121.14 gm of Tris is dissolved in Milli-Q water and volume is adjusted up to 1 litre. It is filtered and autoclaved.
c) 1M HEPES: 238.3 gm of HEPES is dissolved in Milli-Q water and volume is adjusted up to 1 litre
d) 0.5M EDTA: 186.1 gm of disodium EDTA (Na_2EDTA) is dissolved in Milli- Q water at pH 8.0 and volume is made up to 1 litre (pH is adjusted with NaOH)Acrylamide preparation: 30% Acrylamide is prepared by mixing 29% acrylamide and 1% N, N' methylene bisacrylamide.

v. Affinity resins:

GST (Glutathione S-transferase), Ni-NTA (Nickel-Nitrilotriacetic acid), and Amylose resin were obtained from GE Healthcare and Novagen. Pre-packed gel filtration columns Superdex 200 and 75 of 16/60 were purchased from GE Healthcare.

vi. Instruments

FPLC (Fast protein liquid chromatography) system AKTA purifier and (Isothermal titration Calorimetry) ITC-200 GE Healthcare (Sweden), MWCO (Molecular Weight Cut Off) filter unit-Millipore (USA), SDS-PAGE apparatus- Bio-Rad (USA), Circular Dichroism spectrophotometer-JASCO Japan). Spectrofluorometer- Horiba (Japan), High-speed

centrifuge-Thermo scientific. Ultrasonic Homogenizer (Model 300 V/T, Biologics, Inc.-200 C incubator- Sanyo (Japan), -800C deep freezer Thermo Fisher (USA), Vibration-free cooling incubator- Sanyo (Japan), X-Ray Diffractometer (Generator-Bruker, Image Plate-MAR 345 (Germany), Stereo microscope-Olympus (Japan).

vii. Crystallization chemicals

Crystallisation screens: Crystal screen 1/2, PEG/Ion screen, Salt Rx, Natrix and additive screens, Ammonium sulphate, 2-(N-morpholino) ethane sulfonic acid (MES), Cobalt chloride, Crystallization buffers, salts, precipitants, organic solvents were procured from Hampton's Research (USA). Synthetic peptides are from USV Biotech (India). Buffer prepared for crystallization were filter sterilized using 0.45 μM filter. Most of the chemicals and buffer components used during purification were sterilized or made in sterile distilled water to avoid any contamination.

viii. Other chemicals

Protease cocktail kit and trypsin were procured from Sigma-Aldrich. Glutaraldehyde, PMSF, Ethidium Bromide, Coomassie Brilliant Blue, Methanol, Isopropanol, Sodium hydroxide, Hydrochloric acid, Acetic acid, Ammonium persulphate, ortho-Nitrophenyl-β-galactoside, Sodium Carbonate, Bromophenol blue, Acrylamide, Bis-acrylamide, Bradford reagent, Protein ladder was purchased from Sigma (USA), Merck (Germany), Fluka (Germany).

2.2 Methods

2.2.1 Gene Cloning

The method of producing multiple identical copies of a gene/ fragment is known as gene cloning. Cloning of DNA is a prerequisite for a variety of experiments such as protein expression and purification, gene regulation studies, confocal microscopy and siRNA studies. The cloning of the gene of interest or its fragment from cDNA is achieved by designing a

defined sequence oligonucleotide known as a primer. The sequence of these primers is chosen to be complementary to a stretch of cDNA segment to be amplified. The primer sequence also provides the opportunity of introducing additional features (restrictions enzyme site, protease codon sequence or small tags like His tag) that would be helpful in the cloning and protein purification process. The restriction enzyme cleavage sites in the primer and the vectors are identical, so as to anneal the insert in the vector in a specific direction, and this is known as directional-cloning.

Gene cloning is carried out in 4 steps: PCR, Digestion, Ligation, and Transformation.

a. Polymerase chain reaction (PCR):

PCR is based on using the ability of DNA polymerase to synthesize a new complementary strand to the given DNA template, in 5'-3' direction. Obviously, DNA polymerase requires a continuous supply of nucleotides and a set of primers to start its reaction **(Table 2.1)**. PCR comprises of three basic steps: **denaturation -** the two strands of double helical template DNA are separated, **annealing -** the oligonucleotide primers anneal to the denatured strands at complementary sites and **elongation** - DNA polymerase elongates the annealed primer on complementary DNA strand. Under appropriate conditions **(Table 2.2)**, in each PCR cycle, the number of molecules of the target gene doubles and further it increases exponentially. Thus after n cycles of PCR, 2^n copies of the target gene are produced.

***Table 2.1**: Typical PCR reaction*

Sr.No.	Components	Volume (µl)
1	100 ng/µl Template	1.5
2	10 pm/µl Forward primer	1
3	10 pm/µl Reverse primer	1
4	5X HF Buffer	10
5	10 mM dNTPs	2
6	Phusion DNA Polymerase	1
7	Autoclaved D/W	34
8	Total reaction volume	50

The typical program for PCR reaction is mentioned in **Table 2.2**. Different conditions have been modified depending on the requirement. The elongation time depends on the rate of the reaction of the polymerase enzyme and the length of the DNA to be amplified. The annealing and elongation conditions used in the case of site-directed mutagenesis are comparatively different from those of normal PCR reaction.

***Table 2.2**: Typical condition used for PCR setup*

Steps	Temperature (°C)	Duration (min:sec)
1.Initial denaturation	98	3:00
2.Denaturation	95	00:30
3.Annealing	60	00:30
4.Elongation	72	00:50
Run steps 2-4 for 32 more cycles		
5.Final Elongation	72	10:00

b. Digestion of vector and PCR product:

The amplified PCR product and the expression vector are digested using two separate restriction enzymes. Restriction enzymes cut the vector and the target gene at specific sequence thereby generating complementary sticky ends. The digestion reaction is carried out at 37º C for approximately 1.5 hr - 2 hr **(Table 2.4 and 2.5)**

***Table 2.3:** Digestion reaction for target gene*

Sr.no	Components	Volume (µl)
1	PCR product	20
2	10X Buffer	3
3	EcoR 1	1
4	BamH 1	1
5	Autoclaved D/W	5
6	Total reaction volume	30

Table 2.4: *Digestion reaction for vector*

Sr.no	Components	Volume (µl)
1	Vector DNA	10
2	10X Buffer	1
3	EcoR1 enzyme	1
4	BamH1 enzyme	1
5	Autoclaved D/W	7
6	Total reaction volume	20

c. Ligation of vector and PCR product:

DNA ligation refers to the formation of phosphodiester bonds between the acidic 5' phosphate group and alcoholic 3' hydroxyl group of the deoxyribose sugar. The sticky and complementary ends which are formed during restriction digestion of vector and PCR product are ligated using DNA ligase. In general, during ligation there should be excess of insert DNA over vector DNA (3:1 or 6:1). The following types of annealed products may be formed: 1. self-ligation of the vector, 2. ligation of multiple copies of insert and vector and 3. ligation of vector and insert. Setup for ligation reaction is shown below. As different types of ligation products are possible, digestion with corresponding restriction enzyme and DNA sequencing is necessary to confirm the clone.

Table 2.5: *Ligation reaction of vector and insert*

Sr.no	Components	Volume (µl)
1	(50 ng/µl)Double digested Vector DNA	1
2	(50 ng/µl)Double digested Insert/ DNA	3
3	2X Quick Ligation Buffer	5
4	Quick Ligase	1
5	Total reaction volume	10

d. Transformation and plasmid isolation:

The properly ligated plasmid is then transformed into competent bacterial strain to increase the copy number, and then further plasmid isolation is carried out. This isolated plasmid is submitted to DNA sequencing facility to confirm for the sequences of positive clones.

2.2.2 Bacterial strains used for cloning and purification

E. coli is a preferred bacterial strain used for gene cloning and protein expression as it has a short doubling time of about 20 minutes requires low-cost media for growth and, can be easily genetically manipulated. The *E. coli* strains used for recombinant DNA technology are engineered so as to possess the following properties:

a. Insert expression under control of T7 system

The target gene is cloned downstream of T7 promoter on a plasmid which is introduced into the DE3 strain. This targeted gene controlled by T7 promoter is recognized only by T7 RNA polymerase and not by *E. coli* RNA polymerase. Therefore, the *E. coli* is engineered so as to contain a chromosomal copy of the phage T7 RNA polymerase gene & such expression hosts are known as DE3 strains or T7 Express strains. T7 RNA polymerase gene is under the control of the lacUV5 promoter which is transcriptionally activated by lactose analog IPTG. Thus when IPTG is added to the culture, T7 RNA polymerase is expressed and the gene of interest which is present under the control of T7 promoter is transcribed.

b. Elimination of protease activity

The most important characteristics in BL21 are the deletions of OmpT outer membrane protease and ATP-dependent Lon protease. OmpT is a housekeeping protease that degrades foreign peptide material [120]. Lon is an ATP-dependent serine protease that mediates the selective degradation of mutant and abnormal proteins as well as certain short-lived regulatory proteins. Inactivation of Lon protease increases the yield of some recombinant proteins. Hence, Rec A positive BL21 strains should be used for expression, as recombination may alter the exogenous plasmid DNA during cloning.

c. Improved recovery of purified target protein

Lemo21 (DE3) strain can improve the expression and solubility of membrane proteins. The

NiCo21 (DE3) strain has been recently engineered to improve the recovery of the His-tagged target protein.

Thus, there are different kinds of engineered strains of *E. coli* which can be used depending on the kind of protein to be expressed. Few of the frequently used strains are:

i. BL21(DE3)

It is derived from B834 which is a methionine auxotroph; a relatively "wild" B strain of *E. coli*. It is deficient in Lon and OmpT proteases. BL21 (DE3) Star Ta derivative of BL21 has a mutation in RNAse E making it less capable of degrading mRNA and this results in enhanced expression levels.

ii. Rosetta(DE3)

These strains supply additional tRNA genes for codons AGG/AGA (Arg), AUA (Ile), CUA (Leu), CCC (Pro) and GGA (Gly) on a compatible chloramphenicol-resistant plasmid. Thus, the Rosetta strains are for "universal" translation which is otherwise limited by the codon bias in *E. coli* [115].

iii. BL21 (DE3) pLysS

pLysS plasmid encodes the gene for T7 lysozyme. T7 lysozyme inhibits T7 RNA polymerase function. The pLysS plasmid is used to minimize basal expression of the target genes by inhibiting the function of minimal amounts of T7 RNA polymerase formed due to leaky expression. It is mostly used for expressing proteins that are toxic to the cell.

iv. DH5-α

In this strain, the nonspecific endonuclease I (endA1) is inactivated and rec A is mutated. Inactivation of nonspecific endonuclease I (endA1) prevents plasmid degradation and Rec A mutation reduces the chance for deletions and plasmid multimerisation. The chromosome of this strain contains an hsdR17 mutation in the restriction endonuclease of

the EcoKI restriction-modification system. This mutation helps in the efficient transformation of unmethylated DNA derived from PCR, cDNA and many other sources.

2.2.3 Competent cells preparation and Transformation

The ability of bacterial cells to take up the foreign DNA from the environment is known as competence, and the cells that have taken up the foreign DNA are called as transformed cells. Natural competence is being witnessed in bacteria as evidenced by their ability to use DNA as a food source, and also to derive benefits of genetic diversity and lateral gene transfer. Bacterial cells can also be artificially induced to become competent by physical (electroporation) and chemical (use of divalent cations like Mg^{2+} and Ca^{2+}) methods. Under normal conditions, the restriction modification system (RMS) in the bacterial cells degrades the incoming foreign DNA. But the strains used in labs have defective RMS system, and hence can be transformed [121,122].

Competent cells preparation protocol

Competent cells preparation protocol for 500 ml culture. The buffers used are mentioned below:

Transfer Buffer 1	Transfer Buffer 2
30 mM Potassium acetate	10 mM MOPS (3-(N-morpholino)
50 mM Magnesium chloride	propanesulfonic acid)
100 mM Potassium chloride	10 mM Potassium chloride
10 mM Calcium chloride	75 mM Calcium chloride
15% Glycerol (v/v)	15% Glycerol (v/v)

a) A single isolated colony from LB-agar plate is inoculated in 10 ml of Luria-Bertani (LB) broth (without antibiotic, except Rosetta 2 DE3 strain). It is incubated at 37°C in shaker incubator for about 12 hr. 1% of this culture is diluted into 500 ml LB broth and allowed to grow at 37°C until OD_{600} reaches to 0.6. The culture is then chilled on ice.

b) Cells are then harvested at 5000 rpm for 15 min in precooled rotor and then resuspended in pre-chilled 100 ml transfer buffer 1.

c) Centrifuge the suspension at 5000 rpm for 8 min and resuspend the cell pellet in pre-chilled 10 ml transfer buffer 2.

d) Resuspended cells are then aliquoted, each 50 µl, in 1.7 ml Eppendorf tubes and snap-frozen in liquid nitrogen. The entire preparation is performed aseptically inside laminar hood and is maintained in the chilled condition. For long term storage aliquots are kept in -80°C freezer.

Transformation of the construct (a vector with the desired insert) into the competent host bacterial cells

a) Thaw the competent cells on ice. DNA to be cloned/ expressed is mixed with these competent cells and incubated on ice for 30 minutes

b) Heat shock is given at 42^0C for 90 sec in a water bath.

c) 800 µl of LB broth is added and incubated at 37^0C for 45 min on shaker incubator so as to activate the growth of bacterial cells, as the agar plates containing antibiotics may kill such biochemically inactive cells.

d) 100 µl of these cells is plated onto the plate containing LB agar with appropriate antibiotic. The plate is incubated at 37˚C overnight.

2.2.4 Protein purification methods

Protein purification involves a series of processes aimed at separating the protein of interest from cell debris and other proteins. Separation of protein, from other constituents of the cell, usually exploits differences in physicochemical properties like protein size, net charge, solubility, binding affinity and biological activity [123]. Protein purification can be either on an analytical or preparative scale:

1. Analytical: Analytical methods aim to detect and identify a protein in a mixture. 2. Preparative: Preparative methods aim to produce a large quantity of protein for different purposes such as crystallization or industrial use.

Proteins present inside the cells can be extracted by repeated freeze and thaw method, sonication, homogenization by high pressure, osmotic shock and lysozyme treatment etc. Protein released in the buffer can be purified in crude/ advanced ways depending on the requirement. Several chromatographic and other methods used for purification are:

a. Ammonium sulfate precipitation

It is based on salting out effect, which means precipitating the protein from the soluble fraction with the addition of increasing amounts of ammonium sulfate. High molecular weight proteins precipitate faster than low molecular weight proteins. As the salt concentration is increased the hydrophobic groups from the proteins get exposed to the solvent thereby attracting hydrophobic groups of other protein for the aggregation and subsequent precipitation.

b. Ultracentrifugation

It is a process of separation of molecules based on densities with the use of high-speed centrifuge. The centrifugal force separates mixtures of particles of varying masses or densities suspended in a liquid. If samples are centrifuged long enough, the particles in the tube will reach equilibrium where in the particles accumulate specifically at a point in the tube where their buoyant density is balanced with the centrifugal force. Such an "equilibrium" centrifugation can allow extensive purification of a given particle.

c. Size exclusion chromatography

It separates proteins based on the size and conformation. The column is packed with porous inert material like Sephacryl HR, Superdex, Superose, Sephadex etc. Superdex is a composite

matrix of highly cross-linked agarose beads covalently attached to dextran molecules. Agarose provides physical and chemical stability while dextran governs the gel filtration properties of this aqueous gel. Superdex 75 and Superdex 200 have an average bead size of 13 µm. When protein is passed through such calibrated column it is sieved based on its molecular size. The smaller molecules travel through a number of pores in the porous matrix, and hence they will elute later whereas larger will elute earlier from the column. Size exclusion chromatography is used for group separations e.g. for desalting the protein.G-10, G-25, and G-50 columns are used for group separations. The high-resolution fractionation is performed using Superdex 75 and 200 to separate multimeric forms of protein, for estimating molecular weight and so on. The fractionation range for globular proteins in Superdex 200 is 10,000-600,000 Dalton (Da) and for Superdex 75 are 3,000-70,000 Da. The upper exclusion limit for globular proteins of Superdex 200 and 75 is 1300,000 and 100,000 Da respectively.

d. Hydrophobic Interaction Chromatography (HIC)

HIC media is amphiphilic with both hydrophobic and hydrophilic regions, allowing separation of proteins based on their surface hydrophobicity. The interaction between the resin and the hydrophobic regions of the protein is enhanced by applying a protein sample to HIC resin under conditions of high ionic strength. The ionic strength of the buffer is then reduced gradually to elute proteins in order of decreasing hydrophobicity [124].

e. Ion exchange chromatography

Ion exchange chromatography separates compounds according to the nature and degree of their ionic charge. Anion exchange resins have a positive charge and are used to retain and separate negatively charged compounds while cation exchange resins have a negative charge and are used to separate positively charged molecules. Few examples of anion exchangers are

Quaternary ammonium (Q), Diethylaminoethyl (DEAE) and cation exchanger are Sulfopropyl (SP), and Carboxymethyl (CM) cellulose.

f. Affinity Chromatography

Affinity Chromatography is a separation technique based on a reversible interaction between a tagged protein and a specific ligand coupled resins. To purify a protein using affinity chromatography, proteins are expressed with a specific tag, which is incorporated in cloning vector itself. These tagged proteins are then passed through the resins and due to the affinity of the tag towards the resin, the protein binds to the resin. Specificity and reversibility are the two basic criteria that are well considered in affinity purification. The most commonly used tags for protein purification are maltose binding protein (MBP), glutathione-S-transferase (GST) [125] and 6-His tag with six Histidine residues.

i. GST (Glutathione S-transferase)

Glutathione is a tri-peptide, which consists of the amino acids glutamic acid, cysteine, and glycine. GST is a 26 kDa dimeric protein derived from Schistosoma japonicum. GST tag is capable of binding to its substrate, glutathione. This property thus, can be used to capture pure GST or GST-tagged proteins via the enzyme-substrate binding reaction. Reduced glutathione through its sulfhydryl group can be used to crosslink glutathione with agarose beads. The bound GST-fusion protein can be recovered by the addition of excess reduced glutathione since the affinity of GST for free glutathione is higher than the affinity for immobilized glutathione. The free glutathione replaces the immobilized glutathione and releases the GST-tagged protein from the matrix which can be further eluted from the column [126,127].

ii. IMAC (immobilized metal ion affinity chromatography)

IMAC technique is based on the interactions between histidines and transition metal cations (Cu^{2+}, Ni^{2+}, $Zn^{2+}$and $Co2^{+}$). His-tagged proteins can be purified easily using IMAC

chromatography under specific buffer conditions. The transition metal (chelating agent) nickel-nitrilotriacetic acid (Ni-NTA) and cobalt-carboxymethylasparate (Co-CMA) is immobilized to a cross-linked agarose matrix. The bound metal ion forms the coordination complex with the His-tagged protein (**Figure 2.1)** [128,129]. This His-tagged protein immobilized on the chelating agent can be eluted by introducing a competing agent like imidazole or an additional metal chelating agent (EDTA). The major advantage of using His tag is its small size, affinity and its ability to bind to metal ions under denaturing conditions, which becomes important for purification of proteins that are found even in inclusion bodies [128,130].

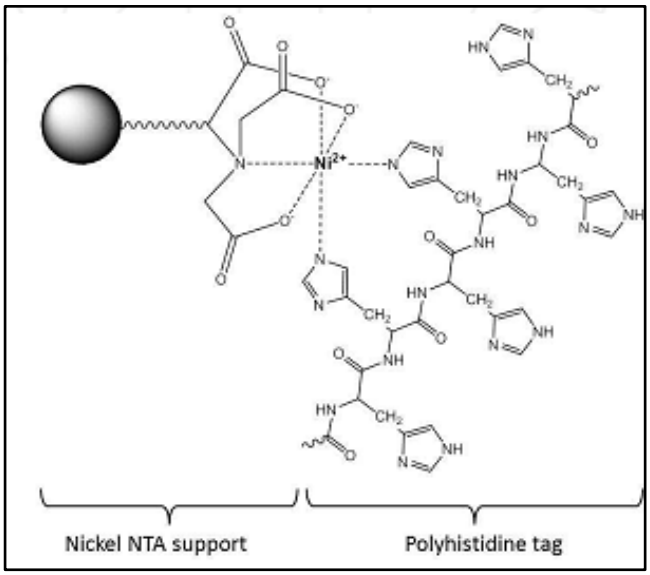


Figure 2.1: *Formation of complex between the poly-histidine tag and nickel NTA support*

2.2.5 Protein Estimation

A. Bradford assay

The Bradford assay is a protein estimation method that involves the binding of the dye to proteins. Bradford reagent is a combination of ethanol (95%), phosphoric acid (85%), and Coomassie Brilliant Blue G 250. The dye Coomassie Brilliant Blue G 250 exists in three forms: cationic (red), neutral (green) and, anionic (blue). The dye binds to basic amino acids of the

protein (lysine and arginine), and is converted to a stable non protonated blue form. This blue colored protein-dye complex formed is detected at λ=595 nm in the assay using a spectrophotometer or microplate reader.

Standard preparation: A stock solution of Bovine serum albumin (BSA) at a concentration of 10 mg/ml is prepared in 2 ml autoclaved distilled water. BSA dilutions of 1, 0.5, 0.25, 0.125 and 0.0625 mg/ml in distilled water were made from stock of 10 mg/ml to make a standard graph. 10 µl of sample is added in each well of Elisa plate in duplicates, from BSA dilutions carefully, to avoid any bubble or pipetting error. To these wells, 300 µl of Bradford reagent is added. The reading is recorded using Elisa plate reader. A plot of absorbance at 595 nm versus different BSA concentration is made and a regression value for the straight line equation is calculated.

From this straight line equation, concentration of protein sample is then extrapolated using straight line equation: $y = mx + c$

Where y = average absorption, λ=595 nm, m = slope of the line, x = concentration of the protein sample, c = y-intercept

B. NanoDrop

Nanodrop instrument has a fiber optic cable connected to a sample loading port, and a second fiber optic cable is brought near the first cable leaving a small gap in between them. The sample loaded onto this port bridges the gap between the fiber optic ends. The gap is controlled to both 1 mm and 0.2 mm paths depending upon the concentration of the sample used. The xenon lamp is used as a source of light and a linear CCD array detector is used to measure the transmitted light from the sample. The instrument is controlled by PC based software. Nanodrop is based on the principle of Beer-Lambert's law. This equation is used to correlate the calculated absorbance with concentration: $A = \varepsilon\ b\ c$. A = the absorbance of sample (blank corrected), ε = extinction coefficient with units of litre $mol^{-1}cm^{-1}$, b = the path

length in cm, c = the analyte concentration in moles/liter or molarity (M). The transmitted light intensity for both the sample and of the blank are required to calculate the absorbance. Protein concentration can be estimated using Nano drop at λ=280 nm. Nano Drop absorbance value and molar extinction coefficient give the value of c in moles.

2.2.6 Gel Electrophoresis

Gel electrophoresis is a technique for separation of macromolecules (proteins, DNA, and RNA) and their fragments, under the influence of an electric field. Under the influence of electric field positively charged (cations) species migrate towards the cathode and negatively charged (anions) species migrate towards the positively charged anode. Gel electrophoresis method can be used to separate a mixture of DNA and RNA fragments of different length to estimate the size of DNA and RNA fragments or to separate proteins by their charge and size. The separation of small molecules from larger ones is brought about by simple sieving process which occurs through the use of the porous material known as a gel. Agarose and acrylamide are the commonly used materials for making the gel. Different concentration of these materials is used for producing different sized mesh networks. Electrophoresis is performed in buffer solutions to reduce pH changes due to the electric field, which is important because the charge on DNA, RNA and protein depends on the pH. Gel electrophoresis is usually performed for analytical and preparative purposes e.g. to check for PCR amplification, mass spectrometry, RFLP, DNA sequencing, Southern/Western blotting and for protein visualisation.

A. 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

the sizes of DNA fragments. The DNA can be visualized by adding a fluorescent dye like ethidium bromide (EtBr), which binds to DNA through intercalation between bases. The EtBr stained DNA fragments are visualized using UV light and documented by Gel documentation system (Fisher Scientific, UK).

B. Polyacrylamide Gel Electrophoresis

(a) Native Polyacrylamide Gel Electrophoresis (Basic proteins)

Table 2.6: *Native Polyacrylamide Gel Electrophoresis (Basic proteins)*

Stock solutions	Volume added (ml)
Resolving	
1.5M Acetate KOH pH 4.3	1.6 ml
50% glycerol	1.5 ml
30% Acrylamide	1.6ml
H_20	1.65ml
10% APS	80 µl
TEMED	10 µl
Stacking gel for 3ml	
0.25M Acetate-KOH PH 6.8	1.25 ml
30% Acrylamide	0.5 ml
H_20	3.2 ml
10%APS	50 µl
TEMED	5 µl
Sample loading buffer	
Glycerol 50%	1.45 ml
0.25 M Acetate KOH Ph 6.8	0.5 ml
Methyl green	Traces
Electrode buffer (1litre- 1X)	
B-Alanine	
Acetic acid	
Adjust pH to 4.3 and makeup volume with water (200 ml)	

The native gel can reveal information about the protein oligomeric characteristics, analysis of molecular structure and conformation of the protein. Here separation of proteins is dependent

on charge and shapes under native conditions. As SDS and β-mercaptoethanol are not used, the proteins retain their native charge and conformation. Use of only β-mercaptoethanol can reveal if any disulphide linkages are involved in oligomerisation/ quaternary structure formation in protein. **Table 2.6** shows the different reagents used for casting the gel. In case of native gels for basic proteins, the gel is pre-run for half an hour for activation. The electrodes are connected in reverse i.e. in power pack connect the positive electrode to the negative and the negative electrode to a positive end.

(b) Native polyacrylamide Gel Electrophoresis (Acidic proteins)

Before electrophoresis, all the apparatus is cleaned properly so as to get rid of any traces of SDS that may interfere with the experiments. The resolving components are mixed and poured between the sealed glass plates and allowed to polymerize. Subsequently, stacking gel mix is added (**Table 2.7).** The samples were mixed with sample loading buffer and directly loaded on the gel for electrophoresis without heating. Electrophoresis is done using electrode buffer (**Table 2.7)** at 25 mA (constant current) and the gel is run at 4^0C. Finally, protein bands on the gels are detected by staining solution.

Table 2.7: *Native Polyacrylamide Gel Electrophoresis (Acidic proteins)*

Stock solutions	Volume added from stock for 8% Gel
Resolving	
30% Acrylamide	2.7 ml
1.5M Tris pH8.8	2.5 ml
10% APS	100 µl
TEMED	40 µl
Water	4.6ml
Stacking gel for 3ml	
30% Acrylamide	0.5ml
1M Tris pH6.8	0.380 ml
10% APS	30 µl
TEMED	3 µl
Water	2.13 ml
Sample loading buffer	
Water	22 ml
0.5M Tris pH-6.8	5 ml
Glycerol	10 ml
β-mercaptoethanol	2 ml
(0.25%) Bromo Phenol blue	800 µl
Electrode buffer (1litre- 1X)	
Tris-chloride	3.02 g
Glycine	18.8 g

(c) SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE)

SDS-PAGE is performed as per the method described by Lammeli 1970 [131,132][132] Separation of proteins by SDS-PAGE can provide information about the molecular weight of primary structure, presence of disulphide bonds and subunit composition. In SDS-PAGE, proteins are denatured in the presence of an anionic detergent SDS (Sodium Dodecyl Sulfate) so as to separate the protein on the basis of mass. Generally, 1.4 gm of SDS binds

per gram of protein imparting an overall negative charge to every protein molecule. Hence, migration of SDS coated proteins in the gel is related to its size and not its charge or shape. β-mercaptoethanol added is responsible for breaking the disulphide linkage formed in protein molecule itself or in the oligomer. All the resolving gel component are mixed and added between sealed glass plates with spacers. It is allowed to set for 30 min

Table 2.8: *SDS Polyacrylamide Gel Electrophoresis (SDS-PAGE)*

Stock solutions	Volume added from stock for 12% Gel
Resolving	
30% Acrylamide	4 ml
1.5M Tris pH8.8	2.5 ml
10% SDS	100 µl
10% APS	100 µl
TEMED	40 µl
Water	3.3 ml
Stacking gel for 3ml	
30% Acrylamide	0.5 ml
1M Tris pH6.8	0.380 ml
10% SDS	30 µl
10% APS	30 µl
TEMED	3 µl
Water	2.1 ml
Sample loading buffer	
Water	22 ml
0.5M Tris pH-6.8	5 ml
10% SDS	8 ml
Glycerol	10 ml
β-mercaptoethanol	2 ml
Electrode buffer (1litre- 1X)	
(0.25%) Bromo Phenol blue	800 µl
Tris-chloride	3.02 gm
SDS	1 gm
Glycine	18.8 gm

A stacking gel is further overlaid on the polymerized resolving gel along with the combs in it. **Table 2.8** shows the different reagents used for casting the gel. The protein samples are mixed and boiled for 5 minutes in sample loading buffer. Electrophoresis is done using electrode buffer at a constant voltage of 110 V. Finally, protein bands on the gels for native and SDS polyacrylamide gel electrophoresis were detected by staining with staining solution containing methanol-45 ml, acetic acid-10 ml, Coomassie Brilliant Blue-0.25 g and water-45 ml.

2.2.7 Mass Spectrometry

MALDI is the abbreviation for "Matrix Assisted Laser Desorption/Ionization. TOF MS is the abbreviation for Time of Flight Mass Spectrometry. MALDI is a method of generating oligopeptide into ionic form for separation through TOF. Since MALDI-TOF/TOF MS/MS data can determine the molecular weight to an accuracy of 0.1 Da, one can sequence oligopeptides using MS. Protein, when digested with site-specific proteases, produce signature oligopeptides for that particular protein. The peptides generated are processed to estimate their molecular weights through MS. The collection of masses so obtained are analysed for protein identification. Here, protein identification is based on the principle of peptide mass fingerprinting. The process of identification of sample in MALDI can be divided into three parts.

i. Mixing of sample and matrix: The protein digest is mixed with the matrix material and spotted on to a metal plate. The solvents from this mixture vaporize forming co-crystals of the matrix with the analyte. The most commonly used matrix material are 3,5-dimethoxy-4-hydroxycinnamic acid (sinapinic acid), α-cyano-4-hydroxycinnamic acid (CHCA, alpha-cyano or alpha-matrix) and 2,5-dihydroxybenzoic acid (DHB).

ii. Laser irradiation: The metal plate is then inserted into the vacuum chamber of the mass spectrometer and the sample is pulse irradiated with the laser. The instrument uses Nd: YAG (neodymium-doped yttrium aluminium garnet) laser at 337 nm to desorb and ionize the molecules.

iii. Ionization and desorption: The laser transfers energy to the matrix, promotes the ionization and transition of the matrix molecules and finally, the analyte molecules are ionized by being protonated or deprotonated in the hot plume of ablated gases, and is then accelerated. This process produces positive and negative ions usually in the forms of $(M+H)^+$, $(M+Na)^+$ and $(M-H)^-$. MALDI-TOF instrument can detect mass of protein at very low concentration of picomole to femtomole amount. These ionized analyte molecules are accelerated under the influence of electric field until they reach the detector. The time required for the ion to reach detector differs according to the mass-to-charge ratio, (m/z) value of the ion. The technique of mass spectrometry that exploits this phenomenon is called Time of Flight Mass Spectrometry.

iv. Peptide mass fingerprinting: The peptide map (peak list) generated due to different mass-to-charge ratio (m/z) is used to search in database such as Swiss-Prot, NCBI nr, EMBL to find a good match in Mascot analysis software. The software applies the cleavage rules of protease (e.g. trypsin) on proteins in the database; the mass of these peptide fragments is then calculated and compared to the peak list of measured peptide masses. The possible matches are statistically analysed [133]. The signal intensities of different peptides cannot be used to quantitate the amount of protein present, as these signals are suppressed because of competition for a charge or an optimal position in the matrix. The higher the molecular weight the more sample is needed as sensitivity decreases with increase in mass [134,135]. Applications of MALDI are: protein identification; measuring mass of the intact protein;

identification of microorganisms. MALDI-TOF/TOF mass spectrometers can be used to reveal amino acid sequence of peptides from high-energy collision-induced dissociation.

5. In-gel digestion with trypsin (Coomassie stained) [136]

NH_4HCO_3-50 mmol/l, Dithiothreitol (DTT)-10mmol/l (freshly prepared), Iodoacetamide (IAA) - 55mmol/l (light sensitive), and Trypsin: 20-25 ng used for about 20 µg of protein.

a. Excision of protein bands from polyacrylamide gels:

Wash the gel thoroughly with water and excise the band of interest from the gel. Cut the excised gel piece into smaller pieces. Wash the gel particles with water and 50 mM NH_4HCO_3/ acetonitrile 1:1(v/v) for 15 minutes. Aspirate the liquid and add enough acetonitrile to cover the gel particles. The gel pieces will shrink and stick together. Remove the acetonitrile and rehydrate the gel pieces in 50 mM NH4HCO3. After 5 min add an equal volume of acetonitrile and incubate further for 15 minutes. Discard the supernatant and add enough acetonitrile to cover the gel particles. After the gel pieces have shrunk remove the acetonitrile and dry the gel particles in a vacuum centrifuge.

b. Reduction and alkylation:

Swell the gel particles in 10 mM DTT /50 mMNH_4HCO_3. Incubate it for 45 min at 56°C. Remove the liquid and replace it with freshly prepared 55 mM IAA in 50mM NH_4HCO_3. Incubate for 30 min at room temperature in the dark. Remove IAA solution. Wash the gel particles with 50 mM NH_4HCO_3 and acetonitrile (1+1; v/v), one or two changes each, 15 minutes per change. Add enough acetonitrile to dehydrate and dry down the gel particles in a vacuum centrifuge.

c. In-gel trypsin digestion:

Add freshly prepared enzyme solution (in 25 mM NH_4HCO_3 and 9% acetonitrile) to cover the gel and incubate further at 37°C overnight.

d. Extraction of peptides:

To this overnight digested gel plugs apply extraction buffer enough to cover the gel pieces completely. Support the extraction by ultrasonication for 30 minutes at room temperature.

The three extraction buffers used are

1. 50% acetonitrile + Trifluroacetic acid (TFA) 5%
2. 1% TFA
3. 100% acetonitrile

Pool all the extracted peptides from above three extraction buffer and speed vac to dry it. The peptide was finally resuspended in 50% acetonitrile and 1% TFA.

2.2.8 Interaction study

The interactions between molecules are very important in a diverse and complex set of biological processes such as signal transduction, regulatory mechanisms, gene expression, immune response etc. The different types of interactions are protein-protein, protein-DNA, antibody-antigen, enzyme-substrate and ligand-receptor interactions. The studies of interactions aim to understand stoichiometry of the interaction, the association constant, thermodynamic parameter like a change in free energy, enthalpy and entropy on binding and detailed 3D structures of the complexes. The different methods used to characterize biologically important interactions include ELISA (enzyme-linked immunosorbent assay), surface plasmon resonance, EMSA, ITC and pull down assay etc. Brief description of techniques used in this thesis is given below. While EMSA and pull down assay is a crude method to study interactions, ITC is a quantitative method for measuring the thermodynamic parameters of an interaction.

A. Electrophoretic Mobility Shift Assay

Electrophoretic mobility gel shift assay (EMSA) is an electrophoretic separation technique to study the protein-DNA, protein-protein and protein-ligand interaction *in vitro*. A mixture of the molecules, interactions between which is being explored, is electrophoresed on a

polyacrylamide or agarose gel. One of the interacting molecules in the sample is called a probe and its amount is held constant in all the lanes of a gel. It is the probe that is detected to know the position of a band in the gel. The mobility of any molecule depends on its molecular mass. For example, the complex of protein: nucleic acid migrates slowly as compared to nucleic acid alone. Hence, a relative shift is observed in the positions of bands corresponding to nucleic acid alone and nucleic acid-protein complex. The gel matrix that surrounds the protein prevents diffusion of protein away from the probe (nucleic acid) promoting prompt re-association of the complex & this effect is known as "caging" effect. To further enhance the shifting of the complex band, antibody-antigen interaction is exploited to decrease mobility by creating a ternary complex, in what is called as a "supershift assay" [137,138,139]. The bands corresponding to bound and free nucleic acid on the gel reflect the number of nucleic acid molecules in the free and bound form. In the saturated titration data (that has no free probe); EMSA can be utilized quantitatively to measure thermodynamic and kinetic parameters of binding. EMSA can also be used to analyze the specific sequence of nucleic acid that is recognized by the protein, and also to detect sequence-specific DNA binding proteins from crude lysates. EMSA can also be used to resolve complexes of different stoichiometry or conformation. In order to study the interaction of the protein and DNA, constant amount of the DNA is titrated with increasing amounts of protein. The sample is incubated along with reaction buffer **(Table 2.9)** in ice for 30 min and loaded onto native or non-denaturing PAGE.

Table 2.9: *Composition of reaction buffer (2.5X)*

Sr. No.	Components	Concentration
1	Tris base	20 mM pH 6.8
2	Sodium chloride	50 mM
3	Triton-X100	0.1%
4	Glycerol	10%
5	Zinc chloride	0.01 mM

To prevent dissociation of the complex due to heat generated during electrophoresis, the gel is cast to have larger pore sizes. The percentage of gel used is dependent on 1) size of the target DNA, 2) size, number and charge on the interacting protein. The percentage of the gel should be such that it should allow the DNA complex to enter the gel. Polyacrylamide gels in the range of 4-8% and Agarose gels (0.7-1.2%) can be used to resolve very large complexes. Non-denaturing Polyacrylamide gel (5%) is prepared as shown in **(Table 2.10)**.

Table 2.10: *Preparation of 5% and 8% native PAGE*

Sr. No.	Components	5% PAGE	8% PAGE
1	30% acrylamide, bis acrylamide	4.15 ml	6.67 ml
2	10X TBE	1.25 ml	1.25 ml
3	10%APS	25 µL	25 µL
4	TEMED	250 µL	250 µL
5	Autoclaved water	19.6 ml	17.08 ml

Composition of (10X) TBE buffer is Tris base - 89 mM, boric acid - 89 mM, EDTA - 0.5 M pH 8.0. The gels were pre-run in 0.5X TBE at 200 V for 30 minutes until the current no longer varies with time in the cold room. Pre-running the gel removes all traces of ammonium persulfate and ensures a constant gel temperature. Samples are loaded and further electrophoresed at 200 V. Bromophenol blue is used as a tracking dye and it migrates 10-20 mm ahead of double stranded DNA probe of ~30 bp. Staining is carried out by the addition of 1.5 µL (10mg/ml) of ethidium bromide in 20 ml of distilled water and the gel is kept on the rocker for 5 minutes. Later, the gel is washed thoroughly in water and observed under UV light in gel doc and the images are captured using Vision Works LS system.

B. Pull-down assay

The pull-down assay is an *in vitro* method used to determine a heterogeneous or homogenous (oligomerisation) physical interaction between molecules. It can also predict the region that may be probably involved in the interaction. In a pull-down assay, a bait protein is expressed with affinity tag and is immobilized on corresponding affinity resin. This immobilised protein acts as a support for purifying other proteins that interact with the tag or bait protein. This bait protein is then incubated with cell lysate that is a source of putative "prey" proteins. A control reaction with only tag protein is also set up to rule out any possibility of tag protein interacting with the prey. The method of protein elution depends on the affinity tag used and ranges from using competitive analytes to low pH or reducing buffers. The final determination of interacting proteins often involves mass spectrometric identification of tryptic digests of the protein band. If pull down is performed using pure protein, western blot analysis of the band can also be performed for confirmation.

C. Isothermal Titration Calorimetry (ITC)

ITC is the quantitative means for measurement of the thermodynamic parameters of the interaction. An ITC instrument consists of two identical cells (sample and reference cell) made of a highly efficient thermal conducting material and is surrounded by an adiabatic jacket. A feedback power system is available to maintain these two cells at the same temperature. Heat evolved or heat absorbed on the association of ligand with its binding partner is measured in this calorimeter. When the reaction is exothermic (heat evolving), the temperature of sample cell increases, and the feedback power will be turned off, to maintain equal temperatures between the two cells. While, in case of endothermic reactions, the feedback circuit will increase power supply to the sample cell so as to achieve equal temperature as the reference cell. The heat absorbed or evolved is directly proportional to the

fraction of the bound ligand during titration. As the macromolecule in the sample cell gets saturated with the ligand the amount of heat change becomes constant on further addition of ligand [140]. The heat change on ligand binding can be represented by the equation [141]

$Q = V_0 \Delta H_b [M]_t K_a [L]/ (1+K_a [L])$ where V_0 is the volume of the cell, ΔH_b is the enthalpy of binding per mole of ligand, [M]t is the total macromolecule concentration including bound and free fractions, K_a is the binding constant, and [L] is the free ligand concentration.

The thermodynamic measurements provide insight about the binding mechanisms. $\Delta G = \Delta H - T\Delta S$ is the equation that determines the favorability and mechanism of interaction. The reaction is allowed only if the total entropy change of the universe is zero or positive and hence when ΔG is negative. Polar interactions contribute favorably to the enthalpic component, whereas entropically favored interactions tend to be more hydrophobic. The negative or favorable binding enthalpy (ΔH) and entropy factor ($T\Delta S$) is dependent on hydrogen bonding and hydrophobic interactions. The values of the binding constant, the stoichiometry, and the enthalpy of binding can be determined from single saturated data of titration. Association constant value is used to determine free energy and entropy of binding. The data of association constant and number of binding sites can also be used in assessing the protein quality. ITC not only provides information about the stoichiometry of reaction, but it also reveals information about cooperativity and affinity of the ligand towards the two binding site [142]. Dimer dissociation model helps us to calculate the heat of association and affinity amongst monomer in a multimer [143,144].

2.2.9 Structural characterization

Biophysical characterization of different functional domains of BRCA1 has been carried out using following methods

A. Circular Dichroism Spectroscopy

"Dichroism" means the differential absorption of polarised light when passed through a medium. Circular dichroism (CD) spectroscopy is a method that has been used in structural biology to examine secondary structures of proteins, polypeptides, and DNA. It is a non-destructive technique used for the evaluation of conformational stability of proteins in several environmental conditions like temperature, ionic strength, and the presence of solutes or small molecules.

a. Principle

Circular Dichroism (CD) relies on the principle of differential absorption of left and right-handed circularly polarised light by chromophores, which either possess intrinsic chirality or are placed in a chiral environment. Superposition of two linearly polarized perpendicular lights that have the same direction, amplitude, and wavelength but are out of phase by 90^0 results in electric field vector that rotates in a circle. The asymmetry in the molecular structure is responsible for absorption of the left circularly polarized light to a different extent than it does for right circularly polarized light. When this light passes through such an optically active sample, the amplitude of the stronger absorbed component will be smaller than that of the less absorbed component. Thus, the resultant yields an ellipse instead of the line. This phenomenon is called circular dichroism.

b. Secondary structure:

The asymmetric peptide bonds in proteins, molecules without a plane of symmetry and other chromophores in proteins can give rise to CD signals. Amide (chromophore) of the peptide bond have two electronic transitions n-π^* and π_0- π^*. In the CD spectrum of the α-helix the n-π^* and π_0- π^* transitions are primarily responsible for the negative band at 222 nm and 208 nm respectively. The n-π^* and π_0- π^* transitions form a negative band in the λ range of 216-218 nm and a positive band at ~190 nm, a signature for β-sheet spectrum. The spectrum for a

random coil has a negative band around 200 nm. Aromatic residues also exhibit circular dichroism with π_0- π^* transition in near UV range from 300-250 nm. The peak at ~ 260 nm is related to n-σ^* transition that is contributed from disulfide bonds and is generally wider than peaks for an aromatic residue. Pyridoxal-5-phosphate, flavins, and haem are some of the non-protein chromophores that exhibit the property of circular dichroism. The stacking of bases in DNA helix gives rise to very large CD signals due to exciton coupling between the bases. The CD spectrum of B-form DNA is composed of four major peaks, negative peaks around 214 and 245 nm, and positive peaks around 225 and 280 nm [145,146]. The positive maximum around 280 nm is attributed to base stacking and the negative peak at 245 nm is attributed to DNA helicity [147]. The relative intensity of CD signal at 275 nm is often taken as a signature for DNA geometry. An increase in its intensity suggests unwinding of DNA and/or displacement of base-pairs away from helix axis.

CD has also been used extensively to give useful information about protein structure, the extent and rate of structural changes on ligand binding. CD spectrum can be analyzed to give the content of regular secondary structural features such as α-helix, β-sheet, coils, and disorder. CD is also used to study the secondary structural changes in the variant protein due to the presence of the mutation. It also can reveal the environments of the aromatic amino acid side chains and thus gives information about the tertiary structure of the protein [148,149].

In the protein design field, CD is used to assess the structure and stability of the designed protein fragments. It is a valuable tool to study secondary structural changes as a function of temperature/ denaturing agents. However, CD spectroscopy is a quick method that does not require large amounts of proteins or extensive data processing [150].

B. Fluorescence Spectroscopy

Fluorescence is an electronic phenomenon that occurs in molecules having conjugated π-electrons. The π-electrons can exist either in the ground state or in the excited state, each with vibrational energy levels in it. Absorption of photons leads to excitations of electrons to a higher energy level known as the excited state. In the excited-state lifetime, a non-radiative relaxation occurs in which the excitation energy is dispersed as vibrations or heat to the solvent and no photon is emitted (**Figure 2.2)**.

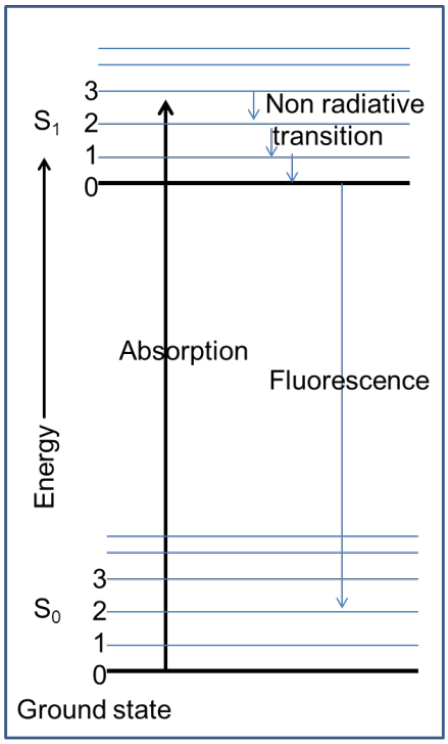


Figure 2.2: *Jablonski diagram. Horizontal lines indicate vibrational energy level in S_0 and S_1 phase. The diagram illustrates the creation of an excited electronic singlet state by energy absorption and subsequent transfer of energy by emission of fluorescence to ground state.*

The difference in energy of incident and emitted photons is due to energy dissipation, which leads to emission of longer wavelength light, and this is called as 'Stokes shift'. The electrons now jump from the lowest vibrational energy level of the excited state to the ground state via the emission of the specific quantum of energy in the form of light. This light thus emitted results in fluorescence. The fluorescence quantum yield is defined as the ratio of photons

emitted to photons absorbed. The fluorescence quantum yield of 0.10 or higher is considered as fluorescent [151].

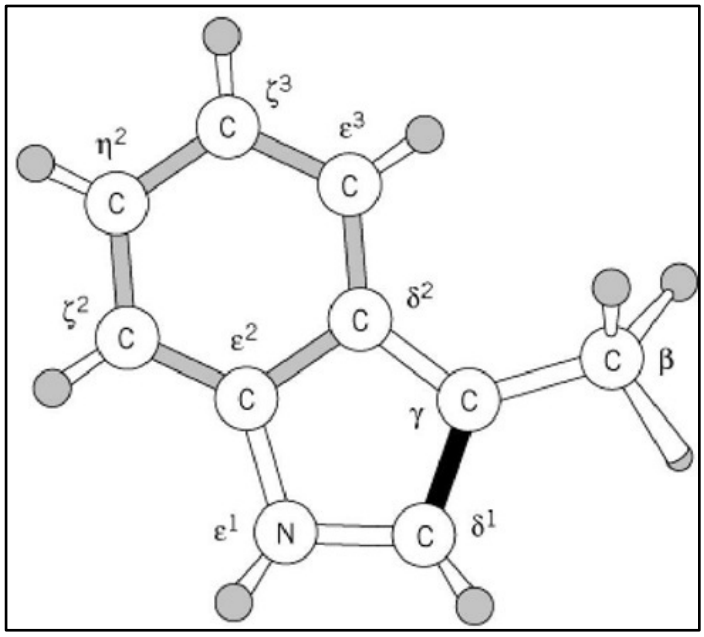


Figure 2.3: *Indole ring structure in Tryptophan*

There are many compounds in the biological system that exhibit fluorescence and proteins are amongst one of them. Phenylalanine (Phe), tyrosine (Tyr) and tryptophan (Trp) are the amino acids with intrinsic fluorescence properties. The selective excitation of Trp residues occurs at wavelengths of λ=295-305 nm and both Trp and Tyr will be excited at a wavelength of 280 nm. The large redistribution of electron density in the asymmetric indole ring of the Trp residue after the excitation makes tryptophan fluorescence sensitive to the environment. Quantum mechanical studies showed that during excitation much electron density is directed from the $N\varepsilon_1$ and $C\gamma$ atoms to $C\varepsilon_3$, $C\zeta_2$, and $C\delta 2$ atoms of the indole ring (**Figure 2.3).** This dipole moment in the indole ring in the excited state creates a local non-equilibrium in the tryptophan environment. The structural property of the tryptophan environment leads to various interactions between atoms of the indole ring and protein atoms, and/or with water molecules. These interactions affect the positions of maxima of the fluorescence spectrum and other fluorescence properties [152].

Owing to the characteristics of tryptophan, tryptophan fluorescence is used as a tool to monitor changes in proteins with regards to the local structure and dynamics. The properties studied are fluorescence intensity, emission maxima, anisotropy, fluorescence lifetimes, and energy transfer. When Tyr and Trp are present in a hydrophobic environment high fluorescence intensity is observed due to high quantum yield and vice versa for the hydrophilic environment. When tryptophan is present in the hydrophobic environment in protein the emission maxima shifts towards lower wavelength (blue shift) while if exposed the emission maxima shifts towards higher wavelength (red shift). In conclusion, if Trp is used as a probe to monitor structural change, fluorescence intensity, as well as the maximum emission wavelength (λ max), will change upon unfolding. Quenching studies are also performed using fluorescence spectroscopy which can be analyzed for study of the thermodynamic properties [153]. Extrinsic fluorescent dyes including 8-Anilinonaphthalene-1-sulfonic acid (ANS), Nile Red, Thioflavin T are used to characterize folding and understand the structural alteration due to ligand interaction. It is also used to study intermediates, detect aggregation and measure surface hydrophobicity [154].

2.2.10 Bioinformatics analyses

Mutations discovered in a cohort of Indian patient have been characterized for their pathogenecity using different online *in silico* tools

1. Align GVGD

Align-GVGD uses multiple sequence alignment (MSA) to characterize the biochemical properties viz. composition (C), polarity (P), and volume (V) of the wild type and variant amino acids at each position. The values of these properties C, P, and V are plotted on a graph. From this graph, the biochemical variation between the wild type and variant is calculated to check for the exchangeability of two amino acids. A Grantham Variation scores

(GV) and a Grantham Difference score is generated (GD). These values are used as a measure of how likely the substitution is to be deleterious or neutral on a classification spectrum. The prediction classes forms a spectrum (C0, C15, C25, C35, C45, C55, C65) with C65 being the most likely and C0 least likely, to interfere with function [155].

2. MutPred

MutPred classifies an amino acid substitution as disease-associated or neutral, based on three classes of attributes: a) the evolutionary conservation of the protein sequence, b) the protein structure and dynamics and c) functional properties like secondary structure, solvent accessibility, stability, intrinsic disorder, B-factor, transmembrane helix, catalytic residues and others. MutPred utilizes the SIFT method [156] for defining the evolutionary attributes, along with PSI-BLAST, transition frequencies and Pfam profiles [157]. MutPred provides two score values g and p: g (general)score for the probability that the substitution is deleterious, and p (property)score for the indication of the structural and functional property impacted, for instance, gain of helical propensity or loss of a phosphorylation site [158]. Certain combinations of g scores and of p scores are referred to as hypotheses.

Table 2.11: *MutPred scores for variant classification*

	Scores	
Hypothesis	g	p
Actionable hypotheses	> 0.5	< 0.05
Confident hypotheses	> 0.75	< 0.05
Very confident hypotheses	> 0.75	< 0.01

3. PhD-SNP (Predictor of human Deleterious Single Nucleotide Polymorphisms)

PhD-SNP is Support Vector Machine (SVM) based prediction method that uses protein sequence information to predict whether a mutant is disease-associated, based on a supervised training algorithm. Two SVMs are incorporated in this program, single-sequence SVM and profile-based SVM. The single-sequence SVM (SVM Sequence) classifies the

missense variant to be pathogenic or neutral based on the nature of the substitution and physico-chemical properties of the neighboring amino acid environment in the sequence. The profile-based SVM (SVM-Profile) utilizes sequence profile information taken from MSA and the conservation index is calculated for the position involved, and classifies the variant according to the ratio between the frequencies of the wild-type and the substituted residue [159].

4. PolyPhen2 (Polymorphism Phenotyping v2)

PolyPhen 2 predicts the effect of mutation using Bayesian classifier. It utilizes a combination of sequence, phylogenetic and structure based characteristics for the description of an amino acid substitution. The PSIC scores (Position-Specific Independent Counts), MSA properties and position of the mutation in relation to domain boundaries as defined by Pfam (38) are included in sequence-based classification. The structure-derived features are: change in solvent accessibility, and crystallographic B-factor. The conservation of amino acid in the MSA and the deleterious effect on the protein structure results in the Position-Specific Independent Count (PSIC) score that ranges from 0 to 1. The mutant is classified as either Possibly Damaging (PSIC > 0.5) or Benign (PSIC < 0.5) [160].

5. SIFT (Sorting Intolerant From Tolerant)

The SIFT tool uses a sequence homology based on the MSA conservation approach to classify the mutant as tolerated or damaging to the protein. SIFT makes inferences from sequence similarity using mathematical operations. The SIFT score is the normalized probability that the amino acid change is tolerated. The score ranges from 0 to 1 with a cut-off score of 0.05. Amino acids substitutions with less than 0.05 are predicted to be deleterious, and those greater than or equal to 0.05 are predicted to be tolerated [161].

6. SNAP (Screening for Non-Acceptable Polymorphisms)

SNAP is a neural network-based tool for the prediction of the effect of a missense variant. The method utilises evolutionary information from PSIBLAST [162], PSIC score, transition frequencies of the mutations, biophysical characteristics of the substitution, secondary structural information, and relative solvent accessibility values. The training sets for the neural network were constructed from Protein Mutant Database (PMD) [163]. The SNAP network uses protein sequences, lists of mutants and provides a score for each substitution, which can then be translated into binary predictions as a neutral or non-neutral effect [164].

7. PSIPRED

PSIPRED is a secondary structure prediction method. Two feed-forward neural networks are incorporated to perform an analysis on the output obtained from PSI-BLAST (Position Specific Iterated-BLAST). The prediction method or algorithm is split into three stages: generation of a sequence profile, prediction of initial secondary structure and filtering of the predicted structure.

PSIPRED normalizes the sequence profile generated by PSIBLAST and by using neural networking, the initial secondary structure is predicted (helix, sheet, coil). A second neural network is used for filtering the predicted structure of the first network. These three final outputs deliver a score for each secondary structure element. Using the secondary structure with the highest score, PSIPRED generates the secondary structure prediction. The Q3 value is the fraction of residues predicted correctly in the secondary structure states, namely helix, strand, and coil [165].

8. SABLE (Solvent AccessiBiLitiEs)

The SABLE server is used for predicting real-valued relative Solvent AccessiBiLitiEs of amino acid residues in proteins. SABLE server also predicts secondary structures using evolutionary profiles [166].

Chapter 3

A systematic clinical and in silico assessment of the missense mutations of BRCA1 identified in the Indian Population

3.1 Introduction

Hereditary breast cancer accounts for 5-10% of all breast cancers and occurs as a part of Hereditary Breast and Ovarian Cancer Syndrome (HBOC), Li-Fraumeni Syndrome and a few other cancer predisposition syndromes. Of these, HBOC is the most common breast cancer predisposition syndrome. HBOC is characterized by early onset of disease, manifestation in multiple generations, bilaterality, and the absence of expression of ER, PR and Her2/neu receptors ('Triple Negative' disease). Deleterious germline mutation in BRCA1 gene, account for 40-45% cases of cancer in HBOC families [167,168] and, in females, confer a very high life time risk of 60-80% for breast cancer and 25-40% risk of ovarian cancer by age 70 [169]. The cloning of BRCA1 gene allowed direct estimation of germline mutations which can be attributed for its role in cancer. A combination of single-strand conformation polymorphism (SSCP), heteroduplex analysis, and chemical cleavage of mismatch was used to screen the BRCA1 gene for mutations [170]. As missense variants could be either neutral polymorphisms or disease causing pathogenic mutations, Greenman et al, [170] for the first time, used five different criteria of pathogenicity, to assess the variants in BRCA1. These include 1) co-segregation, 2) absence in ethnically matched controls, 3) non-conservative amino acid substitutions, 4) residue conservation in the murine and canine homolog of BRCA1, and 5) occurrence of variant in a conserved region and a functional motif [170]. Statistical analysis using Bayesian approach to calculate posterior probability and risk assessment based on family history of proband was also one of the methods used for variant classification [171]. However analysis of a family with few and all affected members, or their availability for study is non-random, the posterior probability value estimation could be affected. [172]. The full pedigree likelihood was proposed later, which contained complete pedigree information with a clear distinction made between unaffected individuals and those of unknown genotype/phenotype [173]. Classification of inherited from sporadic breast cancer based on cytologic and architectural features of these tumours was also a

distinctive criterion later considered [174]. Goldgar et al for the first time classified the variant based on multifactorial likelihood-ratio (LR) model. Classification of variants was based on co-segregation, family history of the disease, amino acid conservation, severity of amino acid change, and an evidence from functional assays [12],[175]. Structure based assessment of variants has also been used as a mode of classification. Residues near the phosphopeptide binding pocket or those involved in binding with phosphorylated peptide or residues responsible for maintaining the integrity of the phosphopeptide binding pocket are all shown to have the deleterious effect. Protein stability and integrity were also used to understand the effect of the variant on protein folding and its functional activity [115,176,177]. Classifying BRCA1 variants using functional assay can be a useful adjunct to multifactorial analysis. Functional analysis of BRCA1 variants has been done through transactivation assay, small colony phenotype, ubiquitin ligase activity, rescue of radiation resistance and embryonic stem cell-based assay [178]. Using the large dataset contained in the largest BRCA testing laboratory, the Myriad Genetics, 1,433 sequence variants of unknown clinical significance were evaluated. The odds ratio in favour of causality was calculated from odds ratio for co-occurrence, co-segregation and family history.

The final odds /likelihood ratio was used to classify the variant as deleterious or neutral [179]. Use of data from a range of functional assays may be required to identify variants which are associated with a low to moderate risk of cancer [180]. In 2008, the International agency for research on Cancer (IARC) working group standardised a method for classification of BRCA1 variants based on the odds ratio. Based on the likelihood of pathogenicity a system of five classes of variants was recommended. A specific recommendation for clinical management of at-risk relatives was suggested for each of the five classes shown in **(Table 3.1)**.

Table 3.1: *Clinical recommendations for variants of different class*

Class	Definition of the variant	Posterior Probability	Clinical Testing for variant	Surveillance recommendations
5	Definitely pathogenic	>0.99	Test at-risk relatives	Full high-risk surveillance
4	Likely pathogenic	0.95-0.99	Test at-risk relativesf	Full high-risk surveillance
3	Uncertain	0.05-0.949	Do not use as predictive testing in at-risk relatives	Counsel based on family history and other risk factors
2	Likely not pathogenic	0.001-0.049	Do not use as predictive testing in at-risk relatives	Counsel as if no mutation detected
1	Not Pathogenic	<0.001	Do not use as predictive testing in at-risk relatives	Counsel as if no mutation detected

A computational statistical approach for inferring the disease relevance of variant from data derived from an *in vitro* functional assay is also under process. It is based upon a Bayesian hierarchical model that accounts for sources of experimental heterogeneity. Classification of variants based on clinical and functional data will lead to improved classification of VUS and will aid in the clinical decision making for the carriers of variants [181]. **(Table 3.1)** shows clinical recommendation made based on posterior probability value calculated from clinical likelihood ratio. A consortium called the **E**vidence-based **N**etwork for the **I**nterpretation of **G**ermline **M**utant **A**lleles (ENIGMA) was formed as an international initiative to evaluate risk and clinical significance of sequence variants in BRCA1/2 genes. It currently includes more than 100 research scientists and clinicians from 19 countries [182].

Here, we have carried out systematic *in silico*, genetic and clinico-pathological evaluation of the germline missense BRCA1 variants that were identified in a cohort of 48 Indian HBOC families. The BRCA1 variants were identified through Sanger sequencing of BRCA1 gene for the cancer

affected proband. Their cancer-affected or healthy relatives were also evaluated for deleterious mutations or VUS identified in the family. Genetic testing was always done after detailed pre-test genetic counselling and written informed consent. Of the 19 missense variants identified, 10 were already reported in the BIC/ ClinVar database, 5 were synonymous variants and the remaining 4 (BRCA1 C1697Y, S1722P, F1124I and, Q210H) were novel unreported variants. The posterior probability value of variants S1722P and C1697Y belong to the definitely pathogenic IARC class 5. Expectedly, these residues are at positions which are conserved across the species. For one of the variant, E809G, reported in ClinVar database, even though not a conserved residue has high posterior probability value and is classified into IARC class 4, suggesting thus a likely pathogenic variant. F1124I and Q210H are at non-conserved positions and belong to IARC class 2, which indicates little clinical significance. Therefore S1722P, C1697Y and, E809G mutants were further pursued for functional study. Study of variant S1722P and C1697Y is described in chapter 4and E809G and P871L (polymorphic variant) in chapter 6.

3.2 Materials and Methods

3.2.1 Identification of missense BRCA1 mutations and pedigree analysis

HBOC families registered at the Cancer Genetics Clinic in Tata Memorial Centre are offered genetic counselling, and individuals who give consent for genetic testing are screened for germ-line mutations in BRCA1, BRCA2 and TP53 genes. Using genomic DNA extracted from peripheral blood from forty eight members of HBOC families, Sanger sequencing of either the mutation hot spots or complete BRCA1 and BRCA2 genes was carried out using amplicons of all the exons amplified by PCR (Gen Bank accession number BRCA1 U14680; BRCA2 U43746). Clinico-pathological information such as the age of cancer onset, type of cancer, unilateral or bilateral cancer, and pathological characteristics of tumor (grade, ER/PR/Her2 status) were ascertained. This information is used in IARC classification of the variant using multifactorial likelihood ratio calculation as described below.

3.2.2 Multifactorial Likelihood Ratio (LR) calculation

The calculation of final posterior probability based on clinical and pedigree information is one of the fastest and reliable methods for variant classification. Co-segregation, Co-occurrence, family and proband LR and pathological features are used for multifactorial LR calculation.

a. Cosegregation

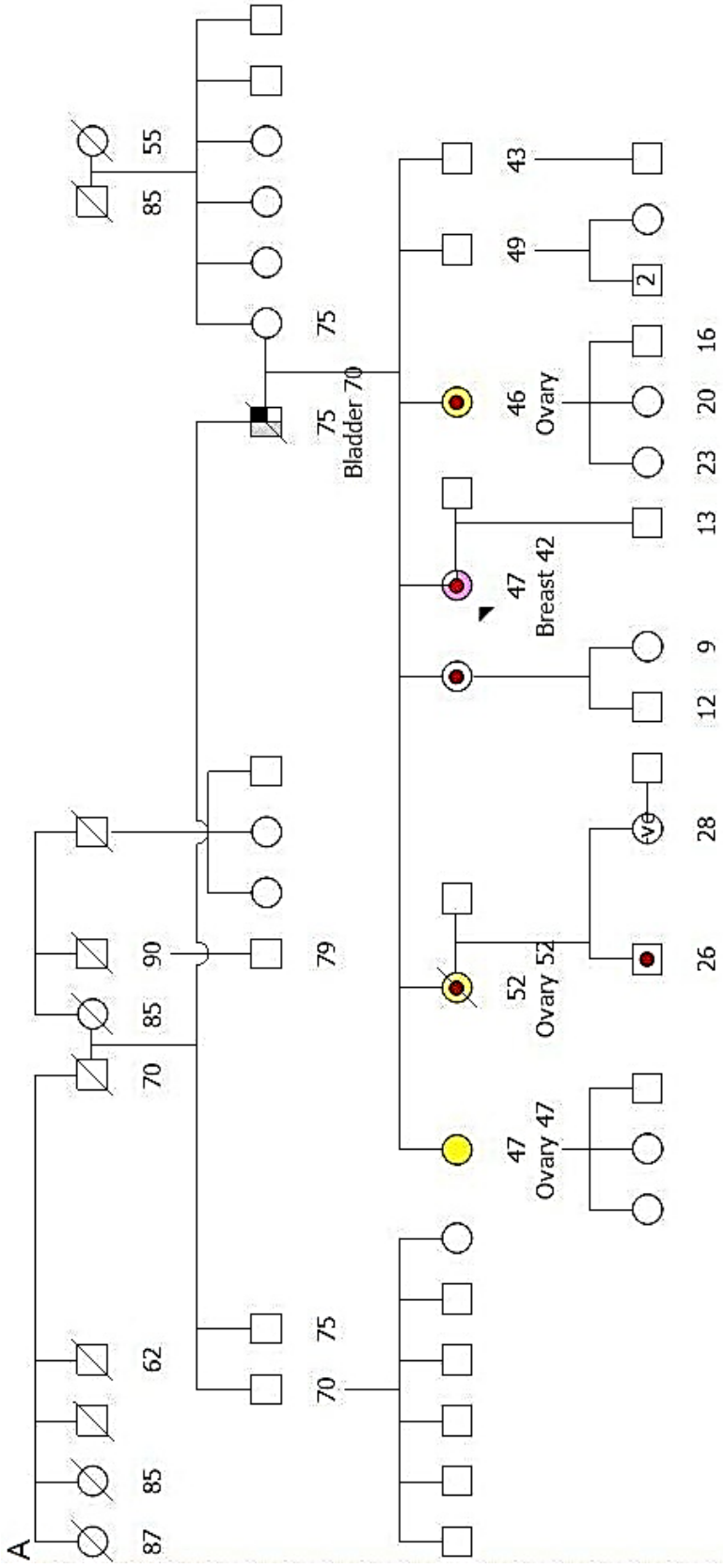
"Cosegregation analysis" calculates the likelihood ratio for inheritance of BRCA1 variant and causation of disease. The LR for co-segregation is calculated from penetrance as a function of age of onset. LR for co-segregation is obtained using the algorithm available online at http://www.msbi.nl/cosegregation [183]. The computational algorithm used to determine the likelihood ratio for pathogenicity within the families takes into account all members with available/missing genotypic information is

$$LR = \frac{\sum_{G_u} P_d(P_f|G_u,G_0,G_P=1)P(G_u,G_o|G_P=1)}{\sum_{G_f} P_d(P_f|G_f)P(G_f|G_P=1)} = \frac{\sum_{G_u} P_d(P_f|G_f)P(G_f|G_P=1)}{\sum_{G_f} P_d(P_f|G_f)P(G_f|G_P=1)}$$

Where, G stands for the genotype (VUS in the family), and is equal to 1 if the VUS is present and is equal to 0 if not. d, n, o, u, denote deleterious, neutral, observed, and unobserved variant for family (f) and proband (p) respectively. Phenotypic information is denoted by P.

Co-segregation analysis of the variant with cancer was done in two families harbouring the BRCA1 VUS, C1697Y (**Figure 3.1A)** and S1722P (**Figure 3.1B)**. **Tables 3.2 and 3.3** show the information used to calculate the likelihood ratio for co-segregation of S1722P and C1697Y respectively. The co-segregation calculation is performed by assigning number 0, 1 and, 2 based on available information which includes 1 for proband and 0 for others, gender (1 male, 2 female, 9 unknown), genotype (0 non-carrier, 1 carrier and 2 for unknown genotype), current age (if ages are unknown, estimations can be made for each generation, based on the mean age of

that generation), age of onset of the first and second breast cancer and age of onset of the ovarian cancer.



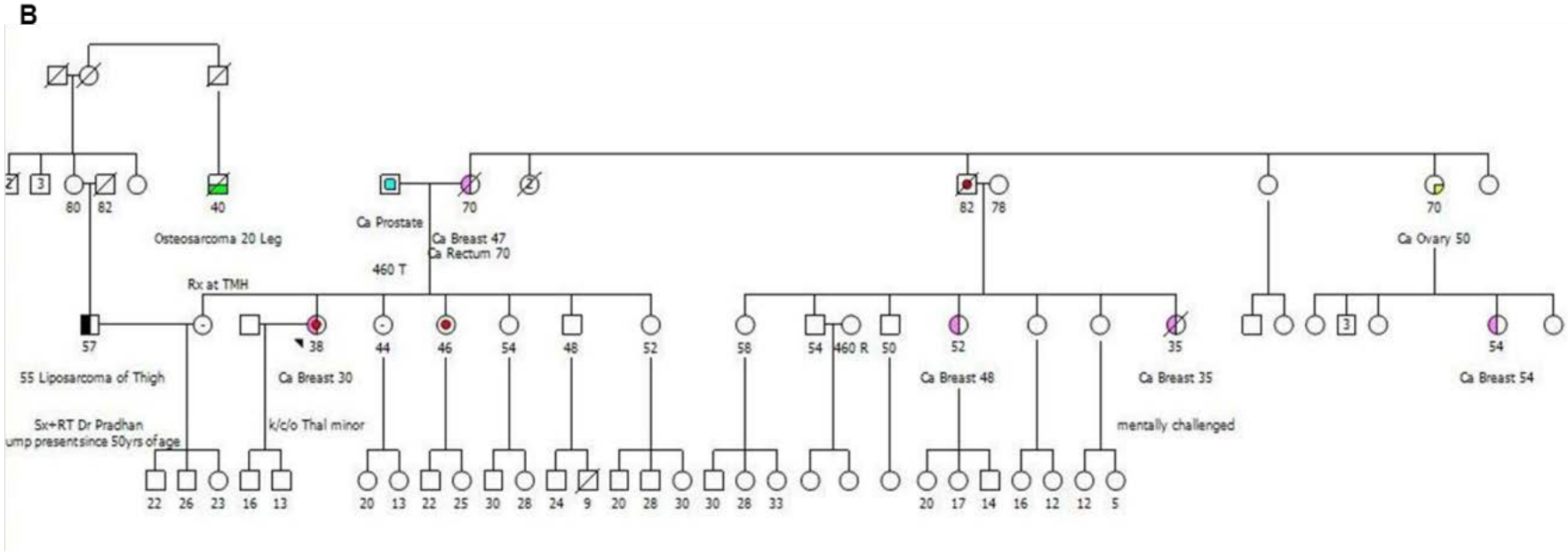


Figure 3.1: Pedigree of family with HBOC syndrome*. Females are denoted by circles and males by squares. Number in the pedigree indicates age of the individual.* *Ovarian cancer with variant,* *breast cancer with variant,* *carrier of variant, empty circle indicates absence of BRCA1 variants (A) C1697Y (B) S1722P*

Table 3.2: *Ped file for BRCA1 S1722P*

Person ID	Father ID	Mother ID	Proband	Age	Gender	Age of onset BC	Age of onset 2nd BC	Age of onset OC	Geno-type
1	3	2	1	38	2	30	0	0	1
2	11	10	0	70	2	47	0	0	1
3	0	0	0	73	1	0	0	0	2
4	3	2	0	44	2	0	0	0	0
5	3	2	0	54	2	0	0	0	0
6	3	2	0	52	2	0	0	0	0
7	3	2	0	50	2	0	0	0	0
8	3	2	0	48	1	0	0	0	1
9	3	2	0	46	2	0	0	0	1
10	0	0	0	85	2	0	0	0	2
11	0	0	0	90	1	0	0	0	2
12	11	10	0	70	2	50	0	0	2
13	0	0	0	75	1	0	0	0	2
14	11	10	0	73	1	0	0	0	1
15	0	0	0	72	2	0	0	0	2
16	14	15	0	54	1	0	0	0	1
17	14	15	0	50	1	0	0	0	1
18	14	15	0	52	2	48	0	0	1
19	14	15	0	42	2	0	0	0	0
20	14	15	0	35	2	35	0	0	2
21	14	15	0	58	2	0	0	0	0
22	14	15	0	38	2	0	0	0	0
23	13	12	0	54	2	54	0	0	2

OC-Ovarian cancer, BC-Breast cancer

The first HBOC family with BRCA1 S1722P variant belongs to the Hindu Sindhi community. Of the 15 members in this family, 7 first degree and 7 third degree relatives of the proband were assessed for this variant. In the second HBOC family, belonging to Oriya Brahmin community, 6 first-degree relatives of the proband were tested for the BRCA1 variant C1697Y.

Table 3.3: *Ped file for BRCA1 C1697Y*

Person ID	Father ID	Mother ID	Proband	Age	Gender	Age of onset BC	Age of onset 2nd BC	Age of onset OC	Genotype
1	2	3	1	47	2	42	0	0	1
2	0	0	0	75	1	0	0	0	2
3	0	0	0	68	2	0	0	0	2
4	2	3	0	47	2	0	0	47	2
5	2	3	0	48	2	0	0	0	1
6	2	3	0	46	2	0	0	46	1
7	2	3	0	52	2	0	0	52	1
8	0	0	0	60	1	0	0	0	2
9	8	7	0	26	1	0	0	0	1

OC-Ovarian cancer, BC-Breast cancer

b. Co-occurrence

The likelihood-ratio for co-occurrence of a VUS in trans with a deleterious mutation is calculated from the following equation [179]

$$(P_2)^k * (1-P_2)^{n-k} / P_1^k * (1-P_1)^{n-k}$$

where n is the number of times the variant was observed, k is the number of times it co-occurred in trans with a known deleterious mutation, P_1 is the total frequency of pathogenic mutations in the data set, and P_2 is the probability of being a compound heterozygote for two pathogenic mutations i.e. 0.0001. P_1 value for co-occurrence which is 0.07 was calculated from the cohort of patient in Tata Memorial centre.

c. Personal and Family history of the proband

Family LR calculations were obtained from odds ratio calculated from logistics regression analysis. As we did not have a large dataset, to calculate odds ratio for classification of variant based on personal and family history of proband, we used the odds ratio calculated from myriad genetics database of 47,335 probands [179]. Odds ratio for a proband is calculated based on gender, age of breast or ovarian cancer diagnosis in the variant carrier and, unilateral or bilateral breast cancer. The family history odds ratio calculation considers number of first degree and second degree relatives affected with cancer and their age at breast and ovarian cancer onset.

d. Pathological characteristics

LR in favour of mutation being pathogenic can be estimated from pathological characteristics of tumours with and without BRCA1 mutation. To calculate the likelihood ratio for histopathology, Chenevix-Trench, et al studied these features in 600 tumours from patient with BRCA1 mutations as against age matched 258 patients without mutation in BRCA1/2. The odds in favour of causality for variant being pathogenic were obtained based on the patient ER/PR status and grade of tumour [184,185]. Odds ratio for pathology of breast cancer with BRCA1 variant carrier in our study was obtained from this previous published information. The tumour grade, ER/PR status, cytokeratin 5, 6, 14 and HER2/neu expression are characteristics in tumors from carriers and non-carriers of BRCA1/2 pathogenic mutations.

3.2.3 Posterior probability calculation

The final posterior probability of a variant is calculated from combined likelihood ratio and prior probability. To determine the combined LR for a particular variant the values of each individual LR of clinical and pedigree information is multiplied

Likelihood of pathogenicity = LR (Co-occurrence × Pathology × Segregation × Cancer history).

where LR is likelihood ratio or odds ratio of pathogenicity.

The prior probability value is obtained from sequence conservation information of BRCA1 obtained from Align-GVGD. These combined likelihood ratio and prior probability is used to obtain posterior odds which is used to calculate the final posterior probability. The final posterior probability obtained was used to classify the variants into five IARC classes based on five tier IARC classification system [13]. This definition of variant being pathogenic or likely not pathogenic is used to counsel patients for genetic testing, disease management and surveillance recommendation.

Posterior Odds = Likelihood ratio × [prior probability/ (1-prior probability)]

Posterior Probability of Pathogenicity = Posterior Odds / (Posterior Odds + 1)

3.2.4 *In silico* analysis

As few of these missense mutations were not reported earlier and their biological effects are unknown, we performed a comprehensive sequence and structure based *in silico* analysis using online tools. The sequence based analysis was performed using Align GVGD [155], MutPred [158], PhD-SNP [186], and multiple sequence alignment using ClustalW [187]. The sequence and structure based analysis was performed using PMUT [188], SNAP [164], and PolyPhen-2[160].

3.3 Results

The systematic evaluation was carried out for the germline missense BRCA1 variants identified in 48 HBOC families through Sanger sequencing. Of the 19 missense variants identified, 10 were already reported in the BIC or ClinVar databases, 5 were synonymous variants and the remaining 4 (C1697Y, S1722P, F1124I and, Q210H) were novel unreported variants. One of these variant (E809G) has been reported recently as a VUS. The likelihood ratio for the variants identified in Indian HBOC families as being deleterious was calculated by examining 1) co-

occurrence, 2) co-segregation 3) proband and family history of cancer and, 4) pathological characteristics.

3.3.1 Clinical analysis of identified variants

Multifactorial likelihood ratio was calculated for Cosegregation, co-occurrence, family history and pathology based on available information.

a. Cosegregation

If majority of the individuals carrying a particular gene variant develop the gene associated disease e.g. HBOC associated cancers with BRCA1, the variant is likely to be pathogenic. Co-segregation of the unclassified variant with cancer cases within the family helps to classify the variant as pathogenic. Such studies in large families are highly informative.

In the family with novel variant C1697Y, 4/5 females were tested for inheritance of variant. 3/4 female carriers developed epithelial ovarian cancer or triple negative infiltrating duct carcinoma of the breast. The only female carrier of this variant in the family who has not yet developed any cancer is 30 years old and has substantial risk of developing cancer later in life (**Figure 3.1A).**

The novel variant S1722P was found in 4/11 females and 4/4 males tested for this variant. Of the 4 female carriers, 3 have developed epithelial ovarian cancer or triple negative breast cancer. The only female carrier of this variant in the family who has not yet developed any cancer is 46 years old and has substantial risk of developing cancer later in life. The 7 females without this variant have not developed any cancer with their ages being in the range of 43-63 years. None of the four male carriers of this variant has developed any cancer at 53, 55, 59 and, 87 years of age (**Figure 3.1B)**. The calculated co-segregation LR values are 5.2 for C1697Y and 32.34 for an S1722P variant.

The assumption made in this algorithm includes selection of the youngest cancer affected VUS carrier in the family as the proband and assuming that the penetrance for the “Causal" VUS's

will be same as a deleterious mutation. However, the power of this analysis depends on the access to genotype data from a large number of individuals.

b. Co-occurrence

In Mendelian autosomal dominant conditions such as HBOC, co-occurrence of an unclassified genetic variant along with a known deleterious mutation, in the same gene, occurring in trans suggests that the unclassified variant is very unlikely to be a deleterious mutation. This is explained by the embryonic lethality of two deleterious mutations in trans in tumour suppressor genes like BRCA1 [189,190]. Forty eight families were assessed for co-occurrence of VUS with the deleterious variant. The LR for co-occurrence depends on the frequency of occurrence VUS in the population and the frequency of co-occurrence of VUS in trans with known deleterious mutation. The likelihood ratio calculated for the different identified VUS is shown in (**Table 3.4).** As expected, VUS which were reported multiple times in this cohort and also co-occurred with a deleterious mutation several times had a low likelihood ratio of co-occurrence. This indicates that the VUS is least likely to be pathogenic. The novel variants reported (C1697Y, Q210H, S1722P and, F1124I) did not co-occur with any deleterious mutation resulting in a high LR for co-occurrence. The absence of co-occurrence in these novel missense variants with deleterious mutations in HBOC cases indicates their pathogenicity.

Although it is assumed that two deleterious variants in trans will be embryonically lethal, existence of “hypomorphic variants” with subtle effects on protein function which may not result in embryonic lethality or Fanconi anaemia is a hurdle in using co-occurrence LR alone for VUS classification.

Table 3.4: *Co-occurrence of VUS in TMC cohort.*

BRCA1 variant	No. of times co-occurred with deleterious mutation	No of times reported	Likelihood ratio
C64R	0	1	1.076318622
E809G	1	0	0.001408451
*F1124I	0	1	1.076318622
*C1697Y	0	5	1.444455849
*S1722P	0	8	1.801054418
M1652I	1	1	1.076318622
*Q210H	1	2	0.001515942
R504C	1	1	0.001408451
P871L	10	31	1.43938E-28
M1008I	1	1	0.001408451
E1038G	9	28	8.82168E-26
S1040N	1	1	0.001408451
K1183R	9	30	1.02196E-25
S1613G	10	36	2.07912E-28

* indicates novel variants

c. Family history

There are certain characteristics observed in a family with true deleterious mutation as compared to a family with sporadic cancer harbouring a benign or polymorphic variant. In families harbouring a deleterious mutation, the cancers generally occurs at the young age, several first and second degree relatives are affected and bilateral disease or breast and ovarian cancer in an individual is seen. The final family likelihood ratio calculated for the VUS is shown in **(Table 3.5).** The odds ratio value is dependent on type of cancer breast or ovarian, age of onset, number of affected relatives with age, number of affected males and bilateral cancer.

Table .3.5*: Family likelihood ratio prediction (TMC cohort).*

BRCA1 variant	Personal LR	Family LR	Final Family LR
C64R	15.3	5.6	85.68
E809G	416.88	3.33	1388.21
*F1124I	1.67	0.56	0.9352
*C1697Y	11.8	36.963	436.1634
*S1722P	9.65	21.2736	205.2902
M1652I	1.67	15.8175	26.41523
*Q210H	3.4	8.31	28.254
R504C	3.4	0.7544	2.56496
P871L	15.3	7.1928	110.0498
M1008I	15.3	8.31	127.143
E1038G	8.16	39.4725	322.0956
S1040N	18	36.963	665.334
K1183R	9.65	2.7306	26.35029
S1613G	9.65	3.33	32.1345

* indicates novel variants

The major hurdle of classifying a VUS based on family data is access to family information of all alive and dead members of the family, reasons for early death if any and information regarding the carrier status of the family members.

d. Tumor Pathology;

Breast cancer pathology is also an important criterion for variants classification. The distinct pathological characteristics of the tumour with and without BRCA1 pathogenic mutation are considered for classification. Tumors with pathogenic BRCA1 mutations are mostly of high grade and negative for estrogen and progesterone receptor, and HER2/neu expression [191].

***Table 3.6**: Scores for pathological characteristics of proband.*

BRCA1 variant	**Pathological Characteristic**	**Score**
C64R	ER, PR -ve, IDC III	2.95
E809G	ER / PR -ve, IDC III, bilateral breast + ovary	2.95
*F1124I	IDC grade 1, Triple negative	0.97
*C1697Y	ER, PR -ve ,IDC III	2.95
*S1722P	ER, PR -ve ,IDC III	2.95
M1652I	IDC grade III, ER, PR +ve, erb2 –ve	0.18
*Q210H	ER, PR -ve ,IDC III	2.95
R504C	ER, PR -ve ,IDC III	2.95
P871L	ER, PR -ve ,IDC III	2.95
M1008I	ER, PR -ve ,IDC III	2.95
#E1038G	ER, PR -ve ,IDC III	2.95
S1040N	Only ovarian cancer in family	NA
#K1183R	ER, PR -ve ,IDC III	2.95
#S1613G	ER, PR -ve ,IDC III	2.95

*Novel variants; #variant co-occurring with a known deleterious mutation in this individual

ER positive and high grade with serous histology is observed for ovarian tumors with pathogenic mutation. Based on the histopathological characteristics, scores are assigned to the variants in our cohort (**Table 3.6)**.

3.3.2 IARC classification of identified variants of BRCA1

The Posterior Probability of Causality for each VUS is estimated from combined likelihood ratio and the prior probability value. The combined likelihood ratio value obtained from multiplication of each LR value (LR for family history, co-segregation, co-occurrence, pathology) is shown in (**Table 3.7).** The posterior probability value, calculated from posterior odds and prior probability is shown in (**Table 3.8)**. The identified variants were categorized into the different class based on posterior probability value.

***Table 3.7**: Combined likelihood ratio from clinical data.*

BRCA1 variant	Co-segregation LR (A)	Co-Occurrence LR (B)	Pathology LR (C)	Final Family LR (D)	Final LR (AxBxCxD)
C64R		1.076318622	2.95	85.68	272.04599
E809G		1.076318622	2.95	1388.2104	4407.76228
*F1124I		1.076318622	0.97	0.9352	0.97637598
*C1697Y	5.199	1.444455849	2.95	436.1634	1858.55538
*S1722P	32.3395	1.801054418	2.95	205.29024	1090.72974
M1652I		0.001408451	0.18	26.415225	0.00669682
*Q210H		0.001515942	2.95	28.254	0.12635268
R504C		0.001408451	2.95	2.56496	0.01065723
P871L		1.43938E-28	2.95	110.04984	4.6729E-26
M1008I		0.001408451	2.95	127.143	0.52827021
E1038G		8.82168E-26	2.95	322.0956	8.3822E-23
S1040N		0.001408451	NA	665.334	0.93709014
K1183R		1.02196E-25	2.95	322.0956	9.7105E-23
S1613G		2.07912E-28	2.95	322.0956	1.9755E-25

* novel variants. Co-Segregation values calculated only for families with variant information in several members of family

The variants S1722P and C1697Y are classified in definitely pathogenic IARC class 5. Expectedly, these residues are strictly conserved across the species. Even though E809 is not a conserved residue, its high posterior probability value classifies it into IARC class 4, a likely pathogenic variant. F1124I and Q210H are not conserved and belong to IARC class 2, which indicates VUS of little clinical significance. The co-occurrence of Q210H variant with a deleterious mutation suggests that this variant is very unlikely to be a deleterious mutation (**Table 3.9)**. All other variant belonging to class 2 and 1 are likely not pathogenic variant are not considered for further functional study.

***Table 3.8**: Posterior probability value of identified variants of BRCA1.*

BRCA1 variant	Prior probability	(prior prob/1-prior prob)	Posterior odds	Posterior probability
C64R	0.81	4.263157895	1159.775009	0.999138507
E809G	0.02	0.020408163	89.95433224	0.989005471
*F1124I	0.02	0.020408163	0.01992604	0.01953675
*C1697Y	0.81	4.263157895	7923.315061	0.999873806
*S1722P	0.81	4.263157895	4649.953088	0.99978499
M1652I	0.03	0.030927835	0.000207118	0.000207075
*Q210H	0.02	0.020408163	0.002578626	0.002571994
R504C	0.02	0.020408163	0.000217494	0.000217447
P871L	0.02	0.020408163	9.53652E-28	9.53652E-28
M1008I	0.02	0.020408163	0.010781025	0.010666034
E1038G	0.02	0.020408163	1.71065E-24	1.71065E-24
S1040N	0.02	0.020408163	0.019124289	0.018765413
K1183R	0.02	0.020408163	1.98173E-24	1.98173E-24
S1613G	0.03	0.030927835	6.10991E-27	6.10991E-27

* Novel variants. Posterior Odds = Final LR × [prior probability/ (1-prior probability)]; Posterior Probability of Pathogenicity = Posterior Odds / (Posterior Odds + 1)

Table 3.9: IARC classification

BRCA1 variants	Posterior probability	IARC class	IARC classification
C64R	0.999138507	5	definitely pathogenic
E809G	0.989005471	4	likely pathogenic
*F1124I	0.01953675	2	little clinical significance
*C1697Y	0.999873806	5	definitely pathogenic
*S1722P	0.99978499	5	definitely pathogenic
M1652I	0.000207075	1	not pathogenic
*Q210H	0.002571994	2	little clinical significance
R504C	0.000217447	1	not pathogenic
P871L	9.53652E-28	1	not pathogenic
M1008I	0.010666034	2	little clinical significance
E1038G	1.71065E-24	1	not pathogenic
S1040N	0.018765413	2	little clinical significance
K1183R	1.98173E-24	1	not pathogenic
S1613G	6.10991E-27	1	not pathogenic

3.3.3 *In-silico* study to identify pathogenicity of BRCA1 variants

The tools used in this study which include Pmut, SNAP, PhD-SNP, Mutpred, Polyphen 2 and Align-GVGD use a combination of factors like sequence conservation across the species, alterations in the biochemical property and structure to predict the pathogenicity of the missense variants. All these tools perform analysis of amino acid, based on their charge, hydrophobicity, bulkiness and solvent accessibility. Analysis of crystal structure, change in active site & structural change are also used by Polyphen2, PMut and Mutpred to evaluate the pathogenicity of the variant. Polyphen2 and PMut also perform alanine scanning and massive mutation analysis. Comprehensive *in silico* analysis predicted the effect of identified variants based on sequence, structure and sequence plus structure based analysis **(Table 3.10).**

In silico analysis was also performed on M1775R which is a well-documented pathogenic variant and is used as a negative control in our yeast functional assay (data shown in chapter 4). In light of the analysis of our and previous data, *in silico* analysis does not unambiguously define pathogenicity of VUS.

Table 3.10: In silico analysis

BRCA1	**A-GVGD**	**PHD-SNP**	**POLYPHEN**	**SIFT**	**Mutpred**
Variant	Hum-Frog	Effect	Prediction	Class	deleterious
C64R	Class C65	DAMAGING	DAMAGING	affect	deleterious
E809G	Class C0	DAMAGING	BENIGN	tolerated	deleterious
*F1124I	Class C0	DAMAGING	BENIGN	affect	deleterious
*C1697Y	Class C65	DAMAGING	DAMAGING	affect	deleterious
*S1722P	Class C65	DAMAGING	DAMAGING	affect	deleterious
M1652I	Class C0	Neutral	BENIGN	affect	Neutral
*Q210H	Class C0	Neutral	DAMAGING	tolerated	Neutral
R504C	Class C0	DAMAGING	DAMAGING	tolerated	deleterious
P871L	Class C0	Neutral	BENIGN	tolerated	Neutral
M1008I	Class C0	Neutral	BENIGN	tolerated	Neutral
E1038G	Class C0	Neutral	DAMAGING	tolerated	Neutral
S1040N	Class C0	Neutral	DAMAGING	tolerated	Neutral
K1183R	Class C0	Neutral	BENIGN	tolerated	Neutral
S1613G	Class C0	Neutral	BENIGN	tolerated	Neutral
M1775R	Class C65	DAMAGING	DAMAGING	affect	deleterious

Red-deleterious variant, green-neutral variant

Different servers do not predict the pathogenicity of VUS in concordance with each other and also ambiguity is observed for *in silico* and functional data which can create a problem in the classification of moderate risk variants.

3.4 Summary

Here, we have performed comprehensive *in silico*, genetic and clinicopathological evaluation of the germline missense BRCA1 variants identified in a cohort of 48 Indian HBOC families. Full gene sequencing of 48 samples was carried out to identify the variants responsible for inherited breast and ovarian cancer. Of the 19 missense variants identified, 10 were already reported in the BIC/ ClinVar databases, 5 were synonymous variants and the remaining 4 (C1697Y, S1722P, F1124I and Q210H) were novel unreported variants. Classification of VUS is necessary for further surveillance recommendation. In novel variants reported, S1722P and C1697Y belong to the definitely pathogenic IARC class 5. F1124I and Q210H were classified into IARC class 2, which indicates the class of little clinical significance. The high posterior probability value of variant E809G, recently reported in ClinVar database, classifies it into IARC class 4, a likely pathogenic variant. *In silico* analysis using different online algorithms has characterised the VUS to be possibly damaging/neutral. BRCA1 variants C1697Y, S1722P, C64R and M1775R were consistently classified as a damaging variant by all online servers used. The P871L, M1008I, K1183R and S1613G variants were classified as the neutral variant. All other variants were classified differently by different servers creating ambiguity in the use of online tools for variant classification. The variants S1722P, C1697Y and E809G which belonged to the IARC class 4 or 5 were further validated with functional and biophysical assay, the results of which are described in chapter 4.

Classification of variant into deleterious or benign requires a panoramic view. Presymptomatic testing cannot be offered in cases of intronic or missense variants based on statistically

insufficient clinical data, family history or *in silico* analysis. Plon et al based on the degree of likelihood of pathogenicity classified the variant into five different classes [13]. The desirable feature of this classification system is it takes into consideration grades of risk and clinical management for that risk. Classification of variant with comprehensive analysis may prevent the patients from unnecessary prophylactic surgeries (mastectomy & oophorectomy), therapeutic treatments and intensive screening. Similarly, cancer patients carrying these mutations could be advised about the benefit of BRCA1 specific targeted therapy like platinum agents or PARPi. The outcome of this definitive information will help avoid misinterpretation & prevent from undesirable stress, anxiety and fear in patients and their relatives. It will also help clinicians to use the available manpower and resources in an economical way. So classifying a variant as pathogenic, based on unreliable methods could create greater uncertainties and challenges in counselling families harbouring such VUS. Therefore, comprehensive analysis, as performed in our study, could help classify most VUS and is a very useful approach for classification of variants. As novel or rarely reported variants may be more common in specific geo-ethnic groups, it is important to perform such comprehensive analysis in diverse populations. This is the first such study in the Indian population.

Chapter 4

Structural and Functional basis to classify novel cancer risk mutations identified in BRCA1 transactivation domain

4.1 Introduction

Mutations and deregulation of BRCA1 is well documented to be responsible for hereditary and sporadic breast and ovarian cancer. Inherited cancers are caused due to missense mutations, frame shift mutations, and indels. Missense mutations of uncertain significance (VUS) can be classified either as polymorphic or disease associated variants. If the missense variants are observed in large numbers among a population, such variants are classified as polymorphic and the classification would be definite. However, if the variants observed are rare and ethnic specific, the classification becomes challenging and not certain. In such a situation, further validation of the effect of variant by structural and functional studies is necessary to be able to effectively classify the variant. In this chapter, we have structurally and functionally characterized two rare IARC class 5 variants C1697Y and S1722P of BRCA1 for which the available clinical data is limited.

4.2 Materials and Methods

4.2.1 Cloning of BRCA1 TAD domain

The Trans-Activation Domain (TAD) of BRCA1 (1560-1859) was amplified and cloned in pGEX-KT and pGBKT7 vectors for protein purification and yeast transactivation assay respectively. The PCR was performed from full length BRCA1 cDNA using the following forward and reverse primers **(Table 4.1).** The PCR were performed in 50 µl total volume containing pfu polymerase, 0.2 mM of each dNTP, 10 pM of both primers, and 100 ng FL BRCA1 DNA. The conditions for various steps used in the amplification of amplicons from cDNA construct are mentioned in **Table 4.2**.

Table 4.1: *List of primers used for BRCA1 TAD cloning*

Region cloned	**Forward and Reverse primers (5'→3')**
BRCA1(1560-1859)	BRCA1 1560 F GTCGGATCCGAGAACCTGTACTTTCAGGGTGGAACCCCTTACCTGGAATC
	BRCA1 1859 R GTAGAATTCCTATTAGTAGTGGCTGTGGGGGATCTGGGG
BRCA1 (1560-1863) (yeast assay)	BRCA1 1560 F GTCGAATTCGGAACCCCTTACCTGGAATCTGGAATC
	BRCA1 1863 R GTCGGATCCCTATTAGTAGTGGCTGTGGGGGATCTGGGGTATCAGGTA

Table 4.2: *Program setup for PCR reaction of TAD*

Sr. nos.	**Steps**	**Temperature (^{o}C)**	**Time (seconds)**
1	Denaturation	95	180
2	Denaturation	95	30
3	Annealing	60	30
4	Extension	72	60
Go to step 2, 32 times			
5	Final Extension	72	600

The PCR product and vector were then subsequently digested with BamH1 and ECoR1 restriction enzymes to generate cohesive ends. The digested vector and PCR product was ligated using ligase enzyme at 16^{o}C for 6 hr. The ligated product was then transformed into *Escherichia coli* DH5α cells and plated onto the LB-antibiotic plates (pGEX-KT ampicillin, pGBKT7-kanamycin). The resistant clones were further checked by insert release and were sequenced to confirm for the correct DNA sequence.

4.2.2 Site-directedmutagenesis of BRCA1 C1697Y, S1722P and M1775R

Site directed-mutagenesis for BRCA1 C1697Y, S1722P and M1775R variants were made in BRCA1 TAD/pGEX-KT and TAD/pGBKT7 as a template. Primers used for the SDM are listed in **Table 4.3**. To the PCR reaction mixture forward primer was added and reaction was continued for 3 cycles, and then reverse primer was added, and further 18 cycles of PCR were performed as mentioned in **Table 4.4**. After the last cycle, the reaction mixture was incubated at 72°C for an additional 10 minutes.

Table 4.3: *Primers for site directed mutagenesis*

Gene Name /(residue position of SDM)	**Forward and Reverse primers (5'→3')**
BRCA1 C1697Y	C1697Y F GCTGAGTTTGTGTATGAACGG
	C1697Y R CCGTTCATACACAAACTCAGC
BRCA1 S1722P	S1722P F GACCCAGTCTATTAAAGA
	S1722P R TCTTTAATAGACTGGGTC
BRCA1 M1775R	M1775R F GACCCAGTCTATTAAAGA
	M1775R R TCTTTAATAGACTGGGTC

M1775R was used as negative control for the transactivation assay. All the constructs were PCR amplified, digested with Dpn1 to remove methylated template. Each of the clones was sequenced for confirmation.

Table 4.4: *Program setup for SDM reaction*

Sr. nos	Steps	Temperature (^oC)	Time (seconds)
1	Denaturation	95	180
2	Denaturation	95	30
3	Annealing	52	30
4	Extension	72	800
Go to step 2, 18 times			
5	Final Extension	72	600

4.2.3 Expression of BRCA1 TAD and variants

The cloned constructs of BRCA1 were transformed in *E. coli* BL21 (DE3) bacterial strain. Single colony was inoculated in Luria Bertani broth containing ampicillin at a concentration of (100 μg/ml), and the culture was allowed to grow for 12-14 hours. These cells were then diluted 100 fold by adding LB media, and were grown till the OD_{600} reached a value in the range of 0.6-0.8. This culture was then induced with 0.4 mM IPTG at 18^0C for 16 hours. Cells were then harvested by centrifugation at 6000 rpm for 10 minutes, and the pellets were stored at -80^0C till further processing.

4.2.4 Purification of BRCA1 TAD and variant S1722P

Cells were resuspended in buffer A (containing 50 mM Tris pH 7.5, 300 mM NaCl, 0.1% Triton X-100, 20 mM BME, 1mM EDTA, 5% glycerol, and protease inhibitor cocktail). Cells were lysed, using sonicator at 65 pulser rate with 1 min duty cycle for 5 times. Cell debris was removed by centrifugation at 18,000 rpm at 4^0C for 50 minutes. The soluble part of protein was allowed to bind on GST resin for 1-2 hours at 4^oC. Washed the GST beads with 5-6 column volume of buffer A to remove non-specifically bound proteins. On-column cleavage was done with TEV for 3 hr so as to separate protein from GST tag and get protein

in native form. Further, protein was eluted from beads and concentrated using Amicon Millipore tube and passed through Ni-NTA resin to remove his-tagged TEV.

The steps of purification were similar for S1722P variant protein, except that the wash buffer used after binding to remove non-specifically bound proteins was buffer A with 10 mM ATP and 15 mM $MgCl_2$. S1722P bound beads were incubated with wash buffer for 30 min to get rid of contaminating chaperon protein. Beads were further washed with buffer A to remove any traces of ATP present and were preceded for TEV cleavage to remove the tag.

BRCA1 TAD and BRCA1 S1722P proteins were finally concentrated up to 2 ml and centrifuged at 4^0C for 10 min and passed through size exclusion chromatography. Superdex 75 and 200 column were used for TAD and S1722P purification respectively, so as to remove other contaminating band and get the proteins in homogenous form.

4.2.5 *In silico* analysis

The molecular models for the mutants were generated using the crystal structure of BRCA1 BRCT (PDB ID 1T15) as a template. These structures were subjected to molecular dynamics simulation for a period of 8 ns under standard temperature pressure conditions using the desmond module [192] of the software suite Schrodinger. The structures were subjected to force field, Optimised Potential for Liquid Simulated force field (OPLS) [193]. Protein structure was solvated using orthorhombic box in Monte-Carlo simulated water model. The molecular models were energy minimised using the steepest descent algorithm until 2000 maximised energy step. The minimum energy structure of PDB ID 1T15 extracted after simulation for BRCA1 BRCT, S1722P and C1697Y were used to analyse weak molecular interactions viz., hydrogen bonds and hydrophobic interactions [194].

4.2.6 Circular Dichroism Spectroscopy

Far UV CD scans were obtained from λ= 195-250 nm at data interval of 0.1 nm with a scanning speed of 20 nm/minutes Protein concentration determined using the molar extinction coefficient and absorbance value at λ=280 nm was 7μm,. All spectra were measured at 25°C. Thermal denaturation experiments were performed from 25^{o}C to 95^{o}C with a temperature interval of 2^{o}C. Samples were equilibrated at each temperature for 180 sec before recording the spectrum. Spectral scans were taken in far-UV CD range from λ= 240 to195 nm at data intervals of 0.5 nm with a scanning speed of 50 nm/minutes

4.2.7 Fluorescence emission spectroscopy

The fluorescence scan was measured in the emission wavelength range λ_{em} 310–400 nm when excited at 295 nm. The emission slit width was 5 nm. Protein concentration calculated using the molar extinction coefficient and absorbance at λ=280 nm was 2 μM. The spectral data were collected using FluorESSCENCE software.

4.2.8 Isothermal Titration Calorimetry

The interactions of BRCA1 TAD and S1722P protein with Abraxas synthetic peptides were studied using isothermal titration calorimetry. While Abraxas peptide [NH2-GFGEYSR (pS406) PTF-COOH] served as a positive control, the differently phosphorylated Abraxas peptide [NH2-GFGEY (pS404) RSPTF-COOH] served as a negative control in the titration experiments. The peptides were diluted in 25 mM Tris, pH 7.5, 150 mM NaCl which was used for BRCA1 TAD protein preparation. The protein concentration was 20 μM and the concentration of peptide used to achieve saturation was 400 μM. All experiments were carried out at 25^{o}C, with reference power of 10 μcal/ sec, with a stirring speed of 1000 rpm. Binding was recorded for 19 injections, with each injections of 2 μl volume, spaced 180 sec apart. Data

was analyzed using software Origin (Origin Lab, Northampton, MA) by assuming one set of binding sites.

4.2.9 Transformation of SFY526 yeast strain using lithium acetate method

A single colony from YPD (Yeast peptone dextrose) agar plate was inoculated in 5ml YPD broth for pre-culture, and was then incubated at 30^0C with shaking at 200 rpm for 12-16 hr. 50 µl of this pre culture was then inoculated in 50 ml of fresh YPD media and incubated at 30^0C till OD_{600} of 0.5 was attained. Culture was then harvested at 3000 rpm for 3 min and the cell pellet was resuspended in 20 ml sterile YPD broth. The resuspended cells were again centrifuged at 3000 rpm for 3 min for an effective wash. This cell pellet was resuspended in 500 µl of 0.1 M Lithium Acetate (LiAc) and further incubated at 30^0C in water bath for 15 minutes. These cells are ready for transformation. 50 µl of competent yeast cells, 100 ng of recombinant plasmid, single stranded calf thymus DNA (used at a concentration of 1mg/ml) and 300 µl PEG LiAc solution were mixed gently. This mixture was then further incubated at 30^0C for 30 min in water bath and then finally at 42^0C for 15 minutes. These cells were then plated on to YNB trp- plate and incubated at 30^0C for 48 hr.

Note: In order to prepare calf thymus ss DNA for use, it was boiled at 99^0C for 10 minutes and incubated in ice immediately for 1-5 minutes. Do not heat for more than 10 minutes Recombinant plasmid used should be approximately 100 ng in 5ul or less for efficient transformation.

4.2.10 Western blotting

S.cerevisiae yeast strain (SFY526) was transformed with vector (pGBKT-7), BRCA1 TAD wild-type, and also with BRCA1 C1697Y, S1722P and M1775R mutants. Each transform ants were grown at 30°C for 16 h. Equal number of cells were taken and lysed in buffer containing 10 mM Tris-HCl pH 7.4, 150 mM NaCl, 5 mM EDTA, 1% Triton X-100, 0.1%

sodium dodecyl sulphate, 0.1% sodium deoxycholate and protease inhibitor cocktail. Cell lysates were sonicated and centrifuged at 13,000 g for 10 min at 4°C. Equal volumes of cell lysates from each cell type were mixed with Lammeli buffer, boiled and loaded on to 12% SDS PAGE, and electrophoresed. Proteins from gel were blotted on PVDF membrane. The blotted membrane was blocked with 5% skimmed milk in 0.1% TBS-T containing (500 mM NaCl, 10mM Tris-HCl pH7.4 containing 0.1% Tween 20) at room temperature for 1h and incubated overnight at 4°C with c-myc antibody (Santacruz Biotechnology.Inc,USA (1:1000 dilution) in 1% BSA containing 0.1% TBS-T). Blots were subsequently incubated with anti-mouse HRPO (1:1000) in 1% BSA containing 0.1% TBS-T for 1h at room temperature. Blot was developed using enhanced chemiluminiscent plus reagent (GE Healthcare, Amersham, UK).

4.2.11 Cell permeabilization assay for transactivation study

The effect of C1697Y and S1722P variant on the BRCA1 transactivation activity was studied using the modified Miller protocol. All experiments were performed in triplicates with two clones. The S. cerevisiae strain SFY526 was transformed with BRCA1 TAD, C1697Y, S1722P and M1775R (negative control) plasmid using lithium acetate method. Single colony of the transformed yeast is inoculated in 5 ml growth media at 30^0C till OD_{600} reached to 0.6. Cells were harvested by centrifugation at 2000 rpm for 5 minutes Cell pellet was washed with 1 ml of sterile milli Q water and resuspended in 5 ml of 60 mM phosphate pH 7 buffer, containing 10 mM KCl, 2 mM of magnesium sulphate, 80 mg/ml CTAB, 40 mg/ml sodium deoxycholate and 50 mM βME. To 1 ml of these cells 60 µl of chloroform and 40 µl of 10% SDS were added, and the mixture was vortexed at top speed for 10 sec for permeabilization. Preincubate the cells at 28^0C for 5 min in water bath. To check for β- galactosidase activity in these cells, 200 µl of substrate solution containing 60 mM phosphate buffer pH 7, 4mg/ml of ONPG (o-nitrophenyl-β-D-Galactoside) was added, and the mixture was monitored for

development of yellow color incubated at 37^0C. When pale yellow color developed, reaction was stopped by addition of 250 µl of sodium carbonate from a 2M stock solution and simultaneously the time taken for the color to develop was recorded. Suspension was centrifuged at 13,000 rpm for 10 minutes to remove any suspended particles or cell debris. Absorbance of o-Nitrophenol was measured at λ=420 nm. The β-Galactosidase activity was calculated in Miller units using the formula: Miller units = $1000*(Abs_{420\ nm})/ ((Abs_{600\ nm})*(v)*(t))$ where Abs_{420} is the absorbance of the yellow o-nitrophenol, Abs_{600} reflects cell density, t = reaction time in minutes, v = volume of culture assayed in millilitres.

4.3 Results

4.3.1 Cloning of BRCA1 TAD domain

The PCR amplified transactivation domain (amino acid 1560-1859) of BRCA1 was cloned in vectors pGEX-kT and pGBKT7 (**Figure 4.1A)**. The fragments released on digestion with BamH1 and ECoR1 are shown in (**Figure 4.1B)**. The size of the insert is consistent with the expected value of 897 bp. The clone was further confirmed by DNA sequencing.

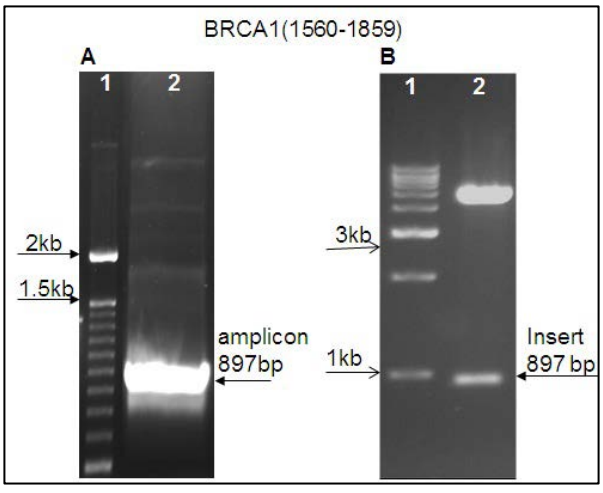


Figure 4.1: Screening clones of BRCA1 *transactivation domain (amino acid 1560-1859) with BamH1 and ECoR1 enzyme in 1% agarose gel. Lane1 DNA ladder (A): PCR product. (B): positive clone with 897 bp insert release.*

4.3.2 Functional characterization of novel VUS

The transactivation activity of the C-terminal region is very important for the tumor suppressor function of BRCA1. Hence pathogenic mutations in the BRCA1 TAD are expected to inhibit the tumor suppressor function of BRCA1. Structural and functional characterization of missense mutations in the BRCA1 TAD can provide insight into pathogenicity of mutations discovered in this domain.

The BRCA1 protein expressed along with Gal 4 DNA binding domain binds to the upstream activating sequence of lac Z gene promoter present in the yeast chromosome. Lac Z gene is a reporter gene which encodes for β-galactosidase enzyme **(Figure 4.2A)**.

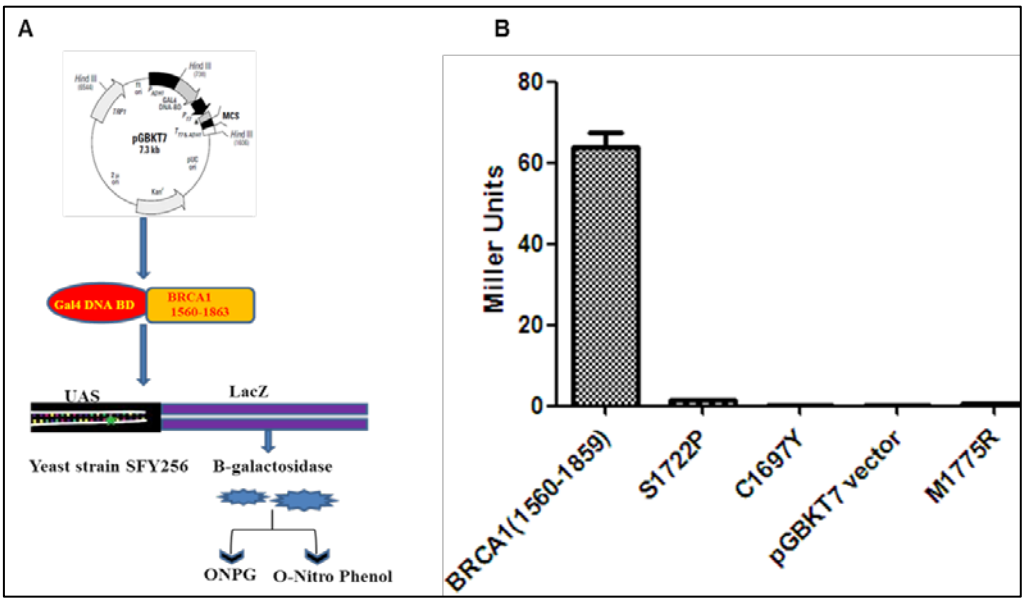


Figure 4.2: Quantitative Functional assay to determine the effect of variants in BRCA1 TAD: *Cell permeabilization assay to check the B galactosidase activity were performed on yeast transformed with TAD wild type (positive control), M1775R and pGBKT7 vector (negative control), C1697Y and S1722P (test sample) construct. Graph represents β-galactosidase activity in Miller units against TAD and its variants. The data is mean value and standard deviation of triplicate samples of two independent assays.*

The yeast strain SFY526 transformed separately with BRCA1 TAD wild type, two novel variants (BRCA1 C1697Y and S1722P) and also with a known pathogenic missense mutation

(M1775R) was grown in YNB trp$^-$ medium. Chromogenic substrate X-gal and ONPG were used to evaluate the β-galactosidase activity.

In quantitative ONPG assay, cells were permeabilized and β-galactosidase activity was spectrophotometrically measured at 420 nm for yellow colored o-nitrophenol formation from ONPG (**Figure 4.2B)**. BRCA1 TAD showed pale yellow color formation within 4-5 minutes of ONPG addition showing transactivation activity. For BRCA1 S1722P, C1697Y and M1775R even extended incubation of cells with ONPG did not show development of color suggesting absence of β-galactosidase activity. Thus, BRCA1 C1697Y and S1722P variants showed total loss in transactivation activity, as was also the case with the known pathogenic mutation M1775R.

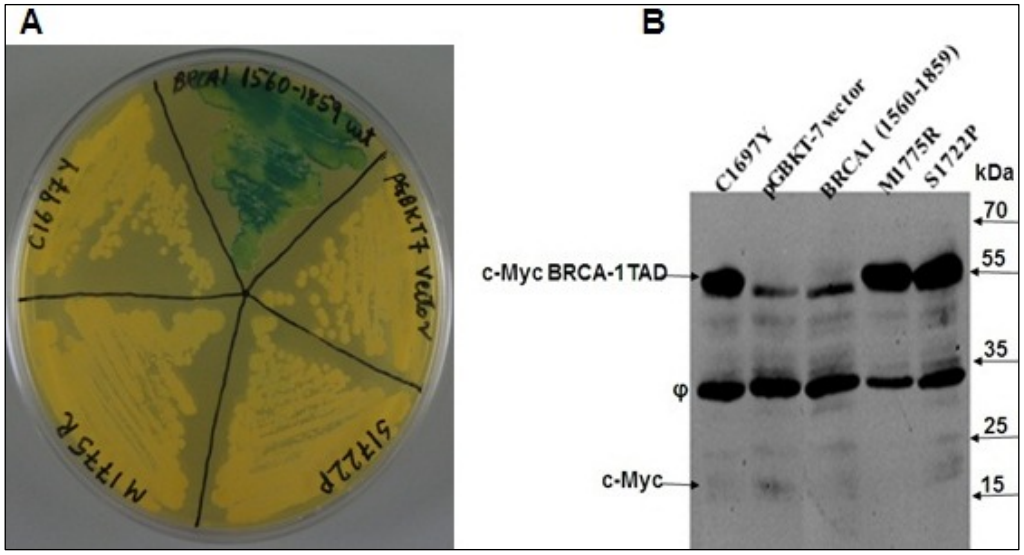


Figure 4.3: Qualitative X-gal assay: *(A): All yeast transformed strains were streaked in YPD plates containing X-gal (40 μg/ml). Blue coloured colony was observed for TAD wild type construct containing cells. (B) Western blot data indicating expression of wild type and mutant TAD. Φ indicates nonspecific band.*

In the qualitative X-gal assay, yeast strains were streaked on agar plate containing X-gal (40 μg/ml). Colonies of BRCA1 TAD construct developed blue coloration on incubation, whereas, BRCA1 M1775R, C1697Y, S1722P variants and vector control did not show any colour

formation (**Figure 4.3A)**. Western blot analysis confirmed the expression of the protein in wild type as well as in all the 3 mutants studied. Hence, it can be concluded here that the loss in transcriptional activity in these variants is not due to lack of expression, but due to their non-functionality (**Figure 4.3B)**.

4.3.3 *In-silico* analysis of BRCA1 TAD novel variants

Pathogenicity of mutations has been analysed with different online servers like SNAP, Mutpred, Polyphen 2 & Align-GVGD. All these tools predict the pathogenicity of variant on the basis of one or a combination of criteria like alanine scanning, massive mutation analysis, multiple sequence alignment (**Figure 4.4)**, and biophysical characteristics such as charge, hydrophobicity, bulkiness and solvent accessibility of amino acids.

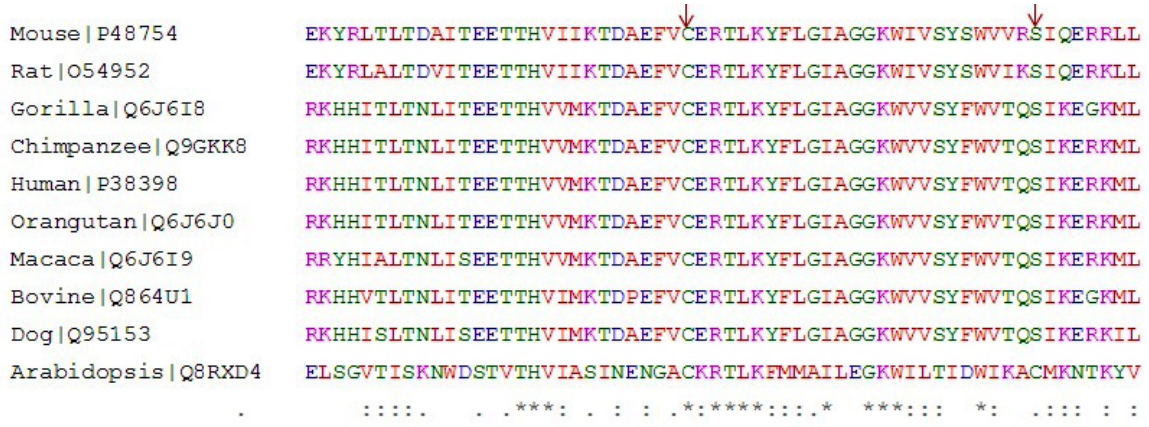


Figure 4.4: *Multiple sequence alignment of BRCA1BRCT domain. Arrow at the top indicates residue S1722P and C1697Y.*

Massive mutation analysis using Align-GVGD algorithm, classified all possible alterations at 1697 (**Table 4.5A)** and 1722 (**Table 4.5B)** position under grade C65 which is suggestive to have deleterious effect. Grantham deviation suggests these residues to be a hot spot residue with clinical relevance. It has been observed that residue at 1722 and 1697 are conserved across species (**Figure 4.4)**. Therefore, all the servers which were used here for *in silico* predictions classified the BRCA1 C1697Y and S1722P modification, with a good score value, to have pathogenic effect on the carrier without any contradiction.

Table 4.5: In silico analysis for mutations at 1697 and 1722 position in BRCA1 *

A				
nsSNP	SNAP Reliability index	MUTPRED Probability	Polyphen Score	Align GVGD Grade (human to frog)
C1697F	4	0.824	0.997	C65
C1697G	3	0.836	0.819	C65
C1697R	4	0.856	0.915	C65
C1697S	3	0.845	0.756	C65
C1697Y	4	0.827	0.984	C65
C1697W	5	0.819	1	C65

B				
nsSNP	SNAP Reliability index	MUTPRED Probability	Polyphen Score	Align GVGD Grade (human to frog)
S1722P	4	0.821	0.983	C65
S1722T	1	0.661	0.480	C65
S1722A	0	0.644	0.111	C65
S1722C	3	0.693	0.112	C65
S1722Y	3	0.778	0.578	C65
S1722F	5	0.785	0.972	C65

** BRCA1 at 1697 and 1722 position was analysed for all probable mutation in Align-GVGD, Mutpred, Polyphen and SNAP for their effect. Grantham deviation suggests the residue to be a hot spot residue with clinical relevance. (A) C1697, (B) S1722.*

4.3.4 Protein expression and purification of BRCA1 TAD and variants

BRCA1 recombinant TAD, BRCA1 S1722P and C1697Y were expressed in BL21 (DE3) bacterial strain. Protein was purified using GST-affinity chromatography. BRCA1 TAD, S1722P and C1697Y showed co-purification and interaction with chaperon, indicative of

aggregated/ misfolded protein [195]. However, there was a difference in strength of interaction with the chaperon which led to separation of chaperon from BRCA1 TAD in gel permeation chromatography (**Figure 4.5A)**, but for BRCA1 S1722P, washing with ATP was essential to get the purified protein. On-beads cleavage with TEV protease was performed to get protein in native form. It has been observed that missense mutation BRCA1 S1722P has drastically altered the protein solubility, and increased its propensity for aggregation & nonspecific interactions with other proteins. BRCA1 C1697Y fusion protein was soluble and also bound to GST beads. Nonetheless, we could not elute the protein from beads after TEV cleavage (**Figure 4.5B).** Addition of different detergents & additives like Emphigen, Triton X-100, NP-40, Sarkosyl, Tween-20, PEG 8000, sucrose and glycerol also was of no use in solubilizing the protein. No beneficial effect of additives in solubilizing the protein suggests irreversible denaturation and precipitation of protein on beads. This suggests the BRCA1 C1697Y mutation has a detrimental effect on protein folding which destabilizes the entire domain leading to loss of function.

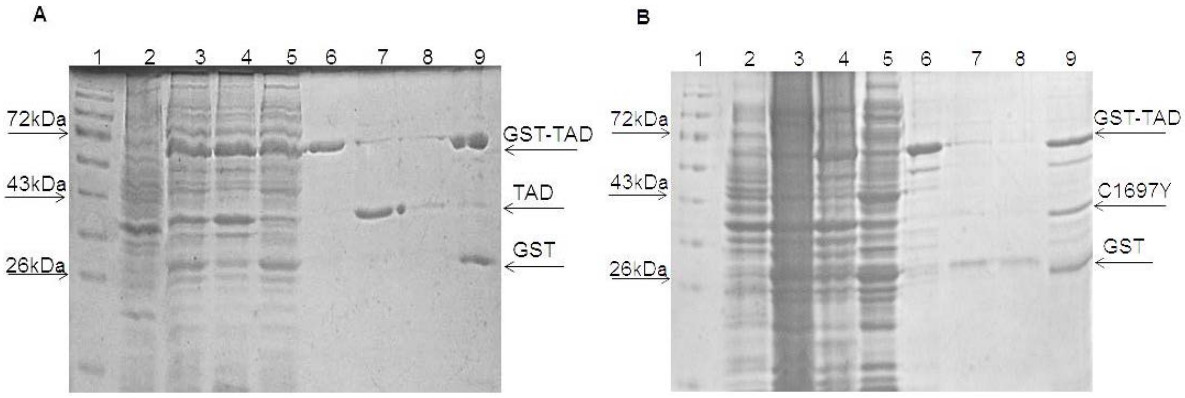


Figure 4.5: Affinity purification of C1697Y mutant*: SDS analysis of TAD construct (A) and C1697Y (B) after affinity purification. lane M -molecular weight marker, 2- un induced whole cell after sonication , 3-induced whole cell after purification, 4- cell debris re suspended and loaded, 5- induced soluble fraction, 6-GST bound protein , 7 & 8- elution fraction after on beads cleavage, 9- beads after cleavage*

4.3.5 Oligomer characterization of BRCA1 TAD and variant S1722P

In gel permeation chromatography BRCA1 S1722P has eluted in void volume, indicating misfolding and aggregation of protein, unlike TAD that eluted at 60 ml (**Figure 4. 6).**

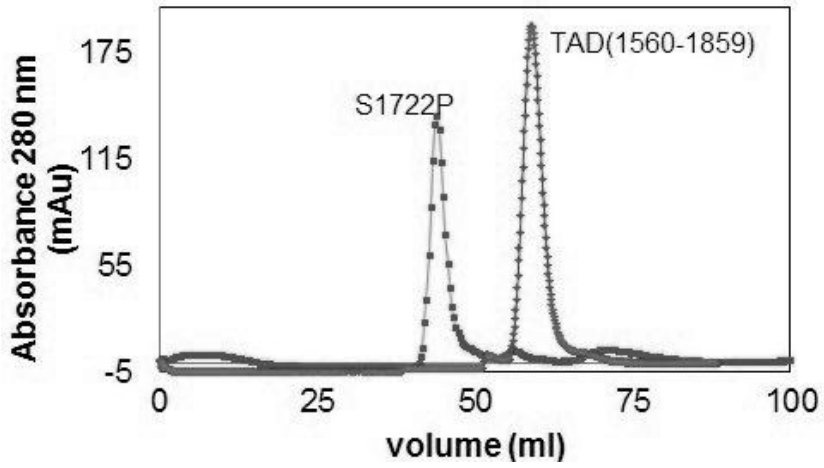


Figure 4.6: Oligomeric characterisation of BRCA1 TAD and S1722P variant:

S1722P elutes in void volume of column indicating oligomerization while TAD elutes around 60 ml in Superdex 75 column.

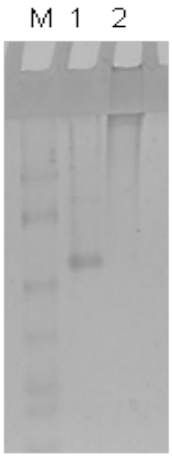


***Figure 4.7: Native gel electrophoresis: Lane** M-marker, 1-TAD, 2-S1722P*

4.3.6 Conformational characterization of BRCA1 TAD and S1722P variant

Circular dichroism and fluorescence spectroscopy studies revealed the effect of variant on protein folding pattern for secondary and tertiary structures. Fluorescence scan has shown the emission maxima at λ=341 and 337 nm for BRCA1 TAD and S1722P variant respectively

(**Figure 4.8)**. Decrease in fluorescence intensity of BRCA1 S1722P variant with a blue shift of 4 nm indicates structural alterations in mutant.

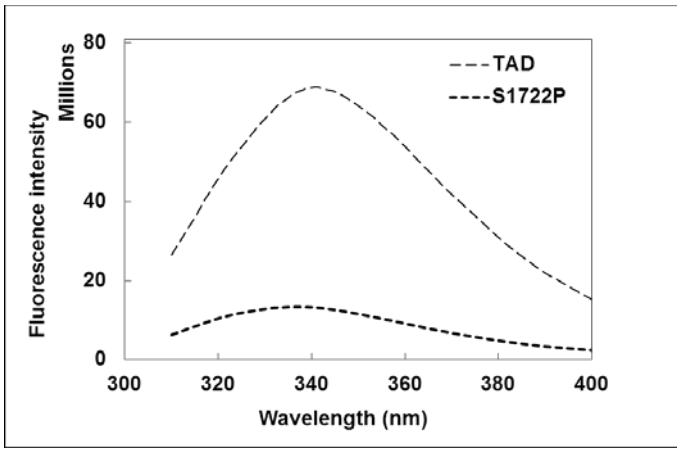


***Figure 4.8**: **Fluorescence spectra of BRCA1 TAD and S1722P,** 2 μM of protein in 25 mM Tris pH 7.5, 150 mM NaCl excited at 295 nm.*

Far-UV CD from λ= 250-195 nm of wild-type and variant has revealed the alterations in secondary structure composition. BRCA1 TAD domain has negative ellipticity at λ= 208 and 222 nm and positive ellipticity at λ=195 nm, both signatures for alpha-helical structure. Comparison of mean residue ellipticity of BRCA1 TAD and S1722P variant indicates decrease in helical content in S1722P variant. There was a 9.5% and 7.6% decrease in helicity, estimated using measurements at λ= 222 nm and 208 nm respectively, for S1722P variant compared to the BRCA1 TAD (**Figure 4. 9).**

Stability of BRCA1 TAD and S1722P was assessed by secondary structural changes as a function of temperature. Ellipticity was monitored at λ=222 nm & normalized to plot for the fraction folded against the temperature. The melting temperature (Tm) for BRCA1 TAD was ~53^{o}C.

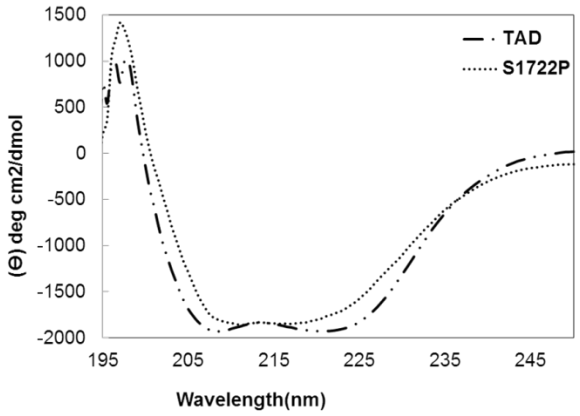


Figure 4.9: Far UV CD scans *of TAD and S1722P indicating BRCA1 TAD with predominantly helix secondary structure & decrease in helicity in BRCA1 S1722P.*

However, no reduction in ellipticity at λ=222 nm was observed up to around 67^{o}C for BRCA1 S1722P (**Inset in Figure 4. 10B),** the probable reason could be the formation of aggregates and thermodynamically stable intermediates. To resolve this, temperature both as a function of high tension voltage (HT) and mean residue ellipticity at λ= 222 nm were plotted (**Figure 4. 10 A and 10B).**

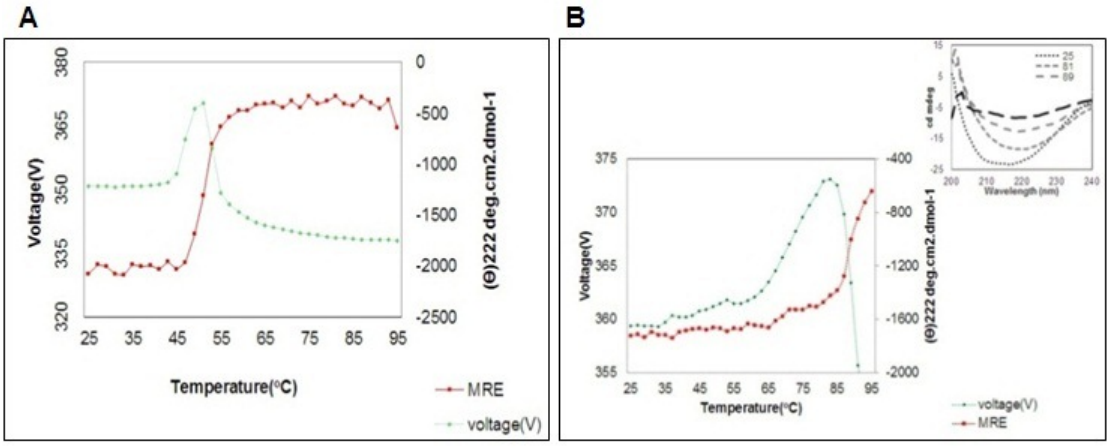


Figure 4.10: Thermal stability study of TAD and S1722P*: Plot of ellipticity at 222 nm (red) and voltage (green) Vs temperature at each 2^{o}C from 25-95^{o}C (**A**): BRCA1 TAD and (**B**): BRCA1 S1722P. The inset shows Far UV CD scan of S1722P at different temperature indicating formation of denatured intermediate that has similar secondary structure conformation like its native form.*

To our conclusion, a heat induced aggregation was observed in BRCA1 TAD and S1722P at ~ 53°C and 67°C respectively. The increase in HT voltage is due to scattering of light that depends on particle size & aggregation in the sample used to record the scan. Gel permeation chromatography and fluorescence spectroscopy also indicated aggregation in the protein

4.3.7 BRCA1 S1722P abrogates the interaction between BRCA1 TAD and phospho specific binding partners

BRCA1 C-terminus is known to interact with proteins containing consensus sequence motif pSer-X-X-F (pSer-phosphorylated serine, F-phenylalanine, X-aromatic/hydrophobic residue). It also has been reported earlier that BRCA1 BRCT domain interacts with BRIP1, CtIP and CCDC98 via pS-X-X-F motif [53,196,197]. Mutations that affect the integrity of BRCA1 BRCT domain are known to impair the interaction between BRCA1 C-terminal domain with this motif [176]. Since S1722P is in the BRCT domain of BRCA1TAD, interaction study using Isothermal Titration Calorimetry with phosphorylated Abraxas peptide was undertaken

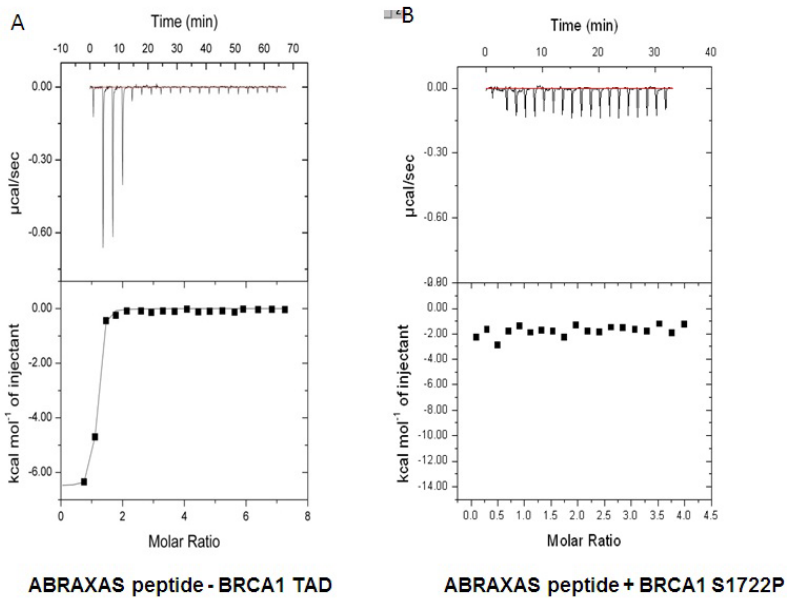


Figure 4.11: Interaction study of BRCA1 TAD and S1722P variant with pS-X-X-F motif containing peptide: *(A) BRCA1 TAD domain interaction with phosphorylated peptide. (B)Interaction was totally abolished in variant BRCA1 S1722P*

BRCA1 TAD interacts with Ser 406 phosphorylated Abraxas peptide (**Figure 4.11A).** However, titration of negative control peptide (NH2-GFGEY (pS404) RSPTF-COOH) and BRCA1 TAD did not show any interaction (data not shown). The interaction between Abraxas peptide and BRCA1 S1722P variant was totally abrogated (**Figure 4.11B) and (Table 4.6)**

***Table 4.6**: Thermodynamics parameters of interaction of BRCA1 TAD with phosphopeptide*

Protein	RNA binding	ABRAXAps	BRCA1	Nup153	Kif1B
N Sites	0.939±0.169	0.590 ± 0.0054	1.12 ± 0.0238	1.27± 0.0139	1.21±0.0498
K(M-1)	8.43E4	2.51E7	1.16E6	1.19E6	1.28E5
H(cal/mol)	-1.552E4	-1.138E4	-2.300E4	-9627	-9939
S(cal/mol/deg)	-29.5	-4.30	-49.4	-4.48	-9.95
Kd(μM)	11.8	0.0398	0.862	0.840	7.80

4.3.8 Structural analysis of the variants S1722P and C1697Y

To investigate the structural changes in the variant proteins, the mutations (S1722P and C1697Y) were introduced into the three dimensional structure of BRCA1 BRCT (PDB ID 1T15) [116]. These molecular models were subjected to molecular dynamics simulation, and the lowest potential energy and Ramchandran plot validated structure, was used for analysing the structural alterations. Superposition of simulated wild type structure with C1697Y and S1722P structures yielded RMSD values of 0.943 and 0.803 Å respectively.

In the wild type structure, the -SH group of Cys 1697 residue is interacting with the π electron cloud of residue Tyr 1703 thereby stabilising the position of that residue. Replacement of Cys by Tyr at 1697 positon leads to a loss of these interactions. Further, the CE atom of Tyr 1697 residue would make a steric contact of 1.8 Å with the hydroxyl group of Tyr 1703. Because of these two factors the Tyr 1703 residue has shifted by 1.9Å in the mutant structure. The side chain of Arg 1699 which is present in the vicinity of 1697 is also

displaced by 2.5 Å from its position in the wild type structure. In the crystal structure of BRCT complexed with phosphopeptides containing the signature sequence pS-X-X-F, Arg 1699 forms a strong hydrogen bond with carbonyl oxygen of phenyalanine residue in the +3 position of the signature sequence. This interaction ensures proper positioning of hydrophobic side chain of Phe in the hydrophobic pocket from BRCT domain [115]. The displacement of Arg 1699 by 2.5 Å in the mutant can have a deleterious effect on positioning of Phe residue leading to loss of phosphopeptide binding and molecular recognition. Valines at 1736 & 1696, Asp 1739, Met 1689, and Glu 1698 are also moved from their positions due to incorporation of this bulky hydrophobic residue. Even though Lys 1702, which is known to interact with phosphorylated Ser of phosphopeptide, is present in the vicinity of 1703 residue, it has not moved away from its position. Thus, overall structural distortions due to mutation at 1697 may interfere with phosphopeptide binding (**Figure 4.12A)**.

The Ser residue at 1722 is present in the α-3 helix of the protein and its substitution by Pro has resulted in kinking of the alpha helix as can be seen in the superposed structure (**Figure 4.13).** As a consequence, the side chains of many conserved residues have been shifted by substantial amounts. For example, Ile at 1723 position is flipped away by 2.9 Å, while Arg at 1726 position is shifted by 4.8 Å. Similarly, Phe at 1668 and Trp at 1718 positions are also displaced by 1.4 and 1.8 Å respectively (**Figure 4.12B)**. Assuming that conserved residues play essential biological roles, the movements of conserved residues observed here are likely to interfere in the biological functioning of BRCT domain, and this may be a reason why S1722P mutation is associated with cancer.

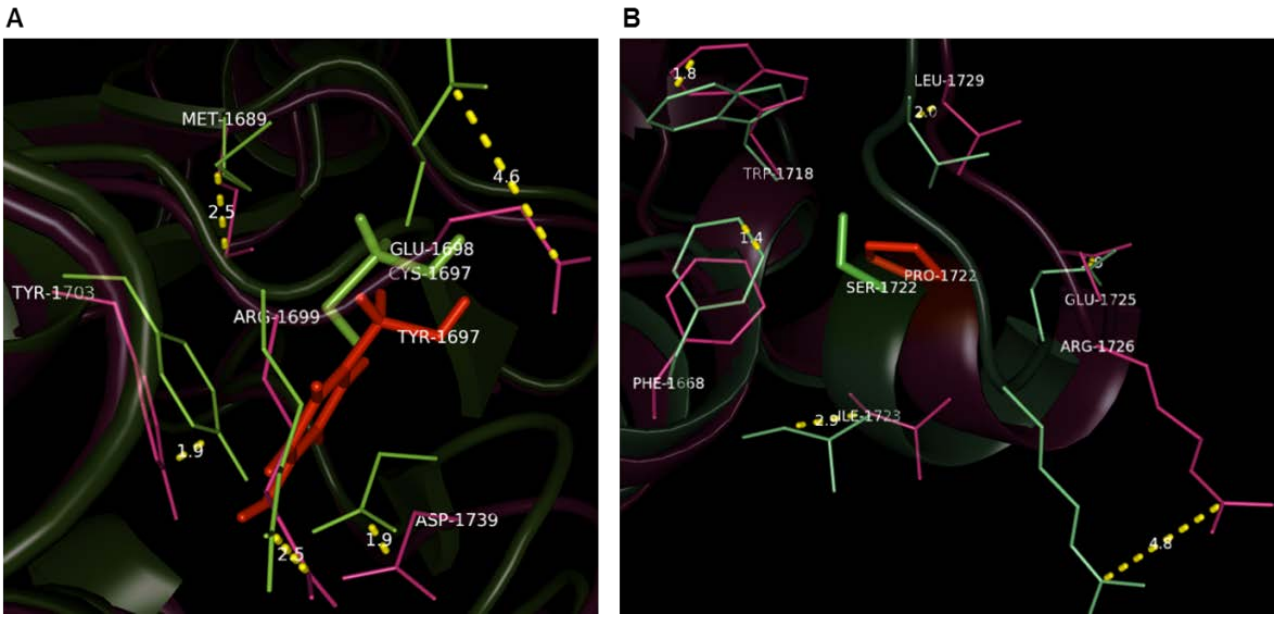


Figure 4.12: Structural analysis of C1697Y and S1722P variants: Superposed structure

1T15 wild type with (A) Cysteine 1697 and (B) Serine 1722. Red color sticks indicate variant residue and green indicate wild type residue.

4.4 Summary

BRCA1 is a pleiotropic tumour suppressor protein. The major challenges that prevail with non-synonymous missense variant is less information on pedigree data, ethnic specificity, segregation profile & expressivity that pose a hindrance for classification of VUS in to low, high or moderate risk. Therefore classification of missense mutations with statistically insufficient clinical data and co-segregation pattern, additional information becomes utmost important to categorize mutations into pathogenic or benign. Pedigree information of variants from patient families has helped clinicians & genetic counsellors for pre symptomatic counselling, risk prediction, disease management and prevention. But the hurdles in obtaining clinical data make it necessary to validate the variants with structural and functional studies.

A germ-line mutation BRCA1 C1697Y was diagnosed in proband at age of 42 with breast cancer. The clinico-pathological examination revealed that patient had triple negative, intra ductal carcinoma grade III cancer. The siblings who harboured the variant were also affected

except for a carrier sister. The missense variant S1722P in BRCA1 shows a variable penetrance. The proband was diagnosed with cancer at age of 30, her mother had breast cancer at 47y, while her brother an obligate male carrier was unaffected at age 82y, and however his daughter was diagnosed with breast cancer at 35y and passed away due to the disease. In order to examine whether the novel missense germ line variant C1697Y (Oriya family) and S1722P (Sindhi family) has a role in early onset of breast and ovarian cancer multidisciplinary *in silico, in vitro* studies was performed. Analysis of all possible alteration at C1697 and S1722 position using Align-GVGD indicates it as a hotspot residue that can be damaging for any possible change. The Grantham-deviation value above 60 & Align-GVGD grade C65 for all possible alterations is suggestive of measurable damaging biochemical difference observed at this position. Previous report of BRCA1 C1697R in Scandinavian breast and ovarian cancer families is a well characterised pathogenic mutation. BRCA1 S1722F variant is also reported to have compromised peptide binding activity, transcriptional activity and protein folding defect [198]. Structural and functional studies of BRCA1 C1697R and S1722F variant, indicated loss in transcriptional activity which is very much in support to our yeast transcriptional activation data [175,198,199,200,201]

To further validate the pedigree and *in silico* studies, yeast transactivation assay was performed to determine the transactivation activity of BRCA1 TAD and its variant. BRCA1 C1697Y and S1722P missense variant is located in exon 18 and 19 respectively, which is a part of BRCT-N of BRCA1 BRCT. The C-terminal region of BRCA1 has maximum number of pathogenic mutations reported, many of which are well characterized to be deleterious with transactivation assay. In this assay, C1697Y and S1722P were tested for their ability to bring about transcriptional activation. M1775R was used as a negative control [202], BRCA1 TAD as a positive control. According to the yeast transactivation assay, BRCA1 TAD

exhibited the transactivation activity but BRCA1 C1697Y and S1722P showed total abrogation of activity like in M1775R, suggesting these to be devastating changes.

In order to understand effects of these variants, secondary and tertiary structural characteristics and protein-protein interactions studies were explored. We have purified TAD domain of BRCA1 in native form and BRCA1 S1722P in oligomer form. However, C1697Y could not be successfully purified due to aggregation and precipitation. CD and fluorescence studies indicated reduction in alpha helical content with aggregation of S1722P. S1722P variant being present in an α- helix, we speculate it to disrupt the helix. Proline, being unable to donate an amide for hydrogen bonding, and steric interference due to its side chain, incorporates bend if present in the middle of a helix.[203] To study the effect of incorporation of Proline, in terms of helical content in BRCA1 TAD domain AGADIR algorithm [204] was used. Decrease in helicity was observed in BRCA1 S1722P as compared to TAD. Circular dichroism spectroscopy along with Mutpred also supported decrease in alpha helical content of protein as indicated by decrease in negative ellipticity at λ=222 nm (**Figure 4.12).** Kink was also visible in the molecular dynamics simulated structure with a lateral displacement of 2.3 Å.

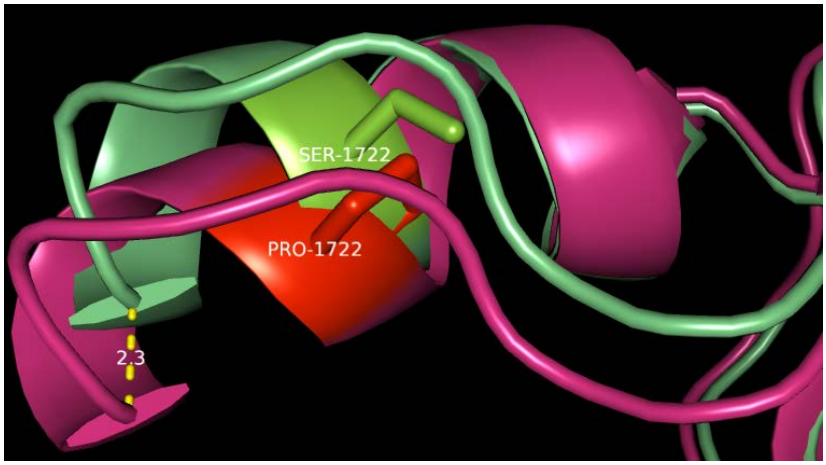


Figure 4.13: *Kink in α-helix in S1722P variant*

BRCA1 S1722P variant not only indicated loss of secondary structural integrity, but we also witness aggregation of protein as indicated by thermal denaturation, native gel and size exclusion profile.

Isothermal Titration Calorimetry showed complete abrogation of interactions with pS-X-X-F motif in variant S1722P. In conclusion, the segregation of variant with the disease in carrier, conservation of residue in question across species, loss in transactivation activity and interaction ability, alteration in secondary and tertiary structural conformation, instability of protein and crystal structure analysis, suggest novel missense variants BRCA1 C1697Y and S1722P are probably high risk variants with high chance of carrier to manifest the disease. There are a lot of missense variants reported with unknown clinical significance and failure in assignment of these variants in to low risk, high risk or moderate risk becomes a major impediment in genetic counselling and preclinical testing.

Chapter 5

Structural characterization of N-terminal extended BRCA1 BRCTs

5.1 Introduction

The C-terminal region of BRCA1, BRCA1 BRCT, is well documented to interact with proteins like RAP80, CCDC98, JunB, AhR, p53, $p27^{Kip1}$ and CtIP [61,80,205] and play a central role in DNA repair and transcription regulation. BRCA1 BRCT recruits DNA repair proteins to the damage site to form a DNA repair complex. Germ-line mutations in BRCA1 BRCT reported in patients with breast or ovarian cancer, interfere with transcription activity, strongly supporting the role of BRCA1 in transcription [206].The C-terminal region of BRCA1 can activate transcription both in yeast and mammalian cells. BRCA1 BRCT residues from 1760 to 1863 form the minimal region required for transactivation function, but reports suggest that the two regions AD1 (amino acids 1293–1558) and AD2 (amino acids 1560-1863) present upstream of BRCT significantly enhance the transcriptional ability of BRCT region. AD1 functions cooperatively with the BRCT and performs its function in a tissue specific manner while AD2 activates transcription in a ubiquitous manner [4,62]. In BRCA1, negative charge along with hydrophobic residues is crucial for performing the transactivation function [207].

Therefore, to explore at the molecular level structural effects of upstream region on BRCT domain, BRCT with three different N-terminal extended regions were cloned, expressed and purified. Crystallisation attempts, *in silico*, biophysical and structural studies were performed on these purified constructs. Results of these experimental studies are described in this chapter.

5.2 Materials and Methods

5.2.1 Gene-Cloning

The following N-terminal extended BRCTs, residues (1314-1863), (1396-1863) and (1560-1859) were cloned in pGEX-KT vector by using BRCA1 cDNA as the template. Cloning of

BRCA1 (1560-1859) has already been described in chapter 4. Forward primers used for BRCA1 1314 - 1863 is 5'-GTCGGATCCGAGAACCTGTACTTTCAGGGTGATCCTTTCTT GATTGGTTCT-3' that for BRCA1 1396 – 1863 is 5'-GTCGGATCCGAGAACCTGTACTTT CAGGGTCAGAGGGATACCATGCAACA-3' and the common reverse primer used is of sequence5'GTCGAATTCCTATTAGTAGTGGCTGTGGGGGATCTGGGG 3' to amplify the different constructs. Templates along with forward and reverse primers were mixed with the PCR reaction mixture and Pfu polymerase. The PCR condition used for the reaction is as mentioned in chapter 2. The amplified PCR product and the vector were digested with restriction enzymes BamH1 and ECoR1, and further ligated. The colonies obtained after ligation were screened for positive clones by checking the insert release and DNA sequencing to confirm the presence of correct DNA sequence.

5.2.2 Expression of N-terminal extended BRCTs constructs

For protein expression, the clones BRCA1 (1314-1863), (1396-1863), (1560-1859) and (1646-1859) (kind gift from Dr. Dilip Badgujar) were transformed in *E. coli* bacterial strain BL21 (DE3). Single transformed colonies for each of the construct were inoculated in Luria Bertani broth with ampicillin. Overnight-grown cultures were diluted 1:100 and were then further grown at 37^0C, until OD_{600} was 0.6. These cultures were then induced with 0.4 mM IPTG and incubated at 18^0C for 16 h after which the cells were harvested by centrifugation for 10 min at 6000 rpm.

5.2.3 Purification of N-terminal extended BRCTs constructs

A. Purification of BRCA1 (1314-1863)**:**

Bacterial pellet from two litre culture was resuspended in 80 ml of buffer **B (**50 mM Tris pH 7.5, 300 mM NaCl, 1 mM EDTA, 5% Glycerol, 0.1% Triton X-100, 20 mM BME and protease inhibitor cocktail). After sonication and centrifugation protein was affinity purified using GST resin. The non-specifically bound proteins were removed by washing the resin with wash buffer containing 1M NaCl in buffer B. On-beads cleavage with TEV was performed and protein was eluted from the affinity column and concentrated.

B. Purification of BRCA1 (1396-1863):

The harvested bacterial pellet was resuspended in buffer C (50 mM Tris pH 7.5, 300 mM NaCl, 1 mM EDTA, 2.5% Glycerol, 0.1% TritonX-100, 2 mM DTT and protease inhibitor cocktail). The cell suspension was sonicated 3 times with 1 minute duty cycle at a pulse rate of 60 to lyse the cells. Cell debris was removed by centrifugation at 18,000 rpm for 40 minutes at 4^{o}C. The supernatant was then passed through the pre-equilibrated glutathione resin for affinity purification. The non-specifically bound proteins were removed by washing the resin with buffer C containing enriched amount of NaCl, Triton X- 100 and DTT (1M NaCl, 0.5% Triton X- 100, 4 mM DTT). This washing was essential to remove the non-specifically bound protein of 60 kDa. On-bead cleavage using TEV protease was performed to remove the fusion GST tag. The protein eluted after TEV cleavage using lysis buffer was further purified using ammonium sulphate precipitation. Ammonium sulphate in powder form was gradually added till its concentration reached 20%, and the mixture was then incubated on ice for 30 minutes with intermittent gentle mixing. The precipitated protein, separated by centrifugation for 30 minutes at 18,000 rpm, was resuspended in 1 ml of lysis buffer and was centrifuged again at 13,000 rpm for 10 min to remove aggregated protein.

C. Purification of (1560-1859) and (1646-1859):

The protocol followed for purification of (1560-1859) and (1646-1859) is similar, and is mentioned in chapter 4.

The purified protein obtained after affinity chromatography was confirmed using MALDI-TOF and peptide mass fingerprinting. In solution and in-gel trypsin digestion were carried out to obtain the peptides. Peptide mass fingerprinting under following search parameters: mass tolerance of 50-200 ppm, and missed cleavage of 0-1, Swiss Prot/NCBInr database, has confirmed the identity of the protein.

5.2.4 Gel permeation chromatography

The purified proteins obtained were further purified to homogeneity using Superdex 200 gel filtration column. The sample and standards (BSA, lysozyme, phosphorylase B, and carbonic anhydrase) were chromatographed at a flow rate of 0.5 ml/min, and 1 ml fractions were collected. From the calibration plot of elution volume versus standard protein molecular weight the molecular size and oligomeric characteristics of the sample were estimated.

5.2.5 Circular Dichroism and Fluorescence spectroscopy

Far UV CD scans for the wavelength range λ=250-195 nm were acquired for BRCA1 (1314-1863), BRCA1 (1396-1863), BRCA1 (1560-1859) and BRCA1 (1646-1859). CD scans, averaged over 3 accumulations, were recorded at 20^0C, with a scanning speed of 20 nm/ minutes Protein concentration was calculated using measured Nano drop absorption at λ= 280 nm and the known molar extinction coefficient. For tertiary structural study fluorescence emission was measured in the wavelength range λ_{em} 310–500 nm, when excited at 295 nm. The emission slit width was 5 nm. All the samples were baseline corrected. The spectral data were collected using FluorESSCENCE software.

5.2.6 Crystallisation Screening and Optimisation

Screening of crystallization conditions for BRCA1 (1314-1863), (1396-1863) and (1560-1859) were carried out using Hampton Crystal Screen 1 and 2, Salt Rx. Additive and PEG ion screens were also used to screen the crystallisation condition for BRCA1 (1560-1859). FPLC purified protein at a concentration of 6-25 mg/ml was used for setting up the trial. Screening trials were set using both hanging and sitting drop vapour diffusion techniques, and the crystallisation plates were placed in vibration-free incubator maintained at 22°C. The droplet volume was 2 µl, obtained by mixing 1 µl drop of protein with 1 µl of reservoir solution. The volume of the reservoir solution was varied from 200 to 500 µl.

Crystals were obtained for only the (1560-1859) protein, and the crystallisation conditions are as under: 2 M Ammonium Sulfate, 0.1M MES pH 6.5 and 10 mM Cobalt Chloride. The crystals obtained were confirmed to be protein crystals by staining with Izit dye. Crystallization condition was further optimized to increase the crystal size and obtain diffraction quality crystals by methodically varying the pH and concentration of precipitant and salt. The ratio of volume of well solution and drop size was also varied in an attempt to obtain larger crystals.

5.2.7 Diffraction data collection and processing

The BRCA1 (1560-1859) crystals were screened for diffraction analysis using the rotating anode X-ray generator available at ACTREC. 25-30 % glycerol performed the best as cryo-protectant. The radiation damage to the protein crystals is also minimized by bathing the crystal in a continuous stream of cooled nitrogen gas. Diffraction data were collected by the oscillation method using mar dtb mounted on a microStar rotating anode X-ray generator operated at a voltage of 45 kV and current of 60 mA. The crystal to detector distance was 200 mm and 180 frames were collected, each for 1^{o} oscillation.

5.2.8 Degradation assay of BRCA1 (1560-1859)

To understand the stability and /or lysis of BRCA1 (1560-1859) protein, purified BRCA1 (1560-1859) alone, BRCA1 (1560-1859) with 10X protease inhibitor cocktail (Sigma) and BRCA1 (1560-1859) in presence of 6 M urea was incubated at 37°C. The samples were incubated for durations of 0, 0.5, 2, 5, 8, 12, 22, 25, 28, 31, 94 and 100 h, and were then electrophoresed on 10% SDS gel.

5.2.9 Limited proteolysis

All the four purified proteins were incubated with trypsin at 37°C for periods of 0, 5, 15, 30, 60 and 120 minutes 360 µl of protein at a concentration of 6-12 µM and 3 µl of trypsin at a concentration of 20 ng/ µl was used for the assay. 60 µl aliquot of the reaction mixture were taken at each time point and the reaction was stopped by the addition of 1µl of 1mM PMSF. Proteolysis digests were analysed on 10% SDS/PAGE.

5.2.10 *In silico* analysis

In-silico analyses of sequence and structure composition were performed for disorder prediction. The softwares used for this purpose are: PONDR [208], Composition profiler and PSIPRED **[209,210]**. The databases used are PDB Select 25 and SwissProt 51.

5.3 Results

5.3.1 Cloning of BRCA1 (1314-1863)

Digestion of recombinant vector, isolated from selected colonies, with BamH1 and ECoR1 gave a fragment of 1647 base pair **(Figure 5.1A and Figure 5.1B)**. These clones obtained were further confirmed by DNA sequencing

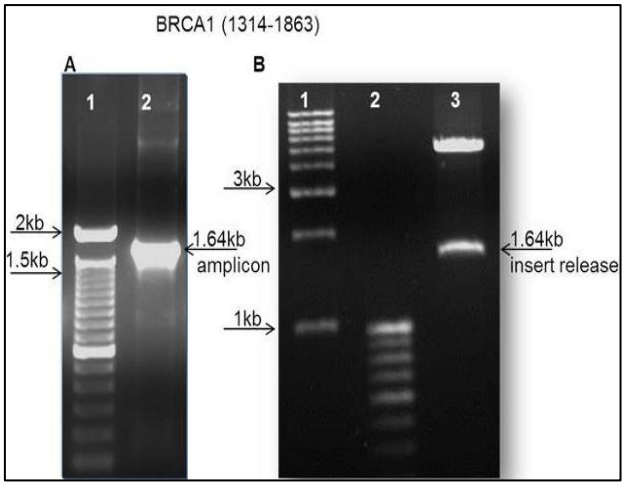


Figure 5.1: Screening of clone of BRCA1 1314-1863 *with BamH1 and ECoR1 enzyme in 1% agarose gel. Lane1: DNA ladder, Lane 2 (A): PCR product. (B): Positive clone showing insert release of 1647bp.*

5.3.2 Purification and confirmation of N-terminal extends of BRCTs

BRCA1 (1314-1863), (1560-1859) and (1646-1859) were expressed in bacterial system and were purified using GST affinity chromatography. BRCA1 (1396-1863) was purified using ammonium sulphate precipitation and affinity chromatography methods. These fusion proteins were cleaved with TEV protease to obtain the protein in native form, and this was further FPLC purified (**Figure 5.2).**

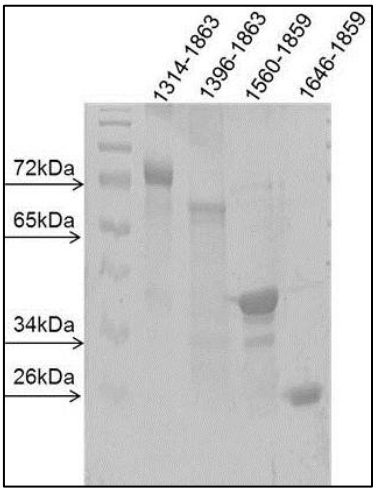


Figure 5.2: *Purified N-terminal extensions of BRCA1 BRCT loaded on SDS-gel*

The purified proteins obtained were confirmed for their identity using MALDI-TOF. The mass spectrum and spectral analysis has confirmed the identity of each regions (**Figure 5.3)**

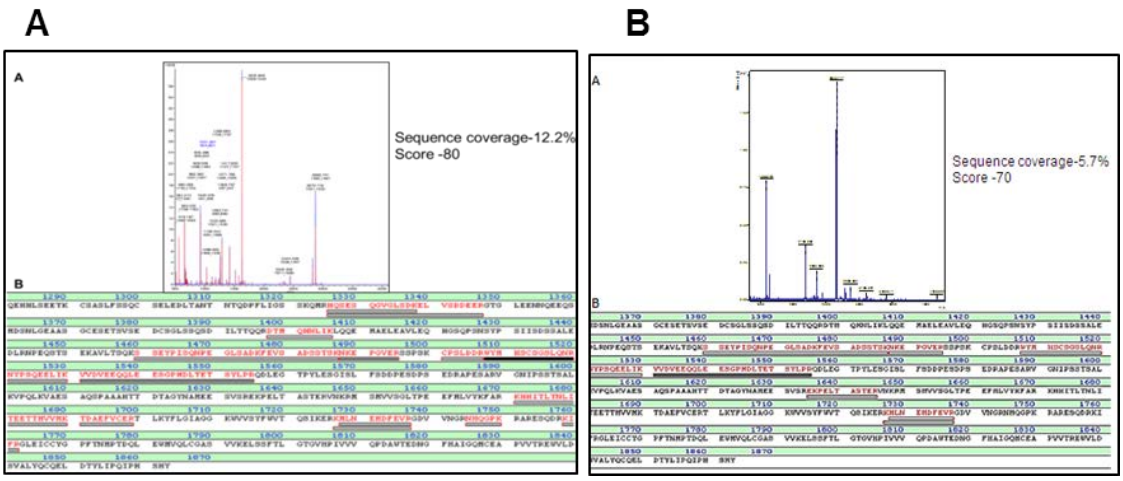


Figure 5.3: Identification of N-terminal extends of BRCA1 BRCT *(A): BRCA1 1314-1863, (B): BRCA1 1396-1863. A-Peptide mass fingerprinting peaks, B- Spectrum analysis report.*

5.3.3 Oligomeric characterisation of N-terminal extended BRCTs fragments.

The retention volumes obtained from size exclusion chromatography were used to estimate the sizes of these purified proteins (**Figure 5.4)**. The molecular weights derived by extrapolation from standard plot are not matching with expected monomeric molecular weights, suggesting either oligomerisation or non-compact three dimensional structures. From the estimated molecular weight data all the constructs appear to be homo dimers except for BRCT **(Table.5.1)**

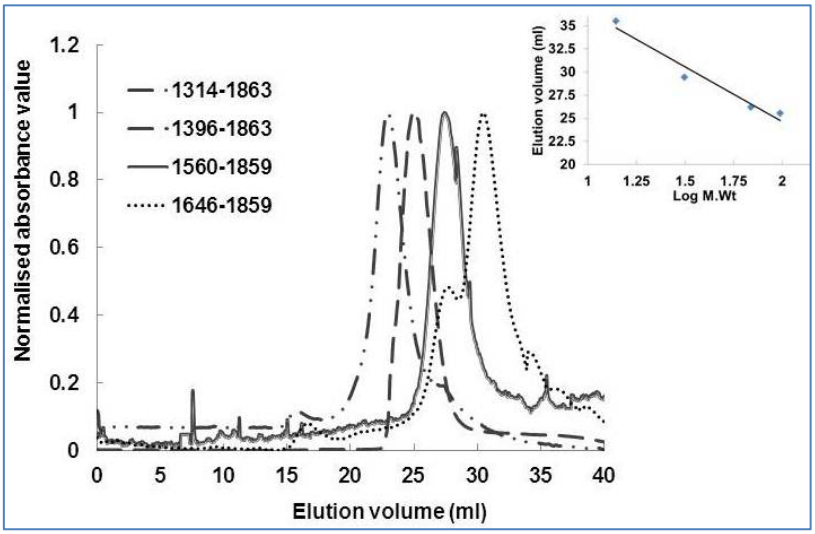


Figure 5.4: Oligomeric characterisation*: Elution profile with inset showing plot of standards used to estimate the molecular weight of N-terminal extended BRCTs fragments.*

When these purified protein samples were run on a native PAGE, BRCA1 (1646-1859) and (1560-1859) band at positions corresponding to a monomer and dimer respectively, while BRCA1 (1314-1863) and (1396-1863) displayed a smeared pattern in the gel indicating the proteins have open conformation.

Table 5.1: *Molecular weight estimation from plot M.wt and elution volume of standards*

BRCA1 fragments	Theoretical molecular weight	Observed molecular weight	Ratio to monomer molecular weight
1646-1859	24.63	32.15	1.30
1560-1859	33.64	58.06	1.72
1396-1863	52.52	91.65	1.74
1314-1863	62.21	137.44	2.21

5.3.4 Structural characterisation of the N-terminal extended BRCTs

Far UV CD scans of N-terminal extended BRCTs revealed the proteins to have alpha helical characteristics (**Figure 5.5A)**. This is probably expected as the crystal structure shows that BRCT domain contains alpha helices. All the N-terminal extended BRCTs of BRCA1 proteins have demonstrated negative ellipticity at λ = 222 and 208 nm, except for BRCA1 (1314-1863) which showed a dip at near λ = 200 nm, which is characteristic of disordered protein. Interestingly, far–UV CD signal due to BRCT domain progressively decreases as the number of N-terminal residues added to BRCA1 (1646-1859) increases. These data suggest that the CD signal is majorly due to BRCT domain while the added amino acids do not have an ordered structure.

Tertiary structural studies using fluorescence has revealed structural influence due to residues added to the N-terminal of the BRCT region (**Figure 5.5B)**. For example, incorporation of around 80 residues upstream to BRCT, as happens in BRCA1 (1560-1859) construct, demonstrates a red shift of 9 nm; however addition of 250 residues to BRCT as in (1396-1863) has revealed emission maximum almost similar to BRCT. Again when 332 amino acid

residues are added to BRCT region, there is a red shift of 8 nm indicating tryptophans are in an exposed environment as compared to BRCT **(Table 5.2).**

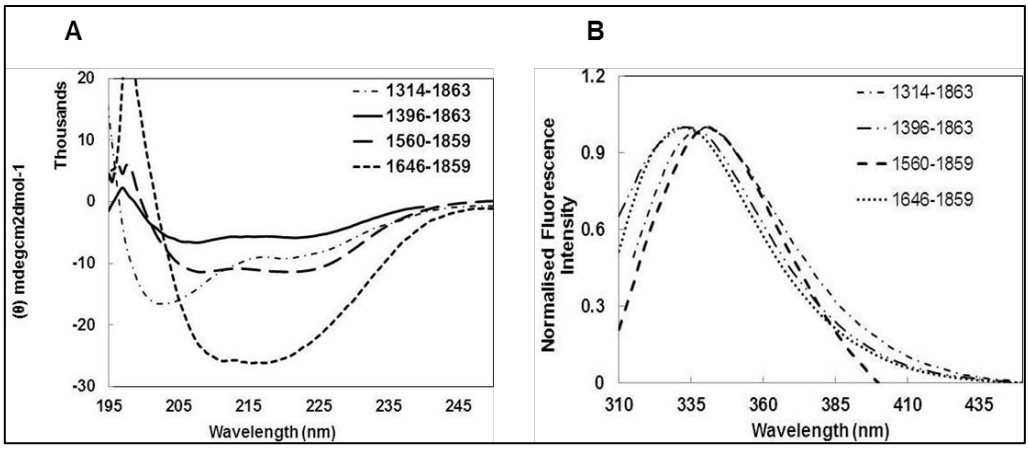


Figure 5.5: Structural characterisation of the N-terminal extended BRCTs: *(A) Far UV CD scans, (B) Emission spectrum of N-terminal extended BRCTs fragments.*

In conclusion, addition of residues upstream to BRCT results in comparatively exposed status of hydrophobic patches. **Table 5.2** shows the emission maxima and number of tryptophan residues present in these different fragments.

Table 5.2: *Fluorescence emission maxima for different N-terminal extended BRCTs*

N-terminal extended BRCTs	Tryptophan residues	λ max (nm)
1314-1863	6	340
1396-1863	6	333
1560-1859	5	341
1646-1859	5	332

5.3.5 Crystallisation of BRCA1 (1560-1859)

Different crystallisation conditions were screened using various commercially available crystal screens from M/s Hampton research (USA). The N-terminal extended BRCTs proteins screened for crystallisations were BRCA1 (1314-1863), (1396-1863) and (1560-1859). BRCA1 (1646-1859) has already been crystallised and its structure is reported (PDB ID:

1JNX).Crystals were obtained only for BRCA1 (1560-1859). Microcrystals were obtained in several different conditions, but they could not be grown to bigger size even after several systematic trials varying droplet size, protein concentration, ionic strength, pH etc. Crystals were obtained under following conditions: Ammonium sulphate (1.2-1.6 M), 0.1M MES (pH 6.2-6.5) and 10 mM Cobalt chloride, protein (25 mM Tris pH 7.5, 150 mM NaCl) at concentration of 6-8 mg/ml. Spindle shaped and hexagonal crystals for BRCA1 (1560-1859) were obtained within six days of setting the trials at 22^0C (**Figure 5.6)**.

Figure5.6: *Crystal of BRCA1 (1560-1859)*

5.3.6 Data processing of BRCA1 (1560-1859) crystal

The BRCA1 (1560-1859) crystals were screened for diffraction analysis on the rotating anode X-ray generator. Majority of the crystals screened, diffracted to the low resolution of around 4 Å (**Figure 5.7).** The post crystallization treatments such as dehydration and cryo-annealing were tried to improve the diffraction quality but these treatments could not improve the diffraction resolution [252].

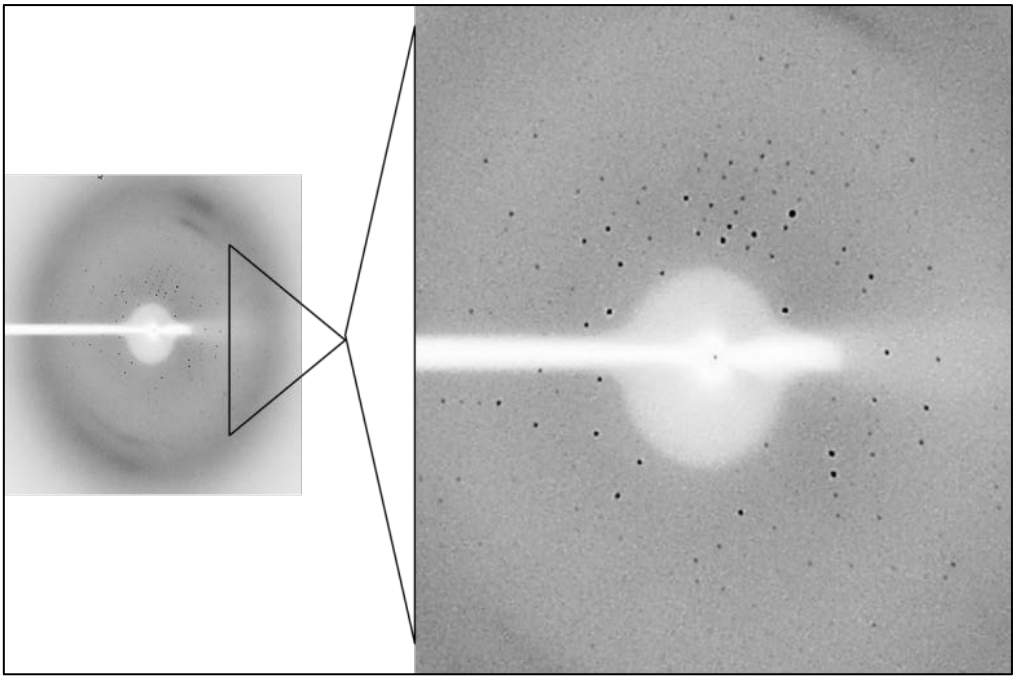


Figure 5.7: *Diffraction pattern of BRCA1 (1560-1859)*

Diffraction data were collected by the oscillation method under cryo conditions with a rotation of 1^0 for each frame. A total of 180 frames were collected with a crystal to detector distance of 200 mm. Diffraction images were processed using iMosflm. The crystals were found to belong to the space group P 6_1 2 2 with unit cell parameters: a=b= 114.2 Å and c=120.9 Å. Cell content analysis calculates the Matthews coefficient (V_M) value to be 3.40 $Å^3Da^{-1}$ which corresponds to a solvent content of about 63 %, if presence of one protein molecule of 33 kDa per asymmetric unit was assumed. Molecular replacement calculations performed by PHASER using the structure with PDB ID 1T29 as a search model confirmed both the space group and the presence of single molecule in the asymmetric unit.

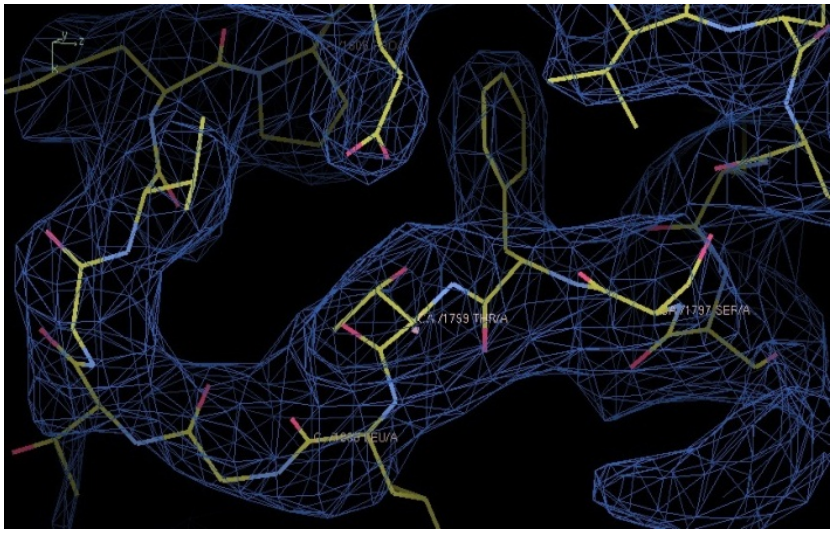


Figure 5.8: *Electron density map contoured at 1.2σ level*

Good electron density in the 2Fo-Fc map was observed only for residues (1649-1859) suggesting that only this region of BRCA1 TAD is ordered consistent with other results described before in this chapter **(Figure 5.8)**.

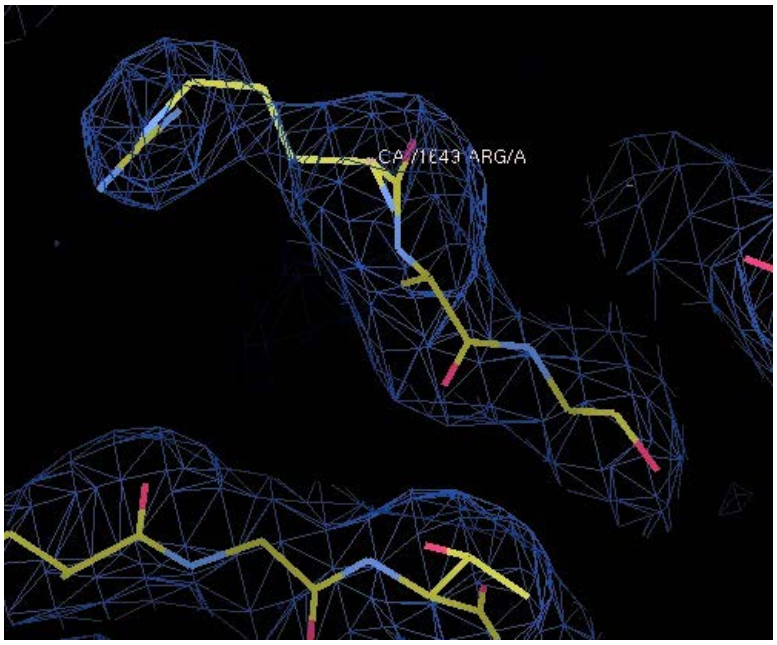


Figure 5.9: *Electron density map contoured at 1.3σ level at residue 1649*

Figure 5.9 shows electron density for a few residues from 1649 to 1652. Even on decreasing the contour level no density developed for residues N-terminal to 1649. Therefore, one of the following possibilities can exist: 1) the region upstream of 1649 is

totally disordered as already indicated by other biophysical studies mentioned above, and 2) the region upstream for 1649 is proteolytically removed. The proteolysis must have happened, if at all, in crystallisation drops because molecular weight of the protein was established as 33 kDa just before setting up for crystallisation**.** Further to check for the integrity of protein in the crystals, crystals were dissolved and loaded onto SDS-PAGE **(Figure 5.10)**. While a majority of experiments showed a protein band at 24 kDa (size of BRCA1 BRCT) on few occasions a protein band was observed at 33 kDa expected for a BRCA1 (1560-1859). So this was further investigated and results are described in the next section.

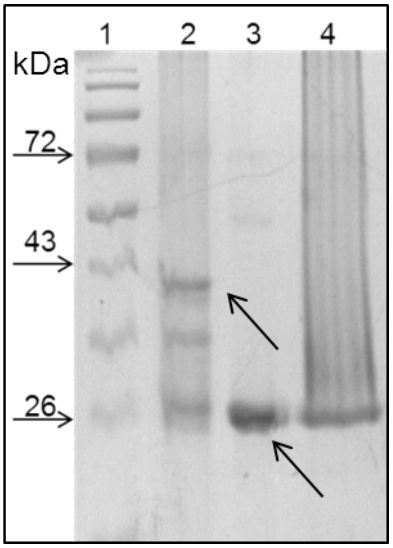


Figure 5.10: SDS-gel electrophoresis of crystal*: Lane1-Marker, 2- 3-crystals, 4-clear drop*

5.3.7 Self-degradation of BRCA1 (1560-1859)

During the purification process, degradation of the protein was noticed even when enough quantity of protease inhibitor cocktail was used. To further examine this phenomenon BRCA1 (1560-1859) was incubated at 37^{o}C under following different conditions: 1) purified protein alone, 2) protein in the presence of the cocktail of protease inhibitors at 10X concentration, and 3 protein in the presence of 6 M urea. Each sample was electrophoresed and the results are shown in **Figure 5.11.**

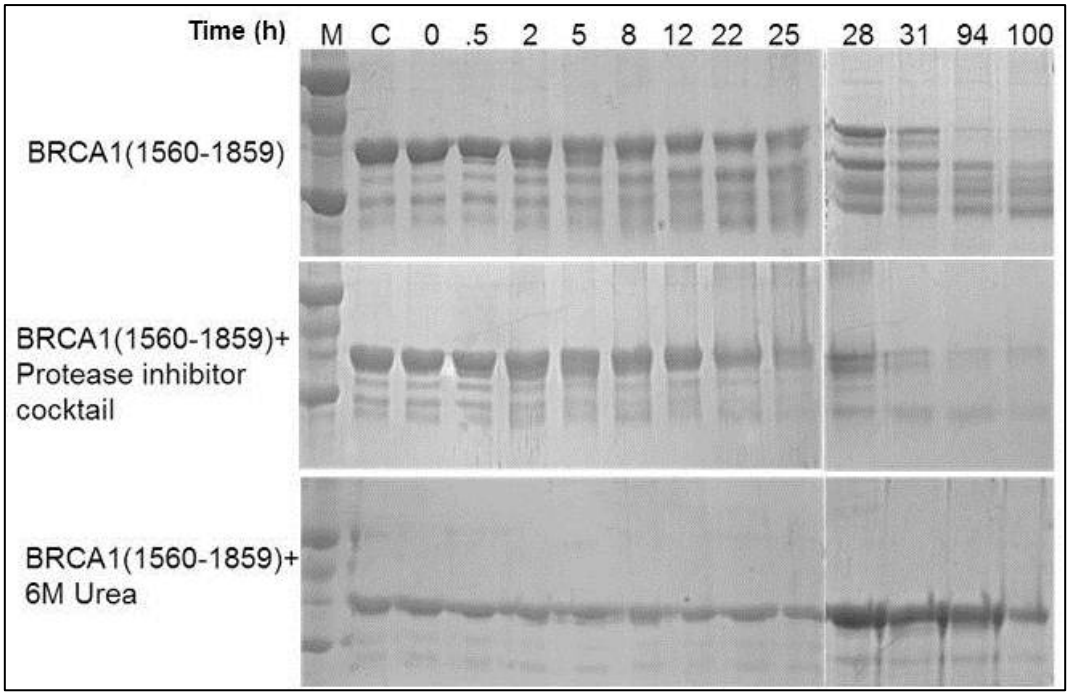


Figure 5.11: Degradation study* of BRCA1 (1560-1859): Lane M-marker, C-control, 0-100: time in hours*

These results show that the protein degrades within 28 h of incubation, both in the presence and in the absence of protease inhibitor cocktail. No degradation was observed when protein was incubated with 6M urea. In conclusion, these experiments suggest three possibilities: 1).the preparation is contaminated by a molecule which has protease activity that is destroyed in the presence of urea. 2) The size exclusion purified protein has protease contamination in it and that this protease is not inhibited by the added protease inhibitor cocktail used and 3) BRCA1 TAD has autolytic activity which is responsible for its self-degradation. In view of the exhaustive purification process adopted here the third alternative is a likely reality. It also means that the added inhibitor cocktail does not inhibit the autolytic activity of BRCA1 TAD. Since, this is the first report of such an observation it is necessary to do further investigation.

5.3.8 Limited proteolysis of BRCA1 (1314-1863), (1396-1863) (1560-1859) and (1646-1859)

Limited proteolysis is often used to explore the ordered structure in proteins. Proteins that are unstructured are sensitive to degradation even at very low protease concentration. To check

the sensitivity, all the N-terminal extended BRCTs were subjected to trypsin digestion. All the purified proteins were degraded within 30 min of incubation to either BRCT or to a larger intermediate size (~33 kDa) and ultimately to BRCT, unlike BRCT which under identical conditions was not degraded even after 2h of incubation (**Figure 5.12)**.

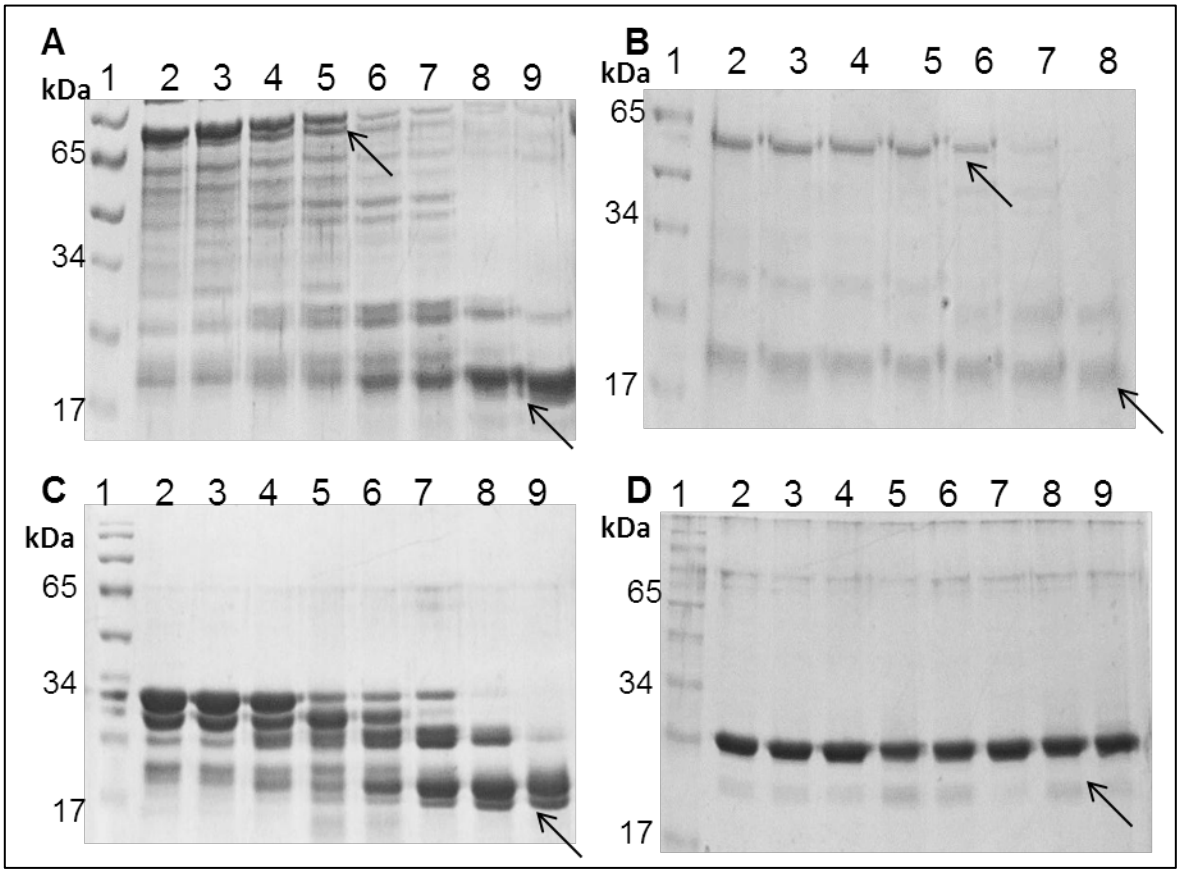


Figure 5.12: Limited proteolysis of N-terminal extended BRCTs constructs*. (A): BRCA1 (1314-1863), (B):BRCA1 (1396-1863), (C):BRCA1 (1560-1859), (D):BRCA1 (1646-1859). Lane1-marker, 2-Control, 3-0min, 4-5min, 5-15min, 6-30min, 7-60min, 8-120min.*

This sensitivity of protein to such a low concentration of protease indicates BRCA1 C-terminal region, except for BRCT, has a disordered structure accessible for protease degradation.

5.3.9 *In silico* analysis of C-terminal region of BRCA1 (1314-1863)

In silico analysis of the region of 550 residues from the extreme C-terminal end of BRCA1 (1314-1863) protein using Composition profiler, Swiss Prot51 and PDB Select 25 databases,

revealed that this region is rich in residues that promote disorder, and is also depleted of order promoting residues. **Figure 5.13A** illustrates the composition profiler output with disorder-promoting residues (red), order-promoting residues (blue), while neutral residues (grey). Graph with bars on lower side indicates depletion and on upper side indicates enrichment of corresponding residues. Analysis of plot of absolute mean net charge versus mean scaled hydropathy of N-terminal extended BRCTs from the software PONDR revealed presence of disorder.

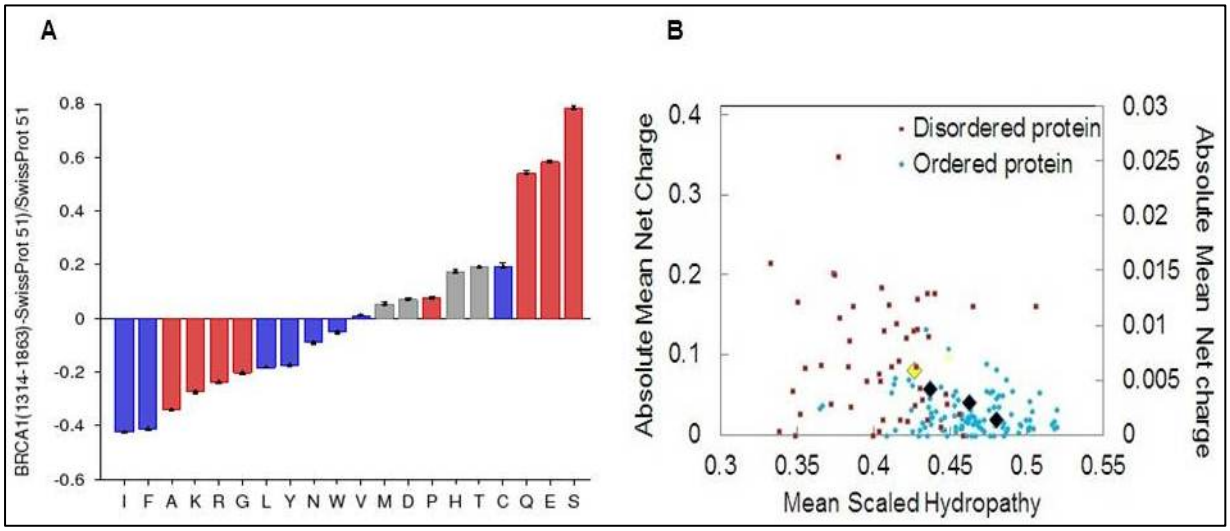


Figure 5.13: Composition analysis of N-terminal extended BRCTs *(A) Composition profiler output. (B) PONDR black squares is N-terminal extended BRCTs; yellow is region from amino acid (1314-1863).*

The plot further shows a decrease in hydrophobicity and increase in net charge of the protein with addition of more and more residues upstream to BRCT, as in regions from BRCA1 (1314-1863), (1396-1863) and (1560-1859) ultimately leading to increases in disorder. Prediction of secondary structural content in the region of BRCA1 (1314-1863) using PSIPRED server indicates the structure content is concentrated in the BRCT region of BRCA1. The upstream region of BRCT up to 1314 residue has maximum coil like structure with only few stretches of short alpha helices and one long alpha helix.

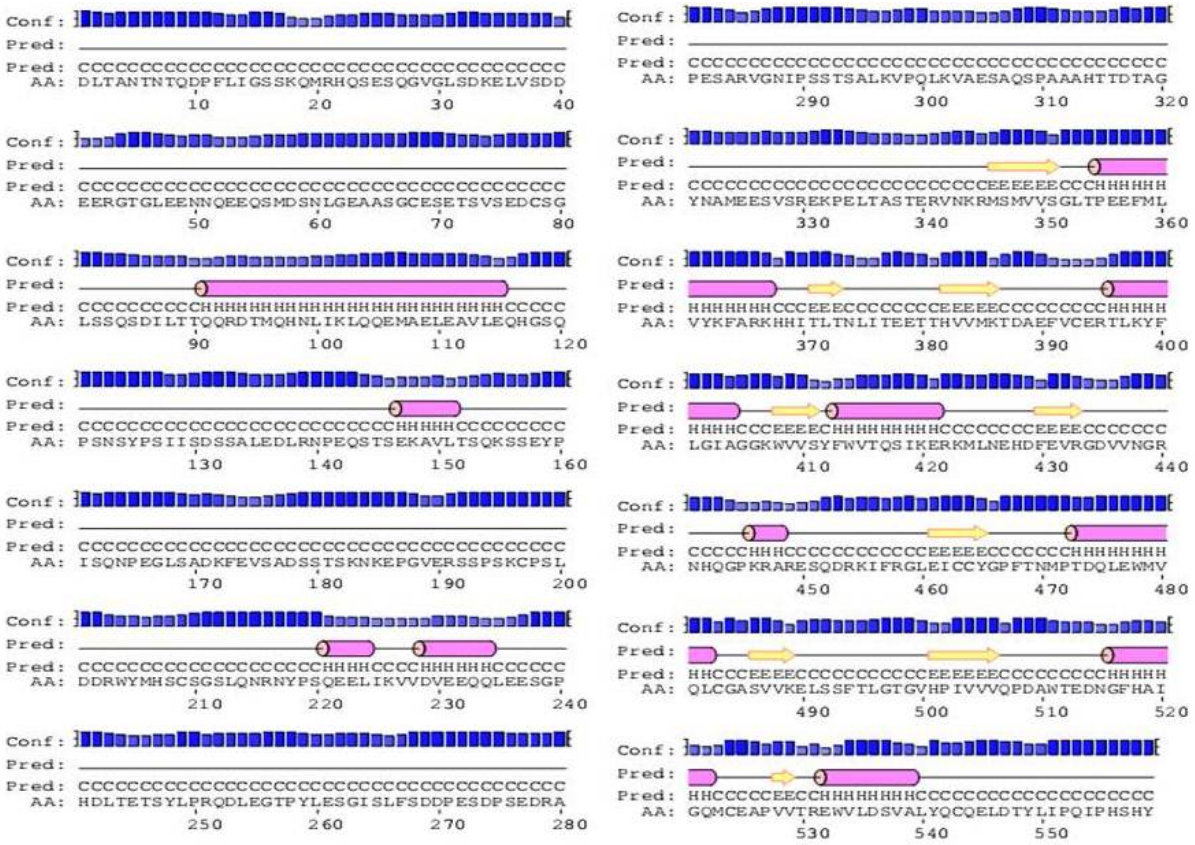


Figure 5.14: Secondary structural content predicted using PSIPRED*. Pink-helix (H), yellow-sheet (E), black line- coil (C), AA-target sequence, conf-prediction confidence*

5.4 Summary

N-terminal extended BRCTs with functional relevance were cloned, expressed and purified from *E. coli*. The homogenous and purified proteins were used to carry out biophysical, biochemical and crystallisation experiments. The elution profile from SEC column suggested all the purified proteins to be homo dimers or non-compact structures except for BRCA1 BRCT. However, the native PAGE profiles for the fragments BRCA1 (1314-1863) and (1396-1863) displayed characteristics of open structure. Secondary structural characterisation by CD revealed the purified N-terminal extended BRCTs to have predominantly alpha helical content in them. The proportion of alpha helix decreased as more and more residues are incorporated upstream to BRCT, and the purified protein BRCA1 (1314-1863) demonstrated behaviour of an unstructured protein. Tertiary structure profiles derived from fluorescence

spectroscopy demonstrate similar behaviour of decrease in compactness of protein with addition of residues upstream to BRCT. To reveal the structures at atomic level, attempts were made to crystallize these N-terminal extended BRCTs by hanging and sitting drop vapour diffusion method. However, crystals were obtained only for the BRCA1 (1560-1859) fragment and these crystals diffracted to a low resolution of around 3.5 Å on the home source of X-rays. Processing of the diffraction images classified the crystals to space group P 6_1 2 2 with unit cell parameters a=b= 114.2 Å and c=120.9 Å. Cell content analysis calculates the Matthews coefficient (V_M) value of 3.40 $Å^3$Da-1 which corresponds to a solvent content of 63%, and suggests presence of one protein molecule of 33 kDa per asymmetric unit. Molecular replacement was performed using software PHASER and the structure with PDB ID: 1T29 as a search model. Interestingly, good electron density was observed only for residues BRCA1 (1646-1859) which forms the BRCT domain. Since Matthews number shows that there is enough space to accommodate extra 86 amino acid upstream of BRCT it is likely that the polypeptide is disordered. As the cell parameters matched exactly with those of BRCA1 BRCT crystals (PDB ID: 1JNX), SDS-gel electrophoresis was employed to investigate the actual protein content of the crystals and the mother liquor. A band corresponding to 33 kDa was observed in some dissolved crystal samples although majority of samples showed band at 24 kDa suggesting sample degradation to BRCT domain. As stringent conditions were maintained to avoid any protease contamination from external source the degradation of protein was astonishing. The reason for degradation of BRCA1 (1560-1859) was explored by incubation of only BRCA1 (1560-1859) , BRCA1 with protease inhibitor cocktail and BRCA1 with 6 M Urea at 37^{o}C.The protease inhibitor cocktail used can inhibit serine proteases, amin peptidases, cystein and metalloproteases. Even in presence of protease inhibitor cocktail at 10X concentration total degradation of protein resulted within 31 hr. This observation suggests the interesting possibility that BRCA1

(1560-1859) has autolytic activity. Limited proteolysis study also shows that while BRCT domain was resistant, all others were susceptible to quick tryptic digestion yielding 24 kDa bands corresponding to BRCT. *In silico* studies using servers PSIPRED, Composition Profiler and PONDR have shown that the regions upstream of BRCT (1646-1859) have no defined structure, and are enriched with disorder promoting residues. These results are consistent with biophysical studies. Thus, in conclusion in the C-terminal region of BRCA1 from 1314-1863 it is only the BRCT that has ordered structure. The intrinsically unstructured nature of this segment, like the central region of BRCA1, may have been deliberately designed to confer multi-functionality to BRCA1 through interaction with other binding partners.

Chapter 6

Multimodal approach to explore intrinsic disorder in central domains of BRCA1

6.1 Introduction

BRCA1 comprises of 1863 amino acids [211] with an N-terminal ring finger domain (RING) [212], the C-terminal BRCT domain and a central region. The central region of BRCA1 (250-1200) interacts non-specifically with DNA, and has strong affinity for cruciform DNA [73,95,104]. It is also known that this region interacts with many different proteins, including ZBRK1, p53, RAD50 and RAD51 [58,213,214].

The structure of the central region of BRCA1 has been investigated previously [17]. These authors have subdivided the central region into 21 fragments of varying lengths, and have expressed each fragment with a histidine-tag. Out of the 21 different constructs generated, 16 were designed such that their N and C termini did not occur in a region of predicted secondary structure. The remaining five were selected based on the regions reported in the literature as binding to other proteins or DNA. These tagged constructs were expressed, purified and used for further biophysical studies without removing the tag. Their experimental studies suggested that these tagged protein fragments were intrinsically unstructured. They further suggested that this disorder may help to encode multi-functionality in the BRCA1 sequence. However, more recently presence of polyhistidine tag is shown to perturb the conformation and compaction in intrinsically disordered proteins, probably because of interaction between terminal histidine residue & other parts of the protein [215]. Further secondary structure prediction is not fool-proof, and it is conceivable that the boundaries between secondary structural elements were not correctly identified [216].

We have identified different set of fragments of BRCA1 that are reported to bind DNA and specific proteins. These fragments are: BRCA1(260-553) which binds to HP1γ involved in gene silencing [85]; BRCA1(341-748) which binds to ZBRK1, RAD50 and RAD51 [57,217]; BRCA1(502-802) which interacts with transcriptional activation domain of STAT1a [218]

and BRCA1(758-1064) which interacts with RAD51 [58]. These fragments have been abbreviated as CRC-1, CRC-2, CRC-3 and CRC-4 respectively. Here, we report cloning, expression, purification and biophysical studies on all these fragments of BRCA1. Further, we have removed the tags introduced to aid in the purification process. Our results also indicate that these fragments are intrinsically unstructured.

6.2 Materials and methods

6.2.1 Cloning of CRCs of BRCA1

The four CRCs were PCR amplified from full length BRCA1 cDNA and cloned in different expression vectors. The forward and reverse primers used in PCRs are mentioned (**Table 6.1).**

Table 6.1: *List of primers used for cloning of central region constructs*

BRCA1 fragments	**F→ Forward, and R→ Reverse, primers (5'→3')**
BRCA1(260-553) CRC-1	BRCA1 260 F GTCGGATCCGAGAACCTGTACTTTCAGGGTAAGTATCAGGGTAGTTCTGTT
	BRCA1 553 R GTCCTCGAGCTATTAATGACCACTATTAGTAATATT
BRCA1 (341-748) CRC-2	BRCA1 341F GTCGGATCCGAGAACCTGTACTTTCAGGGTGATCTGAATGCTGATCCCCTG
	BRCA1 748 R GTCGAATTCCTATTATTTGGGGTCTTCAGCATTATT
BRCA1 (502-802) CRC-3	BRCA1 502 F GTCGGATCCGAGAACCTGTACTTTCAGGGTTTAAAGCGTAAAAGGAGACCT
	BRCA1 802 R GTCGAATTCCTATTACACACATTTATTTGGTTCTGT
BRCA1(758-1064) CRC-4	BRCA1 758 F GTCGGATCCGAGAACCTGTACTTTCAGGGTCAAACTGAAAGATCTGTAGAG
	BRCA1 1064 R GTCGAATTCCTATTAACTGGAACCTATTTCATTAA

The PCRs were performed in 50 µl total volume containing Pfu polymerase (Fermentas), 0.2 mM of each dNTP, 10 pM of both primers, and 100 ng BRCA1 cDNA as template. These samples were incubated for 5 min at 98°C and then the program was set for 32 cycles. The PCR amplification was carried out under following conditions for denaturation, annealing and extension: 30 s at 98°C, 30 s at 65°C, and 1min at 72°C. After the last cycle, the reaction mixture was incubated for an additional 10 min at 72°C.The PCR products were purified using Qiagen gel elute kit to remove excess primers and then digested with BamH1 and ECoR1 to generate cohesive ends for cloning in suitable vectors. The digested products were separated by electrophoresis, and were extracted from the agarose gel and further purified.The digested vector DNA and PCR products were then ligated for 15 minutes using the quick ligase. The ligated samples were transformed into the *Escherichia coli* DH5α cells and plated onto the LB-antibiotic plates. Resistant colonies were picked and screened for positive clones, which were then sequenced for the presence of the correct DNA sequence. Table below lists different vectors and bacterial strains used for cloning and protein expression.

Sr.No	BRCA1 region	Vector	Resistant marker	Protein expression strain
1	BRCA1 (260-553)	pGEX-4T	Ampicillin	BL21(DE3)
2	BRCA1 (341-748)	pET41 a+	Kanamycin	Rosetta2 (DE3)
3	BRCA1 (502-802)	pGEX-KT	Ampicillin	Rosetta2 (DE3)
4	BRCA1 (758-1064)	pGEX-KT	Ampicillin	BL21(DE3)

6.2.2 *In silico* analysis for disorder prediction

In-silico analysis was performed to evaluate the sequence and structure composition of BRCA1 fragments. Composition Profiler software was used to detect the amino acid differences between BRCA1 CRCs and the proteins belonging to the PDB Select 25 reference database, for purposes of predicting disorder propensity [209,210]. PONDR

analysis was also performed to check the positioning of CRCs in the hydropathy plot [208]. Structural disorder in the four CRCs of BRCA1 was also analysed by submitting the amino acid sequences to different online servers: primary servers like Cspritz [219], Anchor [220], and Spine D [221], and the two meta servers MFDp [222] and MetaDisorder [223]. The prediction algorithms of these servers take into account sequence complexity, intramolecular interactions, solvent accessibility, torsion angle fluctuation and secondary structure.

6.2.3 Expression of BRCA1 central region constructs

For protein expression, CRC-1 and CRC-4 were successfully transformed in *E. coli* bacterial strain BL21 (DE3), while CRC-2 and CRC-3 were transformed in *E. coli* Rosetta2 (DE3) strain. A single transformed colony was inoculated in Luria Bertani broth with appropriate antibiotics. Bacterial cultures were grown at 37^0 C until $O.D_{600}$ reached a value between 0.6-0.8 and insert expression was induced with 0.4 mM IPTG. CRC-2 and CRC-3 were incubated for 16 h at 18^0 C whereas CRC-1 and CRC-4 were incubated for 3 h at 37^0 C, after which the cells were harvested.

6.2.4 Protein purification

1. Purification of BRCA1 {260-553} (CRC-1)

Harvested bacterial pellet of CRC-1 was suspended in buffer D (50 mM Tris pH 7.5, 300 mM NaCl, 1 mM EDTA, 5% Glycerol, 0.1% Triton X-100, 20 mM β-mercaptoethanol and protease inhibitor cocktail). Resuspended cells were lysed by sonication for 3 times with 1 minute duty cycle at 60 pulse rate. Cell debris was removed by centrifugation at 18,000 rpm for 40 minutes at 4^0C. The supernatant was passed through the pre-equilibrated glutathione resin for affinity purification. Bound resin was washed with wash buffer (50 mM Tris pH 7.5, 1 mM EDTA, 5% Glycerol, 1M NaCl, 1% Triton X-100 and protease inhibitor cocktail). To get protein in native form, on-beads cleavage using TEV protease was performed. Proteins

were further passed through pre-equilibrated Ni-NTA column to remove His-tagged TEV and flow through fraction was then concentrated.

2. Purification of BRCA1 {341-748} (CRC-2)

Cell pellet of two litres culture was suspended in 80 ml of buffer D. Cells were lysed by sonication at 50 pulse rate, 1min duty cycle for 3 times at 4^0C. Cell lysate was centrifuged to remove the insoluble proteins and cell debris, at 15,000 rpm for 35 min at 4^0C. Supernatant containing the protein in soluble form was allowed to bind with the pre-calibrated GST beads for an hour. To remove non-specifically bound protein, 8 ml beads were washed with 200 ml of wash buffer containing 500 mM sodium chloride, 20 mM magnesium chloride, 50 mM potassium chloride, 20% glycerol, 5 mM ATP, 1mM PMSF and 20mM β-ME in buffer D. To remove traces of ATP, resin was washed with 10 ml of lysis buffer. To remove the GST tag from the protein, TEV protease was added on-beads and incubated at room temperature for 2 h. 30 ml of elution fraction was collected after cleavage. This fraction was then concentrated to 10 ml and passed through 500 µl of Ni-NTA resin to get rid of the TEV protease which is His-tagged.

3. Purification of BRCA1 {502-802} (CRC-3)

For 2 litres cell pellet, 100 ml of lysis buffer E (20 mM phosphate buffer pH 6.5, 500 mM sodium chloride, 2% glycerol, 1 mM EDTA, 20 mM β-mercaptoethanol and protease inhibitor cocktail) was used for resuspension. Cells were lysed by sonication at 60 pulse rate with 1min duty cycle for 3 to 4 times. Centrifuge the cell lysate at 15,000 rpm for 35 min at 4^0C. The supernatant was collected and passed through pre-calibrated GST column containing 5 ml beads. Cell lysate was incubated with resin for an hour. Beads are then washed with 200 ml wash buffer containing buffer E and 500 mM sodium chloride. To remove the GST tag from the protein, TEV protease was added and incubated at room temperature for 2 h. 30 ml of fraction was passed through column to elute native protein from

beads. Lysis buffer containing 1mM dithiothretol was used for elution. This fraction was then concentrated to 25 ml and passed through 500 µl of GST and Ni-NTA resin to get rid of TEV protease and GST tag.

4. Purification of BRCA1 {758-1064} (CRC-4)

Bacterial cells of three litre culture were suspended in 100 ml of buffer D. Cells were lysed by sonication at 60 pulse rate with 1 min duty cycle for 4 times and were centrifuged at 15,000 rpm for 35 minutes at 4^0C. Supernatant was collected and incubated with a pre-calibrated 5ml GST beads, for an hour. Beads were washed with 100 ml buffer D, containing 500 mM sodium chloride, 0.5 % Triton X-100, 20 mM $MgCl_2$, and 5 mM ATP [224]. On-column TEV cleavage was performed and the eluted protein was passed through Ni-NTA resin to remove His tagged TEV protease.

6.2.5 Size exclusion chromatography

BRCA1 CRCs were purified to homogeneity by passing concentrated protein solution through Superdex 200 column. To prepare the standard plot for molecular weight estimation and calculation of Stoke's radii, protein standards (BSA, lysozyme, and carbonic anhydrase) were eluted from the same column under similar conditions. The sample was chromatographed at a flow rate of 0.5 ml/min, and 1 ml fractions were collected. Standard plot of $(-\log K_{av})^{1/2}$ versus Stoke's radii was used to estimate elution volume of a protein from Stoke's radius, for a given molecular weight. The empirical equation used to calculate Stoke's radius is

$$\text{Log } (Rs)^N = 0.369 \log (MW) - 0.254$$

$$\text{Log } (Rs)^U = 0.533 \log (MW) - 0.682$$

where $(RS)^N$ and $(RS)^U$ are Stoke's radii of the native and completely unfolded states, of a globular protein and MW (Da) is the molecular mass [225]. Molecular weights of CRCs were

estimated from elution parameters using the equation K_{av}= (Ve -Vo)/ (Vt-Vo). Ve is elution volume, Vt is the total volume and Vo is the void volume of the SEC column [226,227].

6.2.6 Limited proteolysis assay

Limited proteolysis of BRCA1 CRCs with trypsin was carried out to probe the stability of the purified protein. Protein concentration of 8-12 μM was used, whereas 3ul of trypsin at a concentration of 20 ng/ μl was used for the assay. The ratio of trypsin to protein used was 1:4000. Protein was incubated with trypsin at 37^{o}C and 60 μl of the reaction mixture was taken out at time intervals of 0, 5, 15, 30, 45, 60 and 120 minutes. The reaction was stopped by the addition of 1μl of 1mM PMSF. Proteolysis digests were analysed on 10% SDS/PAGE.

6.2.7 Circular Dichroism

Far UV CD spectra of the BRCA1 CRCs were recorded. The spectra were recorded at 20^{0}C for the wavelength range of λ=250-195 nm with 3 accumulations and scanning speed of 20 nm/minutes All the samples were baseline corrected. Thermal stability studies for all the BRCA1 proteins were carried out in the temperature range of 20 to 95 ^{o}C at a wavelength range of λ=240-195 nm. All the samples were incubated for 200 s and scans were recorded at a data pitch of 5^{o}C. Plot of θ_{222} nm as a function of temperature was plotted to check for helicity changes.

6.2.8 Native gel electrophoresis

To check the homogeneity and the folding pattern of the protein, purified protein sample was loaded on acidic (CRC-1, CRC-2 and CRC-4) and basic (CRC-3) native polyacrylamide gel. 5 μM of purified protein sample was mixed with loading dye and loaded on the gel. Protein concentration was estimated from molar extinction coefficient and absorption at 280 nm. For basic native polyacrylamide gel electrophoresis, 7% polyacrylamide gel was prepared in potassium acetate buffer. β-alanine acetic acid buffer at pH 4.3 was used for electrophoresis.

Methyl green was used as a tracking dye. For acidic protein 10% polyacrylamide gel was prepared in Tris-HCl buffer. Tris glycine buffer was used for electrophoresis.

6.3 Results

6.3.1 Cloning of BRCA1 central region fragments

Figure 6.1 shows electrophoresis profile of the PCR products and screening of clones done by insert release for the four CRC fragments. The clones with insert release were further confirmed by sequencing. This construct is then further used for protein expression.

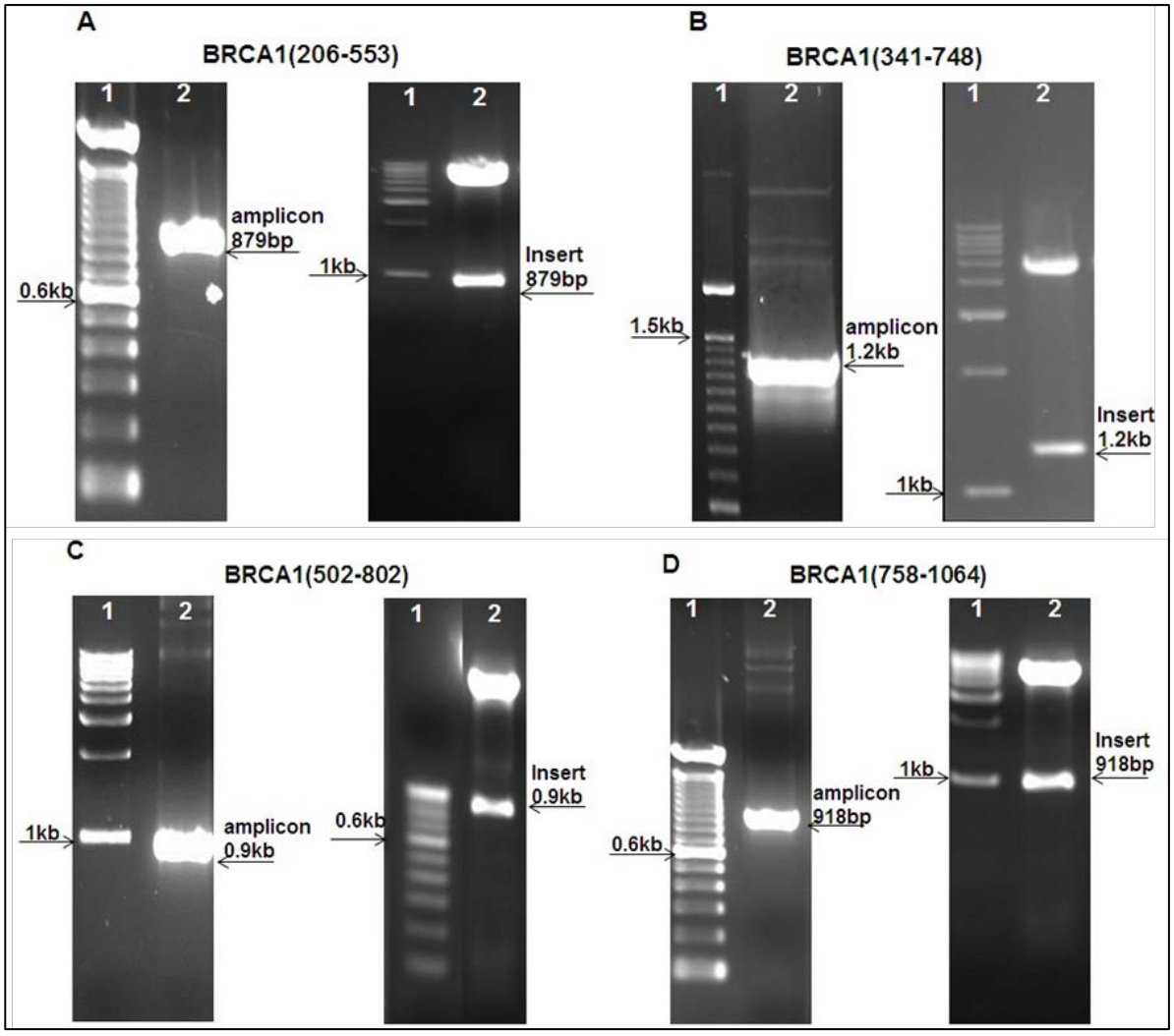


Figure 6.1: Screening of clones of BRCA1 central domains *with BamH1 and ECoR1 enzyme in 1% agarose gel. Lane 1-DNA ladder, 2-Amplicons and insert release (A): CRC-1, (B): CRC-2, (C): CRC-3 and (D): CRC-4.*

6.3.2 Purification of different central fragments of BRCA1

Figure 6.2 shows the SDS profiles at different steps of protein purification for the four CRCs. A contaminating band present at approximately 60 kDa was co-eluted with all the four CRCs. To remove this contaminating band washing with high concentration of sodium chloride, Triton X-100, ATP, magnesium chloride and β-mercaptoethanol was performed. The final purified protein was obtained using gel permeation chromatography.

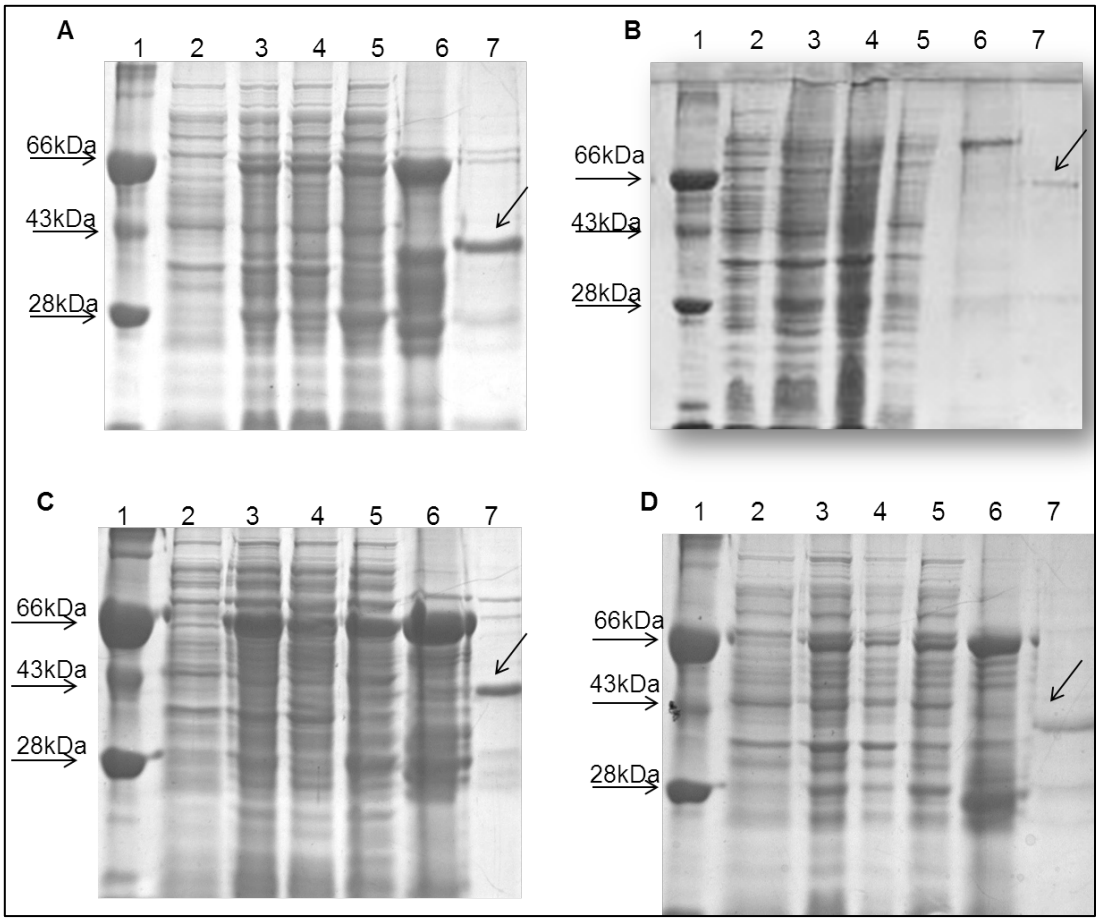


Figure 6.2: Purification of central fragments of BRCA1: A: CRC-1, B: CRC-2, C: CRC-3, D: CRC-4. *Lane 1: Marker, 2: Uninduced whole cell, 3: Induced Whole cell, 4: Pellet, 5: Induced Soluble Fraction, 6: Beads after binding, 7: purified elution fraction.*

To confirm for the protein of interest, trypsin digested peptides were subjected to time of flight mass spectrometry (**Figure 6.3).** MS spectrum of peptide products was analyzed in Flex analysis 3.0 and Bio-tools software.

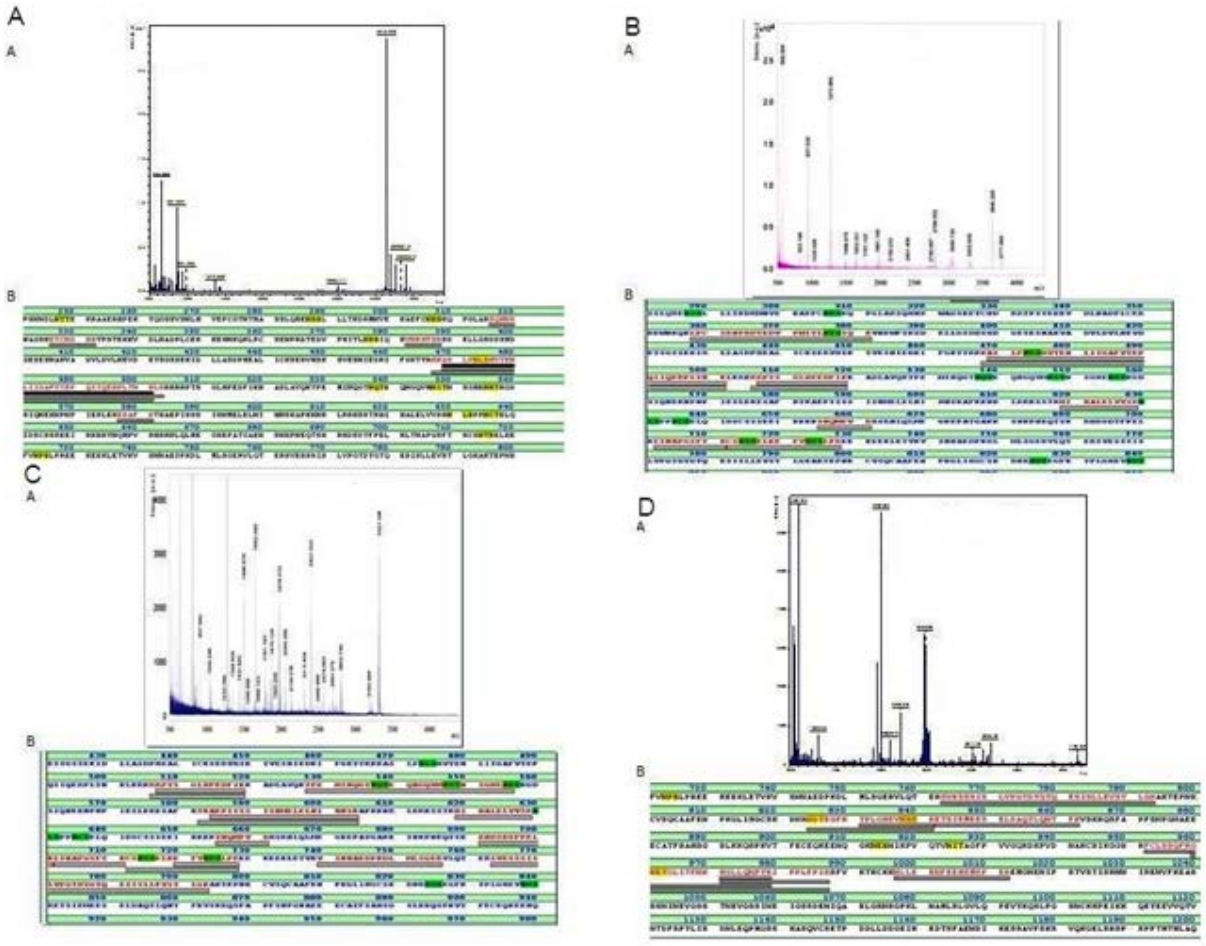


***Figure 6.3: Peptide mass fingerprinting* of CRCs.** *(A) Mass spectrum data, (B) Matched peptide peaks. (A) CRC-1, (B) CRC-2, (C) CRC-3, (D) CRC-4.*

The MS peaks were further analyzed by searching against SwissProt/NCBI nr database using MASCOT search engine. The score shown is with value of $p \leq 0.05$. Table below shows the sequence coverage and score value of confirmed proteins of central region of BRCA1.

Sr.No.	BRCA1 fragments	Sequence Coverage (%)	Score
1	CRC-1	5.3	58.4
2	CRC-2	8	63
3	CRC-3	9	83
4	CRC-4	6.5	82

6.3.3 Oligomeric characterization of the four purified CRCs

The FPLC elution profiles of the four different CRC's loaded on Superdex 200 column are shown in **Figure 6.4**. The retention volumes for all four CRCs are higher than the void volume suggesting the size of all the fragments are smaller than the exclusion limit of the column. The elution peaks are also sharp and symmetrical suggesting homogeneity of molecular species. The lower retention volumes for CRC-1, CRC-2, and CRC-3 suggest them to form higher order oligomers while, for CRC-4, the high retention volume suggests it to be a monomer. A standard plot of $\log_{10}$ (molecular weight) versus (elution volume) was used to estimate the apparent molecular weight of the eluted CRCs. These calculated molecular weights are 126, 153, 105 and 28 kDa respectively.

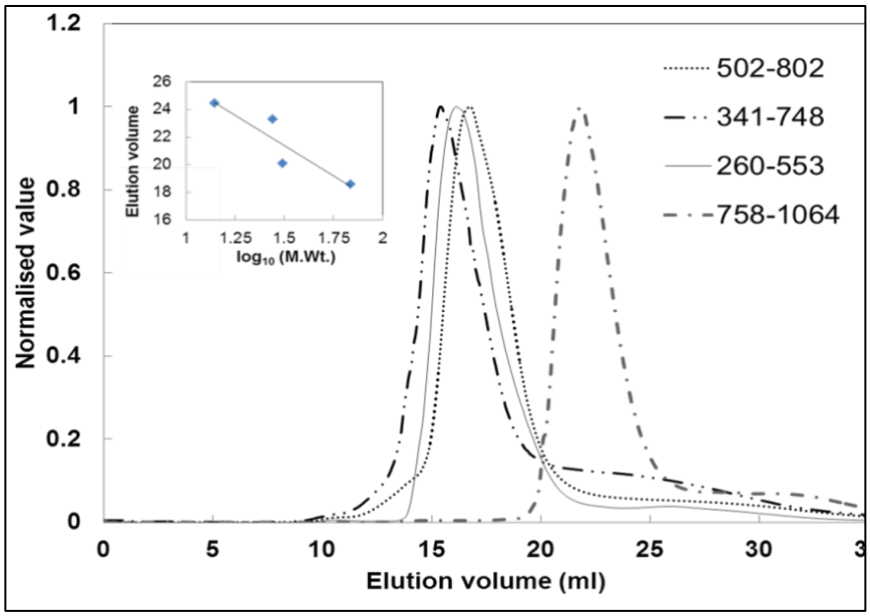


Figure 6.4: Oligomeric characterisation: *Gel filtration profile for the central fragments, molecular weight displayed estimated from standards. Inset shows a standard plot of* log_{10} *(molecular weight) versus (elution volume).*

This estimation of molecular weights is based on the assumption that the molecules are globular (**Inset in figure 6.4).** In the case of native folded/ unfolded proteins following

empirical relationships have been established [225]between the size of the protein and its Stoke's radius

$$\text{Log (Rs)}^{N} = 0.369 \log (\text{MW}) - 0.254$$

$$\text{Log (Rs)}^{U} = 0.533 \log (\text{MW}) - 0.682$$

The Stoke's radius so obtained was used to estimate the elution volume of protein in both, native and unfolded states, by using the standard plot of (–log Kav) ^1/2 versus Stoke's radii (Å) [226,227].

Table 6.2: Characterization of protein conformation from elution volume and Stoke's radii*:*

Da.-Daltons, Ve -elution volume, M.wt.-molecular weight

CRCs	**M.Wt. (Da)**	**Native folded**	**Unfolded**	**Actual Ve**
Monomer (M. Wt. sequence)				
260-553	32910	21.20926	12.7177	16.12
341-748	45990	19.80187	11.2425	15.39
502-802	33620	21.08785	12.5593	16.797
758-1064	33719	21.08783	12.5593	21.878
Dimer (M. Wt. sequence)				
260-553	65820	18.09611	10.3512	16.12
341-748	91980	16.53137	10.0706	15.39
502-802	67240	17.99734	10.3221	16.797
758-1064	67438	17.98372	10.3182	21.878
Trimer (M. Wt. sequence)				
260-553	98730	16.1995	10.0465	16.12
341-748	137970	14.66598	10.0041	15.39
502-802	100860	16.09973	10.0408	16.797
758-1064	101157	16.086	10.0401	21.878
M. Wt. (calculated from Ve of size exclusion column)				
260-553	126000.4	15.0733	10.0086	16.12
341-748	153000.2	14.21338	10.0016	15.39
502-802	105000.8	15.9121	10.0316	16.797
758-1064	27000.82	21.8573	13.6675	21.878

The elution volume so calculated for the four CRCs is in the range of 19.7 to 21 ml when molecule is assumed to be globular, and is in the range of 11.2 to 12.6 ml if protein is assumed to be completely unfolded. However the experimental elution volumes of the CRCs are in the range of 15.3-21.8 ml suggesting that the three CRCs (1-3) have partially folded structures, except for CRC-4 which has demonstrated a folded conformation (**Table 6.2)**. The discrepancy between the observed and calculated elution volumes could be due to oligomerisation of the proteins, so we recalculated the elution volume by considering dimeric and trimeric molecular weights. The match with the observed elution volume was the best when the molecular weight used was that corresponding to a folded trimer (**Table 6.2).** This expectation on the trimerisation of the molecules was not supported by presence of discrete band in native gel electrophoresis. In this gel, each protein gave a smeared pattern with no definite bands suggesting the absence of discrete oligomeric species (**Figure 6.5B**). Hence, looking at the lower retention volume on the size exclusion column, smear pattern of native gel and calculated elution volume from Rs value, it can be concluded that all the four CRCs do not have globular structure. The lower retention volumes in FPLC may be because of skewed shapes.

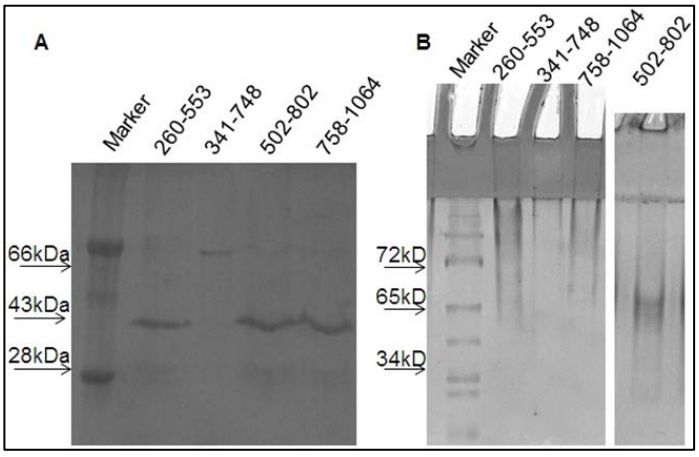


Figure 6.5: Oligomeric characterisation using native-PAGE gel (*A): FPLC pure protein loaded on 10% SDS gel. (B): Native gel profiles of the BRCA1 central domain construct.*

6.3.4 Intrinsic disorder in central regions of BRCA1

The amino acid composition of a protein can be used to predict whether, the protein is in natively folded or unfolded states [26, 27] since, folding of protein is based on the principle that the hydrophobic amino acids are to be buried away from aqueous solvent and hydrophilic amino acids are to be exposed to the solvent. Composition profiler is a web based tool to study for enrichment and depletion of residues in a given sample, when compared to standard dataset of proteins with known physico-chemical and structural properties. It plots data for fractional differences in amino acid residues in the sample and reference dataset for the selected property. The property selected may be any of the following: 1). ability to form α-helical and β-sheet structures, 2). tendency towards surface exposure, 3). positional flexibility, 4). disorder propensity etc (16).

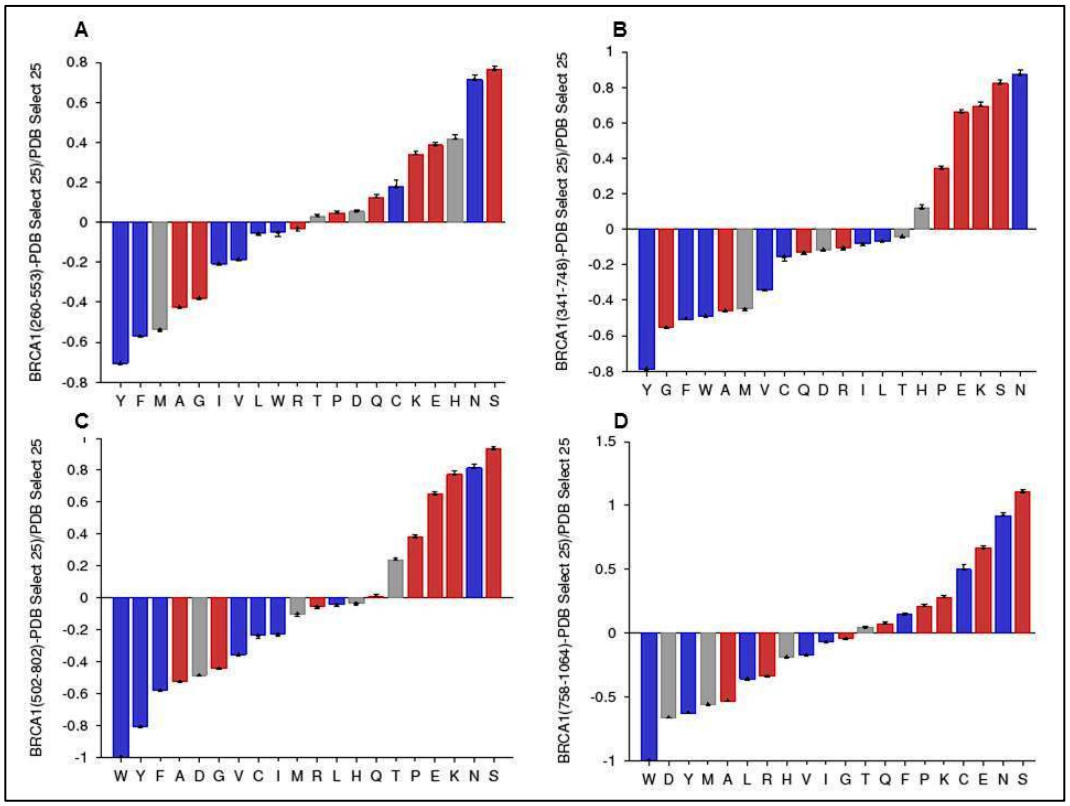


Figure 6.6: Composition profiler output*. Disorder-promoting residues are colored red, order-promoting residues are colored blue, and disorder-order neutral residues are colored black. Graph with bars on lower side indicate depletion and on upper side indicates enrichment of those residues*

In the present analysis of CRCs the reference dataset chosen was PDB Select 25, and the property selected was disorder propensity. A bar plotted on positive y axis denotes enrichment of that amino acid residue while bar on negative y axis denotes depletion of thatamino acid residue **(Figure 6.6).** It may be seen that for all the four CRCs the number of blue bars are more in negative region while number of red bars is more on the positive region. Taken together these results predict that the CRCs are intrinsically unstructured proteins.

This inference is also supported by the analysis using software PONDR, which analyses the net charge and hydrophobic profile [14,15,16] (Figure 6.7). The probability of disorder as predicted by Cspritz primary disorder predictor server is given in **Table 6.3**. Cspritz uses information from X-ray crystal structure and Disprot database for prediction. It is significant that the predicted degree of disorder is substantially high for all the four constructs, ranging from 50 to70%.

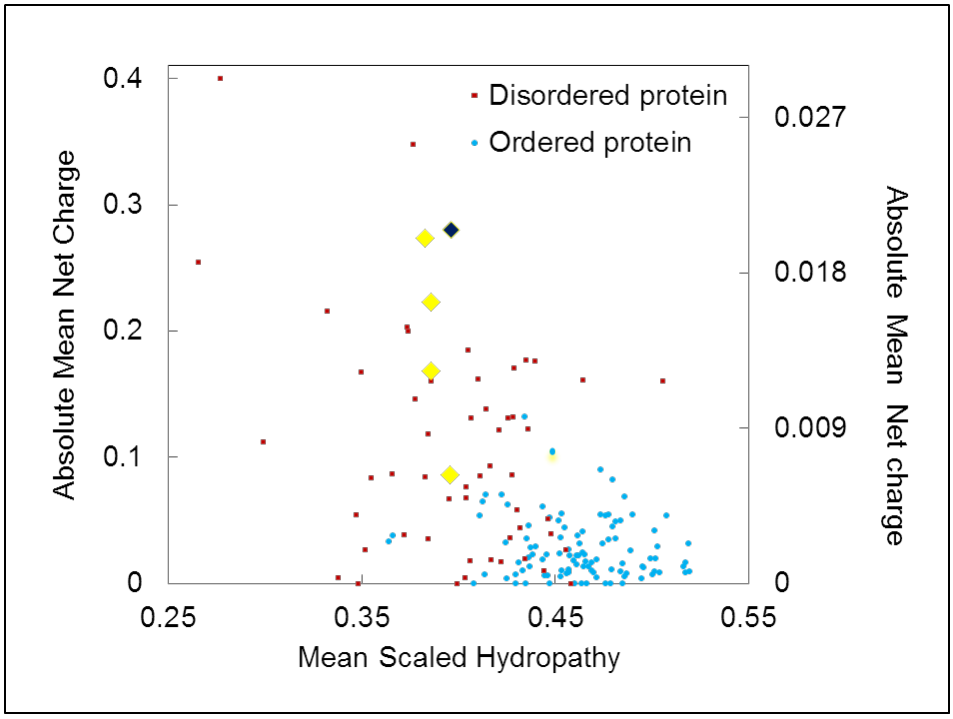


Figure 6.7: PONDR plot: *The plot of absolute mean net charge versus mean scaled hydropathy. Yellow colored squares are the four BRCA1 CRCs; green color is the central region from amino acid 260-1064.*

Table 6.3: Sequence analysis of different domains of BRCA1 from ExPASy, ProtParam tool *to probe for the feature that may lead to disordered protein*

BRCA1 domain	Length	pI	Hydrophobicity (GRAVY)	Disordered residue (%) Cspritz	ligand	Negatively charged residues	Positively charged residues
260-553	293	5.97	-0.934	60.40	DNA, p53	45	39
341-748	408	6.33	-1.033	71.07	ZBRK1	66	62
502-802	301	8.97	-1.038	87.70	DNA	42	48
758-1064	307	5.82	-0.764	51.46	STAT1	40	35

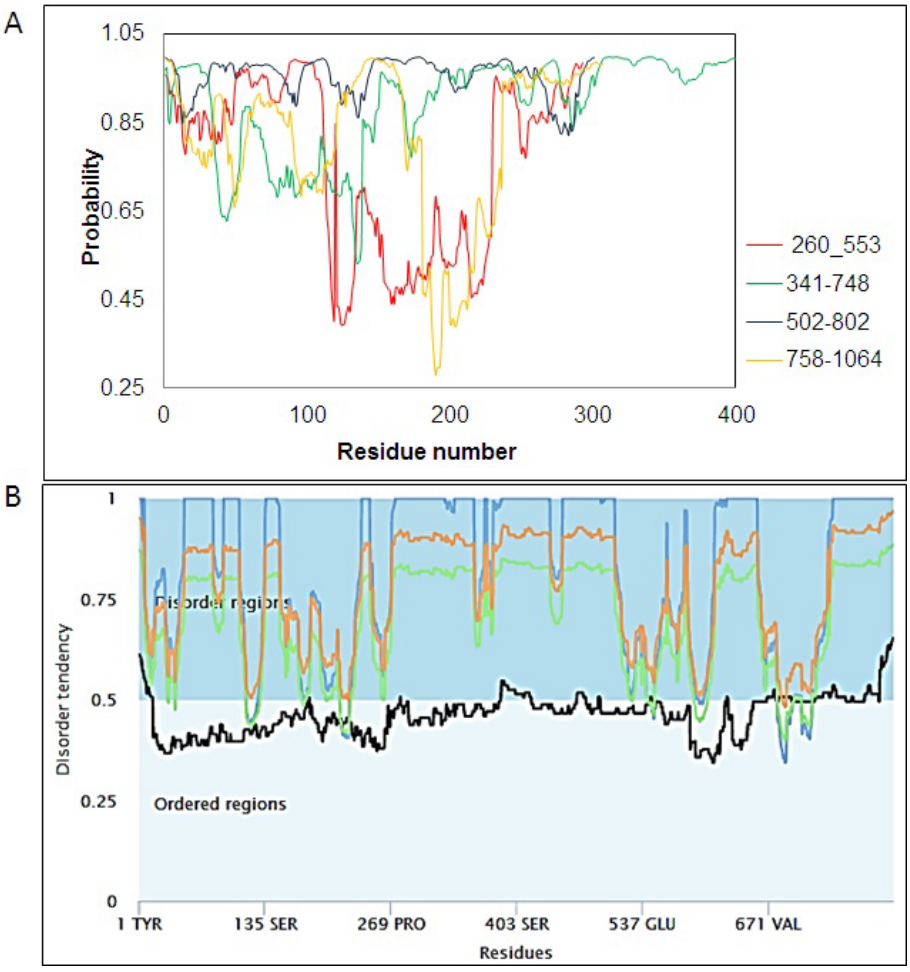


Figure 6.8: In silico prediction *(A)**:** MFDp of the four constructs CRC (1-4), (B): MetaDisorder server analysis considering the entire central region from 260-1064. Here x axis residue 1 corresponds to residue number 260 on the BRCA1 protein.*

The probability of disorder, analyzed from Meta servers (MFDp and Meta disorder) also calculates the probability of disorder to be high. Meta server MFDp is an ensemble of 3 support vector machine which predicts short, long and generic disordered region. It predicts disorder based on three complementary disorder predictors, sequence information, predicted secondary structure, solvent accessibility, backbone dihedral torsion angles, residue flexibility and B-factors **(Figure 6.8A)**. MetaDisorder server combines prediction from 13 different servers and uses the combined output to improve the prediction accuracy **(Figure 6.8B).** Thus the analysis from primary and Meta servers predicts the central region to be intrinsically disordered.

6.3.5 Secondary structural analysis of CRCs

Far UV CD spectra of CRC-1 to CRC-4 are shown in (**Figure 6.9A)**. It may be seen that these proteins show negative ellipticity at λ_{200} nm. It has been reported that random coils or intrinsically-disordered proteins exhibit a negative ellipticity near λ_{200} nm and ~zero ellipticity around λ_{218} nm [228,229,230]. Central regions of BRCA1 thus show CD spectra characteristic of random coil structures.

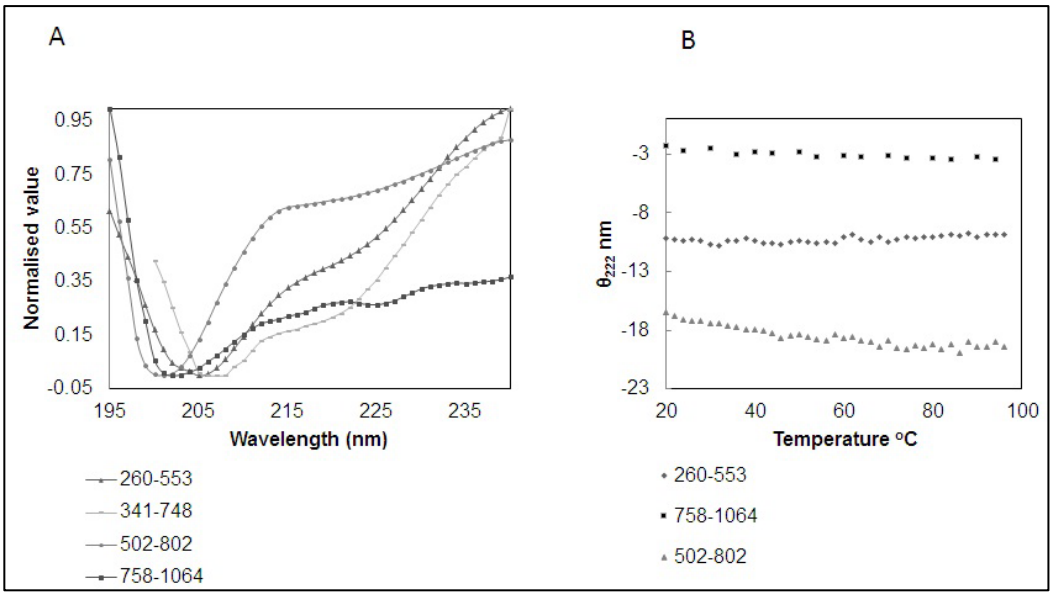


Figure 6.9: Circular Dichroism Spectroscopy :(A)*: Far UV CD scans of all the central region of BRCA1.* ***(B)****: Plot of ellipticity at θ_{222} nm versus temperature*

Further, to confirm for absence of secondary structural characteristics, temperature induced changes in ellipticity was monitored at λ=222 $_{nm}$. A slight increase in negative ellipiticty was observed at around 70^{o}C (**Figure 6.9B)**. This effect may be attributed to increased strength of the hydrophobic interaction at higher temperatures, which may act as a driving force for folding [231,232,233]. This slight increase in negative ellipticity due to increase in temperature is defined as "turned out response" as witnessed with other proteins like α-synuclein [234], caldesmon [235], and phosphodiesterase γ-subunit [236].

6.3.6 Protease sensitivity assay for structural characterisation

Limited proteolysis technique relies on accessibility of the protease target site, enhanced backbone flexibility and local unfolding at the proteolytic site. This is often used to explore disordered regions in the proteins. Proteins with disordered structure are sensitive to degradation even at very low protease concentrations [237,238].

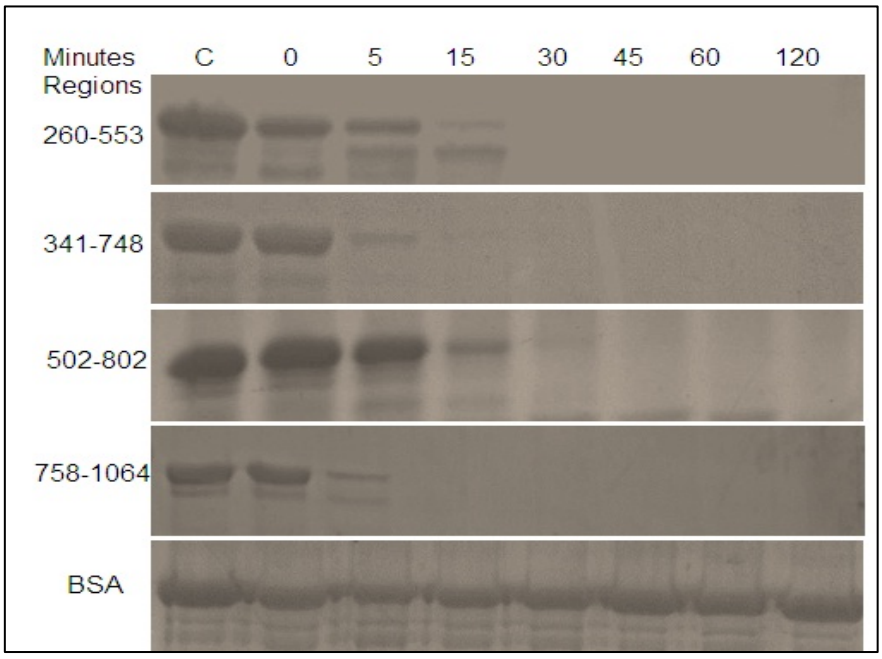


Figure 6.10: Protease sensitivity assay*: Central purified domains treated with trypsin at concentration of 4000:1. All the different domains were incubated at 37^{o}C for different time points of 0, 5, 15, 30, 45, 60, 120 m. C-is untreated protein sample used for assay.*

In fact, molecular simulation studies have suggested that more than 10 residues around the cleavage site are required to be in an unfolded conformation to provide accessibility to trypsin to its active site [239]. To check the sensitivity of this central region of BRCA1, purified proteins were subjected to trypsin degradation. The ratio of trypsin to protein was 1:4000. The four purified CRCs were degraded within 15 min of incubation at 37^{o}C, unlike bovine serum albumin, which, under identical conditions, was not degraded even after 2h of incubation (**Figure 6.10)**. This sensitivity of CRC's to such a low concentration of protease indicates BRCA1 central region has a disordered structure and accessible protease cleavage sites.

6.4 Summary

Initial preliminary study was done to determine the structural elements presents in these different constructs. The central regions of BRCA1 co-purified with another protein of approximate molecular weight of 60 kDa. This contaminant was identified as a bacterial chaperon, because removal of that protein required extensive wash of affinity resin with wash buffer containing Triton X-100, NaCl and ATP with $MgCl_2$. Co-purification of chaperons, especially Hsp70 is reported to assist in solubilising the aggregation-prone, intrinsically disordered proteins [240]. Hsp70 is known to bind proteins in their extended form and protect them from proteases thereby helping in folding [241]). This co-purification of chaperons & purification with $MgCl_2$ and ATP suggests that the central regions of BRCA1 have an extended form which is stabilised by bacterial chaperons. All the four CRC's were extremely prone to degradation supporting their unstructured nature.

Molecular weight estimated from elution volumes on the size exclusion column, suggested the protein to be monomeric with partially folded conformation. Interestingly, the Stoke's radii predicted all the CRCs to be trimeric except for CRC-4 which was monomeric.

However, native gel profile suggested the protein to have a partial structured conformation. This was further confirmed by limited proteolysis. Limited proteolysis assay also showed high sensitivity of the four CRCs to trypsin digestion. However, there appears to be a difference in the rates at which these four proteins are hydrolyzed by trypsin. The CRC-1 and CRC-3 are more stable compared CRC-2 and CRC-4.suggesting presence of residual secondary structure in CRC-1 and CRC-3. Since, binding to DNA involves specific structural elements of protein this observation suggests that CRC-1 and CRC-3 are the regions utilised by BRCA1 for direct interaction with DNA. This expectation is also supported by experimental finding of previous published report [242].

Bioinformatics analysis of sequences of BRCA1 central regions carried out using primary and Metadisorder servers, PONDR and Composition profiler classified them as intrinsically unstructured proteins based on their mean-hydrophobicity, mean charge at pH 7, presence of flexible loops, predicted secondary structure and missing coordinates in crystal structures. All these data suggest the central region of BRCA1 to be intrinsically unstructured

6A: Conformational characterisation of BRCA1 (502-802) in presence of DNA

6A.1 Introduction

The region in between the ring finger & BRCT domains of BRCA1 is known as the central region that interacts with transcriptional repressors, activators, DNA repair proteins and also with DNA [57,68,105,243,244,245]. BRCA1 central region from amino acid 504-802 was responsible for the inhibition of nucleolytic activity of enzyme Mre, involved in double-strand break repair, in the Mre11/Rad50/Nbs1 complex [104]. Competition gel retardation assay has revealed that BRCA1 has the ability to discriminate between four-way junction DNA, mismatch containing DNA, bulge containing DNA and linear DNA. Central region of BRCA1 is reported to bind to DNA non-specifically and to have preference for cruciform DNA/ supercoiled DNA [242,246]. BRCA1 protein when bound to DNA forms two types of complexes: Type1 complex which has very strong affinity for nonspecific DNA which bind without any structural requirement and Type II complex shows cooperative binding, and which is affected by the structure and the length of the DNA molecule [104]. NMR and biophysical studies on the central region of BRCA1 established the central region to be intrinsically unstructured/ disordered [17,112].

Disordered state of a protein is only thermodynamic state, unlike ordered protein that is both in thermodynamic and structural state [247]. The proteins which are intrinsically unstructured are involved mostly in signalling, regulation or control e.g. BRCA1, p53, Calmodulin. The interfacial and contact free energy of ligand binding to unstructured protein is used to bring about the disorder to order transition [14]. This association results in a highly specific interaction with low net free energy of association, which is useful for reversible processes like signalling. Due to flexibility in structure, disordered protein can act as scaffold to bind multiple partners forming a hub [248,249]. The globular protein structure is restricted to maintain long range interaction and a nonpolar interior unlike disordered protein; hence the

rate of evolution is constrained in globular protein. The pattern of amino acids substitution is therefore chemically restrained in globular proteins [250]. Disordered proteins are reportedly known to fold upon binding to their interacting partners [251,252,253]. Phospho tyrosine-binding domain of Numb protein [254] and the intrinsically disordered calcium binding domain of calmodulin undergo disorder-to-order transitions upon Ca2+ binding [255]. The release of solvent molecules is supposed to provide a favourable entropic contribution to the binding. This structural plasticity leads to functional plasticity of protein e.g. Calmodulin, Numb [256,257].

BRCA1 is indispensable for ATM and ATR-dependent phosphorylation of downstream targets e.g. p53, Nbs1 and Chk2. The intrinsic disorder in central region of BRCA1 enables it to act as a scaffold for the binding of the two kinases ATM and ATR, thereby facilitating phosphorylation of downstream components required for apoptosis and checkpoint activation [258]. Protein phosphorylation sites were found to be in regions of the protein that is structurally characterized as natively disordered [15,259,260]. The central region of BRCA1 is also believed to be involved in direct interaction with DNA. However, the minimal region required for DNA binding is not yet clearly established. [17,104,242]. The fragment (452-1079) of BRCA1 is reported to bind DNA and multimerise on DNA binding [104]. No binding was reported with DNA for the fragment BRCA1 (1021-1552). Naseem et al suggested the BRCA1 (340-554) is the minimal DNA binding region of BRCA1 [111]. Further work by Brazda et al suggested BRCA1 (498-663) as the region sufficient for effective binding. The BRCA1 (219-498) did not establish any binding with DNA [242]. In view of all uncertainties we have overexpressed the region of BRCA1 (502-802), which encompasses most DNA binding regions described above. Our aim has been to examine if order is conferred on the central region of BRCA1 on binding to DNA.

In this study, we have shown that the central region of BRCA1 is intrinsically unfolded, binds effectively to duplex and cruciform DNA as expected. Surprisingly, this region remains unfolded even after interaction with the DNA.

6A.2 Materials and Methods

6A.2.1 Size exclusion chromatography

To monitor the conformational change in BRCA1 (502-802) on binding with DNA, purified protein was incubated along with duplex and cruciform DNA. Sequence for oligonucleotide used for cruciform and duplex DNA formation is as follows [111].

Type of DNA	Sequence 5'---3'
Cruciform DNA	F1: GAATTCAGCACGAGTCCTAACGCCAGATCT
	F2: AGATCTGGCGTTAGGTGATACCGATGCATC
	F3: GATGCATCGGTATCAGGCTTACGACTAGTG
	F4: CACTAGTCGTAAGCCACTCGTGCTGAATTC
Duplex DNA	D1: GTCGACTTATGCCAAGTGGTACGTCTCCGT
	D2: ACGGAGACGTACCACTTGGCATAAGTCGAC

Purified protein (20 μM), protein (16 μM) along with duplex DNA (45 μM) and protein (16 μM) with cruciform DNA (35 μM) was chromatographed using Superdex 200 column. Standard proteins (BSA, lysozyme, phosphorylase B and carbonic anhydrase), DNA (duplex and cruciform) were eluted from the same column under similar conditions to prepare the standard plot for molecular weight estimation, and to get the elution volume of free DNA respectively. 500 μl of sample were injected and resolved at a flow rate of 0.5 ml/minutes. The elution was monitored by absorbance at λ=215 and 280 nm.

6A.2.2 Circular Dichroism

Far UV CD scan for the wavelength range of 240-195 nm was monitored for BRCA1 (502-802) (14μM), BRCA1 (502-802) (14μM) with duplex DNA (4μM), and BRCA1 502-802 (14

μM) with cruciform DNA (2μM). Near UV CD scan from λ=310-230 nm was also recorded to monitor the structural changes in the DNA. Protein concentration used was 1.7 μM with 0.8 μM cruciform DNA and 1.8 μM duplex DNA. All the measurements were baseline corrected. CD scans were recorded at 20^0C, at a scanning speed of 20 nm/min, and 3 accumulations were averaged. Protein concentration was measured using Bradford assay. DNA concentration was calculated based on vendor provided concentration.

6A.2.3 Fluorescence spectroscopy

Fluorescence measurements were made on sample excited at λ=295 nm. Protein concentration was 1.7 μM with 0.8 μM cruciform DNA and 1.8 μM for duplex DNA. The emission slit width was 5 nm. The fluorescence scan was monitored for emission wavelength range λ_{em} 310–500 nm. All the measurements were baseline corrected. The spectral data were collected using FluorESSCENCE software.

6A.2.4 Limited proteolysis

To understand the effect of cruciform and duplex DNA on stability of BRCA1 (502-802), limited proteolysis with trypsin was performed. Protein at a concentration of 20 μM was mixed with DNA (duplex and cruciform 10 μM) and incubated in ice for 30 minutes.. Protein concentration of 20 μM was estimated from Bradford using molar extinction coefficient value. 60 ng of trypsin was used for the assay. The molar ratio of trypsin to protein was 1:7500. Protein was incubated with trypsin at 30^oC, for periods of 0, 1, 2, 4, 6, 8, 12, and 20 minutes. Reaction was stopped by addition of 1mM PMSF in 60 μl aliquots and the mixture was boiled at 100°C for 3 to 5 minutes, and then these samples was loaded on the gel immediately. Proteolysis digests were analysed on 10% SDS/PAGE.

6A.2.5 Electrophoretic Mobility shift assay

To check for the interaction of BRCA1 (502-802) and DNA, cruciform and duplex DNA were titrated with increasing concentrations of protein. Binding was performed in binding

buffer containing 20 mM Tris pH 8.0, 50 mM NaCl, 4 mM DTT, 0.1% Triton X-100, 10% glycerol, 0.01 mM $ZnCl_2$. Cruciform DNA (0.10625) μM was titrated with protein of concentration 0, 0.065, 0.131, 0.196, 0.262, 0.328, 0.393, 0.459, 0.525, 0.590 μM and Duplex DNA (0.225) μM was titrated with protein of concentration 0, 0.372, 0.496, 0.620, 0.744, 0.868, 0.993, 1.117, 1.241 μM. These samples were incubated at 4ºC for 30 min, and were then electrophoresed with 0.5X Trisborate- EDTA buffer onto a 7% pre-electrophoresed polyacrylamide gel. The gel was run in the cold room at 200 V and 40 mA current.

6A.3 Results

6A.3.1 Conformational characterisation of BRCA1 (502-802) and DNA complex

The elution profiles of BRCA1 (502-802), BRCA1 (502-802) with duplex DNA, and BRCA1 (502-802) with cruciform DNA chromatographed on Superdex 200 column are presented (**Figure 6A.1)**. The elution volumes of standards injected was used to estimate the molecular weight of the sample (**Inset 6A.1).**

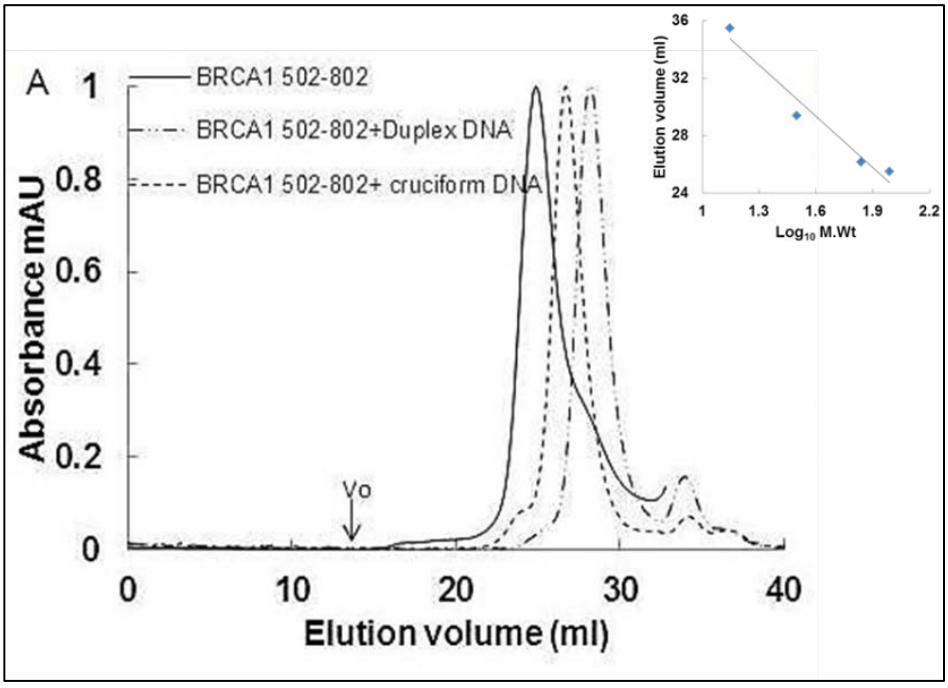


***Figure 6A.1**: **Size exclusion chromatography profile**: BRCA1 purified protein along with complex of protein with duplex and cruciform DNA injected in Superdex 200 column. Inset shows plot of elution volume vs. log molecular weight of standards.*

The BRCA1 (502-802) fraction eluted at 22 mL in Superdex 200 column, that corresponds to a molecular weight of 95 kDa. The protein complexed with duplex and cruciform DNAs eluted at a higher retention volumes that is equivalent to molecular weights of 49 and 67 kDa respectively. The higher retention volume of protein DNA complex suggests that the protein has a smaller compact size when bound to DNA.

6A.3.2 Secondary structural characterisation of BRCA1 (502-802) and DNA complex

Far UV CD scan of (BRCA1 (502-802)) and protein DNA complex were obtained **(Figure 6A.2A).** The overlay of scans of protein and protein DNA complex revealed increase in negative ellipticity at around λ=222 nm, and the negative ellipticity at λ=200 nm, which signifies disorder, showed a shift towards 205 nm. This increase in negative ellipiticity λ=222 nm indicates incorporation of structural characteristics in the protein incubated with DNA. To understand the structural changes in the DNA due to protein binding CD scan was monitored from λ=310-230 nm. Increase in positive ellipticity at 280 nm and negative ellipticity at 245 nm was observed (**Figure 6A.2B).**

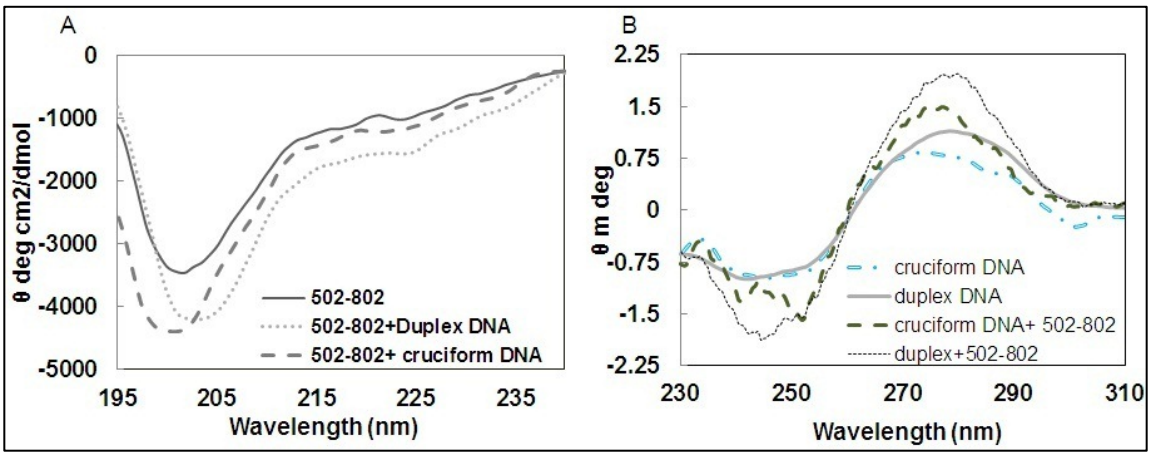


Figure 6 A.2: Secondary structure characterisation*: Secondary structure characteristics of BRCA1 (502-802) studied in presence of DNA: (A) Far UV CD scans of BRCA1 (502-802), and BRCA1 (502-802) with duplex and cruciform DNA. (B) Near UV CD scans to monitor the structural changes in DNA in presence of BRCA1 (502-802)*

The positive maximum around λ=280 nm is attributed to base stacking and the negative peak at 245 nm is attributed to DNA helicity [147]. The increase in ellipticity at 245 and 280 nm suggests unwinding of DNA and/or displacement of base-pairs away from helix axis.

6A.3.4 Tertiary structural study of BRCA1 (502-802) and DNA complex

Fluorescencescans were recorded to see the changes in tertiary structural characteristics of protein and protein DNA complex **(Figure 6A.3)**. Increase in fluorescence intensity in the protein: DNA complex suggests that the protein forms a more compact structure in terms of hydrophobic core when bound to the DNA. No change in emission maximum was observed indicating unaltered hydrophobic environment of the innate fluorophores.

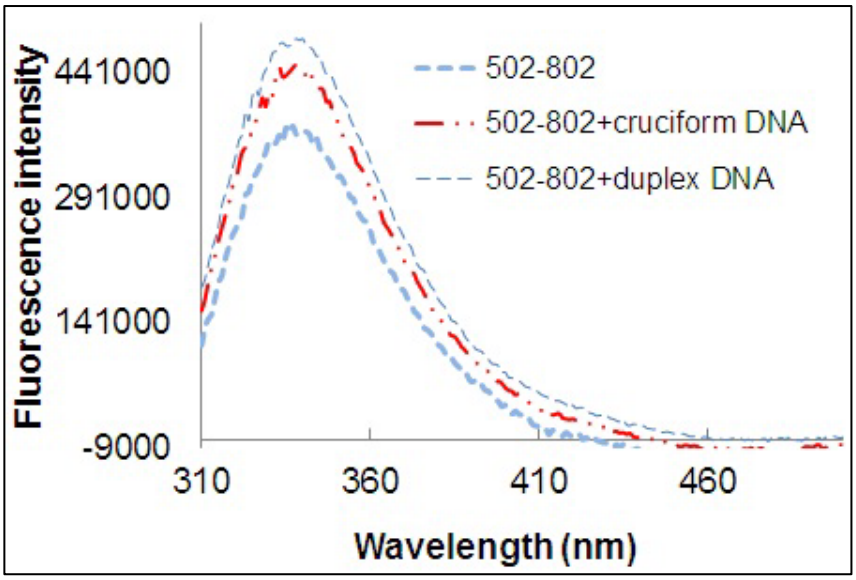


Figure 6 A.3: Fluorescence spectra*: Tertiary structural changes in BRCA1 (502-802) on interaction with duplex and cruciform DNA.*

6A.3.5 BRCA1 (502-802) sensitivity to trypsin in presence of DNA

Comparison of band patterns of protein, with and without DNA, suggests presence of DNA do not enhance the stability of protein (**Figure 6A.4).** This sensitivity of protein in complex with DNA indicates association of DNA does impart structural characteristics that would have protected BRCA1 (502-802) from protease degradation.

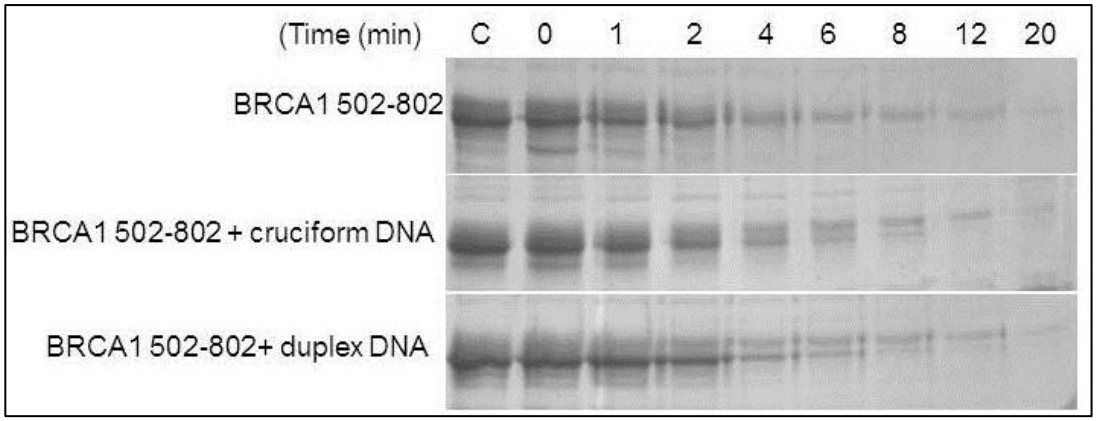


Figure 6A.4: Limited proteolysis assay: *BRCA1 (502-802) and BRCA1 in complex with DNA treated with trypsin at concentration of 1:7500. The protein and complex were incubated at 30^{o}C for different time points of 0, 1, 2, 4, 6, 8, 12, and 20 min.*

6A.3.6 Conformational characterisation of protein: DNA complex

The central region of BRCA1 is reported to bind to different forms of DNA non-specifically. [104,246]. To demonstrate the binding of DNA with BRCA1 (502-802) DNA mobility shift assay with duplex and cruciform DNA was performed. A constant amount of duplex and cruciform DNA was titrated with increasing concentration of protein.

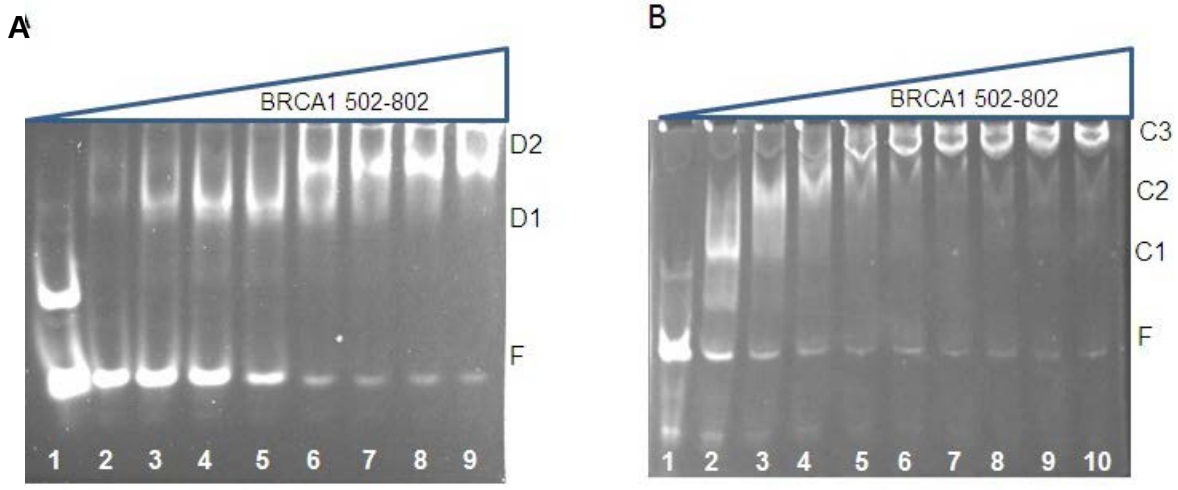


Figure 6A.5*: **Electrophoretic mobility shift assay**: (A) Titration of Duplex DNA with protein. Lane1- Free duplex and cruciform DNA, lane 2-9 has 0.225μM of DNA and increasing amount of protein (0.3-1.2-uM). (B): Titration of cruciform DNA with protein. Lane1- Free cruciform DNA, lane 2-10 has 0.10625 μM of DNA titrated with increasing amount of protein (0.06-0.6-uM).*

Increase in protein concentration resulted in appearance of a band with reduced electrophoretic mobility as compared to free DNA (F) suggesting formation of DNA-protein complex (**Figure 6A.5A & 5B)**. Incubation of BRCA1 (502-802) and cruciform DNA resulted in formation of three types of complexes (C1, C2 and C3) each shifted by different extent on the gel. At low protein concentrations complex I is formed which totally disappears as the protein concentration increases leading to formation of complex II, suggesting co-operativity as reported earlier [104].

A slow migrating C3 complex predominated, when protein concentration was further increased (**Figure 6A.5B).** In the case of interaction with duplex DNA, formations of two shifted bands were observed (D1 and D2). Increase in D1 complex formation was observed with an initial increase in protein concentration. However, after some stage increase in the concentration of protein led to the formation of only slow migrating D2 complex (**6A.5A).**

6A.4 Summary

Studies on the central region of BRCA1 revealed it to be an intrinsically disordered region. It has also been reported by different investigators that BRCA1 interacts with DNA non-specifically. In order to investigate whether interactions with binding partners such as DNA would confer folded structural characteristics, complexes of BRCA1 (502-802) with duplex and cruciform DNA have been studied. These studies revealed an increase in retention volume for BRCA1 (502-802):DNA complex in gel permeation chromatography. Far UV CD scan demonstrated increase in ellipticity at λ=222 nm, and a shift in the peak of negative ellipticity at λ=200 nm to a higher wavelength range, indicating incorporation of secondary structural characteristics. Near UV CD scan demonstrated unwinding of DNA, when bound with BRCA1 protein. This phenomenon could be probably important in unwinding the DNA during the process of repair.

The fluorescence intensity of the BRCA1 (502-802) protein was increased when complexed with DNA, suggesting that the protein forms a more compact structure in terms of hydrophobic core. EMSA study showed two and three different types of complex formation for BRCA1 (502-802) with duplex and cruciform DNA, respectively. Although all the above studies suggest incorporation of some structural component and compaction of protein, no change in susceptibility to protease was observed.

As we are not observing substantial structural alterations in protein on binding to DNA, we presume that unstructured part of BRCA1 left after DNA binding may probably provide space for other binding partners involved in DNA repair and transactivation process, thereby facilitating the role of BRCA1 as a scaffold protein **(Figure 6A.6)**.

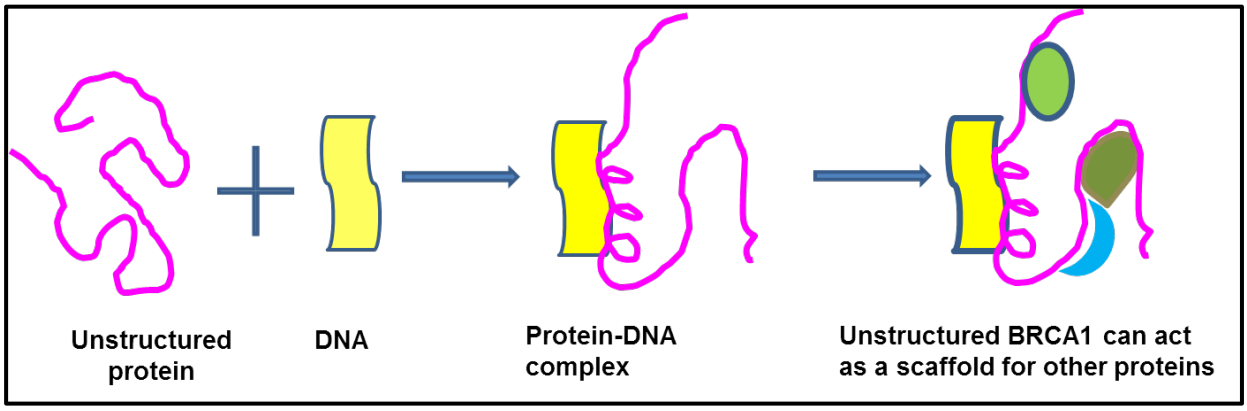


Figure 6A.6: BRCA1 as a scaffold protein: *Cartoon illustrates the unstructured BRCA1 region which binds to DNA and other protein to gain structured form (yellow-DNA, pink-BRCA1 (502-802), green /blue are other interacting protein*

6B: Biophysical characterisation of BRCA1 Glu809Gly and Pro871Leu variants

6B.1 Introduction

Mutations reported in BRCA1 are well recognised for their association with breast and ovarian cancer. Out of the 19 missense variants identified in patients of TMC cohort in our study, BRCA1 E809G and P871L were located in the putative RAD51-binding domain of BRCA1 [58,261]. These mutations may alter the RAD51-binding capacity, and also compromise cell's ability to repair damaged DNA. Though, E809G variant has been reported both in COSMIC (which reports somatic mutations) and ClinVar databases, in ClinVar database, it is reported as a germline variant of uncertain significance [262,263]. To the best of our knowledge no further clinical or functional study is reported till date on the BRCA1 E809G mutant.

The allele frequency of P871L variant, as reported in ClinVar database, is 0.49. This fairly high value characterises this variant as a polymorphic, which is expected not to affect the fitness of the individual. However, certain reports from population based and functional studies present conflicting biological effects for this variant.

A population based study suggests correlation between occurrence of ovarian cancer and existence of P871L variant [264]. Again pathology and family history study of 110 unrelated patients affected by familial breast and/or ovarian cancer, indicated association of P871L with high tumour proliferation [265]. On the other hand, a recent study with proper age and race matched controls in North Carolina, of 312 women with ovarian cancer (76% invasive and 24% borderline) indicated no association of BRCA1 P871L genotype and ovarian cancer risk. It has been suggested that this discrepancy in previous findings might be due to effects of gene-gene or gene-environment interactions e.g. effect of polymorphisms on the penetrance of deleterious BRCA1 mutations, cancer risk in females with high lifetime ovulatory cycles or oral contraceptive use [266]. Altered protein expression of BRCA1 is

known to be responsible for tumor susceptibility. Protein expression is recently shown to be post transcriptionally regulated via micro-RNA:mRNA interactions. Over-expression of miR-638 showed a reduction in BRCA1 P871L protein levels. Interaction of miR-638:BRCA1 P871L mRNA is responsible for this reduction of BRCA1 protein levels, indicating the functional significance of the BRCA1 P871L variant in vivo [267].

A population based case control study, reported the role of BRCA1 P871L in reducing the risk of cervical cancer [268]. Advanced gastric cancer patients with BRCA1 P871L polymorphism, showed a longer progression-free survival and overall survival when treated with taxane and cisplatin [269]. Hence, the dilemma posed with various reports on BRCA1 P871L suggest a need for a detailed investigation and understanding of this variant.

Here, we have attempted to study the variants BRCA1 P871L and E809G using biophysical and *in silico* approach. BRCA1 (758-1064), BRCA1 E809G and P871L were cloned, purified and structurally characterised. Secondary and tertiary structural profiles revealed alteration in structure that leads to increase in the hydrodynamic radius for both the variants. *In silico* study (chapter 3) unanimously classified BRCA1 P871L as neutral variant while BRCA1 E809G was ambiguously classified both as a neutral and a deleterious variant by different servers.

6B.2 Materials and Methods

6B.2.1 Site-Directed Mutagenesis

BRCA1 E809G and P871L variants were incorporated, in BRCA1 (758-1064) construct, in pGEX-kT vector. The forward and reverse primers used for polymerase chain reaction are mentioned in the (**Table 6B.1)**.100 ng of BRCA1 (758-1064) construct, 2.5 mM of each dNTP, 10 pM of forward primer, PCR buffer and Pfu polymerase were mixed in the PCR mix. The samples were pre-heated for 3 minutes at 95°C and then 3 cycles of PCR were performed for 30 s at 95°C, 30 s at 51°C, and 14 minutes at 72°C with only forward primer.

After the 3 cycles, reverse primer was added to the reaction mixture and further 18 cycles of PCR were performed: 30 s at 95°C, 30 s at 51°C, and 14 minutes at 72°C. After the last cycle, the reaction mixture was incubated for an additional 10 minutes at 72°C.

Table 6B.1: Primer used for site-directed mutagenesis of E809G and P871L

Region cloned	**Forward and Reverse primers (5'→3')**
BRCA1 (E809G)	E809G F primer: TGTGCAGCATTTGGAAACCCCAAGGGA E809G R primer: TCCCTTGGGGTTTCCAAATGCTGCACA
BRCA1 (P871L)	BRCA1 P871L F primer: CGCCAGTCATTTGCTCTGTTTTCAAATCCA
	BRCA1 P871L R primer: TGGATTTGAAAACAGAGCAAATGACTGGCG

The PCR products were then digested with Dpn1 to remove methylated template containing the wild type sequence. The digested product was then transformed into the *Escherichia coli* DH5α cells and plated onto the LB-ampicillin plates. Several colonies were then screened for the presence of the variant DNA sequence.

6B.2.2 Purification of (758-1064) E809G variant

Cell pellet obtained after harvesting of 3 litres of cell culture, was resuspended in 100 ml chilled buffer D (50 mM Tris-HCl pH 7.5, 300 mM sodium chloride, 5% glycerol, 20 mM β-ME, 1 mM EDTA, 0.1% Triton X- 100 and protease inhibitor cocktail). Cells were lysed using sonicator with 1 min duty cycle at 60 pulse for 3 times at 4^oC. The cell lysate was centrifuged at 15000 rpm for 45 minutes at 4^oC to remove the insoluble proteins and the cell debris. The soluble protein of interest present in the supernatant was bound for 1-2 hrs to pre-calibrated GST affinity resin. To remove the unbound protein beads were washed with wash buffer (buffer D and 1 M sodium chloride, 20 mM $MgCl_2$, 5 mM ATP). The beads were

washed with 100 ml wash buffer/ 7ml of beads. All the washing was performed by incubating the beads with wash buffer for 5 minutes. Finally beads were washed with 10 ml of lysis buffer to remove any traces of ATP present. On-beads cleavage with TEV protease was performed to get the target protein in the native form. Two hrs after TEV incubation, protein was eluted from the beads with elution buffer containing 50 mM Tris-HCl pH 7.5, 300 mM sodium chloride, 5% glycerol, 20 mM β-ME, 1 mM EDTA, and protease inhibitor cocktail. The eluted fraction is then passed through 700 µl of pre-calibrated Ni-NTA resin to get rid of TEV protease. The size exclusion column was pre calibrated with buffer containing 25 mM Tris-HCl pH 7.5 and 150 mM sodium chloride. The collected eluent is then concentrated to 1.5 ml, and passed through size exclusion chromatography column for purification of protein to homogeneity.

The purification protocol for BRCA1 P871L is same as the protocol used for of its wild type BRCA1 (758-1064) construct mentioned in first part of this chapter.

6B.2.3 Size exclusion chromatography

BRCA1 (758-1064) and its variants BRCA1 E809G and P871L were purified to homogeneity using gel permeation chromatography. The purified proteins were chromatographed on Superdex 200 column. 700 µl of sample were injected and resolved at a flow rate of 0.5 ml/minutes. The elution profile of protein was monitored by absorbance at 280 nm.

6B.2.4 Circular Dichroism

Far UV CD scan at wavelength range of λ=240-195 nm were monitored for BRCA1 (758-1064) (8µM), BRCA1 E809G (10µM) and BRCA1 P871L (10µM). All the measurements were baseline corrected. CD scans were recorded at 20^{o}C, with a scanning speed of 20 nm/min and 3 accumulations were averaged. Protein concentration was measured using Bradford assay and the known molar extinction coefficient. Comparative scans of wild- type

and its variants were plotted as mean residual ellipticity versus wavelength to eliminate the effect of concentration dependent change in ellipticity.

6B.2.5 Fluorescence spectroscopy

The fluorescence scan was measured in the emission wavelength range λ_{em} 310–500 nm, when the sample was excited at 295 nm. The emission slit width was 5 nm. All the data were baseline corrected. Protein concentration measured using BSA as a standard was 3 µM.

6B.3 Results

6B.3.1 Purification of BRCA1 (758-1064) and variants E809G and P871L

BRCA1 (758-1064) was transformed in BL21 (DE3), and its variants (E809G and P871L) were transformed in Rosetta 2 (DE3) for over expression. SDS-PAGE profile at different stages of purification of BRCA1 variants E809G (**Figure 6B.1A)** and P871L (**Figure 6B.1B)** is shown.

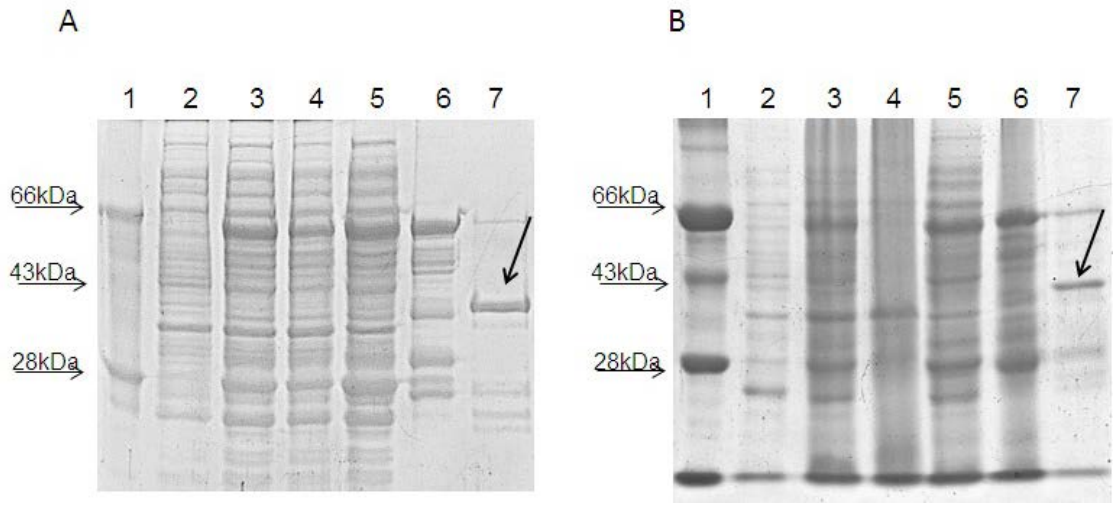


Figure 6B.1: Purification of BRCA1 variants E809G and P871L. (*A) E809G (B) P871L. Lane 1: Marker, 2: Uninduced whole cell, 3: Induced Whole cell, 4: Pellet, 5: Induced Soluble Fraction, 6: Beads after binding, 7: purified elution fraction.*

Protein was analysed for purity using 10% SDS-PAGE. A single protein band at approximately 33kDa was obtained after all purification steps as shown in (**Figure 6B.3).** Two litres of bacterial culture yielded approximately 0.5 mg of BRCA1 758-1064 wild type

and its variant proteins. The solubilities of wild type construct and its variants were almost similar.

6B.3.2 Oligomeric characterisation of BRCA1 (758-1064) and variants E809G and P871L

BRCA1 (758-1064) eluted at approximately 21 ml, while the variants BRCA1 E809G and P871L eluted at approximately 15 ml elution volume in the Superdex 200 size exclusion column (**Figure 6B.2)**. This indicates that the variants have a larger hydrodynamic radius as compared to wild type protein.

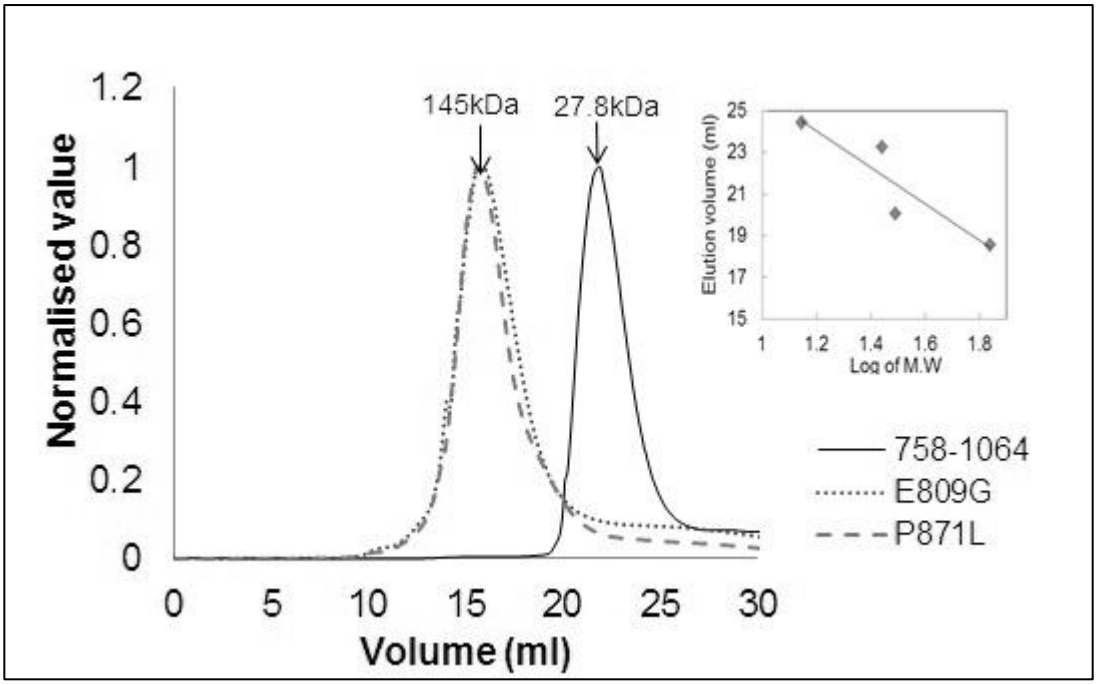


Figure 6B.2: Oligomeric characterization of BRCA1 758-1064 and its variant E809G and P871L. *Size exclusion chromatography profiles of purified proteins. Inset is a plot of log of molecular weight standards versus. elution volumes.*

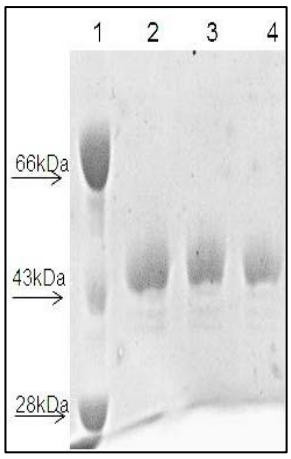


Figure 6B.3: FPLC purified protein loaded on 10% SDS gel.

Lane 1 -marker, 2-BRCA1 758-1064, 3- BRCA1 (E809G), 4- BRCA1 (P871L)

6B.3.3 Structural characterisation of BRCA1 (758-1064) and variants E809G and P871L

To study the secondary structural characteristics of BRCA1 (758-1064) and variants E809G and P871L, far UV CD scan of protein from λ=240-195 nm was monitored (**Figure 6B.4).** The far UV CD scan showed a negative ellipticity at 208 nm and a positive peak at 198 nm, and this is indicative of beta sheet like structure. However the BRCA1 E809G and P871L variants have negative ellipticity at 200 nm, and this indicates a disordered structure.

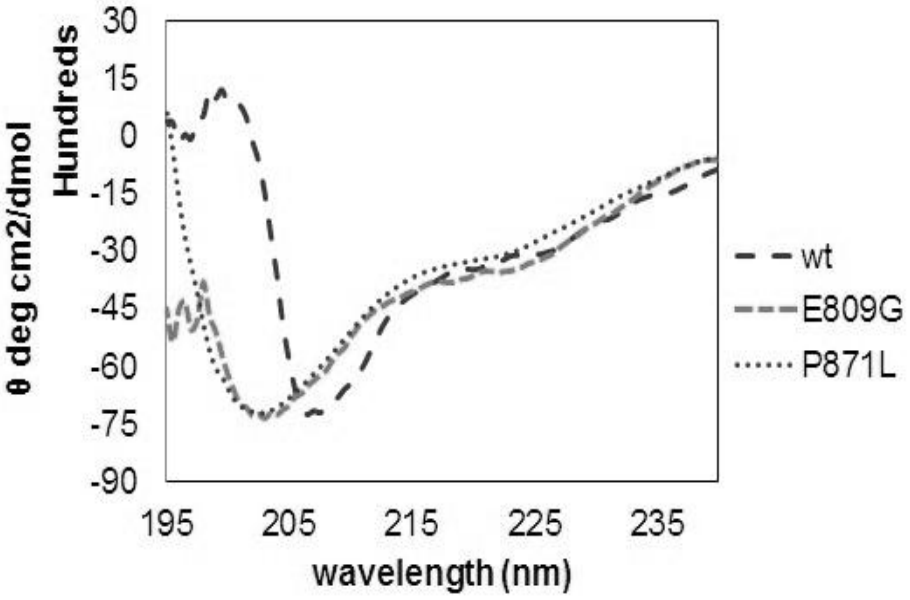


Figure 6B.4: Circular Dichroism spectra of BRCA1 758-1064 and its variants E809G and P871L

Intrinsic fluorescence spectra were recorded from λ=310-500 nm, upon excitation at 295 nm (**Figure 6B.5).** The fluorescence emission maxima for the BRCA1 (758-1064) and its variants BRCA1 E809G and P871L are at 332, 338 and 322 nm respectively. Red shift of 6 nm in BRCA1 E809G variant suggests loss of burial of the residues responsible for the fluorescence signal. On the other hand, in the case of BRCA1 P871L, the blue shift of 10 nm is suggestive of an increase in the burial of the residues responsible for the fluorescence signal

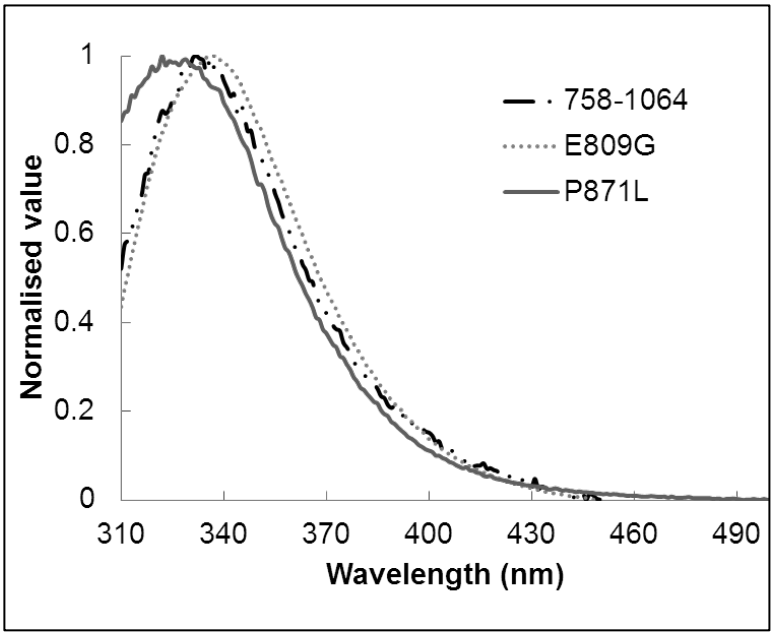


Figure 6B.5: Fluorescence spectra of BRCA1 758-1064 and its variants E809Gand P871L.

6B.4 Summary

Strong inherited predisposition to breast and ovarian cancers in families is attributed to high penetrance of deleterious germ line mutations in BRCA1. Understanding the atomic level detail of multi-functional BRCA1 protein becomes essential, to characterise the variants of uncertain significance. BRCA1 E809G is a missense germ line variant identified in TMC cohort. The proband developed bilateral breast cancer at the age of 42 y and ovarian cancer at 47 y. The pathological study revealed IDC grade III tumour with ER, PR negative breast

cancer. The first and second degree relatives were diagnosed with breast and ovarian cancer and leukaemia at an early age.

Multiple sequence alignment showed that BRCA1 E809 is not a conserved residue, but its high posterior probability value classifies it into IARC class 4, a likely pathogenic variant. Hence, to understand and classify the variant E809G and BRCA1 P871L, biophysical study was performed on the variant classified under likely pathogenic IARC class 4 and not pathogenic class 1 respectively.

BRCA1 E809 residue is present in the central region of BRCA1. The central region of BRCA1 (758-1064) is reported to interact with RAD51 protein. Hence region encoding residues BRCA1 758-1064 was cloned, and variants E809G and P871L were prepared, using site directed mutagenesis. Biophysical study explored the effects of variant E809G and P871L variant. Elution profile in gel permeation chromatography indicates less compact structure for E809G and P871L variant compared to BRCA1 (758-1064) wild type. Far UV CD scan suggests that incorporation of E809G/ P871L variant leads to increase in disorder demonstrated with the negative ellipticity around 200 nm. The emission maxima value indicates that both the variants show opposite effects on tertiary structural property of BRCA1 (758-1064).

In conclusion, central region of BRCA1 (758-1064) demonstrated a change in structure in the variant proteins. Based on the observation of pedigree data and biophysical study it seems that this change is sufficient to alter the normal function of BRCA1 protein. It would be interesting to further investigate the influence of the variants on DNA repair function through cell-based *in vitro* experiments.

Chapter 7

Characterisation of BRCA1 dependent ZBRK1 repressor domain

7.1 Introduction

The majority of Zinc Finger (ZF) proteins belong to the C2H2 class, which contains a consensus amino acid sequence ψ-X-Cys-$X_{(2-4)}$-Cys-X3-ψ-X5-ψ-X2-His-$X_{(3-4)}$-His, where X is any amino acid and ψ is any hydrophobic residue. Zinc Finger (ZF) proteins are known to play a central role in transcriptional regulation, DNA replication and repair, protein translation, cell proliferation and apoptosis [270,271]. C2H2 type containing the KRAB (Kruppel-associated box) domain forms the largest subfamily (KRAB-ZFP) of ZF proteins [272] and, ZBRK1 (Zinc Finger And BRCA1-Interacting Protein with a KRAB Domain) is one member of this family.

Human ZBRK1 consists of 532 amino acids and performs a variety of biological functions through its different domains: the N-terminal KRAB, eight consecutive central C2H2 zinc fingers and a BRCA1-dependent C-terminal transcriptional repression domain (CTRD) [18]. The N-terminal KRAB domain also performs repressor function by interacting with other proteins, such as KAP1. The repressor ZBRK1 binds to a 15 bp consensus nucleotide sequence, GGGXXXCAGXXXTTT (where X is any nucleotide), which is found near promoter regions of many DNA damage inducible genes like p21, ANG1 [273], GADD153 and GADD45 [213,217]. The ZBRK1 region containing residues from 206 to 424 is the zinc finger DNA-binding domain (DBD). ZBRK1 represses cellular invasion and metastasis and, its loss enhances MMP9 transcription in cervical cancer [274]. ZBRK1 represses a high mobility group AT-hook 2 (HMGA2), a DNA architectural protein, activation of which plays a significant role in tumorigenesis and metastasis [275]. Exposure of cells to DNA damaging agents leads to degradation of ZBRK1 via the BRCA1-independent ubiquitin proteasome pathway leading to de-repression [276]. Binding interactions between ZBRK1 with Trim28 and histone deacetylase, results in repression of HIV-1 LTR acting as an intrinsic retroviral defense system. In order to understand the different functions of ZBRK1, we have carried out

structural and binding interactions studies of ZBRK1 with different ligands. The present chapter deals with expression, purification and biophysical characterization of the ZBRK1 Zinc Finger Domain (ZFD) or (ZBRK1 DBD), and binding analyses with zinc and cognate oligonucleotide.

Our results indicate that ZBRK1-DBD domain exists as a homo tetramer in solution and it binds double stranded DNA, even in the absence of zinc ions. The DNA is found to be overwound on binding, while the conformation of the protein is relaxed on DNA binding.

7.2 Materials and methods

7.2.1 Cloning of ZBRK1-DNA binding domain

ZBRK1-DBD (206-424) was sub-cloned at the Bam H1 and HinD III site in pRSET-A vector (Invitrogen). The forward primer has TEV site incorporated in it so as to enable cleavage of the native protein from the His-tagged protein. The forward primer used for amplification of ZBRK1 DBD (206-424) is

5’-GTCTCCGAGAACCTGTACTTTCAGGGTGTGTGCAGTGAATGTGGGAA-3’

and reverse primer used is 5’-GTCAAGCTTCTATTAGTGTATTCTCTTATGCTTAAC-3’. The PCR reaction containing Pfu polymerase, 0.2 mM of each dNTP, 10 pM of both primers, and 100 ng of ZBRK1cDNA were performed in 50 µl of total volume, The samples were preheated for 3 minutes at 95°C and then 32 cycles of PCR were performed under following condition: 30s at 95°C, 30 s at 65°C and 1 minutes at 72°C. After the last cycle, the reaction was incubated for an additional 10 minutes at 72°C. The PCR products were then purified using Sigma gel elute kit to remove excess primers and enzymes. It was then digested with BamH1 and HinD III to generate cohesive ends for ligation into pRSET-A vector. The digested vector and insert were ligated in 1:3 molar ratio using Quick DNA ligase at 16°C for about 30 minutes. The ligated samples were then transformed into *Escherichia coli* DH5α

cells and plated on ampicillin plates. Ampicillin resistant colonies were screened for positive clones and correct DNA sequence.

7.2.2 Protein expression and purification

E. coli bacterial strain Rosetta 2(DE3) was transformed with ZBRK1-DBD construct. Single colony was inoculated in Luria Bertani broth containing ampicillin (100 μg/ml) & chloramphenicol (12 μg/ml), and the culture was allowed to grow over night. This grown culture was then diluted 100 fold and further allowed to grow till the OD_{600} reaches a value between 0.6-0.7. This culture was then induced with 0.4 mM IPTG for 16 h at 18^0C. Cells were harvested & resuspended in the lysis buffer A containing 10 mM HEPES pH 7.0, 500 mM NaCl, 0.1% Triton X-100, 20 mM βME, 1 mM EDTA, 2.5% glycerol, and 10 mM imidazole. Cells were lysed by sonication at a pulse rate of 50, three times with 1 minute's duty cycle. Cell debris was removed by centrifugation at 18,000 rpm for 45 minutes. The induced soluble fraction was allowed to bind on Ni-NTA resin for 1-2 h at 25^oC. To remove the non-specifically bound proteins, resin was washed with 50 mM imidazole made in lysis buffer A. Twenty column volume of wash buffer was used to get rid of the nonspecifically bound contaminant protein. Resin bound protein was eluted with (500 and 750 mM) imidazole made in lysis buffer A at pH 8.0 and further treated with TEV protease to remove histidine tag.

7.2.3 Gel Filtration chromatography

ZBRK1-DBD was further purified to homogeneity using FPLC. TEV-cleaved protein was concentrated up to 1.5 ml using 10 kDa Millipore centricon tubes. This protein was then further centrifuged at 13,000 rpm for 10 min to remove precipitated protein. Concentrated protein was chromatographed using Superdex-75 column, pre equilibrated with buffer (5 mM HEPES pH 7.0, 150 mM NaCl, and 20 mM β-mercaptoethanol). Four proteins (BSA, lysozyme, phosphorylase B, and carbonic anhydrase) were eluted from the same column

under similar conditions to prepare the standard plot for molecular weight estimation. The samples were run at a flow rate of 0.5 ml/ minutes, and 1 ml fractions were collected. FPLC purified protein was further passed through buffer exchange column to remove β-mercaptoethanol that could interfere in other experiments.

7.2.4 Mass Spectrometry

Molecular mass of proteins were determined by using MALDI-MS. Peptide mass fingerprinting was performed for protein identifications. The detailed protocol for mass spectrometry sample preparation has been described in materials and method chapter 2. Trypsin was used for protein digestion. The protein samples were suspended in sinapinic acid matrix and spotted onto the stainless steel plate. MS spectrum obtained was analyzed in Mascot (http://www.matrixscience.com) for peptide mass fingerprinting.

7.2.5 MBP Pull down Assay

MBP-ZBRK1ΔK (amino acid 146-532) and MB`P vector (pMAL c2X) were transformed into Rosetta 2 (DE3) strain. These constructs were then expressed and affinity purified using amylose resin. Purified ZBRK1-DBD was incubated with resin bound MBP-ZBRK1ΔK, and resin bound MBP for 3 h on ice with occasional gentle mixing. Amylose resin was washed with buffer (5 mM HEPES pH 7.0, 150 mM NaCl) to remove non-specifically bound ZBRK1-DBD. The resin was then loaded on to 12% SDS-PAGE gel to check for the pull down results.

7.2.6 Circular Dichroism Spectroscopy

Far-UV CD scan of ZBRK1-DBD (7 µM) was monitored to study secondary structural characteristics. Scan was recorded from λ=240-190 nm at a scanning speed of 20 nm/minutes, data interval of 0.1 nm and 3 accumulations. To study the effect of zinc sulfate and DNA on secondary structural characteristics, ZBRK1-DBD was titrated with zinc sulfate (10-180 µM) and DNA (5-30 µM). To reveal the structural alteration in DNA due to ZBRK1

binding, CD scan was obtained from λ=300-200 nm at a scanning speed of 20 nm/minutes Protein concentration was calculated using OD $_{595}$ nm from Bradford method and a molar extinction coefficient of 12950 M^{-1} cm^{-1}. All the spectra were measured at 25^{o}C and were baseline corrected. α-helical content of the protein was estimated using ellipticity signals at wavelengths of 208 and 222 nm and the standard equation [27,28)

7.2.7 Thermal stability

To measure the thermal stability of ZBRK1-DBD (5 μM), CD spectra in the range λ= 250-200 nm were recorded, as a function of temperature. Far-UV CD scan was monitored at every 5^{o}C increase in temperature, with an equilibration time of 3 minutes and a scanning speed of 50 nm/minutes. Thermal unfolding pathway was also studied to reveal the effect of $ZnSO_4$ (100 μM) and DNA (30 μM) on the stability of ZBRK1-DBD. Ellipticity at λ= 222 nm was plotted as a function of temperature to understand the course of change in secondary structure elements.

7.2.8 Fluorescence spectroscopy

The fluorescence scan was monitored from λ_{em} =310-500 nm, when excited at λ= 295 nm. The emission slit width was 5 nm. Titration experiments with ZBRK1-DBD and different concentrations of DNA (4-60 μM) and $ZnS0_4$ (5-110 μM) were performed to study the tertiary structural alterations [277].

7.2.9 Isothermal Titration calorimetry

Interaction studies of ZBRK1-DBD with $ZnS0_4$ and consensus double stranded DNA oligonucleotide (5'-GGGACGCAGTTTTAT-3') were performed using MicroCal iTC200 calorimeter. To understand the stoichiometry and binding affinity of ZBRK1-DBD, titration experiments were performed under following conditions; (1) 5.2 μM of ZBRK-DBD in cell and 1 mM of ZnS04 in syringe, (2) 20 μM of ZBRK1-DBD in cell and 200 μM of DNA in syringe. Protein concentration was estimated, using BSA as standard. To study the heat

effects from dissociation of ZBRK1 DBD tetramer, ZBRK1-DBD (50 µM) was injected into calorimetric cell containing buffer 5 mM HEPES pH 7.0, 150 mM NaCl. Thirteen injections each of 3 µl, were injected at 25^{o}C into the cell, allowing an equilibration time of 150 seconds between each injections.

7.3 Results

7.3.1 Cloning of ZBRK1 DNA binding domain

The ZBRK1-DBD amplicons (**Figure 7.1A)** were ligated into histidine tag expression vector pRSET-A for affinity purification. The clones obtained from the ligated products were digested with BamH 1 and HinD III to confirm for the presence of insert (**Figure 7.1B)**.

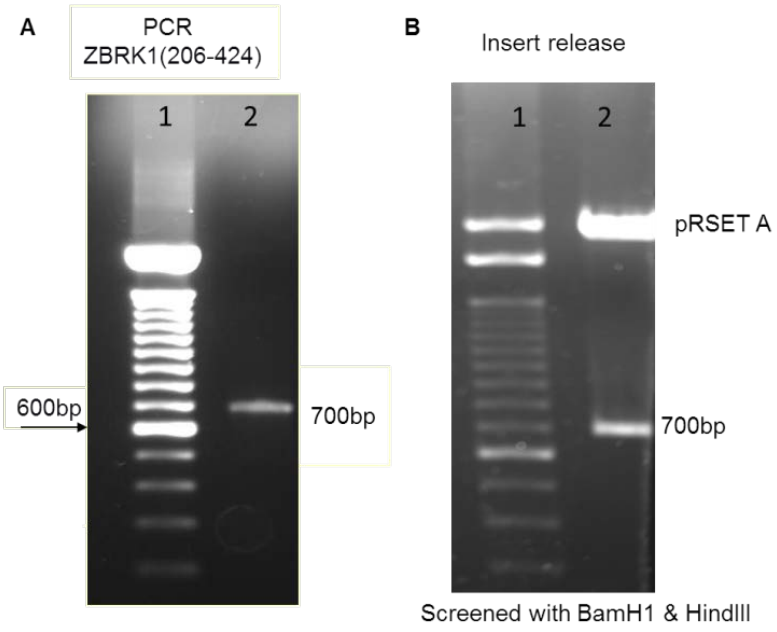


Figure 7.1: Cloning of ZBRK1 DBD*. Lane 1- DNA ruler (A): PCR amplified product, lane 2 - amplicons of 684 bp of ZBRK1 gene, (B): Insert release Lane 2- Screening with BamH1 & HindIII to confirm for clone*

7.3.2 Expression and purification of ZBRK1 DBD

ZBRK1 DNA binding domain encoded by amino acids (206-424) was expressed in Rosetta 2 (DE3) strain and the target recombinant protein with Histidine tag was affinity purified using Ni-NTA resin. Nonspecifically bound proteins on the beads were washed with lysis buffer-A containing 50 mM imidazole. Resin bound protein was eluted with lysis buffer-A containing 500 mM and 750 mM imidazole (**Figure 7.2A).** Final purification of protein was achieved after size exclusion chromatography. FPLC purified protein was further passed through the buffer exchange column to remove β-mercaptoethanol that can interfere in other experiments. Two litres of bacterial culture yielded ~ 0.4 mg of pure ZBRK1-DBD. Protein was analyzed for purity using 10% SDS-PAGE. A single purified band at approximately 24 kDa was obtained after final purification (**Figure 7.2B).**

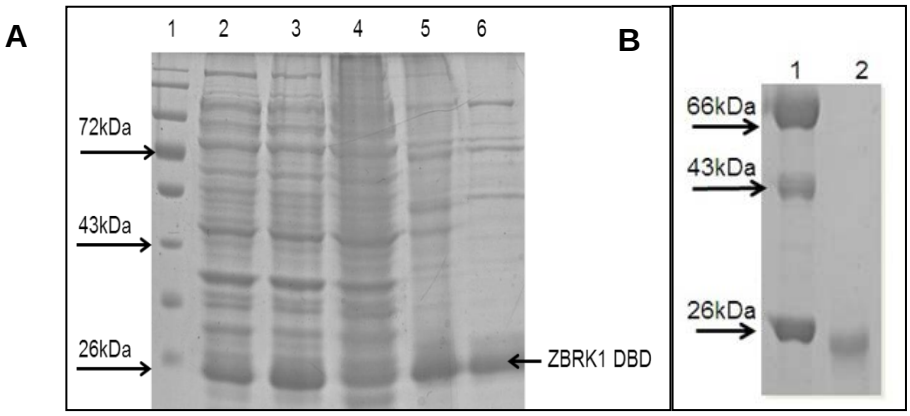


Figure 7.2: Expression and purification of ZBRK1 DNA binding Domain*: (A): 10% SDS gel analysis of ZBRK1 DBD after affinity purification. Lane 1- molecular weight marker, 2-uninduced whole cell, 3-induced whole cell, 4-induced soluble fraction, 5-Ni-NTA bound protein, 6-elution fraction, (B): SDS electrophoresis showing sample purity.*

The molecular weight of ZBRK1 was estimated to be 22.3745 kDa from MALDI-TOF (**Figure 7.3A).** Peptide mass fingerprinting confirmed the presence of 24 kDa band to be indeed, of ZBRK1 (**Figure 7.3B & C).** The search parameters used were mass tolerance of 100 ppm in SwissProt database with sequence coverage of 11% and Score of 72.

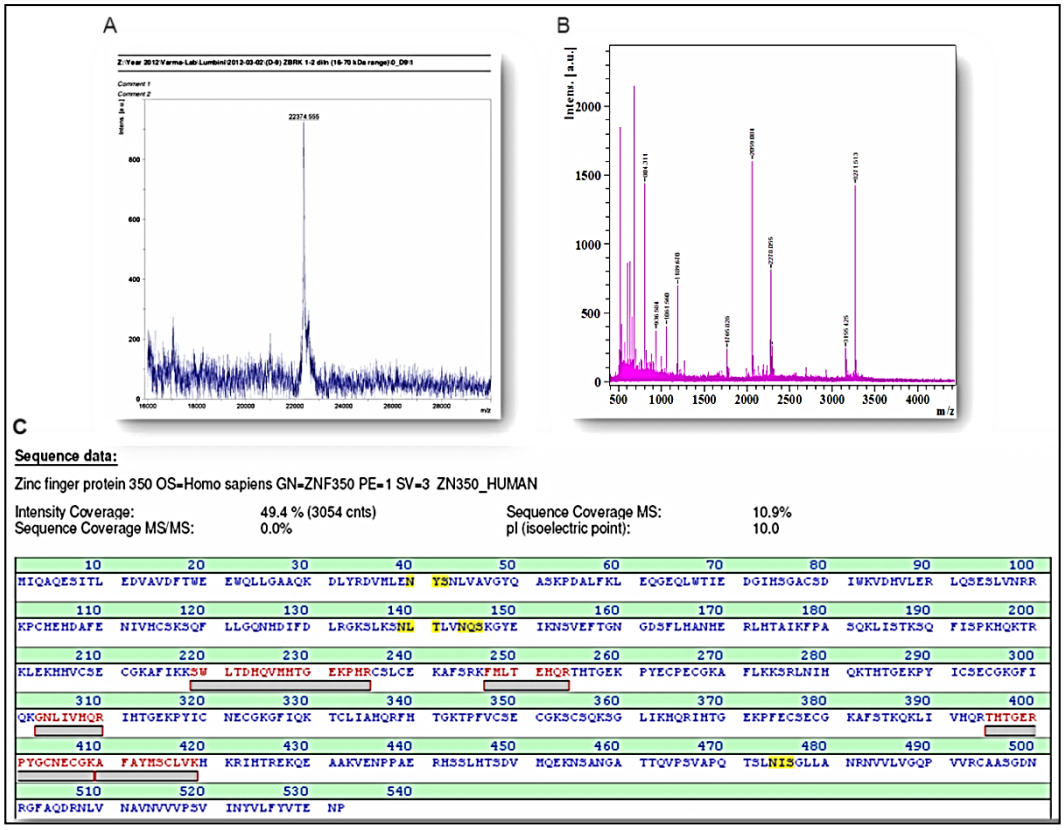


Figure 7.3: Peptide mass fingerprinting: *(A) Molecular weight estimation. (B). Tryptic peptides. (C). Spectral analysis report.*

7.3.3 ZBRK1-DBD forms a tetramer

The elution of ZBRK1-DBD at ~ 58 ml in 16/60 Superdex-75 size exclusion column indicates that the protein forms a higher order oligomer **(Figure 7.4A).** From the standard linear graph of $\log_{10}$ (molecular weight) versus (elution volume) **(Figure 7.4 A inset)**, the size of the eluted sample was estimated to be approximately 94 kDa**.** To explore the ability of self-association of ZBRK1-DBD, MBP pull down assay with MBP tagged ZBRK1ΔK construct was performed. ZBRK1-DBD showed interaction with ZBRK1ΔK construct. No interaction was visible with MBP **(Figure 7.4B).** Thus, based on size exclusion chromatography and pull down assay, it can be concluded that the ZBRK1-DBD forms a homo tetramer in solution.

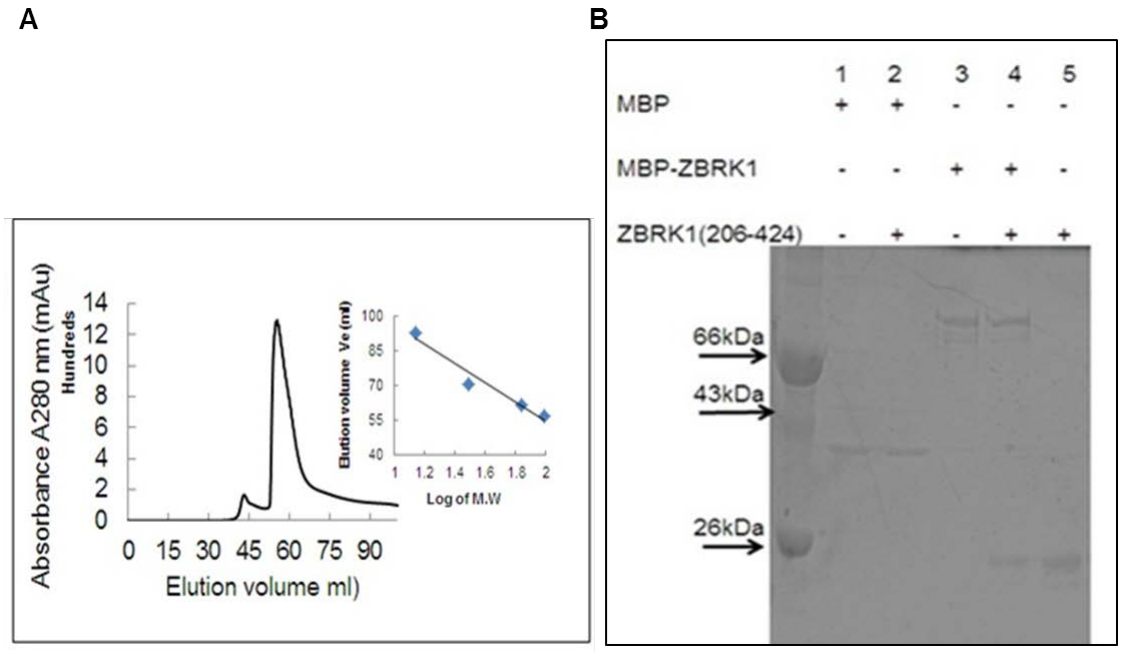


Figure 7.4: Oligomeric characterization of ZBRK1-DBD. *(A) Size exclusion chromatography profile of purified ZBRK1-DBD protein. Inset is a plot of log of molecular weight standards. (B) MBP pull-down (lanes 1, 3, 5-show input), (lane 2, 4 -pull down) samples.*

7.3.4 Dissociation of ZBRK1 DBD tetramer

To further confirm the oligomeric behaviour of ZBRK1 DBD, we have performed dissociation study using ITC.

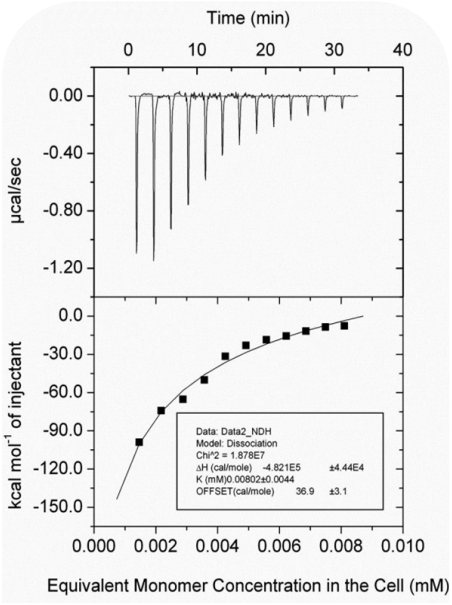


Figure 7.5: *Dissociation study of ZBRK1 DBD.*

Addition of concentrated protein from the syringe into calorimeter cell containing buffer (5 mM HEPES pH 7.0, 150 mM NaCl) resulted in a heat change which gradually decreased as the concentration of the protein inside the cell increased **(Figure 7.5).** This observation suggests that the heat change was due to dissociation of an oligomeric species. The dissociation curve yielded enthalpy of dissociation as -482 kcal/ mol with a dissociation constant of 0.125 μM.

7.3.5 Secondary structure alterations on ligand binding

The CD scan showed peaks at approximately λ=222, 208 and 195 nm, which are characteristic of α- helical conformation [277,278]. The α-helical content of ZBRK1-DBD, estimated using measured ellipticity at λ= 222 and 208 nm, was 12 and 15% respectively. Far- UV CD spectrum of ZBRK1-DBD is shown in (**Figure.7.6A**). The CD spectra of ZBRK1-DBD were recorded also in the presence of varying concentrations of DNA and zinc ions. The ratio $\theta_{222}/\ \theta_{208\ nm}$ of only protein is 1.0, which decreases during titration with DNA, indicating the reduction in coiled coil structure (**Figure.7.6B**). However, interaction with zinc sulfate, the ratio of $\theta_{222}/\ \theta_{208\ nm}$ increases with further addition of ZnS04, indicating the incorporation of quaternary structure (**Figure.7.6C**) [277,279]. (**Figure.7.6D)** shows the CD spectrum over the λ range of 320-230 nm, when DNA was titrated with different amounts of ZBRK1-DBD. The CD spectrum of B-form DNA is composed of four major peaks: negative peaks around λ =214 and 245 nm and positive peaks around λ= 225 and 280 nm [145,146]. The positive maximum peak around λ =280 nm is attributed to base stacking and the negative peak at λ =245 nm is attributed to DNA helicity [147]. The changes in ellipticities at λ =275 nm and λ = 245 nm, on addition of ZBRK1-DBD, indicate that the standard B-geometry of the DNA is being altered on protein binding.

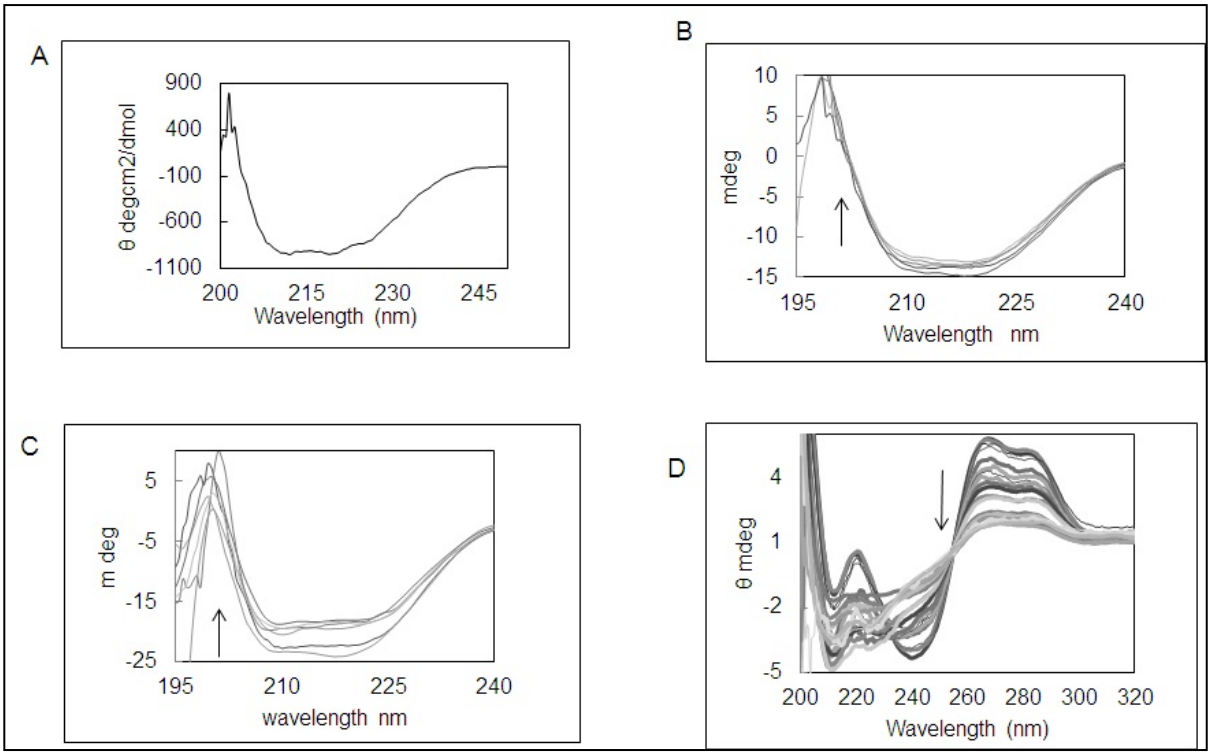


Figure.7.6: Circular Dichroism spectra: *(A) ZBRK1-DBD (B) ZBRK1-DBD titrated with DNA, (C) ZBRK1-DBD titrated with ZnSO4. (D) DNA titrated with ZBRK1-DBD. Arrows indicate directions of decrease in ellipticity.*

7.3.6 Intrinsic tryptophan fluorescence to elucidate local tryptophan environment

Tryptophan emission spectrum was recorded from λ = 310-500 nm, when excited at λ = 295 nm (**Figure 7.7A**). Emission maximum observed at λ =336 nm indicates the presence of single tryptophan residue buried inside of hydrophobic environment [280]. Titration of ZBRK1-DBD with DNA (**Figure 7.7B)** and $ZnSO_4$ **(Figure 7.7C)** showed a decrease in fluorescence intensity without any shift in the position of the maxima. This suggests interactions of ZBRK1-DBD with either zinc sulfate or with DNA does not have any effect on local tryptophan environment

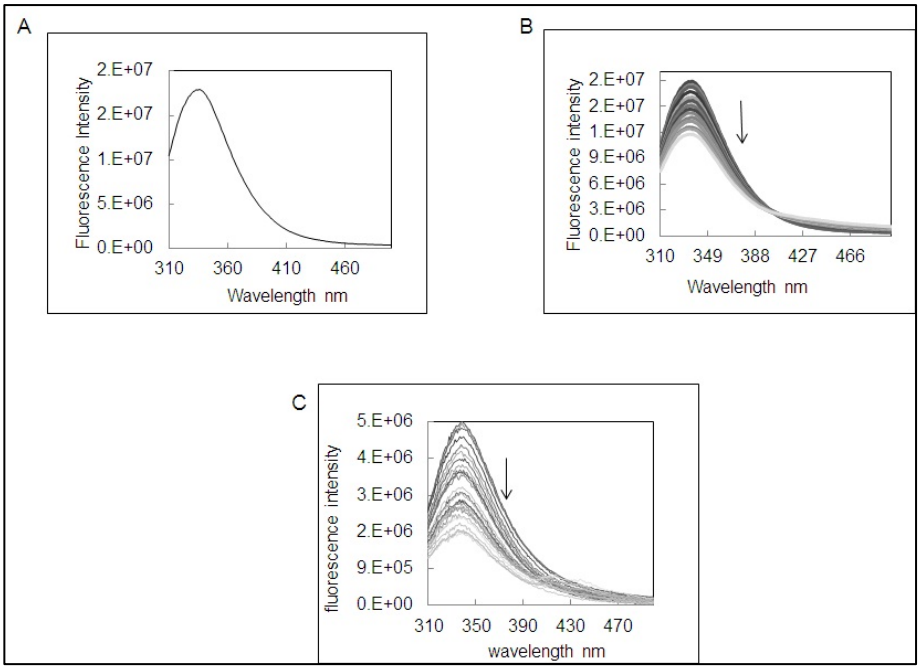


Figure 7.7: Fluorescence spectra*: (A) ZBRK1-DBD. (B) ZBRK1-DBD titrated with DNA. (C) ZBRK1-DBD titrated with ZnSO4. Arrows indicate directions of decrease in fluorescence intensity.*

7.3.7 Thermal stability of ZBRK1-DBD in presence of DNA and $ZnSO_4$

To evaluate the nature of folding and unfolding transitions and to measure the structural stability of ZBRK1-DBD in presence of DNA and $ZnSO_4$, the far UV CD spectra were recorded at different temperatures (**Figure.7.8).** Variations in the ellipticity at λ=222 nm as a function of temperature suggests that ZBRK1-DBD unfolds via three state unfolding transitions with T_m values of 58 and 69^o C. Presence of zinc sulfate or DNA with ZBRK1 increased thermal stability of protein as suggested by increase in melting temperatures from 57^{o}C to 62^{o}C & 68^{o}C respectively. Interestingly, in the presence of either DNA or zinc sulfate, ZBRK1 demonstrated simply a two state unfolding transitions.

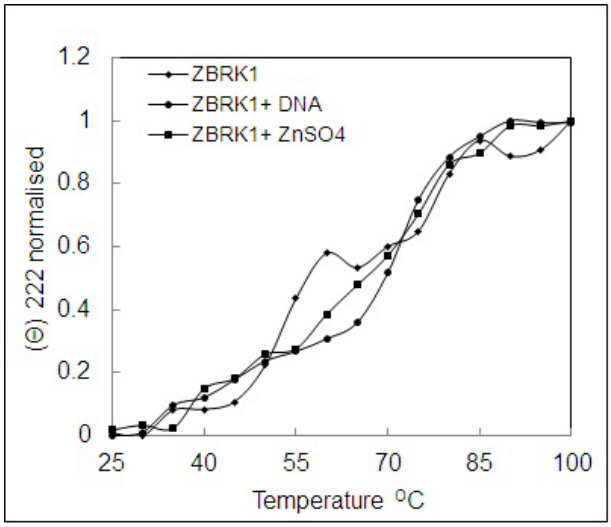


Figure.7.8: Thermal denaturation profile of ZBRK1-DBD*. Plot of fraction unfolded versus temperature at 5°C intervals.*

7.3.8 Thermodynamics of ligand binding to ZBRK1 DBD

Calorimetry experiments have been performed to study the interactions of ZBRK1 with zinc sulfate and DNA (**Figure.7.9)**. Model fitting for ZBRK1-DBD with zinc sulfate with one set of sites showed exothermic reaction with a stoichiometry of 1:42. It demonstrated an affinity of 1.4 μM under these conditions **(Figure.7.9A).**

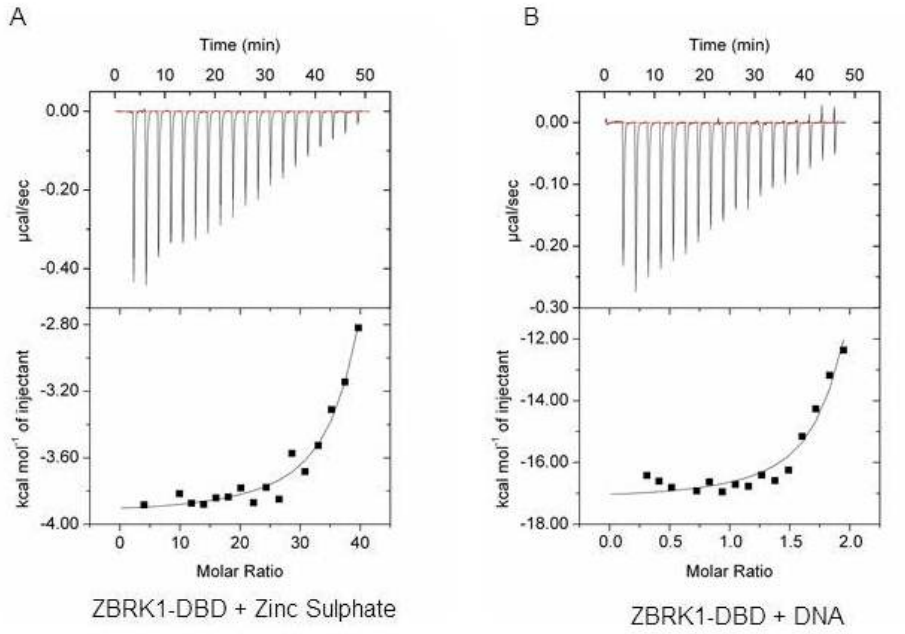


Figure.7.9: Isothermal Titration Calorimetry*: Interaction of ZBRK1-DBD with (A) $ZnSO_4$, (B) DNA.*

Nonlinear curve fitting done, using one set of sites for ZBRK1-DBD with DNA demonstrated a stoichiometry of 1:2 with binding affinity of 0.29 μM **(Figure.7.9B).** ZBRK1-DBD interaction with zinc sulfate and DNA is an enthalpically driven reaction. **Table.7.1** shows the thermodynamic parameters derived from heat change of ZBRK1-DBD interaction with ZnSO4 and DNA on curve fitting in origin software.

Table.7.1: Thermodynamic parameters derived from heat change: *ZBRK1-DBD interaction with ZnSO4 and DNA.*

	$ZnSO_4$	DNA
Interaction model	One set of sites	One set of sites
N sites	42.9 ± 0.0817	2.1 ± 0.0538
K_d (μm)	1.4	0.29
K (M^{-1})	7.06E5 ± 1.91E5	3.44E6 ± 1.26E6,
ΔH (cal/mol)	−3927 ± 29.88	−1.714E4 ± 192.6
ΔS (cal/mol/deg)	13.4	−27.6
Chi^2 value	3510	1.34E + 05

7.4 Summary

ZBRK1 DBD has been cloned, expressed and purified. The purified protein was used for structural and biophysical characterization to study the binding interactions with its potential binding partners. The elution profile of ZBRK1-DBD from the gel filtration column indicates it to be a homo tetramer in solution. MBP pull down assay also revealed the ability of ZBRK1-DBD to oligomerize with ZBRK1ΔK in the MBP tagged ZBRK1ΔK construct. The ITC titration also suggests that dissociation of ZBRK1-DBD is enthalpically favoured. The negative enthalpy change, on dissociation implies that association process will be endothermic and entropically driven. The presence of two transitions in thermal unfolding pathway is also consistent with presence of tetrameric species. The first transition with a Tm of 58^0C would be for the disassembly of the tetramer, while the second transition with a Tm

of 68^0C would be for denaturation of the monomers. The stoichiometry of 1:42 for zinc binding of ZBRK1-DBD, containing eight zinc fingers is also supportive for tetrameric oligomerisation of ZBRK1-DBD. The DNA-binding motif in ZF proteins comprises of two beta strands and an α- helix, and the residues from the helix bind to a triplet of bases in the major groove of the DNA helix [281,282,283]

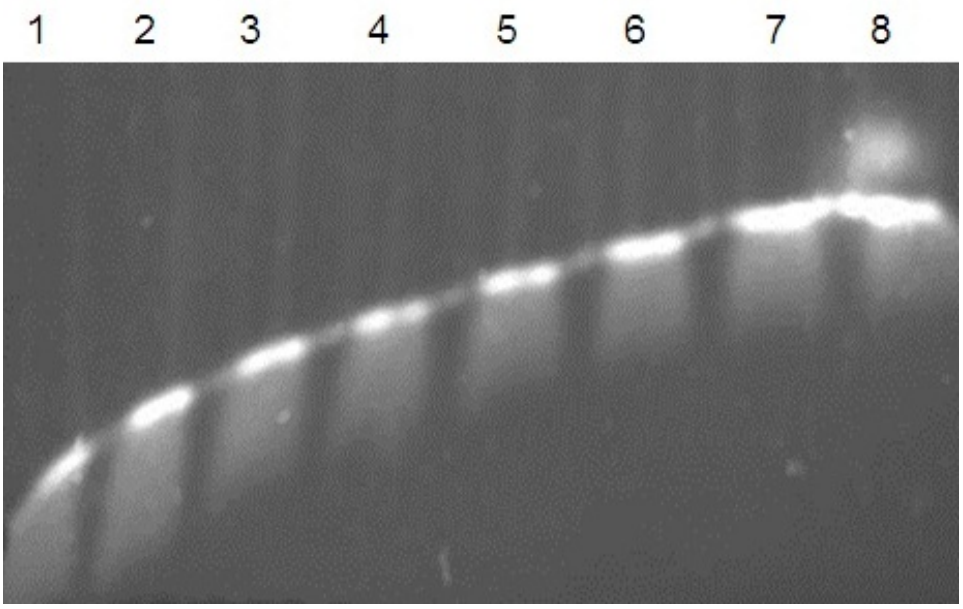


Figure 7.10: Electrophoretic mobility shift assay: *Lane 1: duplex free DNA, lane 2-8: fixed amount of DNA was titrated with ZBRK1-DBD from 0.6 to 1.8μM in absence of zinc ions.*

Formation of the helix and proper positioning of the helical residues for optimal interaction is assumed to be ensured by coordination of zinc ion to conserved cysteine and histidine residues of the motif [284]. Therefore Zn ions are essential for repressor activity of C2H2 type of zinc fingers [285]. Interestingly, in the present study, ZBRK1-DBD binds to 15 bp consensus DNA sequence even in the absence of Zn ions. EMSA (**Figure 7.10)** and ITC (**Figure 7.9A)** studies also revealed binding interactions in absence of Zn ions.

It has been observed that the structure of DNA bound by ZF protein is altered from the standard B-geometry [286]. *In silico* based approach; molecular modeling has revealed that this distortion is necessitated by the requirement for neighbouring zinc fingers to bind

optimally in the major groove. The actual distortions observed in the crystal structures of complex molecules are in agreement with *in silico* based predictions. The relative intensity of CD signal at λ= 275 nm is often taken as a signature for DNA geometry. An increase in its intensity suggests unwinding of DNA and/or displacement of base-pairs away from helix axis. However, in the present study on ZBRK1-DBD: DNA complex, the CD signals at 275 nm is significantly decreased. The geometry of the DNA in the complex is likely to be over wound compared to B-DNA. This probably may help ZBRK1 in repression activity, on interaction with DNA. Although previous reports mention no change occurs in ZF conformation on DNA binding, our results indicate subtle changes. Interestingly, binding of Zn and DNA to ZBRK1-DBD appear to have opposite effects on protein conformation. The n-π* transition at λ=220 nm in CD reflects the α-helical content. The π-π* excitation band at λ= 208 nm is sensitive to whether the α-helix is involved in tertiary contacts [287]. The ratio of $\theta_{222}/\theta_{208\ nm}$ is, therefore, regarded as an indicator of quaternary structure through inter-α-helix coiling/interaction. The binding of ZBRK1-DBD protein to DNA suggests decrease in $\theta_{222}/\theta_{208\ nm}$ ratio indicating decrease in compactness of protein. On the other hand, increase in $\theta_{222}/\theta_{208\ nm}$ ratio for ZBRK1-DBD with Zn-ions suggests incorporation of quaternary structure in presence of Zn as compared to Zn deficient protein.

In conclusion, the results presented here indicate that ZBRK1-DBD forms a tetramer. ZBRK1-DBD binds exclusively to double stranded DNA in the absence of zinc ions. The future goal of this work is to crystallize the protein for understanding the structure, that will help in using it as a drug target and also to engineer it for therapeutic purposes.

Chapter 8

Summary and Conclusion

Breast cancer is one of the leading cancers in women. Along with breast cancer, ovarian cancer also accounts for the majority of cancer deaths in women. Strong inherited predisposition to breast, ovarian and few other cancers is attributed to the high penetrance of deleterious germline mutations discovered in BRCA1/2 genes. BRCA1 is a multi-functional protein involved in maintaining genomic stability, DNA repair in phosphorylation-dependent manner and transcription regulation. BRCA1 interacts with many different cellular partners in executing this variety of functions.

Mutations reported in BRCA1 include truncating mutations (nonsense and frameshift) & missense mutations. Truncating mutations drastically change the primary sequence thereby affecting protein folding necessary for function, whereas missense mutations which substitute a single residue could be either polymorphic with no clinical significance or a jeopardizing variant. Germline missense variants in BRCA1/2 for which the cancer risk has not been determined, are termed as unclassified variants (UCV) or variants of uncertain significance (VUS). In this study aimed at understanding the functionality of BRCA1 at a molecular level, different fragments of BRCA1 were cloned and site-directed mutagenesis were performed for the identified variants. These constructs were expressed in *E.coli* and the purified proteins were used for biophysical and functional studies.

Over thousands of different BRCA1 variants are deposited in BIC database, and of these, 94% of the variants are unclassified variants. The advent of new sequencing technologies may lead to a steep rise in the number of the unclassified variants. Proper and reliable classification of unclassified variants is helpful in counselling cancer patients. Therefore, we have combined clinical, biophysical and computational techniques in an attempt to classify the novel identified variants in Indian HBOC families. The Cancer Genetics Unit, Tata Memorial Centre, Mumbai, has carried out full gene sequencing of 48 samples of BRCA1/2. In this sequence data we have identified 19 missense variants, and of these 19 variants, 10

were already reported in the BIC/ ClinVar database, 5 were synonymous variants and the remaining 4 (BRCA1 C1697Y, S1722P, F1124I and Q210H) were novel unreported variants. The variants BRCA1 S1722P and C1697Y were classified under definitely pathogenic IARC class 5. Mutation at E809G belongs to IARC class 4, a likely pathogenic variant. BRCA1 F1124I, Q210H and other reported variants belong to IARC class 2 and 1, which is the class of little clinical significance. The variants S1722P and C1697Y belonging to definitely pathogenic class impair the transactivation activity of BRCA1. Incorporation of Proline at 1722 position led to a decrease in secondary structure and also distortion in the tertiary structure of the protein. This loss in secondary and tertiary structures affected the binding of the S1722P variant with the phospho-peptide derived from binding partners. BRCA1 C1697Y variant had a deleterious effect on protein folding leading to aggregation and insolubility. *In silico* studies on both these variants demonstrated alterations in weak intramolecular interactions which may affect the structural integrity of the protein. The results presented here for BRCA1 C1697Y and S1722P, provide a rationale for co-segregation of variants with the disease in the HBOC families. The variant present in the RAD51 interacting region of BRCA1, E809G, showed an increase in the hydrodynamic radius of the protein and distortion in its secondary structure. As this variant lies in a region that is intrinsically disordered, mammalian tissue culture-based study is necessary for definitive classification of this variant.

Four central regions of BRCA1 (amino acids 260-553, 341-748, 502-802 and 758-1064) were cloned & purified. Biophysical and *in silico* studies revealed that each of these central regions is intrinsically disordered. All of them show extreme sensitivity to trypsin digestion consistent with the predicted disorder. Bioinformatics analyses considering sequence, structure, amino acid charge, hydrophobicity also supports disorder in all the four constructs. The structure of BRCA1 (502-802) in the presence of its binding partner (cruciform and duplex DNA) was explored using SEC, CD and fluorescence spectroscopy. It was found that

there was a change in the hydrodynamic radius of the protein suggesting the formation of intermolecular complexes. The complexes were found to be of multiple forms when investigated using EMSA. Sensitivity to trypsin digestion, however, did not decrease. Thus, there was only a marginal decrease in the disorder of BRCA1 (502-802) on binding to different types of DNA. These experiments indicate that the central region from amino acid 260-1064 is intrinsically disordered even when complexed with some binding partner molecule such as DNA. The study on the ZBRK1, a binding partner to the central region of BRCA1 was also accomplished. BRCA1 is a co-repressor of ZBRK1: a transcriptional repressor protein that binds to sequence-specific DNA and represses transcription of cell cycle regulatory proteins.

Our study indicates that ZBRK1-DBD forms a homotetramer. It binds exclusively to double-stranded DNA with a stoichiometry of 1:2, even in the absence of zinc ions. The thermal stability of the protein is enhanced on binding to ZnSO4 and DNA. DNA binding led to decrease in compactness of protein as evident from the decrease in the $\theta_{222}/\theta_{208}$ nm ratio in far UV spectra. On the other hand, an increase in $\theta_{222}/\theta_{208}$ nm ratio on titration of protein with zinc suggested incorporation of quaternary structure. In the complex with ZBRK1-DBD, the DNA is distorted from the standard B-geometry.

The C-terminal region of BRCA1 from 1314-1863 is known to be involved in transactivation and DNA repair function. Though the minimal region required for this function is the BRCT domain (1649-1863), its N-terminal extensions are suggested to enhance an efficiency of transactivation. Therefore, segments 1314-1863, 1396-1863, 1560-1859 and 1646-1859 which also correspond to different N-terminal deletion constructs of BRCA1, were expressed and purified from *E. coli* to carry out biophysical studies. Purification was extremely difficult as all these three constructs (1314-1863, 1396-1863 and 1560-1859) were unstable and prone to fast degradation. The experiments indicate that these three constructs were intrinsically

disordered, with order confined only to the 1646-1859 (BRCA1 BRCT) portion. One of the constructs (1560-1859) could be crystallized and its structure has been refined to 3.2 Å resolution. Though cell content analysis indicates the presence of full length construct of molecular weight 33 kDa in the crystals, no electron density was visible for residues upstream of BRCT domain. Purification and structure solution data indicate the interesting possibility that BRCA1 TAD is autolytic; however, this need to be further investigated and confirmed.

While the association of deleterious mutations with the disease is well established, it is not known whether a given VUS missense mutation is only polymorphic or can lead into disease development. Very often the available clinical data on the variant is inadequate to make a definitive classification of the variant. In such situations *in vitro* studies can provide critical information beneficial to predict the potential pathogenicity of missense variants and also to recommend the steps to be taken for correct testing and disease management. In this context, a variety of experimental and theoretical techniques have been employed here to evaluate a number of novel missense mutants discovered in HBOC families in Indian population. The structural properties of two missense mutants have been shown to be altered in a way that affects their ability to bind cellular proteins necessary for proper functioning of BRCA1. The central region of BRCA1 and the region N-terminal to BRCA1 BRCT domain have been shown to be intrinsically disordered and even interaction with the natural binding partner, at least in one case, has not conferred order. This suggests that the intrinsic disorder is a design feature probably to facilitate multi-functionality of BRCA1. A zinc finger protein that binds to the central region of BRCA1, ZBRK1, has been shown to be a homotetramer binding to DNA even in the absence of zinc ions.

Structural and functional analyses are necessary to understand the 3D structure of the protein, and the effect of identified mutations on this structure and how it may hinder the

functionality. X-ray crystal structure determination could be useful in screening of therapeutic compounds capable of restoring normal function in mutant protein structure. Information of structure could be used to design anti-cancer drug molecules using bioinformatics tools. Structural and biophysical analysis could be useful in discriminating between cancer-predisposing mutations and neutral polymorphism. It could also prove useful in the interpretation of the effect of these alterations, even in the absence of clinical data.

References

[1] M.E. Robson, J. Boyd, P.I. Borgen, H.S. Cody, 3rd, Hereditary breast cancer, Curr Probl Surg 38 (2001) 387-480.

[2] N. Rahman, M.R. Stratton, The genetics of breast cancer susceptibility, Annual review of genetics 32 (1998) 95-121.

[3] D.F. Easton, D.T. Bishop, D. Ford, G.P. Crockford, Genetic linkage analysis in familial breast and ovarian cancer: results from 214 families. The Breast Cancer Linkage Consortium, Am J Hum Genet 52 (1993) 678-701.

[4] Y.F. Hu, T. Miyake, Q. Ye, R. Li, Characterization of a novel trans-activation domain of BRCA1 that functions in concert with the BRCA1 C-terminal (BRCT) domain, J Biol Chem 275 (2000) 40910-40915.

[5] J.W. Paterson, BRCA1: a review of structure and putative functions, Dis Markers 13 (1998) 261-274.

[6] J.M. Larner, H. Lee, J.L. Hamlin, S phase damage sensing checkpoints in mammalian cells, Cancer Surv 29 (1997) 25-45.

[7] G.Y. Zhao, E. Sonoda, L.J. Barber, H. Oka, Y. Murakawa, K. Yamada, T. Ikura, X. Wang, M. Kobayashi, K. Yamamoto, S.J. Boulton, S. Takeda, A critical role for the ubiquitin-conjugating enzyme Ubc13 in initiating homologous recombination, Mol Cell 25 (2007) 663-675.

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

[9] J.H. Lee, T.T. Paull, Direct activation of the ATM protein kinase by the Mre11/Rad50/Nbs1 complex, Science 304 (2004) 93-96.

[10] X. Yu, J. Chen, DNA damage-induced cell cycle checkpoint control requires CtIP, a phosphorylation-dependent binding partner of BRCA1 C-terminal domains, Mol Cell Biol 24 (2004) 9478-9486.

[11] B. Xu, S. Kim, M.B. Kastan, Involvement of Brca1 in S-phase and G(2)-phase checkpoints after ionizing irradiation, Mol Cell Biol 21 (2001) 3445-3450.

[12] D.E. Goldgar, D.F. Easton, A.M. Deffenbaugh, A.N. Monteiro, S.V. Tavtigian, F.J. Couch, Integrated evaluation of DNA sequence variants of unknown clinical significance: application to BRCA1 and BRCA2, Am J Hum Genet 75 (2004) 535-544.

[13] S.E. Plon, D.M. Eccles, D. Easton, W.D. Foulkes, M. Genuardi, M.S. Greenblatt, F.B. Hogervorst, N. Hoogerbrugge, A.B. Spurdle, S.V. Tavtigian, Sequence variant classification and reporting: recommendations for improving the interpretation of cancer susceptibility genetic test results, Hum Mutat 29 (2008) 1282-1291.

[14] A.K. Dunker, J.D. Lawson, C.J. Brown, R.M. Williams, P. Romero, J.S. Oh, C.J. Oldfield, A.M. Campen, C.M. Ratliff, K.W. Hipps, J. Ausio, M.S. Nissen, R. Reeves, C. Kang, C.R. Kissinger, R.W. Bailey, M.D. Griswold, W. Chiu, E.C. Garner, Z. Obradovic, Intrinsically disordered protein, J Mol Graph Model 19 (2001) 26-59.

[15] A.K. Dunker, C.J. Brown, Z. Obradovic, Identification and functions of usefully disordered proteins, Adv Protein Chem 62 (2002) 25-49.

[16] V.N. Uversky, J.R. Gillespie, A.L. Fink, Why are “natively unfolded” proteins unstructured under physiologic conditions?, Proteins: Structure, Function, and Bioinformatics 41 (2000) 415-427.

[17] W.Y. Mark, J.C. Liao, Y. Lu, A. Ayed, R. Laister, B. Szymczyna, A. Chakrabartty, C.H. Arrowsmith, Characterization of segments from the central region of BRCA1: an intrinsically disordered scaffold for multiple protein-protein and protein-DNA interactions?, J Mol Biol 345 (2005) 275-287.

[18] L. Zheng, H. Pan, S. Li, A. Flesken-Nikitin, P.L. Chen, T.G. Boyer, W.H. Lee, Sequence-specific transcriptional corepressor function for BRCA1 through a novel zinc finger protein, ZBRK1, Mol Cell 6 (2000) 757-768.

[19] W. Tan, S. Kim, T.G. Boyer, Tetrameric oligomerization mediates transcriptional repression by the BRCA1-dependent Kruppel-associated box-zinc finger protein ZBRK1, J Biol Chem 279 (2004) 55153-55160.

[20] B. Vogelstein, K.W. Kinzler, Cancer genes and the pathways they control, Nat Med 10 (2004) 789-799.

[21] D. Hanahan, R.A. Weinberg, The hallmarks of cancer, Cell 100 (2000) 57-70.

[22] D.J. Slamon, G.M. Clark, S.G. Wong, W.J. Levin, A. Ullrich, W.L. McGuire, Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene, Science 235 (1987) 177-182.

[23] N. Cheng, A. Chytil, Y. Shyr, A. Joly, H.L. Moses, Transforming growth factor-beta signaling-deficient fibroblasts enhance hepatocyte growth factor signaling in mammary carcinoma cells to promote scattering and invasion, Mol Cancer Res 6 (2008) 1521-1533.

[24] N.A. Bhowmick, E.G. Neilson, H.L. Moses, Stromal fibroblasts in cancer initiation and progression, Nature 432 (2004) 332-337.

[25] B. Marinari, F. Moretti, E. Botti, M.L. Giustizieri, P. Descargues, A. Giunta, C. Stolfi, C. Ballaro, M. Papoutsaki, S. Alema, G. Monteleone, S. Chimenti, M. Karin, A. Costanzo, The tumor suppressor activity of IKKalpha in stratified epithelia is exerted in part via the TGF-beta antiproliferative pathway, Proc Natl Acad Sci U S A 105 (2008) 17091-17096.

[26] I. Park, H.K. Son, Z.M. Che, J. Kim, A novel gain-of-function mutation of TGF-beta receptor II promotes cancer progression via delayed receptor internalization in oral squamous cell carcinoma, Cancer Lett 315 (2012) 161-169.

[27] D.L. Vaux, S. Cory, J.M. Adams, Bcl-2 gene promotes haemopoietic cell survival and cooperates with c-myc to immortalize pre-B cells, Nature 335 (1988) 440-442.

[28] S.H. Kaufmann, G.J. Gores, Apoptosis in cancer: cause and cure, Bioessays 22 (2000) 1007-1017.

[29] L. Hayflick, P.S. Moorhead, The serial cultivation of human diploid cell strains, Exp Cell Res 25 (1961) 585-621.

[30] C.Q. Zhu, J.C. Cutz, N. Liu, D. Lau, F.A. Shepherd, J.A. Squire, M.S. Tsao, Amplification of telomerase (hTERT) gene is a poor prognostic marker in non-small-cell lung cancer, Br J Cancer 94 (2006) 1452-1459.

[31] M.J. Karkkainen, T.V. Petrova, Vascular endothelial growth factor receptors in the regulation of angiogenesis and lymphangiogenesis, Oncogene 19 (2000) 5598-5605.

[32] M.B. Sporn, The war on cancer, The Lancet 347 (1996) 1377-1381.

[33] M.G. Vander Heiden, L.C. Cantley, C.B. Thompson, Understanding the Warburg effect: the metabolic requirements of cell proliferation, Science 324 (2009) 1029-1033.

[34] L. Yang, Y. Pang, H.L. Moses, TGF-beta and immune cells: an important regulatory axis in the tumor microenvironment and progression, Trends Immunol 31 (2010) 220-227.

[35] I. Ali, W.A. Wani, K. Saleem, Cancer scenario in India with future perspectives, Cancer Therapy 8 (2011) 56-70.

[36] R. Dikshit, P.C. Gupta, C. Ramasundarahettige, V. Gajalakshmi, L. Aleksandrowicz, R. Badwe, R. Kumar, S. Roy, W. Suraweera, F. Bray, M. Mallath, P.K. Singh, D.N.

Sinha, A.S. Shet, H. Gelband, P. Jha, Cancer mortality in India: a nationally representative survey, Lancet 379 (2012) 1807-1816.

[37] R. Siegel, J. Ma, Z. Zou, A. Jemal, Cancer statistics, 2014, CA Cancer J Clin 64 (2014) 9-29.

[38] J. Paymaster, P. Gangadharan, Some observations on the epidemiology of cancer of the breast in women of western India, International Journal of Cancer 10 (1972) 443-450.

[39] D. Rao, B. Ganesh, P. Desai, Role of reproductive factors in breast cancer in a low-risk area: a case-control study, Br J Cancer 70 (1994) 129.

[40] G. Agarwal, P. Pradeep, V. Aggarwal, C.-H. Yip, P.S. Cheung, Spectrum of breast cancer in Asian women, World J Surg 31 (2007) 1031-1040.

[41] J. Paymaster, P. Gangadharan, Cancer in the Parsi community of Bombay, International Journal of Cancer 5 (1970) 426-431.

[42] M.K. Nair, R. Sankaranarayanan, K.S. Nair, N.S. Amma, C. Varghese, G. Padmakumari, T. Cherian, Overall survival from breast cancer in Kerala, India, in relation to menstrual, reproductive, and clinical factors, Cancer 71 (1993) 1791-1796.

[43] R.P. Dikshit, B.B. Yeole, R. Nagrani, P. Dhillon, R. Badwe, F. Bray, Increase in breast cancer incidence among older women in Mumbai: 30-year trends and predictions to 2025, Cancer Epidemiol 36 (2012) 20.

[44] P.M. Campeau, W.D. Foulkes, M.D. Tischkowitz, Hereditary breast cancer: new genetic developments, new therapeutic avenues, Hum Genet 124 (2008) 31-42.

[45] A.S. Whittemore, G. Gong, J. Itnyre, Prevalence and contribution of BRCA1 mutations in breast cancer and ovarian cancer: results from three U.S. population-based case-control studies of ovarian cancer, Am J Hum Genet 60 (1997) 496-504.

[46] H.A. Risch, J.R. McLaughlin, D.E. Cole, B. Rosen, L. Bradley, I. Fan, J. Tang, S. Li, S. Zhang, P.A. Shaw, S.A. Narod, Population BRCA1 and BRCA2 mutation frequencies

and cancer penetrances: a kin-cohort study in Ontario, Canada, J Natl Cancer Inst 98 (2006) 1694-1706.

[47] Y. Bignon, Y. Fonck, M. Chassagne, Histoprognostic grade in tumours from families with hereditary predisposition to breast cancer, The Lancet 346 (1995) 258.

[48] J.N. Marcus, P. Watson, D.L. Page, S.A. Narod, G.M. Lenoir, P. Tonin, L. Linder-Stephenson, G. Salerno, T.A. Conway, H.T. Lynch, Hereditary breast cancer: pathobiology, prognosis, and BRCA1 and BRCA2 gene linkage, Cancer 77 (1996) 697-709.

[49] J.E. Meza, P.S. Brzovic, M.C. King, R.E. Klevit, Mapping the functional domains of BRCA1. Interaction of the ring finger domains of BRCA1 and BARD1, J Biol Chem 274 (1999) 5659-5665.

[50] R. Kalb, D.L. Mallery, C. Larkin, J.T. Huang, K. Hiom, BRCA1 Is a Histone-H2A-Specific Ubiquitin Ligase, Cell Rep 8 (2014) 999-1005.

[51] L.C. Wu, Z.W. Wang, J.T. Tsan, M.A. Spillman, A. Phung, X.L. Xu, M.C. Yang, L.Y. Hwang, A.M. Bowcock, R. Baer, Identification of a RING protein that can interact in vivo with the BRCA1 gene product, Nat Genet 14 (1996) 430-440.

[52] L. Caleca, A.L. Putignano, M. Colombo, C. Congregati, M. Sarkar, T.J. Magliery, C.B. Ripamonti, C. Foglia, B. Peissel, D. Zaffaroni, S. Manoukian, C. Tondini, M. Barile, V. Pensotti, L. Bernard, L. Papi, P. Radice, Characterization of an Italian founder mutation in the RING-finger domain of BRCA1, PLoS One 9 (2014) e86924.

[53] H. Kim, J. Huang, J. Chen, CCDC98 is a BRCA1-BRCT domain-binding protein involved in the DNA damage response, Nat Struct Mol Biol 14 (2007) 710-715.

[54] I. Drikos, G. Nounesis, C.E. Vorgias, Characterization of cancer-linked BRCA1-BRCT missense variants and their interaction with phosphoprotein targets, Proteins 77 (2009) 464-476.

[55] O.N. Aprelikova, B.S. Fang, E.G. Meissner, S. Cotter, M. Campbell, A. Kuthiala, M. Bessho, R.A. Jensen, E.T. Liu, BRCA1-associated growth arrest is RB-dependent, Proceedings of the National Academy of Sciences 96 (1999) 11866-11871.

[56] Q. Wang, H. Zhang, K. Kajino, M.I. Greene, BRCA1 binds c-Myc and inhibits its transcriptional and transforming activity in cells, Oncogene 17 (1998) 1939-1948.

[57] Q. Zhong, C.F. Chen, S. Li, Y. Chen, C.C. Wang, J. Xiao, P.L. Chen, Z.D. Sharp, W.H. Lee, Association of BRCA1 with the hRad50-hMre11-p95 complex and the DNA damage response, Science 285 (1999) 747-750.

[58] R. Scully, J. Chen, A. Plug, Y. Xiao, D. Weaver, J. Feunteun, T. Ashley, D.M. Livingston, Association of BRCA1 with Rad51 in mitotic and meiotic cells, Cell 88 (1997) 265-275.

[59] D. Cortez, Y. Wang, J. Qin, S.J. Elledge, Requirement of ATM-dependent phosphorylation of brca1 in the DNA damage response to double-strand breaks, Science 286 (1999) 1162-1166.

[60] B. Xu, A.H. O'Donnell, S.T. Kim, M.B. Kastan, Phosphorylation of serine 1387 in Brca1 is specifically required for the Atm-mediated S-phase checkpoint after ionizing irradiation, Cancer Res 62 (2002) 4588-4591.

[61] X. Yu, L.C. Wu, A.M. Bowcock, A. Aronheim, R. Baer, 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 273 (1998) 25388-25392.

[62] Y.F. Hu, R. Li, JunB potentiates function of BRCA1 activation domain 1 (AD1) through a coiled-coil-mediated interaction, Genes Dev 16 (2002) 1509-1517.

[63] Y. Houvras, M. Benezra, H. Zhang, J.J. Manfredi, B.L. Weber, J.D. Licht, BRCA1 physically and functionally interacts with ATF1, J Biol Chem 275 (2000) 36230-36237.

[64] R. Hashizume, M. Fukuda, I. Maeda, H. Nishikawa, D. Oyake, Y. Yabuki, H. Ogata, T. Ohta, The RING heterodimer BRCA1-BARD1 is a ubiquitin ligase inactivated by a breast cancer-derived mutation, J Biol Chem 276 (2001) 14537-14540.

[65] R. Scully, S.F. Anderson, D.M. Chao, W. Wei, L. Ye, R.A. Young, D.M. Livingston, J.D. Parvin, BRCA1 is a component of the RNA polymerase II holoenzyme, Proc Natl Acad Sci U S A 94 (1997) 5605-5610.

[66] S.A. Krum, G.A. Miranda, C. Lin, T.F. Lane, BRCA1 associates with processive RNA polymerase II, J Biol Chem 278 (2003) 52012-52020.

[67] S.F. Anderson, B.P. Schlegel, T. Nakajima, E.S. Wolpin, J.D. Parvin, BRCA1 protein is linked to the RNA polymerase II holoenzyme complex via RNA helicase A, Nat Genet 19 (1998) 254-256.

[68] H. Zhang, K. Somasundaram, Y. Peng, H. Tian, D. Bi, B.L. Weber, W.S. El-Deiry, BRCA1 physically associates with p53 and stimulates its transcriptional activity, Oncogene 16 (1998) 1713-1721.

[69] T. Ouchi, A.N. Monteiro, A. August, S.A. Aaronson, H. Hanafusa, BRCA1 regulates p53-dependent gene expression, Proc Natl Acad Sci U S A 95 (1998) 2302-2306.

[70] T. Ouchi, S.W. Lee, M. Ouchi, S.A. Aaronson, C.M. Horvath, Collaboration of signal transducer and activator of transcription 1 (STAT1) and BRCA1 in differential regulation of IFN-γ target genes, Proceedings of the National Academy of Sciences 97 (2000) 5208-5213.

[71] Q. Wang, H. Zhang, K. Kajino, M.I. Greene, BRCA1 binds c-Myc and inhibits its transcriptional and transforming activity in cells, Oncogene 17 (1998) 1939-1948.

[72] L. Zheng, H. Pan, S. Li, A. Flesken-Nikitin, P.-L. Chen, T.G. Boyer, W.-H. Lee, Sequence-specific transcriptional corepressor function for BRCA1 through a novel zinc finger protein, ZBRK1, Mol Cell 6 (2000) 757-768.

[73] S.B. Cantor, D.W. Bell, S. Ganesan, E.M. Kass, R. Drapkin, S. Grossman, D.C. Wahrer, D.C. Sgroi, W.S. Lane, D.A. Haber, D.M. Livingston, BACH1, a novel helicase-like protein, interacts directly with BRCA1 and contributes to its DNA repair function, Cell 105 (2001) 149-160.

[74] J. Yan, J. Zhu, H. Zhong, Q. Lu, C. Huang, Q. Ye, BRCA1 interacts with FHL2 and enhances FHL2 transactivation function, FEBS Lett 553 (2003) 183-189.

[75] T. Xiang, A. Ohashi, Y. Huang, T.K. Pandita, T. Ludwig, S.N. Powell, Q. Yang, Negative Regulation of AKT Activation by BRCA1, Cancer Res 68 (2008) 10040-10044.

[76] H. Hiraike, O. Wada-Hiraike, S. Nakagawa, S. Koyama, Y. Miyamoto, K. Sone, M. Tanikawa, T. Tsuruga, K. Nagasaka, Y. Matsumoto, K. Oda, K. Shoji, H. Fukuhara, S. Saji, K. Nakagawa, S. Kato, T. Yano, Y. Taketani, Identification of DBC1 as a transcriptional repressor for BRCA1, Br J Cancer 102 (2010) 1061-1067.

[77] D.E. Jensen, M. Proctor, S.T. Marquis, H.P. Gardner, S.I. Ha, L.A. Chodosh, A.M. Ishov, N. Tommerup, H. Vissing, Y. Sekido, BAP1: a novel ubiquitin hydrolase which binds to the BRCA1 RING finger and enhances BRCA1-mediated cell growth suppression, Oncogene 16 (1998) 1097-1112.

[78] H. Wang, N. Shao, Q.M. Ding, J.-q. Cui, E. Reddy, V.N. Rao, BRCA1 proteins are transported to the nucleus in the absence of serum and splice variants BRCA1a, BRCA1b are tyrosine phosphoproteins that associate with E2F, cyclins and cyclin dependent kinases, Oncogene 15 (1997) 143-157.

[79] G. Nadeau, N. Boufaied, A. Moisan, K.M. Lemieux, C. Cayanan, A.N. Monteiro, L. Gaudreau, BRCA1 can stimulate gene transcription by a unique mechanism, EMBO reports 1 (2000) 260-265.

[80] E.A. Williamson, F. Dadmanesh, H.P. Koeffler, BRCA1 transactivates the cyclin-dependent kinase inhibitor p27(Kip1), Oncogene 21 (2002) 3199-3206.

[81] K. Somasundaram, H. Zhang, Y.X. Zeng, Y. Houvras, Y. Peng, G.S. Wu, J.D. Licht, B.L. Weber, W.S. El-Deiry, Arrest of the cell cycle by the tumour-suppressor BRCA1 requires the CDK-inhibitor p21WAF1/CiP1, Nature 389 (1997) 187-190.

[82] E.P. Lamber, A.A. Horwitz, J.D. Parvin, BRCA1 represses amphiregulin gene expression, Cancer Res 70 (2010) 996-1005.

[83] L. Zheng, L.A. Annab, C.A. Afshari, W.H. Lee, T.G. Boyer, BRCA1 mediates ligand-independent transcriptional repression of the estrogen receptor, Proc Natl Acad Sci U S A 98 (2001) 9587-9592.

[84] A.A. Horwitz, B. Affar el, G.F. Heine, Y. Shi, J.D. Parvin, A mechanism for transcriptional repression dependent on the BRCA1 E3 ubiquitin ligase, Proc Natl Acad Sci U S A 104 (2007) 6614-6619.

[85] J.D. Choi, M.A. Park, J.S. Lee, Suppression and recovery of BRCA1-mediated transcription by HP1gamma via modulation of promoter occupancy, Nucleic Acids Res 40 (2012) 11321-11338.

[86] Q. Zhong, C.F. Chen, P.L. Chen, W.H. Lee, BRCA1 facilitates microhomology-mediated end joining of DNA double strand breaks, J Biol Chem 277 (2002) 28641-28647.

[87] Q. Zhong, T.G. Boyer, P.L. Chen, W.H. Lee, Deficient nonhomologous end-joining activity in cell-free extracts from Brca1-null fibroblasts, Cancer Res 62 (2002) 3966-3970.

[88] J.H. Lee, T.T. Paull, Activation and regulation of ATM kinase activity in response to DNA double-strand breaks, Oncogene 26 (2007) 7741-7748.

[89] L. Zou, S.J. Elledge, Sensing DNA damage through ATRIP recognition of RPA-ssDNA complexes, Science 300 (2003) 1542-1548.

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

[91] A. Celeste, S. Petersen, P.J. Romanienko, O. Fernandez-Capetillo, H.T. Chen, O.A. Sedelnikova, B. Reina-San-Martin, V. Coppola, E. Meffre, M.J. Difilippantonio, C. Redon, D.R. Pilch, A. Olaru, M. Eckhaus, R.D. Camerini-Otero, L. Tessarollo, F. Livak, K. Manova, W.M. Bonner, M.C. Nussenzweig, A. Nussenzweig, Genomic instability in mice lacking histone H2AX, Science 296 (2002) 922-927.

[92] M. Stucki, J.A. Clapperton, D. Mohammad, M.B. Yaffe, S.J. Smerdon, S.P. Jackson, MDC1 directly binds phosphorylated histone H2AX to regulate cellular responses to DNA double-strand breaks, Cell 123 (2005) 1213-1226.

[93] H. Kim, J. Chen, X. Yu, Ubiquitin-binding protein RAP80 mediates BRCA1-dependent DNA damage response, Science 316 (2007) 1202-1205.

[94] B. Sobhian, G. Shao, D.R. Lilli, A.C. Culhane, L.A. Moreau, B. Xia, D.M. Livingston, R.A. Greenberg, RAP80 targets BRCA1 to specific ubiquitin structures at DNA damage sites, Science 316 (2007) 1198-1202.

[95] B. Wang, S. Matsuoka, B.A. Ballif, D. Zhang, A. Smogorzewska, S.P. Gygi, S.J. Elledge, Abraxas and RAP80 form a BRCA1 protein complex required for the DNA damage response, Science 316 (2007) 1194-1198.

[96] M. Fabbro, K. Savage, K. Hobson, A.J. Deans, S.N. Powell, G.A. McArthur, K.K. Khanna, BRCA1-BARD1 complexes are required for p53Ser-15 phosphorylation and a G1/S arrest following ionizing radiation-induced DNA damage, J Biol Chem 279 (2004) 31251-31258.

[97] E. Kumaraswamy, R. Shiekhattar, Activation of BRCA1/BRCA2-associated helicase BACH1 is required for timely progression through S phase, Mol Cell Biol 27 (2007) 6733-6741.

[98] R. Scully, S. Ganesan, K. Vlasakova, J. Chen, M. Socolovsky, D.M. Livingston, Genetic analysis of BRCA1 function in a defined tumor cell line, Mol Cell 4 (1999) 1093-1099.

[99] A.J. Davis, L. Chi, S. So, K.J. Lee, E. Mori, K. Fattah, J. Yang, D.J. Chen, BRCA1 modulates the autophosphorylation status of DNA-PKcs in S phase of the cell cycle, Nucleic Acids Res (2014).

[100] B. Xia, Q. Sheng, K. Nakanishi, A. Ohashi, J. Wu, N. Christ, X. Liu, M. Jasin, F.J. Couch, D.M. Livingston, Control of BRCA2 cellular and clinical functions by a nuclear partner, PALB2, Mol Cell 22 (2006) 719-729.

[101] F. Zhang, J. Ma, J. Wu, L. Ye, H. Cai, B. Xia, X. Yu, PALB2 links BRCA1 and BRCA2 in the DNA-damage response, Curr Biol 19 (2009) 524-529.

[102] F. Zhang, Q. Fan, K. Ren, P.R. Andreassen, PALB2 functionally connects the breast cancer susceptibility proteins BRCA1 and BRCA2, Mol Cancer Res 7 (2009) 1110-1118.

[103] N.A. Wallace, K. Robinson, H.L. Howie, D.A. Galloway, beta-HPV 5 and 8 E6 Disrupt Homology Dependent Double Strand Break Repair by Attenuating BRCA1 and BRCA2 Expression and Foci Formation, PLoS Pathog 11 (2015) e1004687.

[104] T.T. Paull, D. Cortez, B. Bowers, S.J. Elledge, M. Gellert, Direct DNA binding by Brca1, Proc Natl Acad Sci U S A 98 (2001) 6086-6091.

[105] Y. Wang, D. Cortez, P. Yazdi, N. Neff, S.J. Elledge, J. Qin, BASC, a super complex of BRCA1-associated proteins involved in the recognition and repair of aberrant DNA structures, Genes Dev 14 (2000) 927-939.

[106] C. Baldeyron, E. Jacquemin, J. Smith, C. Jacquemont, I. De Oliveira, S. Gad, J. Feunteun, D. Stoppa-Lyonnet, D. Papadopoulo, A single mutated BRCA1 allele leads to impaired fidelity of double strand break end-joining, Oncogene 21 (2002) 1401-1410.

[107] D.T. Bau, Y.C. Mau, C.Y. Shen, The role of BRCA1 in non-homologous end-joining, Cancer Lett 240 (2006) 1-8.

[108] Y. Chen, J. Xu, S. Borowicz, C. Collins, D. Huo, O.I. Olopade, c-Myc activates BRCA1 gene expression through distal promoter elements in breast cancer cells, BMC Cancer 11 (2011) 246.

[109] A. De Siervi, P. De Luca, J.S. Byun, L.J. Di, T. Fufa, C.M. Haggerty, E. Vazquez, C. Moiola, D.L. Longo, K. Gardner, Transcriptional autoregulation by BRCA1, Cancer Res 70 (2010) 532-542.

[110] J.C. Rice, H. Ozcelik, P. Maxeiner, I. Andrulis, B.W. Futscher, Methylation of the BRCA1 promoter is associated with decreased BRCA1 mRNA levels in clinical breast cancer specimens, Carcinogenesis 21 (2000) 1761-1765.

[111] R. Naseem, A. Sturdy, D. Finch, T. Jowitt, M. Webb, Mapping and conformational characterization of the DNA-binding region of the breast cancer susceptibility protein BRCA1, Biochem J 395 (2006) 529-535.

[112] L.R. Yadav, S. Rai, M.V. Hosur, A.K. Varma, Functional assessment of intrinsic disorder central domains of BRCA1, J Biomol Struct Dyn (2015) 1-10.

[113] P.B. Mullan, J.E. Quinn, P.M. Gilmore, S. McWilliams, H. Andrews, C. Gervin, N. McCabe, S. McKenna, P. White, Y.H. Song, S. Maheswaran, E. Liu, D.A. Haber, P.G. Johnston, D.P. Harkin, BRCA1 and GADD45 mediated G2/M cell cycle arrest in response to antimicrotubule agents, Oncogene 20 (2001) 6123-6131.

[114] I.A. Manke, D.M. Lowery, A. Nguyen, M.B. Yaffe, BRCT repeats as phosphopeptide-binding modules involved in protein targeting, Science 302 (2003) 636-639.

[115] R.S. Williams, M.S. Lee, D.D. Hau, J.N. Glover, Structural basis of phosphopeptide recognition by the BRCT domain of BRCA1, Nat Struct Mol Biol 11 (2004) 519-525.

[116] J.A. Clapperton, I.A. Manke, D.M. Lowery, T. Ho, L.F. Haire, M.B. Yaffe, S.J. Smerdon, Structure and mechanism of BRCA1 BRCT domain recognition of phosphorylated BACH1 with implications for cancer, Nat Struct Mol Biol 11 (2004) 512-518.

[117] P.S. Brzovic, J.R. Keeffe, H. Nishikawa, K. Miyamoto, D. Fox, 3rd, M. Fukuda, T. Ohta, R. Klevit, Binding and recognition in the assembly of an active BRCA1/BARD1 ubiquitin-ligase complex, Proc Natl Acad Sci U S A 100 (2003) 5646-5651.

[118] P.S. Brzovic, P. Rajagopal, D.W. Hoyt, M.C. King, R.E. Klevit, Structure of a BRCA1-BARD1 heterodimeric RING-RING complex, Nat Struct Biol 8 (2001) 833-837.

[119] D.E. Anderson, A genetic study of human breast cancer, J Natl Cancer Inst 48 (1972) 1029-1034.

[120] B.Y. Hwang, N. Varadarajan, H. Li, S. Rodriguez, B.L. Iverson, G. Georgiou, Substrate specificity of the Escherichia coli outer membrane protease OmpP, J Bacteriol 189 (2007) 522-530.

[121] I. Chen, D. Dubnau, DNA uptake during bacterial transformation, Nature Reviews Microbiology 2 (2004) 241-249.

[122] J.M. Solomon, A.D. Grossman, Who's competent and when: regulation of natural genetic competence in bacteria, Trends Genet 12 (1996) 150-155.

[123] J.C. Samuelson, Recent developments in difficult protein expression: a guide to E. coli strains, promoters, and relevant host mutations, Heterologous Gene Expression in E. coli, Springer, 2011, pp. 195-209.

[124] J. Queiroz, C. Tomaz, J. Cabral, Hydrophobic interaction chromatography of proteins, J Biotechnol 87 (2001) 143-159.

[125] B. Ketterer, A bird's eye view of the glutathione transferase field, Chem Biol Interact 138 (2001) 27-42.

[126] K.L. Guan, J.E. Dixon, Eukaryotic proteins expressed in Escherichia coli: an improved thrombin cleavage and purification procedure of fusion proteins with glutathione S-transferase, Anal Biochem 192 (1991) 262-267.

[127] D.B. Smith, K.S. Johnson, Single-step purification of polypeptides expressed in Escherichia coli as fusions with glutathione S-transferase, Gene 67 (1988) 31-40.

[128] R.C. Cheung, J.H. Wong, T.B. Ng, Immobilized metal ion affinity chromatography: a review on its applications, Appl Microbiol Biotechnol 96 (2012) 1411-1420.

[129] H. Block, B. Maertens, A. Spriestersbach, N. Brinker, J. Kubicek, R. Fabis, J. Labahn, F. Schäfer, Immobilized-metal affinity chromatography (IMAC): a review, Methods in enzymology 463 (2009) 439-473.

[130] E.K. Ueda, P.W. Gout, L. Morganti, Current and prospective applications of metal ion-protein binding, J Chromatogr A 988 (2003) 1-23.

[131] U.K. Laemmli, Cleavage of structural proteins during the assembly of the head of bacteriophage T4, Nature 227 (1970) 680-685.

[132] (!!! INVALID CITATION !!!).

[133] K. Gevaert, J. Vandekerckhove, Protein identification methods in proteomics, Electrophoresis 21 (2000) 1145-1154.

[134] L.F. Marvin, M.A. Roberts, L.B. Fay, Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry in clinical chemistry, Clinica chimica acta 337 (2003) 11-21.

[135] A. Bizzini, G. Greub, Matrix-assisted laser desorption ionization time-of-flight mass spectrometry, a revolution in clinical microbial identification, Clinical Microbiology and infection 16 (2010) 1614-1619.

[136] P.M. Gehrig, A. Krzyzaniak, J. Barciszewski, K. Biemann, Mass spectrometric amino acid sequencing of a mixture of seed storage proteins (napin) from Brassica napus, products of a multigene family, Proc Natl Acad Sci U S A 93 (1996) 3647-3652.

[137] W. Hendrickson, Protein-DNA interactions studied by the gel electrophoresis-DNA binding assay, Biotechniques 3 (1985) 198-207.

[138] M.J. Guille, G.G. Kneale, Methods for the analysis of DNA-protein interactions, Mol Biotechnol 8 (1997) 35-52.

[139] M.G. Fried, D.M. Crothers, Kinetics and mechanism in the reaction of gene regulatory proteins with DNA, J Mol Biol 172 (1984) 263-282.

[140] M.W. Freyer, E.A. Lewis, Isothermal titration calorimetry: experimental design, data analysis, and probing macromolecule/ligand binding and kinetic interactions, Methods in cell biology 84 (2008) 79-113.

[141] M.M. Pierce, C.S. Raman, B.T. Nall, Isothermal titration calorimetry of protein-protein interactions, Methods 19 (1999) 213-221.

[142] M.L. Doyle, Characterization of binding interactions by isothermal titration calorimetry, Curr Opin Biotechnol 8 (1997) 31-35.

[143] A. Velazquez-Campoy, S.A. Leavitt, E. Freire, Characterization of protein-protein interactions by isothermal titration calorimetry, Protein-Protein Interactions, Springer, 2004, pp. 35-54.

[144] S. Leavitt, E. Freire, Direct measurement of protein binding energetics by isothermal titration calorimetry, Current opinion in structural biology 11 (2001) 560-566.

[145] M. Vorlickova, Conformational transitions of alternating purine-pyrimidine DNAs in perchlorate ethanol solutions, Biophys J 69 (1995) 2033-2043.

[146] J. Kypr, M. Vorlickova, Circular dichroism spectroscopy reveals invariant conformation of guanine runs in DNA, Biopolymers 67 (2002) 275-277.

[147] C.A. Sprecher, W.A. Baase, W.C. Johnson, Jr., Conformation and circular dichroism of DNA, Biopolymers 18 (1979) 1009-1019.

[148] S.M. Kelly, N.C. Price, The use of circular dichroism in the investigation of protein structure and function, Current protein and peptide science 1 (2000) 349-384.

[149] Y.-M. Chang, C.K.-M. Chen, M.-H. Hou, Conformational changes in DNA upon ligand binding monitored by circular dichroism, Int J Mol Sci 13 (2012) 3394-3413.

[150] D.H. Corrêa, C.H. Ramos, The use of circular dichroism spectroscopy to study protein folding, form and function, African Journal of Biochemistry Research 3 (2009) 164-173.

[151] J.T. Vivian, P.R. Callis, Mechanisms of tryptophan fluorescence shifts in proteins, Biophys J 80 (2001) 2093-2109.

[152] M. Möller, A. Denicola, Protein tryptophan accessibility studied by fluorescence quenching, Biochemistry and Molecular Biology Education 30 (2002) 175-178.

[153] C. Duy, J. Fitter, How aggregation and conformational scrambling of unfolded states govern fluorescence emission spectra, Biophys J 90 (2006) 3704-3711.

[154] A. Hawe, M. Sutter, W. Jiskoot, Extrinsic fluorescent dyes as tools for protein characterization, Pharm Res 25 (2008) 1487-1499.

[155] S.V. Tavtigian, A.M. Deffenbaugh, L. Yin, T. Judkins, T. Scholl, P.B. Samollow, D. de Silva, A. Zharkikh, A. Thomas, Comprehensive statistical study of 452 BRCA1

missense substitutions with classification of eight recurrent substitutions as neutral, J Med Genet 43 (2006) 295-305.

[156] P.C. Ng, S. Henikoff, SIFT: Predicting amino acid changes that affect protein function, Nucleic Acids Res 31 (2003) 3812-3814.

[157] R.D. Finn, J. Mistry, J. Tate, P. Coggill, A. Heger, J.E. Pollington, O.L. Gavin, P. Gunasekaran, G. Ceric, K. Forslund, L. Holm, E.L. Sonnhammer, S.R. Eddy, A. Bateman, The Pfam protein families database, Nucleic Acids Res 38 (2010) D211-222.

[158] B. Li, V.G. Krishnan, M.E. Mort, F. Xin, K.K. Kamati, D.N. Cooper, S.D. Mooney, P. Radivojac, Automated inference of molecular mechanisms of disease from amino acid substitutions, Bioinformatics 25 (2009) 2744-2750.

[159] E. Capriotti, P. Fariselli, R. Calabrese, R. Casadio, Predicting protein stability changes from sequences using support vector machines, BIOINFORMATICS-OXFORD- 21 (2005) ii54.

[160] I.A. Adzhubei, S. Schmidt, L. Peshkin, V.E. Ramensky, A. Gerasimova, P. Bork, A.S. Kondrashov, S.R. Sunyaev, A method and server for predicting damaging missense mutations, Nat Methods 7 (2010) 248-249.

[161] P. Kumar, S. Henikoff, P.C. Ng, Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm, Nat Protoc 4 (2009) 1073-1081.

[162] S.F. Altschul, T.L. Madden, A.A. Schaffer, J. Zhang, Z. Zhang, W. Miller, D.J. Lipman, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, Nucleic Acids Res 25 (1997) 3389-3402.

[163] T. Kawabata, M. Ota, K. Nishikawa, The Protein Mutant Database, Nucleic Acids Res 27 (1999) 355-357.

[164] Y. Bromberg, B. Rost, SNAP: predict effect of non-synonymous polymorphisms on function, Nucleic Acids Res 35 (2007) 3823-3835.

[165] L.J. McGuffin, K. Bryson, D.T. Jones, The PSIPRED protein structure prediction server, Bioinformatics 16 (2000) 404-405.

[166] R. Adamczak, A. Porollo, J. Meller, Accurate prediction of solvent accessibility using neural networks–based regression, Proteins: Structure, Function, and Bioinformatics 56 (2004) 753-767.

[167] D.F. Easton, D. Ford, D.T. Bishop, Breast and ovarian cancer incidence in BRCA1-mutation carriers. Breast Cancer Linkage Consortium, Am J Hum Genet 56 (1995) 265-271.

[168] D. Ford, D.F. Easton, D.T. Bishop, S.A. Narod, D.E. Goldgar, Risks of cancer in BRCA1-mutation carriers. Breast Cancer Linkage Consortium, Lancet 343 (1994) 692-695.

[169] A.C. Antoniou, A.P. Cunningham, J. Peto, D.G. Evans, F. Lalloo, S.A. Narod, H.A. Risch, J.E. Eyfjord, J.L. Hopper, M.C. Southey, H. Olsson, O. Johannsson, A. Borg, B. Pasini, P. Radice, S. Manoukian, D.M. Eccles, N. Tang, E. Olah, H. Anton-Culver, E. Warner, J. Lubinski, J. Gronwald, B. Gorski, L. Tryggvadottir, K. Syrjakoski, O.P. Kallioniemi, H. Eerola, H. Nevanlinna, P.D. Pharoah, D.F. Easton, The BOADICEA model of genetic susceptibility to breast and ovarian cancers: updates and extensions, Br J Cancer 98 (2008) 1457-1466.

[170] J. Greenman, S. Mohammed, D. Ellis, S. Watts, G. Scott, L. Izatt, D. Barnes, E. Solomon, S. Hodgson, C. Mathew, Identification of missense and truncating mutations in the BRCA1 gene in sporadic and familial breast and ovarian cancer, Genes Chromosomes Cancer 21 (1998) 244-249.

[171] G. Parmigiani, D. Berry, O. Aguilar, Determining carrier probabilities for breast cancer-susceptibility genes BRCA1 and BRCA2, Am J Hum Genet 62 (1998) 145-158.

[172] G.M. Petersen, G. Parmigiani, D. Thomas, Missense mutations in disease genes: a Bayesian approach to evaluate causality, Am J Hum Genet 62 (1998) 1516-1524.

[173] D. Thompson, D.F. Easton, D.E. Goldgar, A full-likelihood method for the evaluation of causality of sequence variants from family data, Am J Hum Genet 73 (2003) 652-655.

[174] S.R. Lakhani, J. Jacquemier, J.P. Sloane, B.A. Gusterson, T.J. Anderson, M.J. van de Vijver, L.M. Farid, D. Venter, A. Antoniou, A. Storfer-Isser, E. Smyth, C.M. Steel, N. Haites, R.J. Scott, D. Goldgar, S. Neuhausen, P.A. Daly, W. Ormiston, R. McManus, S. Scherneck, B.A. Ponder, D. Ford, J. Peto, D. Stoppa-Lyonnet, Y.J. Bignon, J.P. Struewing, N.K. Spurr, D.T. Bishop, J.G. Klijn, P. Devilee, C.J. Cornelisse, C. Lasset, G. Lenoir, R.B. Barkardottir, V. Egilsson, U. Hamann, J. Chang-Claude, H. Sobol, B. Weber, M.R. Stratton, D.F. Easton, Multifactorial analysis of differences between sporadic breast cancers and cancers involving BRCA1 and BRCA2 mutations, J Natl Cancer Inst 90 (1998) 1138-1145.

[175] J. Vallon-Christersson, C. Cayanan, K. Haraldsson, N. Loman, J.T. Bergthorsson, K. Brondum-Nielsen, A.M. Gerdes, P. Moller, U. Kristoffersson, H. Olsson, A. Borg, A.N. Monteiro, Functional analysis of BRCA1 C-terminal missense mutations identified in breast and ovarian cancer families, Hum Mol Genet 10 (2001) 353-360.

[176] M. Tischkowitz, N. Hamel, M.A. Carvalho, G. Birrane, A. Soni, E.H. van Beers, S.A. Joosse, N. Wong, D. Novak, L.A. Quenneville, S.A. Grist, P.M. Nederlof, D.E. Goldgar, S.V. Tavtigian, A.N. Monteiro, J.A. Ladias, W.D. Foulkes, Pathogenicity of the BRCA1 missense variant M1775K is determined by the disruption of the BRCT

phosphopeptide-binding pocket: a multi-modal approach, Eur J Hum Genet 16 (2008) 820-832.

[177] P.J. Rowling, R. Cook, L.S. Itzhaki, Toward classification of BRCA1 missense variants using a biophysical approach, J Biol Chem 285 (2010) 20080-20087.

[178] P. Radice, S. De Summa, L. Caleca, S. Tommasi, Unclassified variants in BRCA genes: guidelines for interpretation, Annals of oncology 22 (2011) i18-i23.

[179] D.F. Easton, A.M. Deffenbaugh, D. Pruss, C. Frye, R.J. Wenstrup, K. Allen-Brady, S.V. Tavtigian, A.N. Monteiro, E.S. Iversen, F.J. Couch, D.E. Goldgar, A systematic genetic assessment of 1,433 sequence variants of unknown clinical significance in the BRCA1 and BRCA2 breast cancer-predisposition genes, Am J Hum Genet 81 (2007) 873-883.

[180] P.K. Lovelock, A.B. Spurdle, M.T. Mok, D.J. Farrugia, S.R. Lakhani, S. Healey, S. Arnold, D. Buchanan, F.J. Couch, B.R. Henderson, D.E. Goldgar, S.V. Tavtigian, G. Chenevix-Trench, M.A. Brown, Identification of BRCA1 missense substitutions that confer partial functional activity: potential moderate risk variants?, Breast Cancer Res 9 (2007) R82.

[181] E.S. Iversen, Jr., F.J. Couch, D.E. Goldgar, S.V. Tavtigian, A.N. Monteiro, A computational method to classify variants of uncertain significance using functional assay data with application to BRCA1, Cancer Epidemiol Biomarkers Prev 20 (2011) 1078-1088.

[182] A.B. Spurdle, S. Healey, A. Devereau, F.B. Hogervorst, A.N. Monteiro, K.L. Nathanson, P. Radice, D. Stoppa-Lyonnet, S. Tavtigian, B. Wappenschmidt, F.J. Couch, D.E. Goldgar, ENIGMA--evidence-based network for the interpretation of germline mutant alleles: an international initiative to evaluate risk and clinical

significance associated with sequence variation in BRCA1 and BRCA2 genes, Hum Mutat 33 (2012) 2-7.

[183] L. Mohammadi, M.P. Vreeswijk, R. Oldenburg, A. van den Ouweland, J.C. Oosterwijk, A.H. van der Hout, N. Hoogerbrugge, M. Ligtenberg, M.G. Ausems, R.B. van der Luijt, C.J. Dommering, J.J. Gille, S. Verhoef, F.B. Hogervorst, T.A. van Os, E. Gomez Garcia, M.J. Blok, J.T. Wijnen, Q. Helmer, P. Devilee, C.J. van Asperen, H.C. van Houwelingen, A simple method for co-segregation analysis to evaluate the pathogenicity of unclassified variants; BRCA1 and BRCA2 as an example, BMC Cancer 9 (2009) 211.

[184] N.M. Lindor, L. Guidugli, X. Wang, M.P. Vallee, A.N. Monteiro, S. Tavtigian, D.E. Goldgar, F.J. Couch, A review of a multifactorial probability-based model for classification of BRCA1 and BRCA2 variants of uncertain significance (VUS), Hum Mutat 33 (2012) 8-21.

[185] G. Chenevix-Trench, S. Healey, S. Lakhani, P. Waring, M. Cummings, R. Brinkworth, A.M. Deffenbaugh, L.A. Burbidge, D. Pruss, T. Judkins, T. Scholl, A. Bekessy, A. Marsh, P. Lovelock, M. Wong, A. Tesoriero, H. Renard, M. Southey, J.L. Hopper, K. Yannoukakos, M. Brown, D. Easton, S.V. Tavtigian, D. Goldgar, A.B. Spurdle, Genetic and histopathologic evaluation of BRCA1 and BRCA2 DNA sequence variants of unknown clinical significance, Cancer Res 66 (2006) 2019-2027.

[186] E. Capriotti, P. Fariselli, R. Casadio, I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure, Nucleic Acids Res 33 (2005) W306-310.

[187] M. Larkin, G. Blackshields, N. Brown, R. Chenna, P.A. McGettigan, H. McWilliam, F. Valentin, I.M. Wallace, A. Wilm, R. Lopez, Clustal W and Clustal X version 2.0, Bioinformatics 23 (2007) 2947-2948.

[188] C. Ferrer-Costa, J.L. Gelpi, L. Zamakola, I. Parraga, X. de la Cruz, M. Orozco, PMUT: a web-based tool for the annotation of pathological mutations on proteins, Bioinformatics 21 (2005) 3176-3178.

[189] P. Hohenstein, M.F. Kielman, C. Breukel, L.M. Bennett, R. Wiseman, P. Krimpenfort, C. Cornelisse, G.-J. van Ommen, P. Devilee, R. Fodde, A targeted mouse Brca1 mutation removing the last BRCT repeat results in apoptosis and embryonic lethality at the headfold stage, Oncogene 20 (2001) 2544-2550.

[190] L.C. Gowen, B.L. Johnson, A.M. Latour, K.K. Sulik, B.H. Koller, Brca1 deficiency results in early embryonic lethality characterized by neuroepithelial abnormalities, Nat Genet 12 (1996) 191-194.

[191] S.R. Lakhani, M.J. Van De Vijver, J. Jacquemier, T.J. Anderson, P.P. Osin, L. McGuffog, D.F. Easton, The pathology of familial breast cancer: predictive value of immunohistochemical markers estrogen receptor, progesterone receptor, HER-2, and p53 in patients with mutations in BRCA1 and BRCA2, J Clin Oncol 20 (2002) 2310-2318.

[192] K.J. Bowers, E. Chow, H. Xu, R.O. Dror, M.P. Eastwood, B.A. Gregersen, J.L. Klepeis, I. Kolossvary, M.A. Moraes, F.D. Sacerdoti, Scalable algorithms for molecular dynamics simulations on commodity clusters, SC 2006 Conference, Proceedings of the ACM/IEEE, IEEE, 2006, pp. 43-43.

[193] W.L. Jorgensen, J. Tirado-Rives, The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin, Journal of the American Chemical Society 110 (1988) 1657-1666.

[194] R.A. Laskowski, M.B. Swindells, LigPlot+: multiple ligand-protein interaction diagrams for drug discovery, J Chem Inf Model 51 (2011) 2778-2786.

[195] P.V. Viitanen, A.A. Gatenby, G.H. Lorimer, Purified chaperonin 60 (groEL) interacts with the nonnative states of a multitude of Escherichia coli proteins, Protein Sci 1 (1992) 363-369.

[196] P.R. Joseph, Z. Yuan, E.A. Kumar, G.L. Lokesh, S. Kizhake, K. Rajarathnam, A. Natarajan, Structural characterization of BRCT-tetrapeptide binding interactions, Biochem Biophys Res Commun 393 (2010) 207-210.

[197] Z. Liu, J. Wu, X. Yu, CCDC98 targets BRCA1 to DNA damage sites, Nat Struct Mol Biol 14 (2007) 716-720.

[198] M.S. Lee, R. Green, S.M. Marsillac, N. Coquelle, R.S. Williams, T. Yeung, D. Foo, D.D. Hau, B. Hui, A.N. Monteiro, J.N. Glover, Comprehensive analysis of missense variations in the BRCT domain of BRCA1 by structural and functional assays, Cancer Res 70 (2010) 4880-4890.

[199] N. Mirkovic, M.A. Marti-Renom, B.L. Weber, A. Sali, A.N. Monteiro, Structure-based assessment of missense mutations in human BRCA1: implications for breast and ovarian cancer predisposition, Cancer Res 64 (2004) 3790-3797.

[200] R.S. Williams, D.I. Chasman, D.D. Hau, B. Hui, A.Y. Lau, J.N. Glover, Detection of protein folding defects caused by BRCA1-BRCT truncation and missense mutations, J Biol Chem 278 (2003) 53007-53016.

[201] J.T. Bergthorsson, B. Ejlertsen, J.H. Olsen, A. Borg, K.V. Nielsen, R.B. Barkardottir, S. Klausen, H.T. Mouridsen, K. Winther, K. Fenger, A. Niebuhr, T.L. Harboe, E. Niebuhr, BRCA1 and BRCA2 mutation status and cancer family history of Danish women affected with multifocal or bilateral breast cancer at a young age, J Med Genet 38 (2001) 361-368.

[202] R.S. Williams, J.N. Glover, Structural consequences of a cancer-causing BRCA1-BRCT missense mutation, J Biol Chem 278 (2003) 2630-2635.

[203] J.S. Richardson, The anatomy and taxonomy of protein structure, Adv Protein Chem 34 (1981) 167-339.

[204] V. Munoz, L. Serrano, Elucidating the folding problem of helical peptides using empirical parameters, Nat Struct Biol 1 (1994) 399-409.

[205] H.J. Kang, H.J. Kim, C.H. Cho, Y. Hu, R. Li, I. Bae, BRCA1 transcriptional activity is enhanced by interactions between its AD1 domain and AhR, Cancer Chemother Pharmacol 62 (2008) 965-975.

[206] G. Gill, I. Sadowski, M. Ptashne, Mutations that increase the activity of a transcriptional activator in yeast and mammalian cells, Proceedings of the National Academy of Sciences 87 (1990) 2127-2131.

[207] A.N. Monteiro, A. August, H. Hanafusa, Evidence for a transcriptional activation function of BRCA1 C-terminal region, Proc Natl Acad Sci U S A 93 (1996) 13595-13599.

[208] P. Romero, Z. Obradovic, C. Kissinger, J. Villafranca, A. Dunker, Identifying disordered regions in proteins from amino acid sequence, Neural Networks, 1997., International Conference on, IEEE, 1997, pp. 90-95.

[209] S. Vucetic, Z. Obradovic, V. Vacic, P. Radivojac, K. Peng, L.M. Iakoucheva, M.S. Cortese, J.D. Lawson, C.J. Brown, J.G. Sikes, C.D. Newton, A.K. Dunker, DisProt: a database of protein disorder, Bioinformatics 21 (2005) 137-140.

[210] V. Vacic, V.N. Uversky, A.K. Dunker, S. Lonardi, Composition Profiler: a tool for discovery and visualization of amino acid composition differences, BMC Bioinformatics 8 (2007) 211.

[211] Y. Miki, J. Swensen, D. Shattuck-Eidens, P.A. Futreal, K. Harshman, S. Tavtigian, Q. Liu, C. Cochran, L.M. Bennett, W. Ding, et al., A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1, Science 266 (1994) 66-71.

[212] Y. Xia, G.M. Pao, H.W. Chen, I.M. Verma, T. Hunter, Enhancement of BRCA1 E3 ubiquitin ligase activity through direct interaction with the BARD1 protein, J Biol Chem 278 (2003) 5255-5263.

[213] V. Garcia, J.M. Garcia, C. Pena, J. Silva, G. Dominguez, R. Rodriguez, C. Maximiano, R. Espinosa, P. Espana, F. Bonilla, The GADD45, ZBRK1 and BRCA1 pathway: quantitative analysis of mRNA expression in colon carcinomas, J Pathol 206 (2005) 92-99.

[214] F. Eisinger, J. Jacquemier, J.M. Guinebretiere, D. Birnbaum, H. Sobol, p53 involvement in BRCA1-associated breast cancer, Lancet 350 (1997) 1101.

[215] J.A. Marsh, J.D. Forman-Kay, Sequence determinants of compaction in intrinsically disordered proteins, Biophys J 98 (2010) 2383-2390.

[216] J.R. Huth, K. Mountjoy, F. Perini, E. Bedows, R.W. Ruddon, Domain-dependent protein folding is indicated by the intracellular kinetics of disulfide bond formation of human chorionic gonadotropin beta subunit, J Biol Chem 267 (1992) 21396-21403.

[217] W. Tan, L. Zheng, W.H. Lee, T.G. Boyer, Functional dissection of transcription factor ZBRK1 reveals zinc fingers with dual roles in DNA-binding and BRCA1-dependent transcriptional repression, J Biol Chem 279 (2004) 6576-6587.

[218] T. Ouchi, S.W. Lee, M. Ouchi, S.A. Aaronson, C.M. Horvath, Collaboration of signal transducer and activator of transcription 1 (STAT1) and BRCA1 in differential regulation of IFN-gamma target genes, Proc Natl Acad Sci U S A 97 (2000) 5208-5213.

[219] I. Walsh, A.J. Martin, T. Di Domenico, A. Vullo, G. Pollastri, S.C. Tosatto, CSpritz: accurate prediction of protein disorder segments with annotation for homology, secondary structure and linear motifs, Nucleic Acids Res 39 (2011) W190-196.

[220] Z. Dosztanyi, B. Meszaros, I. Simon, ANCHOR: web server for predicting protein binding regions in disordered proteins, Bioinformatics 25 (2009) 2745-2746.

[221] T. Zhang, E. Faraggi, B. Xue, A.K. Dunker, V.N. Uversky, Y. Zhou, SPINE-D: accurate prediction of short and long disordered regions by a single neural-network based method, J Biomol Struct Dyn 29 (2012) 799-813.

[222] M.J. Mizianty, W. Stach, K. Chen, K.D. Kedarisetti, F.M. Disfani, L. Kurgan, Improved sequence-based prediction of disordered regions with multilayer fusion of multiple information sources, Bioinformatics 26 (2010) i489-496.

[223] L.P. Kozlowski, J.M. Bujnicki, MetaDisorder: a meta-server for the prediction of intrinsic disorder in proteins, BMC Bioinformatics 13 (2012) 111.

[224] D.V. Rial, E.A. Ceccarelli, Removal of DnaK contamination during fusion protein purifications, Protein Expr Purif 25 (2002) 503-507.

[225] M. Bouvier, W.F. Stafford, Probing the three-dimensional structure of human calreticulin, Biochemistry 39 (2000) 14950-14959.

[226] F. Wolfram, E. Morris, C.W. Taylor, Three-dimensional structure of recombinant type 1 inositol 1,4,5-trisphosphate receptor, Biochem J 428 (2010) 483-489.

[227] C. Dulac, A.A. Michels, A. Fraldi, F. Bonnet, V.T. Nguyen, G. Napolitano, L. Lania, O. Bensaude, Transcription-dependent association of multiple positive transcription elongation factor units to a HEXIM multimer, J Biol Chem 280 (2005) 30619-30629.

[228] O. Schweers, E. Schonbrunn-Hanebeck, A. Marx, E. Mandelkow, Structural studies of tau protein and Alzheimer paired helical filaments show no evidence for beta-structure, J Biol Chem 269 (1994) 24290-24297.

[229] P.H. Weinreb, W. Zhen, A.W. Poon, K.A. Conway, P.T. Lansbury, NACP, a protein implicated in Alzheimer's disease and learning, is natively unfolded, Biochemistry 35 (1996) 13709-13715.

[230] O.B. Ptitsyn, Molten globule and protein folding, Adv Protein Chem 47 (1995) 83-229.

[231] V.N. Uversky, Natively unfolded proteins: a point where biology waits for physics, Protein Sci 11 (2002) 739-756.

[232] I.S. Millett, S. Doniach, K.W. Plaxco, Toward a taxonomy of the denatured state: small angle scattering studies of unfolded proteins, Adv Protein Chem 62 (2002) 241-262.

[233] V.N. Uversky, Intrinsically disordered proteins and their environment: effects of strong denaturants, temperature, pH, counter ions, membranes, binding partners, osmolytes, and macromolecular crowding, Protein J 28 (2009) 305-325.

[234] V.N. Uversky, J. Li, A.L. Fink, Evidence for a partially folded intermediate in alpha-synuclein fibril formation, J Biol Chem 276 (2001) 10737-10744.

[235] S.E. Permyakov, I.S. Millett, S. Doniach, E.A. Permyakov, V.N. Uversky, Natively unfolded C-terminal domain of caldesmon remains substantially unstructured after the effective binding to calmodulin, Proteins 53 (2003) 855-862.

[236] V.N. Uversky, S.E. Permyakov, V.E. Zagranichny, I.L. Rodionov, A.L. Fink, A.M. Cherskaya, L.A. Wasserman, E.A. Permyakov, Effect of zinc and temperature on the conformation of the gamma subunit of retinal phosphodiesterase: a natively unfolded protein, J Proteome Res 1 (2002) 149-159.

[237] L.M. Iakoucheva, A.L. Kimzey, C.D. Masselon, J.E. Bruce, E.C. Garner, C.J. Brown, A.K. Dunker, R.D. Smith, E.J. Ackerman, Identification of intrinsic order and disorder in the DNA repair protein XPA, Protein Sci 10 (2001) 560-571.

[238] M. Bouvier, W.F. Stafford, Probing the three-dimensional structure of human calreticulin, Biochemistry 39 (2000) 14950-14959.

[239] S.J. Hubbard, F. Eisenmenger, J.M. Thornton, Modeling studies of the change in conformation required for cleavage of limited proteolytic sites, Protein Sci 3 (1994) 757-768.

[240] C.A. Kyratsous, S.J. Silverstein, C.R. DeLong, C.A. Panagiotidis, Chaperone-fusion expression plasmid vectors for improved solubility of recombinant proteins in Escherichia coli, Gene 440 (2009) 9-15.

[241] J.P. Hendrick, F.U. Hartl, The role of molecular chaperones in protein folding, FASEB J 9 (1995) 1559-1569.

[242] V. Brazda, E.B. Jagelska, J.C. Liao, C.H. Arrowsmith, The central region of BRCA1 binds preferentially to supercoiled DNA, J Biomol Struct Dyn 27 (2009) 97-104.

[243] O.N. Aprelikova, B.S. Fang, E.G. Meissner, S. Cotter, M. Campbell, A. Kuthiala, M. Bessho, R.A. Jensen, E.T. Liu, BRCA1-associated growth arrest is RB-dependent, Proc Natl Acad Sci U S A 96 (1999) 11866-11871.

[244] A. Folias, M. Matkovic, D. Bruun, S. Reid, J. Hejna, M. Grompe, A. D'Andrea, R. Moses, BRCA1 interacts directly with the Fanconi anemia protein FANCA, Hum Mol Genet 11 (2002) 2591-2597.

[245] C.X. Deng, BRCA1: cell cycle checkpoint, genetic instability, DNA damage response and cancer evolution, Nucleic Acids Res 34 (2006) 1416-1426.

[246] R. Naseem, M. Webb, Analysis of the DNA binding activity of BRCA1 and its modulation by the tumour suppressor p53, PLoS One 3 (2008) e2336.

[247] G.W. Daughdrill, G.J. Pielak, V.N. Uversky, M.S. Cortese, A.K. Dunker, Natively disordered proteins, Protein folding handbook (2005) 275-357.

[248] R.W. Kriwacki, L. Hengst, L. Tennant, S.I. Reed, P.E. Wright, Structural studies of p21Waf1/Cip1/Sdi1 in the free and Cdk2-bound state: conformational disorder mediates binding diversity, Proc Natl Acad Sci U S A 93 (1996) 11504-11509.

[249] A.K. Dunker, Z. Obradovic, The protein trinity--linking function and disorder, Nat Biotechnol 19 (2001) 805-806.

[250] C. Chothia, A. Lesk, The evolution of protein structures, Cold Spring Harb Symp Quant Biol, Cold Spring Harbor Laboratory Press, 1987, pp. 399-405.

[251] A.P. Demchenko, Recognition between flexible protein molecules: induced and assisted folding†, Journal of molecular recognition 14 (2001) 42-61.

[252] C.R. Wobbe, J.P. Lee, S.C. I-larrisoni, K. Struhl, Folding transition in the DNA-binding domain of GCN4 on specific binding to DNA, Nature 347 (1990) 11.

[253] H.J. Dyson, P.E. Wright, Coupling of folding and binding for unstructured proteins, Curr Opin Struct Biol 12 (2002) 54-60.

[254] C. Zwahlen, S.C. Li, L.E. Kay, T. Pawson, J.D. Forman-Kay, Multiple modes of peptide recognition by the PTB domain of the cell fate determinant Numb, EMBO J 19 (2000) 1505-1515.

[255] J.L. Urbauer, J.H. Short, L.K. Dow, A.J. Wand, Structural analysis of a novel interaction by calmodulin: high-affinity binding of a peptide in the absence of calcium, Biochemistry 34 (1995) 8099-8109.

[256] S.-C. Li, C. Zwahlen, S.J. Vincent, C.J. McGlade, L.E. Kay, T. Pawson, J.D. Forman-Kay, Structure of a Numb PTB domain–peptide complex suggests a basis for diverse binding specificity, Nat Struct Mol Biol 5 (1998) 1075-1083.

[257] S.E. Permyakov, I.S. Millett, S. Doniach, E.A. Permyakov, V.N. Uversky, Natively unfolded C-terminal domain of caldesmon remains substantially unstructured after the effective binding to calmodulin, Proteins: Structure, Function, and Bioinformatics 53 (2003) 855.

[258] N. Foray, D. Marot, A. Gabriel, V. Randrianarison, A.M. Carr, M. Perricaudet, A. Ashworth, P. Jeggo, A subset of ATM- and ATR-dependent phosphorylation events requires the BRCA1 protein, EMBO J 22 (2003) 2860-2871.

[259] A.K. Dunker, C.J. Brown, J.D. Lawson, L.M. Iakoucheva, Z. Obradovic, Intrinsic disorder and protein function, Biochemistry 41 (2002) 6573-6582.

[260] L.M. Iakoucheva, P. Radivojac, C.J. Brown, T.R. O'Connor, J.G. Sikes, Z. Obradovic, A.K. Dunker, The importance of intrinsic disorder for protein phosphorylation, Nucleic Acids Res 32 (2004) 1037-1049.

[261] J. Chen, D.P. Silver, D. Walpita, S.B. Cantor, A.F. Gazdar, G. Tomlinson, F.J. Couch, B.L. Weber, T. Ashley, D.M. Livingston, R. Scully, Stable interaction between the products of the BRCA1 and BRCA2 tumor suppressor genes in mitotic and meiotic cells, Mol Cell 2 (1998) 317-328.

[262] M.J. Landrum, J.M. Lee, G.R. Riley, W. Jang, W.S. Rubinstein, D.M. Church, D.R. Maglott, ClinVar: public archive of relationships among sequence variation and human phenotype, Nucleic acids research 42 (2014) D980-D985.

[263] S.A. Forbes, D. Beare, P. Gunasekaran, K. Leung, N. Bindal, H. Boutselakis, M. Ding, S. Bamford, C. Cole, S. Ward, COSMIC: exploring the world's knowledge of somatic mutations in human cancer, Nucleic acids research 43 (2015) D805-D811.

[264] S.A. Janezic, A. Ziogas, L.M. Krumroy, M. Krasner, S.J. Plummer, P. Cohen, M. Gildea, D. Barker, R. Haile, G. Casey, H. Anton-Culver, Germline BRCA1 alterations in a population-based series of ovarian cancer cases, Hum Mol Genet 8 (1999) 889-897.

[265] S. Tommasi, B. Pilato, R. Pinto, A. Monaco, M. Bruno, M. Campana, M. Digennaro, F. Schittulli, R. Lacalamita, A. Paradiso, Molecular and in silico analysis of BRCA1 and BRCA2 variants, Mutat Res 644 (2008) 64-70.

[266] R.M. Wenham, J.M. Schildkraut, K. McLean, B. Calingaert, R.C. Bentley, J. Marks, A. Berchuck, Polymorphisms in BRCA1 and BRCA2 and risk of epithelial ovarian cancer, Clin Cancer Res 9 (2003) 4396-4403.

[267] M.S. Nicoloso, H. Sun, R. Spizzo, H. Kim, P. Wickramasinghe, M. Shimizu, S.E. Wojcik, J. Ferdin, T. Kunej, L. Xiao, S. Manoukian, G. Secreto, F. Ravagnani, X. Wang, P. Radice, C.M. Croce, R.V. Davuluri, G.A. Calin, Single-nucleotide polymorphisms inside microRNA target sites influence tumor susceptibility, Cancer Res 70 (2010) 2789-2798.

[268] X. Zhou, S. Han, S. Wang, X. Chen, J. Dong, X. Shi, Y. Xia, X. Wang, Z. Hu, H. Shen, Polymorphisms in HPV E6/E7 protein interacted genes and risk of cervical cancer in Chinese women: a case-control analysis, Gynecol Oncol 114 (2009) 327-331.

[269] H.J. Shim, J.Y. Yun, J.E. Hwang, W.K. Bae, S.H. Cho, J.H. Lee, H.N. Kim, M.H. Shin, S.S. Kweon, H.J. Kim, I.J. Chung, BRCA1 and XRCC1 polymorphisms associated with survival in advanced gastric cancer treated with taxane and cisplatin, Cancer Sci 101 (2010) 1247-1254.

[270] H.S. Zhang, D. Liu, Y. Huang, S. Schmidt, R. Hickey, D. Guschin, H. Su, I.S. Jovin, M. Kunis, S. Hinkley, Y. Liang, L. Hinh, S.K. Spratt, C.C. Case, E.J. Rebar, B.E. Ehrlich, P.D. Gregory, F.J. Giordano, A designed zinc-finger transcriptional repressor of phospholamban improves function of the failing heart, Mol Ther 20 (2012) 1508-1515.

[271] A. Klug, Towards therapeutic applications of engineered zinc finger proteins, FEBS Lett 579 (2005) 892-894.

[272] M.F. Rousseau-Merck, D. Koczan, I. Legrand, S. Moller, S. Autran, H.J. Thiesen, The KOX zinc finger genes: genome wide mapping of 368 ZNF PAC clones with zinc finger gene clusters predominantly in 23 chromosomal loci are confirmed by human sequences annotated in EnsEMBL, Cytogenet Genome Res 98 (2002) 147-153.

[273] S. Furuta, J.M. Wang, S. Wei, Y.M. Jeng, X. Jiang, B. Gu, P.L. Chen, E.Y. Lee, W.H. Lee, Removal of BRCA1/CtIP/ZBRK1 repressor complex on ANG1 promoter leads

to accelerated mammary tumor growth contributed by prominent vasculature, Cancer Cell 10 (2006) 13-24.

[274] L.F. Lin, C.H. Chuang, C.F. Li, C.C. Liao, C.P. Cheng, T.L. Cheng, M.R. Shen, J.T. Tseng, W.C. Chang, W.H. Lee, J.M. Wang, ZBRK1 acts as a metastatic suppressor by directly regulating MMP9 in cervical cancer, Cancer Res 70 (2010) 192-201.

[275] T. Ahmed, S.M. Kelly, N.C. Price, A.J. Lawrence, Activation of phospholipase A(2) by long chain fatty acyl groups involves a novel unstable linkage, J Biochem 127 (2000) 871-875.

[276] J. Yun, W.H. Lee, Degradation of transcription repressor ZBRK1 through the ubiquitin-proteasome pathway relieves repression of Gadd45a upon DNA damage, Mol Cell Biol 23 (2003) 7305-7314.

[277] S.M. Kelly, N.C. Price, The use of circular dichroism in the investigation of protein structure and function, Curr Protein Pept Sci 1 (2000) 349-384.

[278] R.W. Woody, [4] Circular dichroism, Methods in enzymology 246 (1995) 34-71.

[279] T. Rao, G. Ruiz-Gomez, T.A. Hill, H.N. Hoang, D.P. Fairlie, J.M. Mason, Truncated and helix-constrained peptides with high affinity and specificity for the cFos coiled-coil of AP-1, PLoS One 8 (2013) e59415.

[280] R.W. Alston, L. Urbanikova, J. Sevcik, M. Lasagna, G.D. Reinhart, J.M. Scholtz, C.N. Pace, Contribution of single tryptophan residues to the fluorescence and stability of ribonuclease Sa, Biophys J 87 (2004) 4036-4047.

[281] S.A. Wolfe, L. Nekludova, C.O. Pabo, DNA recognition by Cys2His2 zinc finger proteins, Annu Rev Biophys Biomol Struct 29 (2000) 183-212.

[282] D. Lu, [C2H2 zinc-finger recognition of biomolecules], Yao Xue Xue Bao 48 (2013) 834-841.

[283] P.V. Pedone, R. Ghirlando, G.M. Clore, A.M. Gronenborn, G. Felsenfeld, J.G. Omichinski, The single Cys2-His2 zinc finger domain of the GAGA protein flanked by basic residues is sufficient for high-affinity specific DNA binding, Proc Natl Acad Sci U S A 93 (1996) 2822-2826.

[284] R.N. De Guzman, H.Y. Liu, M. Martinez-Yamout, H.J. Dyson, P.E. Wright, Solution structure of the TAZ2 (CH3) domain of the transcriptional adaptor protein CBP, J Mol Biol 303 (2000) 243-253.

[285] E.E. Zheleznova, J.H. Crosa, R.G. Brennan, Characterization of the DNA- and metal-binding properties of Vibrio anguillarum fur reveals conservation of a structural Zn(2+) ion, J Bacteriol 182 (2000) 6264-6267.

[286] L. Nekludova, C.O. Pabo, Distinctive DNA conformation with enlarged major groove is found in Zn-finger-DNA and other protein-DNA complexes, Proc Natl Acad Sci U S A 91 (1994) 6948-6952.

[287] N.E. Zhou, C.M. Kay, R.S. Hodges, Synthetic model proteins. Positional effects of interchain hydrophobic interactions on stability of two-stranded alpha-helical coiled-coils, J Biol Chem 267 (1992) 2664-2670.

List of publications arising from the thesis

1. Functional assessment of intrinsic disorder central domains of BRCA1. Lumbini R. Yadav, Sharad Rai, M.V. Hosur, & Ashok K. Varma. Journal of Biomolecular Structure and Dynamics, 2015 January.
2. Tetrameric ZBRK1 DNA binding domain has affinity towards cognate DNA in absence of zinc ions. Lumbini R. Yadav, Mahamaya N.Biswal, Vikrant, M.V. Hosur, Ashok K. Varma. BBRC, 2014 June.
3. Structural Basis to Characterise Transactivation Domain of BRCA1. Lumbini R. Yadav, Mahamaya N. Biswal, M.V. Hosur, Nachimuthu Senthil Kumar, Ashok K. Varma. JBSD, 2016 January.
4. Structural and functional characterization of transactivation Domain of BRCA1. Lumbini R. Yadav, Ashok K. Varma. MS01.P19 Acta Cryst. A67, C221-C222, 2011 August.

Biochemical and Biophysical Research Communications 450 (2014) 283–288

Contents lists available at ScienceDirect

Biochemical and Biophysical Research Communications

journal homepage: www.elsevier.com/locate/ybbrc

Tetrameric ZBRK1 DNA binding domain has affinity towards cognate DNA in absence of zinc ions

Lumbini R. Yadav, Mahamaya N. Biswal, Vikrant, M.V. Hosur, Ashok K. Varma *

Advanced Centre for Treatment, Research and Education in Cancer, Kharghar, Navi Mumbai, Maharashtra 410 210, India

ARTICLE INFO

Article history:
Received 12 May 2014
Available online 9 June 2014

Keywords:
ZBRK1
Zinc-finger
Oligomer
Secondary structure

ABSTRACT

Zinc finger transcription regulatory proteins play crucial roles in cell-cycle regulation, DNA damage response and tumor genesis. Human ZBRK1 is a zinc-finger transcription repressor protein, which recognizes double helical DNA containing consensus sequences of 5′GGGXXXCAGXXXTTT3′. In the present study, we have purified recombinant DNA binding domain of ZBRK1, and studied binding with zinc ions and DNA, using biophysical techniques. The elution profile of the purified protein suggests that this ZBRK1 forms a homotetramer in solution. Dissociation and pull down assays also suggest that this domain forms a higher order oligomer. The ZBRK1-DNA binding domain acquires higher stability in the presence of zinc ions and DNA. The secondary structure of the ZBRK1-DNA complex is found to be significantly altered from the standard B-DNA conformation.

1. Introduction

Zinc finger proteins, are known to play pivotal roles, through transcriptional regulation in DNA replication and repair, protein translation, cell proliferation and apoptosis. Regulation of gene expression using customized transcription factors has been reported recently [1,2]. Chimeric ZF proteins have been engineered to obtain customized activities, such as restriction, methylation and integration [3,4]. The erb2 oncogene and peripheral arterial obstructive diseases have been repressed by customized ZF proteins [5]. ZF proteins contain more than one ZF in tandem, and are considered potential targets for drugs. They can also make the genome a potential drug target.

There are different types of ZF proteins, and these are classified based on the environment around zinc ions. The major class is the C2H2 class, which contains a consensus amino acid sequence ψ-X-Cys-X(2–4)-Cys-X3-ψ-X5-ψ-X2-His-X(3,4)-His, where X is any amino acid and ψ is any hydrophobic residue. C2H2 type containing the KRAB (Kruppel-associated box) domain forms the largest subfamily (KRAB-ZFP) of ZF proteins [6] and, ZBRK1 is a member of this family. Human ZBRK1 performs a variety of biological functions through its various domains: the N-terminal KRAB, eight consecutive central C2H2 zinc fingers and a BRCA1-dependent C-terminal transcriptional repression domain (CTRD) [7]. ZBRK1 consists of 532 amino acids, and the region from residues 206 to 424 is the DNA-binding domain (DBD). The N-terminal KRAB domain also performs repressor function by interacting with other proteins, such as KAP1. ZBRK1 represses cellular invasion and metastasis and, its loss enhances MMP9 transcription in cervical cancer [8]. ZBRK1 represses a high mobility group AT-hook 2 (HMGA2), a DNA architectural protein, activation of which plays a significant role in tumor genesis and metastasis [9]. The repressor ZBRK1 binds to a 15 bp consensus nucleotide sequence, GGGXXXCAGXXXTTT (where X is any nucleotide), which is found near promoter regions of many DNA damage inducible genes like p21, ANG1 [10] GADD153 and GADD45 [11,12]. Exposure of cells to DNA damaging agents leads to degradation of ZBRK1 via the BRCA1-independent ubiquitin–proteasome pathway leading to de-repression [13]. Interaction of ZBRK1 with Trim28 and histone deacetylase, results in repression of HIV-1 LTR acting as an intrinsic retroviral defense system [14]. In order to understand the various functions of ZBRK1, in details, we have initiated to study the structure and interactions of ZBRK1 with different ligands.

Here, we report purification and biophysical characterization of the DNA binding domain of ZBRK1 (ZBRK1-DBD), and an analysis of its binding with zinc and the cognate oligonucleotide. It is found that this domain exists as a homotetramer in solution, in contradiction to earlier report that C-terminal residues were essential for tetramer formation [15]. ZBRK1-DBD binds only to double

Abbreviations: CD, circular dichroism spectroscopy; ZBRK1, zinc finger and BRCA1-interacting protein with a KRAB domain 1; ZF, zinc-finger.

* Corresponding author. Fax: +91 22 2740 5085.
E-mail address: avarma@actrec.gov.in (A.K. Varma).

http://dx.doi.org/10.1016/j.bbrc.2014.05.104
0006-291X/

stranded DNA, even in the absence of zinc ions. The DNA is found to be overwound on binding, while the conformation of the protein becomes less compact.

2. Materials and methods

2.1. Reagents

All the enzymes used for cloning were procured from NEB (US). Oligonucleotides were ordered from Sigma Aldrich (US). Chemicals used were from HiMedia, Qualigens and Sigma Aldrich. Bacterial culture media, IPTG and ampicillin were from HiMedia (India).

2.2. Cloning of ZBRK1-DNA binding domain

ZBRK1-DBD from amino acid 206 to 424 was sub-cloned at the BamHI and HinDIII site in pRSET-A vector (Invitrogen). Forward primer 5′-GTCTCCGAGAACCTGTACTTTCAG GGTTGTGTGCAGTGAA TGTGGGAA-3′ and Reverse primer 5′-GTCAAGCTTCT ATTAGTGTAT TCTCTTATGCTTAAC-3′ were used for polymerase chain reaction. The forward primer has TEV site incorporated in it so as to enable cleavage of the native protein from histidine tag. The ligated samples were transformed into *Escherichia coli* DH5α cells and plated on ampicillin plates. Ampicillin resistant colonies were screened for positive clones and correct DNA sequence (Genetic Analyser-Applied Bio Systems).

2.3. Protein expression and purification

E. coli bacterial strain Rosetta 2(DE3) (Invitrogen) was transformed with ZBRK1-DBD construct. Single colony was inoculated in Luria Bertani broth containing ampicillin (100 μg/ml) and chloramphenicol (34 μg/ml), and the culture was allowed to grow for 12–14 h. Cells were then diluted 100 fold and grown till the OD_{600} reached a value of 0.7. This culture was then induced with 0.4 mM IPTG for 16 h at 18 °C. Cells were harvested and resuspended in buffer-A containing 10 mM HEPES pH 7.0, 500 mM NaCl, 0.1% Triton X-100, 20 mM β-ME, 1 mM EDTA, 2.5% glycerol, and 10 mM imidazole. Cells were lysed by sonication at a pulse rate of 50, three times with 1 min duty cycle. Cell debris was removed by centrifugation. Cell lysis was allowed to bind on Ni–NTA resin for 1–2 h at 25 °C. To remove the nonspecifically bound protein, resin was washed with buffer-A addition containing 50 mM imidazole. Fusion protein was eluted with imidazole (500 and 750 mM) and further treated with TEV protease to remove histidine tag.

2.4. Gel filtration chromatography

ZBRK1-DBD was further purified to homogeneity using AKTA explorer system (GE Healthcare). Concentrated protein was chromatographed using Superdex-75 column pre equilibrated with buffer (5 mM HEPES pH 7.0, 150 mM NaCl, and 20 mM β-mercaptoethanol). Four proteins (BSA, lysozyme, phosphorylase B, and carbonic anhydrase) were eluted from the same column under similar conditions to prepare the standard plot for molecular weight estimation. The sample was run at a flow rate of 0.5 ml/min, and 1 ml fractions were collected. FPLC purified protein was further passed through buffer exchange column to remove β-mercaptoethanol. Peptide mass fingerprinting was done for protein confirmation using MALDI-MS (Ultraflex Bruker Daltonics system).

2.5. MBP pull down assay

MBP-ZBRK1ΔK and MBP were expressed and affinity purified using amylose resin. Purified ZBRK1-DBD was incubated with MBP-ZBRK1ΔK and MBP for 3 h on ice with gentle mixing. Amylose resin was washed with buffer-A to remove non-specifically bound ZBRK1-DBD. The resin was then loaded on 12% SDS gel to check for the pull down.

2.6. Circular dichroism spectroscopy

Far UV CD scan of ZBRK1-DBD (7 μM) was monitored to study secondary structural characteristics. Scan was recorded from 240 to 190 nm at a scanning speed of 20 nm/min, data interval of 0.1 nm and 3 accumulations. To study the effect of zinc sulfate and DNA on secondary structural characteristics, ZBRK1-DBD was titrated with zinc sulfate (10–180 μM) and DNA (5–30 μM). To reveal the structural alteration in DNA due to ZBRK1 binding, CD scan was obtained from 300 to 200 nm at a scanning speed of 20 nm/min. Protein concentration was estimated directly by Bradford method assuming a molar extinction coefficient of 12,950 M^{-1} cm^{-1}. All the spectra were measured at 25 °C using (Jasco J-815) spectropolarimeter, and were baseline corrected. α-Helical content of the protein was estimated using ellipticity signals at 208 and 222 nm and the standard equation [16].

2.7. Thermal stability

To measure the thermal stability of ZBRK1-DBD (5 μM), CD spectra in the range 250–200 nm was recorded as a function of temperature. Far UV CD scan was monitored at every 5 °C increase in temperature, with an equilibration time of 3 min and a scanning speed of 50 nm/min. Thermal unfolding pathway was also studied to reveal the effect of $ZnSO_4$ (0.1 mM) and DNA (30 μM) on the stability of ZBRK1-DBD. Ellipticity at 222 nm was plotted as a function of temperature to understand the course of change in secondary structure elements.

2.8. Fluorescence spectroscopy

Spectrofluorometer (JOBIN YVON Horiba Fluorolog 3) was used for fluorescence measurements. The emission slit width was 5 nm. The fluorescence scan was measured in the emission wavelength range of λ_{em} 310–500 nm, and the excitation was at 295 nm. The spectral data were collected using FluorESSENCE software. Protein concentration was measured using molar extinction coefficient as described before. Titration experiments with ZBRK1-DBD and different concentrations of DNA (4–60 μM) and $ZnSO_4$ (5–110 μM) were performed to study the tertiary structural alterations.

2.9. Isothermal titration calorimetry (ITC)

ZBRK1-DBD (50 μM) was injected into calorimetric cell containing buffer to study the heat effects from dissociation. Thirteen injections each 3 μl, were injected at 25 °C, allowing an equilibration time of 150 s between injections. Interaction studies of ZBRK1-DBD with $ZnSO_4$ and consensus double stranded DNA oligonucleotide (5′GGGACGCAGTTTTAT3′) were performed using MicroCal iTC200 calorimeter. (GE Healthcare) The titration experiments were done under following conditions: (1) 5.2 μM of ZBRK-DBD in cell and 1 mM of $ZnSO_4$ in syringe, (2) 20 μM of ZBRK1-DBD in cell and 200 μM of DNA in syringe. Protein concentration was estimated, using BSA as standard.

3. Results

3.1. Expression and purification of ZBRK1-DBD

Two liters of bacterial culture yielded 0.4 mg of >95% of pure ZBRK1-DBD protein (Fig. 1A). A single band corresponding to a

molecular weight of 24 kDa suggests that the preparation is very homogeneous. Peptide mass fingerprinting further confirmed the sample to be ZBRK1-DBD (data not shown).

3.2. ZBRK1-DBD forms a tetramer

ZBRK1-DBD eluted at approximately 58 ml from 16/60 Superdex75 size exclusion column (Fig. 1B). This indicates that the protein forms a higher order oligomer. From the standard linear graph of $\log_{10}$ (molecular weight) versus (elution volume), the size of the eluted sample was estimated to be approximately 94 kDa (Fig. 1B). The partition coefficient Kav, for ZBRK1-DBD was calculated from the elution volume [17,18]. Stoke's radius of ZBRK1-DBD was interpolated from a plot of partition coefficient (Kav) versus Stoke's radii for standard markers. The calculated Stoke's radius of ZBRK1-DBD was 5.6 nm which is closer to Stoke's radius of standard of 97 kDa. To explore the ability of self-association of ZBRK1-DBD, MBP pull down assay with MBP tagged ZBRK1ΔK construct was performed. ZBRK1-DBD showed interaction with ZBRK1ΔK construct. No interaction was visible with MBP (Fig. 1C).

Thus, based on size exclusion chromatography and pull down assay, it is proposed that the ZBRK1-DBD forms a homotetramer in solution.

To confirm the formation of an oligomer, next we performed dissociation study using ITC. Addition of concentrated protein from the syringe into calorimeter cell containing the buffer-resulted in a heat change, which gradually decreased as concentration of the protein inside the cell increased (Fig. 1D). This observation suggests that the heat change was due to dissociation of an oligomeric species. The dissociation curve yielded enthalpy of dissociation as −482 kcal/mol with a dissociation constant of 0.125 μM.

3.3. Secondary structure alterations on ligand binding

Far UV CD spectrum of ZBRK1-DBD is shown in (Fig. 2A). The scan showed peaks at approximately 222, 208 and 195 nm, which are characteristic of alpha-helical conformation [16,19]. The alpha-helical content of ZBRK1-DBD, estimated using measured ellipticities at 222 and 208 nm, was 12% and 15% respectively. The CD spectra of ZBRK1-DBD were recorded at different concentrations of zinc (Fig. 2B) and DNA (Fig. 2C). The $\theta_{222}/\theta_{208\,nm}$ ratio of only protein is 1.0, and this ratio increases with further addition of $ZnSO_4$, indicating incorporation of quaternary structure. In the case of interaction with DNA, this ratio decreased, indicating reduction in coiled coil structure [16,20]. Fig. 2D shows the CD spectrum over the λ range of 320–230 nm, when DNA was titrated with varying

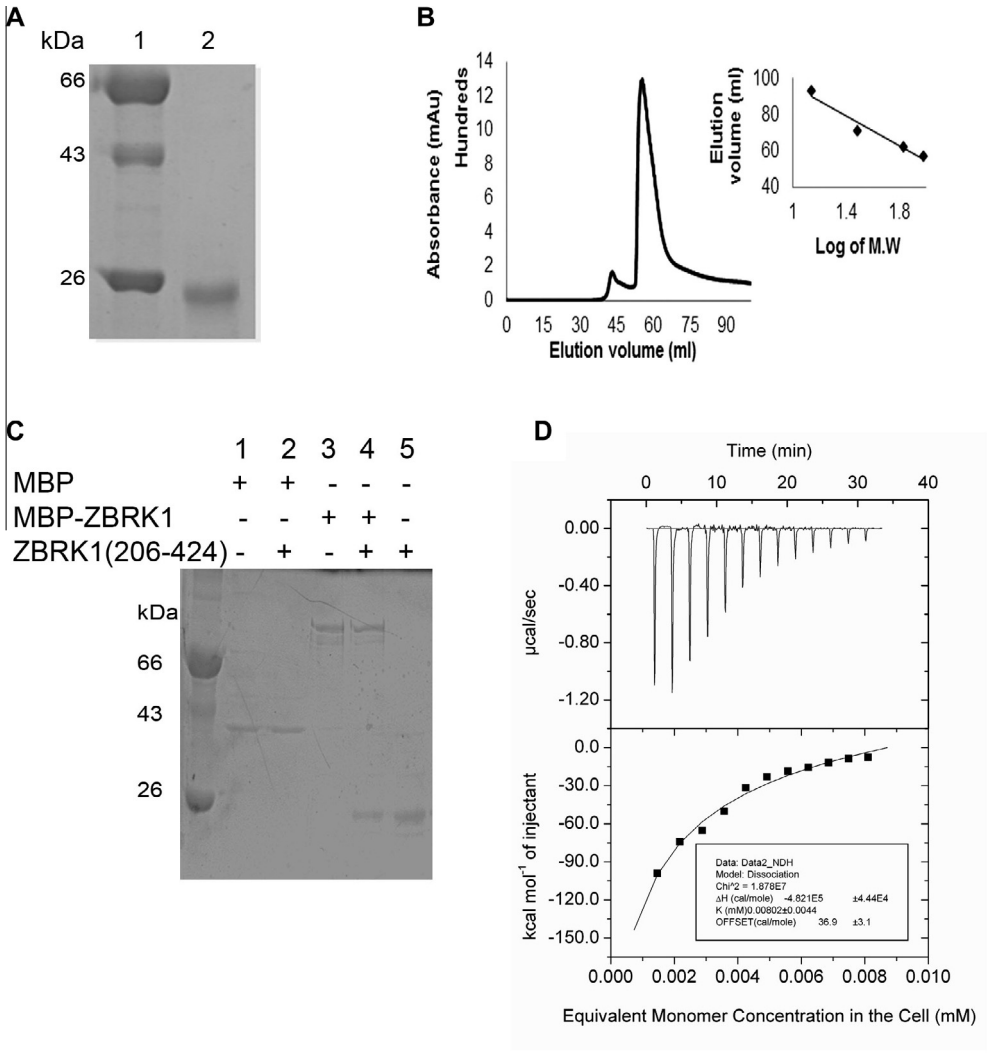


Fig. 1. Oligomeric characterization of ZBRK1-DBD: (A) size exclusion chromatography profile of purified ZBRK1-DBD protein. Inset is a plot of log of molecular weight standards (carbonic anhydrase, lysozyme, phosphorylase B and BSA) vs. elution volumes. (B) SDS electrophoresis showing sample purity. (C) MBP pull-down (lanes 1, 3, 5 – show input), (lane 2, 4 – pull down samples). (D) Exothermic heat pulses for 3 μl injections into buffer of ZBRK1-DBD (50 μM). Lower panel shows integrated heat data after blank correction.

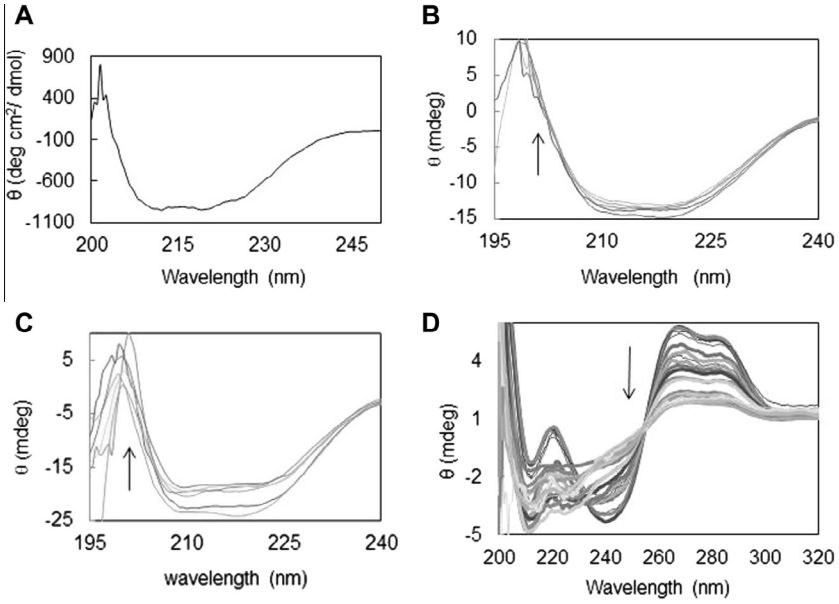


Fig. 2. CD spectra: (A) ZBRK1-DBD. ZBRK1-DBD titrated with (B) DNA, (C) zinc sulfate. (D) DNA titrated with ZBRK1-DBD. Arrows indicate directions of decrease in ellipticity.

amounts of ZBRK1-DBD. The CD spectrum of B-form DNA is composed of four major peaks: negative peaks around 214 and 245 nm and positive peaks around 225 and 280 nm [21,22]. The positive maximum around 280 nm is attributed to base stacking and the negative peak at 245 nm is attributed to DNA helicity [23]. The changes in ellipticities at 275 nm and 245 nm, on addition of ZBRK1-DBD, indicate that the standard B-geometry of the DNA is being altered by protein binding.

3.4. Intrinsic tryptophan fluorescence to elucidate local tryptophan environment

Tryptophan emission spectrum recorded from 310 to 500 nm, upon excitation at 295 nm (supplemental Fig. 1A). Emission maximum observed at 336 nm indicates that the single tryptophan residue is in hydrophobic environment [24]. Titration of ZBRK1-DBD with DNA (supplemental Fig. 1B) and zinc sulfate (supplemental Fig. 1C) showed a decrease in fluorescence intensity without any shift in the position of the maxima. This suggests that interaction of ZBRK1-DBD with either zinc sulfate or with DNA does not have any effect on local tryptophan environment.

3.5. Thermal stability

To evaluate the nature of folding and unfolding transitions, and to measure the structural stability of ZBRK1-DBD in presence of DNA and $ZnSO_4$ the far UV CD spectra, were recorded at different temperatures (Fig. 3). Variation in the ellipticity at 222 nm as a function of temperature suggests that ZBRK1-DBD unfolds via three state unfolding transitions with T_m values of 58 and 69 °C. Presence of zinc sulfate or DNA with ZBRK1 increased thermal stability of protein as suggested by increases in melting temperatures from 57 °C to 62 °C and 68 °C, respectively. Interestingly, in the presence of either DNA or zinc sulfate, ZBRK1 demonstrated simply two state unfolding transitions.

3.6. Thermodynamics of ligand binding to ZBRK1-DBD

Fig. 4 shows the results of calorimetry experiments to study the interactions of ZBRK1 with zinc sulfate and DNA. Model fitting for ZBRK1-DBD with zinc sulfate with one set of sites showed

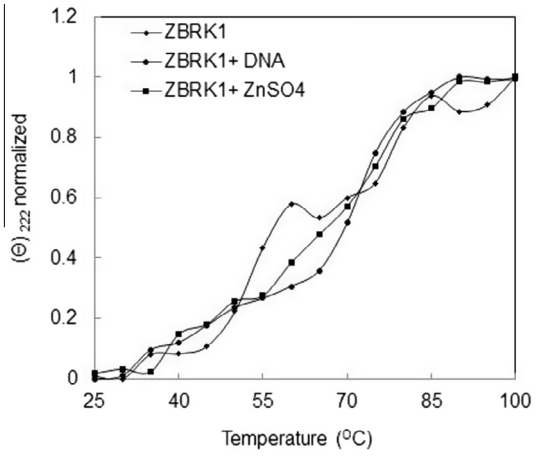


Fig. 3. Thermal denaturation profile of ZBRK1-DBD. Plot of fraction unfolded versus temperature at 5 °C intervals.

exothermic reaction with a stoichiometry of 1:42. It demonstrated an affinity of 1.4 μM under these conditions (Fig. 4A). Nonlinear curve fitting done using one set of sites for ZBRK1-DBD with DNA demonstrated a stoichiometry of 1:2 with binding affinity of 0.29 μM (Fig. 4B). ZBRK1-DBD interaction with zinc sulfate and DNA is an enthalpically driven reaction. Thermodynamic parameters of protein interaction with zinc and DNA, derived from ITC experiments are tabulated (Table 1).

4. Discussion

4.1. Oligomeric DBD

The elution profile of ZBRK1-DBD from the gel filtration column clearly indicates that ZBRK1-DBD was a homotetramer in solution. The absence of any substantial monomeric peak implies a very high affinity for tetramer formation. MBP pull down assay also revealed the ability of ZBRK1-DBD to oligomerize with ZBRK1ΔK in the MBP tagged ZBRK1ΔK construct. The ITC titration also suggests that dissociation of ZBRK1-DBD is enthalpically favored.

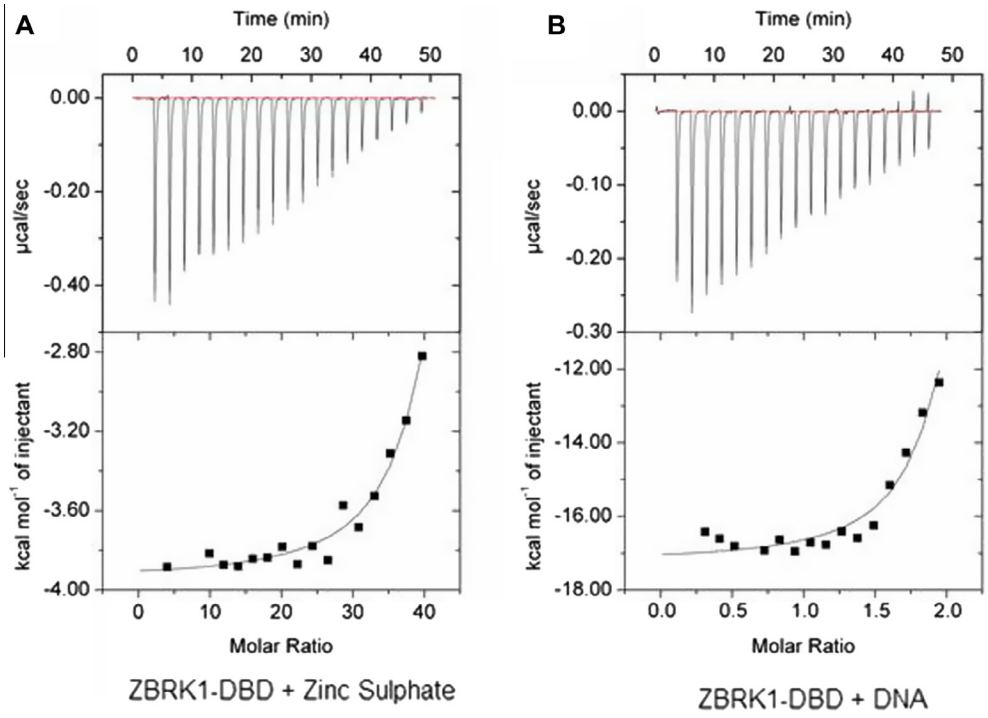


Fig. 4. Interaction of ZBRK1-DBD with: (A) zinc sulfate (B) DNA.

Table 1
Thermodynamic parameters derived from heat change: ZBRK1-DBD interaction with (1) zinc sulfate, (2) DNA.

	$ZnSO_4$	DNA
Interaction model	One set of sites	One set of sites
N sites	42.9 ± 0.0817	2.1 ± 0.0538
K_d (µm)	1.4	0.29
K (M^{-1})	7.06E5 ± 1.91E5	3.44E6 ± 1.26E6,
ΔH (cal/mol)	−3927 ± 29.88	−1.714E4 ± 192.6
ΔS (cal/mol/deg)	13.4	−27.6
Chi^2 value	3510	1.34E + 05

The negative enthalpy change of dissociation implies that association process will be endothermic and entropically driven. The presence of two transitions in thermal unfolding pathway is also consistent with presence of tetrameric species. The first transition with a T_m of 58 °C would be for the disassembly of the tetramer, while the second transition with a T_m of 68 °C would be for denaturation of the monomers. The stoichiometry of 1:42 for zinc binding of ZBRK1-DBD, containing eight zinc fingers is also supportive for tetrameric oligomerisation of ZBRK1-DBD (Table.1). Tetramer formation has also been established by chemical cross-linking experiments [15]. However, in these studies, the molecule was full length ZBRK1ΔK construct, and not just the DNA binding domain used here. In fact experiments conducted by these authors with a variety of C-terminal deletion constructs show that, C-terminal nine residues are absolutely essential for formation of stable tetramers. Against this background our observation that ZBRK1-DBD, which lacks 108 residues from the C-terminal region, is a homo tetramer in solution, is a very novel finding. It is conceivable that the nature of tetramers formed by full length ZBRK1 and ZBRK1-DBD are different. The biological implications of this observation are not yet clear. Tetrameric DNA-binding proteins are known. The transcription factor p53 forms a tetramer. One dimer part of the tetramer binds to half site of consensus DNA sequence & the binding of second dimer to the other half enhances the affinity to DNA almost 50-fold [25]. Mutations in the mostly hydrophobic residues responsible for tetramerization can inactivate the wild type protein suggesting the importance of tetramerization in p53 function [26] OB fold proteins that bind single stranded DNA are tetrameric [27,28]. Bioinformatics and EMSA analysis, however, shows that ZBRK1-DBD does not have the OB fold and does not bind ss DNA. (data not shown).

4.2. Zinc ions for DNA binding

The DNA-binding motif in ZF proteins comprises of two beta strands and an alpha helix, and the amino acid residues from the helix bind a triplet of bases in the major groove of the DNA helix [29–31]. Formation of the helix and proper positioning of the helical residues for optimal interaction is assumed to be ensured by coordination of zinc ion to conserved cysteine and histidine residues of the motif [32]. Therefore Zn ions are essential for repressor activity of C2H2 type of zinc fingers [33]. Interestingly, in the present study, ZBRK1-DBD binds to 15 bp consensus DNA sequence even in the absence of Zn ions (Fig. 2D, supplemental Fig. 1B). EMSA (data not shown) and ITC (Fig. 4A) also revealed interaction in absence of Zn ions.

4.3. DNA and protein conformation

It has been observed that the structure of DNA bound by ZF protein is altered from the standard B-geometry [34]. Molecular modeling has revealed that this distortion is necessitated by the requirement for neighboring fingers to bind optimally in the major groove. The actual distortions observed in the crystal structures of complexes are in agreement with predictions by modeling experiments. [35]. The relative intensity of CD signal at 275 nm is often taken as a signature for DNA geometry. An increase in its intensity suggests unwinding of DNA and/or displacement of base-pairs away from helix axis. However, in the present study on ZBRK1-DBD: DNA complex, the CD signals at 275 nm is significantly decreased. The geometry of the DNA in the complex is likely to be over wound compared to B-DNA. This probably may help ZBRK1

in repression activity, on interaction with DNA. Although previous reports mention no change occurs in ZF conformation on DNA binding, our results indicate subtle changes. Interestingly, binding of Zn and DNA to ZBRK1-DBD appear to have opposite effects on protein conformation. The n–$\pi*$ transition at 220 nm in CD reflects the α-helical content. The π–$\pi*$ excitation band at 208 nm is sensitive to whether the α-helix is involved in tertiary contacts [36]. The ratio $\theta_{222}/\theta_{208\,nm}$ is, therefore, regarded as an indicator of quaternary structure through inter-α-helix coiling/interaction. The binding of protein to DNA (Fig. 2C) suggests decrease in $\theta_{222}/\theta_{208\,nm}$ ratio indicating decrease in compactness of protein. On the other hand, ZBRK1-DBD with Zn ions indicated increase in $\theta_{222}/\theta_{208\,nm}$ ratio (Fig. 2B) suggesting incorporation of quaternary structure in presence of Zn as compared to Zn deficient protein.

In conclusion, our report indicates that ZBRK1-DBD forms tetramer. ZBRK1-DBD binds exclusively to double stranded DNA with stoichiometry of 1:2, even in the absence of zinc ions. Interaction of ZBRK1-DBD with $ZnSO_4$ and DNA leads to higher thermal stability. In the complex with ZBRK1-DBD, the DNA is distorted from standard B-geometry. The future goal of this work is to crystallize the protein for understanding the structure, which will help in using it as a drug target and also to engineer it for therapeutic purposes.

Acknowledgments

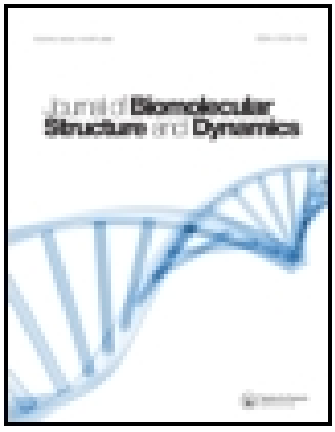
We thank Dr. Thomas G. Boyer, University of Texas Health Science Center at San Antonio, Texas, USA for MBP-ZBRK1ΔK construct. MVH thanks DAE for the award of Raja Ramanna Fellowship. AKV thanks TMC-ACTREC for funding.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.bbrc.2014.05.104.

References

[1] H.S. Zhang, D. Liu, Y. Huang, S. Schmidt, R. Hickey, D. Guschin, H. Su, I.S. Jovin, M. Kunis, S. Hinkley, Y. Liang, L. Hinh, S.K. Spratt, C.C. Case, E.J. Rebar, B.E. Ehrlich, P.D. Gregory, F.J. Giordano, A designed zinc-finger transcriptional repressor of phospholamban improves function of the failing heart, Mol. Ther. 20 (2012) 1508–1515.
[2] A. Klug, Towards therapeutic applications of engineered zinc finger proteins, FEBS Lett. 579 (2005) 892–894.
[3] F.D. Bushman, M.D. Miller, Tethering human immunodeficiency virus type 1 preintegration complexes to target DNA promotes integration at nearby sites, J. Virol. 71 (1997) 458–464.
[4] R.R. Beerli, B. Dreier, C.F. Barbas 3rd, Positive and negative regulation of endogenous genes by designed transcription factors, Proc. Natl. Acad. Sci. U.S.A. 97 (2000) 1495–1500.
[5] P.Q. Liu, E.J. Rebar, L. Zhang, Q. Liu, A.C. Jamieson, Y. Liang, H. Qi, P.X. Li, B. Chen, M.C. Mendel, X. Zhong, Y.L. Lee, S.P. Eisenberg, S.K. Spratt, C.C. Case, A.P. Wolffe, Regulation of an endogenous locus using a panel of designed zinc finger proteins targeted to accessible chromatin regions. Activation of vascular endothelial growth factor A, J. Biol. Chem. 276 (2001) 11323–11334.
[6] M.F. Rousseau-Merck, D. Koczan, I. Legrand, S. Moller, S. Autran, H.J. Thiesen, The KOX zinc finger genes: genome wide mapping of 368 ZNF PAC clones with zinc finger gene clusters predominantly in 23 chromosomal loci are confirmed by human sequences annotated in EnsEMBL, Cytogenet. Genome Res. 98 (2002) 147–153.
[7] L. Zheng, H. Pan, S. Li, A. Flesken-Nikitin, P.L. Chen, T.G. Boyer, W.H. Lee, Sequence-specific transcriptional corepressor function for BRCA1 through a novel zinc finger protein, ZBRK1, Mol. Cell 6 (2000) 757–768.
[8] L.F. Lin, C.H. Chuang, C.F. Li, C.C. Liao, C.P. Cheng, T.L. Cheng, M.R. Shen, J.T. Tseng, W.C. Chang, W.H. Lee, J.M. Wang, ZBRK1 acts as a metastatic suppressor by directly regulating MMP9 in cervical cancer, Cancer Res. 70 (2010) 192–201.
[9] T. Ahmed, S.M. Kelly, N.C. Price, A.J. Lawrence, Activation of phospholipase A(2) by long chain fatty acyl groups involves a novel unstable linkage, J. Biochem. 127 (2000) 871–875.
[10] S. Furuta, J.M. Wang, S. Wei, Y.M. Jeng, X. Jiang, B. Gu, P.L. Chen, E.Y. Lee, W.H. Lee, Removal of BRCA1/CtIP/ZBRK1 repressor complex on ANG1 promoter leads to accelerated mammary tumor growth contributed by prominent vasculature, Cancer Cell 10 (2006) 13–24.
[11] W. Tan, L. Zheng, W.H. Lee, T.G. Boyer, Functional dissection of transcription factor ZBRK1 reveals zinc fingers with dual roles in DNA-binding and BRCA1-dependent transcriptional repression, J. Biol. Chem. 279 (2004) 6576–6587.
[12] V. Garcia, J.M. Garcia, C. Pena, J. Silva, G. Dominguez, R. Rodriguez, C. Maximiano, R. Espinosa, P. Espana, F. Bonilla, The GADD45, ZBRK1 and BRCA1 pathway: quantitative analysis of mRNA expression in colon carcinomas, J. Pathol. 206 (2005) 92–99.
[13] J. Yun, W.H. Lee, Degradation of transcription repressor ZBRK1 through the ubiquitin–proteasome pathway relieves repression of Gadd45a upon DNA damage, Mol. Cell. Biol. 23 (2003) 7305–7314.
[14] H. Nishitsuji, M. Abe, R. Sawada, H. Takaku, ZBRK1 represses HIV-1 LTR-mediated transcription, FEBS Lett. 586 (2012) 3562–3568.
[15] W. Tan, S. Kim, T.G. Boyer, Tetrameric oligomerization mediates transcriptional repression by the BRCA1-dependent Kruppel-associated box-zinc finger protein ZBRK1, J. Biol. Chem. 279 (2004) 55153–55160.
[16] S.M. Kelly, N.C. Price, The use of circular dichroism in the investigation of protein structure and function, Curr. Protein Pept. Sci. 1 (2000) 349–384.
[17] F. Wolfram, E. Morris, C.W. Taylor, Three-dimensional structure of recombinant type 1 inositol 1,4,5-trisphosphate receptor, Biochem. J. 428 (2010) 483–489.
[18] C. Dulac, A.A. Michels, A. Fraldi, F. Bonnet, V.T. Nguyen, G. Napolitano, L. Lania, O. Bensaude, Transcription-dependent association of multiple positive transcription elongation factor units to a HEXIM multimer, J. Biol. Chem. 280 (2005) 30619–30629.
[19] R.W. Woody, Circular dichroism, Methods Enzymol. 246 (1995) 34–71.
[20] T. Rao, G. Ruiz-Gomez, T.A. Hill, H.N. Hoang, D.P. Fairlie, J.M. Mason, Truncated and helix-constrained peptides with high affinity and specificity for the cFos coiled-coil of AP-1, PLoS One 8 (2013) e59415.
[21] M. Vorlickova, Conformational transitions of alternating purine–pyrimidine DNAs in perchlorate ethanol solutions, Biophys. J. 69 (1995) 2033–2043.
[22] J. Kypr, M. Vorlickova, Circular dichroism spectroscopy reveals invariant conformation of guanine runs in DNA, Biopolymers 67 (2002) 275–277.
[23] C.A. Sprecher, W.A. Baase, W.C. Johnson Jr., Conformation and circular dichroism of DNA, Biopolymers 18 (1979) 1009–1019.
[24] R.W. Alston, L. Urbanikova, J. Sevcik, M. Lasagna, G.D. Reinhart, J.M. Scholtz, C.N. Pace, Contribution of single tryptophan residues to the fluorescence and stability of ribonuclease Sa, Biophys. J. 87 (2004) 4036–4047.
[25] K.G. McLure, P.W. Lee, How p53 binds DNA as a tetramer, EMBO J. 17 (1998) 3342–3350.
[26] P. Chène, The role of tetramerization in p53 function, Oncogene 20 (2001).
[27] S. Raghunathan, C.S. Ricard, T.M. Lohman, G. Waksman, Crystal structure of the homo-tetrameric DNA binding domain of *Escherichia coli* single-stranded DNA-binding protein determined by multiwavelength X-ray diffraction on the selenomethionyl protein at 2.9-A resolution, Proc. Natl. Acad. Sci. U.S.A. 94 (1997) 6652–6657.
[28] D.L. Theobald, R.M. Mitton-Fry, D.S. Wuttke, Nucleic acid recognition by OB-fold proteins, Annu. Rev. Biophys. Biomol. Struct. 32 (2003) 115–133.
[29] S.A. Wolfe, L. Nekludova, C.O. Pabo, DNA recognition by Cys2His2 zinc finger proteins, Annu. Rev. Biophys. Biomol. Struct. 29 (2000) 183–212.
[30] D. Lu, C2H2 zinc-finger recognition of biomolecules, Yao Xue Xue Bao 48 (2013) 834–841.
[31] P.V. Pedone, R. Ghirlando, G.M. Clore, A.M. Gronenborn, G. Felsenfeld, J.G. Omichinski, The single Cys2-His2 zinc finger domain of the GAGA protein flanked by basic residues is sufficient for high-affinity specific DNA binding, Proc. Natl. Acad. Sci. U.S.A. 93 (1996) 2822–2826.
[32] R.N. De Guzman, H.Y. Liu, M. Martinez-Yamout, H.J. Dyson, P.E. Wright, Solution structure of the TAZ2 (CH3) domain of the transcriptional adaptor protein CBP, J. Mol. Biol. 303 (2000) 243–253.
[33] E.E. Zheleznova, J.H. Crosa, R.G. Brennan, Characterization of the DNA- and metal-binding properties of *Vibrio anguillarum* fur reveals conservation of a structural Zn(2+) ion, J. Bacteriol. 182 (2000) 6264–6267.
[34] L. Nekludova, C.O. Pabo, Distinctive DNA conformation with enlarged major groove is found in Zn-finger-DNA and other protein–DNA complexes, Proc. Natl. Acad. Sci. U.S.A. 91 (1994) 6948–6952.
[35] M. Elrod-Erickson, M.A. Rould, L. Nekludova, C.O. Pabo, Zif268 protein-DNA complex refined at 1.6 A: a model system for understanding zinc finger-DNA interactions, Structure 4 (1996) 1171–1180.
[36] N.E. Zhou, C.M. Kay, R.S. Hodges, Synthetic model proteins. Positional effects of interchain hydrophobic interactions on stability of two-stranded alpha-helical coiled-coils, J. Biol. Chem. 267 (1992) 2664–2670.



Journal of Biomolecular Structure and Dynamics
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tbsd20

Functional assessment of intrinsic disorder central domains of BRCA1

Lumbini R. Yadav[a], Sharad Rai[a], M.V. Hosur[a] & Ashok K. Varma[a]
[a] Tata Memorial Centre, Advanced Centre for Treatment, Research and Education in Cancer, Kharghar, Navi Mumbai, Maharashtra 410 210, India
Published online: 23 Jan 2015.

Journal of Biomolecular Structure and Dynamics, 2015
http://dx.doi.org/10.1080/07391102.2014.1000973

Functional assessment of intrinsic disorder central domains of BRCA1

Lumbini R. Yadav, Sharad Rai, M.V. Hosur and Ashok K. Varma*

Tata Memorial Centre, Advanced Centre for Treatment, Research and Education in Cancer, Kharghar, Navi Mumbai, Maharashtra 410 210, India

Communicated by Ramaswamy H. Sarma

(*Received 14 October 2014; accepted 16 December 2014*)

The most studied function of BRCA1 is that of tumor suppression through its role in DNA repair and transcription regulation. Germline mutations discovered in a larger cohort of patients, abrogate BRCA1 interactions with reported cellular partners, and are responsible for breast and ovarian cancer. The different functional regions of BRCA1 interact with nearly 30 different cellular partners. Thus, it becomes clinically significant to understand the detailed protein-protein interactions associated with functional regions of BRCA1. Different overlapping central domains of BRCA1 have been characterized using *in silico*, *in vitro* and biophysical approaches. To our conclusions, it has been observed that central domains of BRCA1 are intrinsically disordered and has large hydrodynamic radius with random coil like structures.

Keywords: BRCA1 central domains; intrinsic disorder; protein-protein interactions

1. Introduction

BRCA1 is a cancer susceptibility gene responsible for hereditary breast and ovarian cancer syndrome. The basic characteristic of family with (HBOC) syndrome is early onset of breast and ovarian cancer (Goldgar et al., 2004). Among cancer patients, 3–5% of breast cancer and 10% of ovarian cancer cases are estimated to be due to alterations discovered in *BRCA1* and *BRCA2* gene (Risch et al., 2006; Rubin et al., 1998). However, families with inherited mutations in *BRCA1* has 70–80% risk of developing breast cancer and 50–60% risk of developing ovarian cancer (Roy, Chun, & Powell, 2012).

BRCA1 is a pleiotropic protein that performs a variety of function in the cells. The well-known tumor suppressor function is because of its involvement in DNA repair through homologous recombination and non-homologous end joining mechanisms. (Bau, Mau, & Shen, 2006; Moynahan, Chiu, Koller, & Jasin, 1999). It also helps in transcription regulation, transcriptional repair, cell-cycle regulation, chromatin remodeling and genetic recombination (Hu, Hao, & Li, 1999; Narod & Foulkes, 2004; Wang et al., 2000). Inability of BRCA1 to perform repair function, due to deleterious mutations, leads to chromosomal instability that increases the chances of malignant transformation. So, it becomes necessary to structurally characterize the BRCA1 protein so as to understand the molecular basis of cancer progression. This will also help in genetic counseling and disease management of HBOC cases that predispose the carrier to high risk of cancer.

BRCA1 comprises 1863 amino acids, distributed into distinct functional domains (Miki et al., 1994): the N-terminal, ring finger domain (RING) responsible for ubiquitin ligase activity (Xia, Pao, Chen, Verma, & Hunter, 2003), the C-terminal BRCT domain responsible for recognition of specific phosphorylated binding partners, and a central region. The RING domain and the BRCT domain have been extensively characterized structurally. It has been reported that the central region of BRCA1 (residues 250–1200) interacts nonspecifically with DNA, and has strong affinity for cruciform DNA (Cantor et al., 2001; Paull, Cortez, Bowers, Elledge, & Gellert, 2001; Wang et al., 2007). It is also reported that this region functionally interacts with many different proteins, including ZBRK1, p53, RAD50 and RAD51. However the central region of BRCA1 has not been characterized at the atomic level.

Protein that lacks 3-D structures under physiological conditions *in vitro* are intrinsically disordered protein. Globular proteins exist in four different conformations i.e. folded, molten globule, pre-molten globule and unfolded (Ptitsyn, 1995). Intrinsically disordered proteins have native molten globules, native pre-molten globules and native coils (Dunker & Obradovic, 2001; Uversky, Li, & Fink, 2001; Uversky, 2003; Uversky, 2009). Furthermore, IDPs are enriched with negatively charged amino acids and has less number of hydrophobic residues (Gast et al., 1995; Uversky, Gillespie, & Fink, 2000; Weinreb, Zhen, Poon, Conway & Lansbury 1996) (Uversky, 2002b). However folding pattern of disordered

*Corresponding author. Email: avarma@actrec.gov.in

protein may change when it binds to other cellular partners (Dunker et al., 2001; Dunker & Obradovic, 2001; Wright & Dyson, 1999). These disordered proteins act as hubs for signaling interactions and regulation.

The protein phosphorylation site is also known to be predominantly enriched with disorder-promoting residues (Iakoucheva et al., 2004). Phospho tyrosine-binding domain of numb protein (Zwahlen, Li, Kay, Pawson, & Forman-Kay, 2000) and the intrinsically disorder CaM binding domains undergo disorder-to-order transitions upon Ca^{2+} binding (Urbauer, Short, Dow, & Wand, 1995). The studies on differentially spliced gene and associated proteins demonstrated that regions affected by alternative splicing are often intrinsically disordered. This provides a wide possibility of splicing without affecting the structured part of protein (Romero et al., 2006). Applications of intrinsically disordered proteins in recombinant protein purification, polypeptide nanoparticles for drug delivery, and as thermally sensitive carriers of active peptides and proteins has been well established (Uversky, 2014). Furthermore, intrinsically disordered proteins as a bio-sensing system have been demonstrated using BRCA1 as a model system (Cissell, Shrestha, Purdie, Kroodsma, & Deo, 2008).

The structure of the central region of BRCA1 has been investigated previously (Mark et al., 2005). These authors have subdivided the central region into 21 fragments of varying lengths, and have expressed each fragment with a histidine-tag. Out of the 21 different constructs generated 16 were designed such that their N and C termini did not occur in a region of predicted secondary structure. The remaining five were selected based on the regions reported in the literature as binding to other proteins or DNA. These tagged constructs were expressed and purified and were used for the biophysical studies without removing the tag. The experimental studies suggest that these tagged protein fragments were intrinsically unstructured. They further suggest that this disorder may help to encode multiple-functionality in the BRCA1 sequence. It is conceivable that the samples were found to be unstructured because of the domain boundaries being incorrectly identified (Huth, Mountjoy, Perini, Bedows, & Ruddon, 1992).

We have identified different set of fragments of BRCA1 that are reported to bind DNA and specific proteins. The reported domains of BRCA1 which include (260–553) binds to HP1γ involved in gene silencing (Choi, Park, & Lee, 2012), (758–1064) interacts with RAD51 (Scully et al., 1997), (341–748) bind ZBRK1, RAD50 and RAD51 (Tan, Zheng, Lee, & Boyer, 2004; Zhong et al., 1999), (502–802) interacts with transcriptional activation domain of STAT1a (Ouchi, Lee, Ouchi, Aaronson, & Horvath, 2000). Further, we have removed the tags used to aid in the purification process. Here, we report expression, purification and biophysical studies on these fragments of BRCA1, which are expected to be more native like. Our results also indicate that these fragments are intrinsically unstructured.

2. Results

2.1. *Central regions of BRCA1 is disordered*

The chemical composition of a protein is very often used to predict, whether, the protein is natively folded or unfolded (Gast et al., 1995; Weinreb et al. 1996). Sequence composition analysis for all the four constructs revealed that they have less of amino acids I, L, V, W, F, Y, C, and N and are enriched with E, K, R, G, Q, S, P, and A which may account for low overall hydrophobicity and high net charge characteristics of natively unfolded protein. This has been demonstrated by the plot of absolute mean net charge vs. mean scaled hydropathy in PONDR (Table 1(A)) (Dunker et al., 2001; Dunker, Brown, & Obradovic, 2002; Uversky, Gillespie, & Fink, 2000). For the four constructs (CRC1–4) results of such analysis are given in (Table 1(B)). The pI values were on the extreme basic side for CRC-3 and in the acidic side for all others CRC-1, CRC-2 and CRC-4 domains indicating these regions to have characteristics of intrinsically unstructured proteins. The probability of disorder as predicted by Cspritz primary disorder-predictor server is given in (Table 1(B)). It is significant that the predicted degree of disorder is substantially high for all the four constructs, ranging from 50 to 70%. Composition profiler output on the use of background set PDB select 25: a set of ordered proteins, predicted the presence of more number of residues responsible for disordered proteins (supplementary Figure 1). The same has been confirmed by ANCHOR (Supplementary Figure 2). ANCHOR predicts disordered regions using estimated energy calculations and sensitivity to the structural environment (Dosztanyi, Meszaros, & Simon, 2009;

Table 1A. The plot of absolute mean net charge vs. mean scaled hydropathy. Yellow colored squares are the four BRCA1 CRC proteins; green color is the central region from amino acid 260–1064.

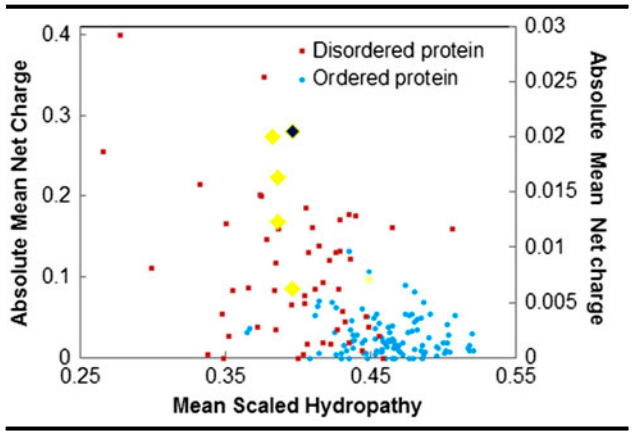


Table 1B. Sequence analysis of different domains of BRCA1 from ExPASy, ProtParam.

BRCA1 region	pI	Hydropathicity	%Disorder (Cspritz)	ligand
(260-553)	5.97	−0.934	60.40	DNA, p53
(341-748)	6.33	−1.033	71.07	ZBRK1
(502-802)	8.97	−1.038	87.70	DNA
(758-1064)	5.82	−0.764	51.46	STAT1

Mészáros, Simon, & Dosztányi, 2009). The probability of disorder as a function of amino acid position in the sequence was investigated using Meta servers (MFDp and Meta disorder). MFDp is an ensemble of 3 support vector machines which predicts short, long and generic disordered regions. Meta Disorder server combines prediction from 13 different servers and uses the combined output to improve the prediction accuracy. These servers take into consideration physicochemical factors such as chemical nature, secondary structure propensity, sequence information, solvent accessibility, backbone dihedral torsion angles, residue flexibility and B-factors if available. The predictions from these meta servers are shown in (Figure 1). It is interesting that these results confirm primary server prediction, that the entire central region of BRCA1 is unstructured.

2.2. Purification of different central regions of BRCA1

Each of the four central regions of BRCA1 was purified using affinity chromatography. SDS-gel

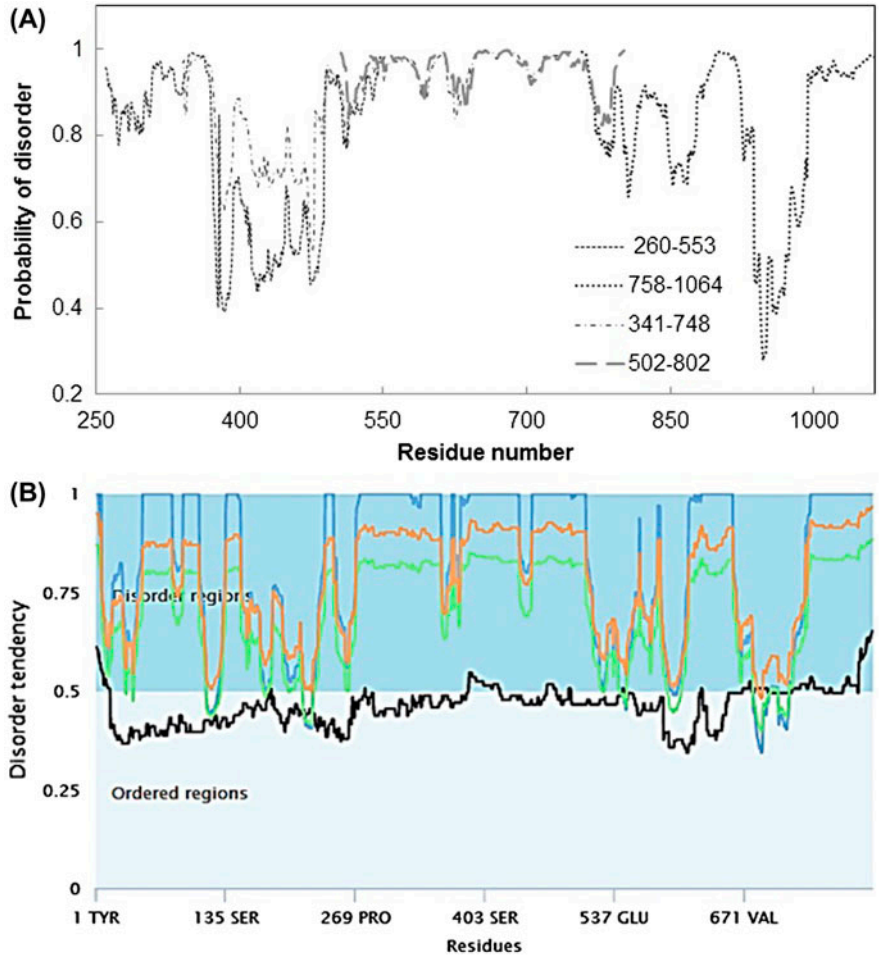


Figure 1. *In-silico* prediction data from (A): MFDp of the four constructs CRC(1–4), (B): Meta Disorder server analysis considering the entire central region from 260 to 1064. Here *x* axis residue 1 corresponds to residue number 260 on and so on BRCA1 protein.

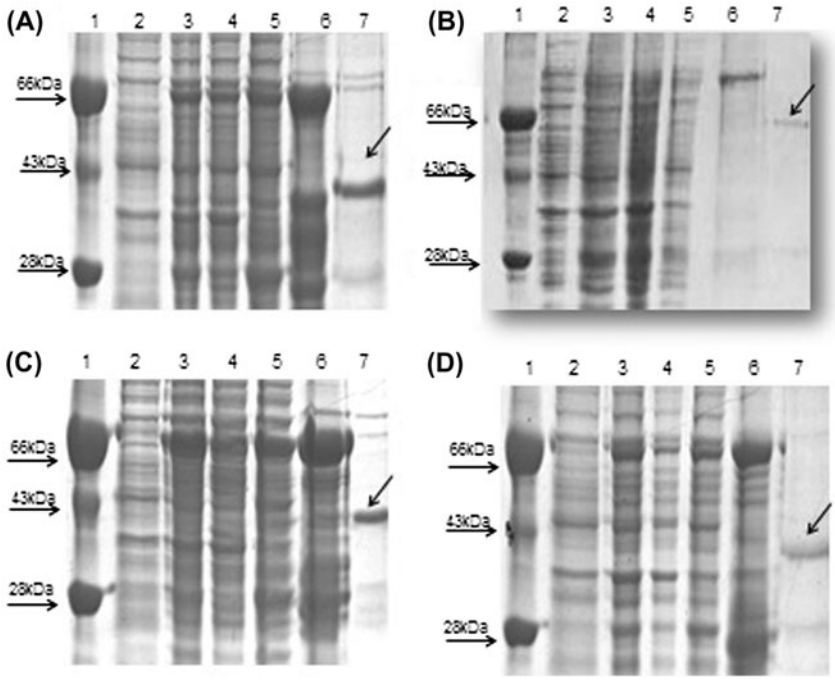


Figure 2. Purification of central regions of BRCA1. (A): CRC-1, (B): CRC-2, (C): CRC-3, (D): CRC-4 Lane 1: Marker, 2: Uninduced whole cell, 3: Induced whole cell, 4: Pellet, 5: Induced Soluble Fraction, 6: Beads after binding, 7: Purified elution fraction.

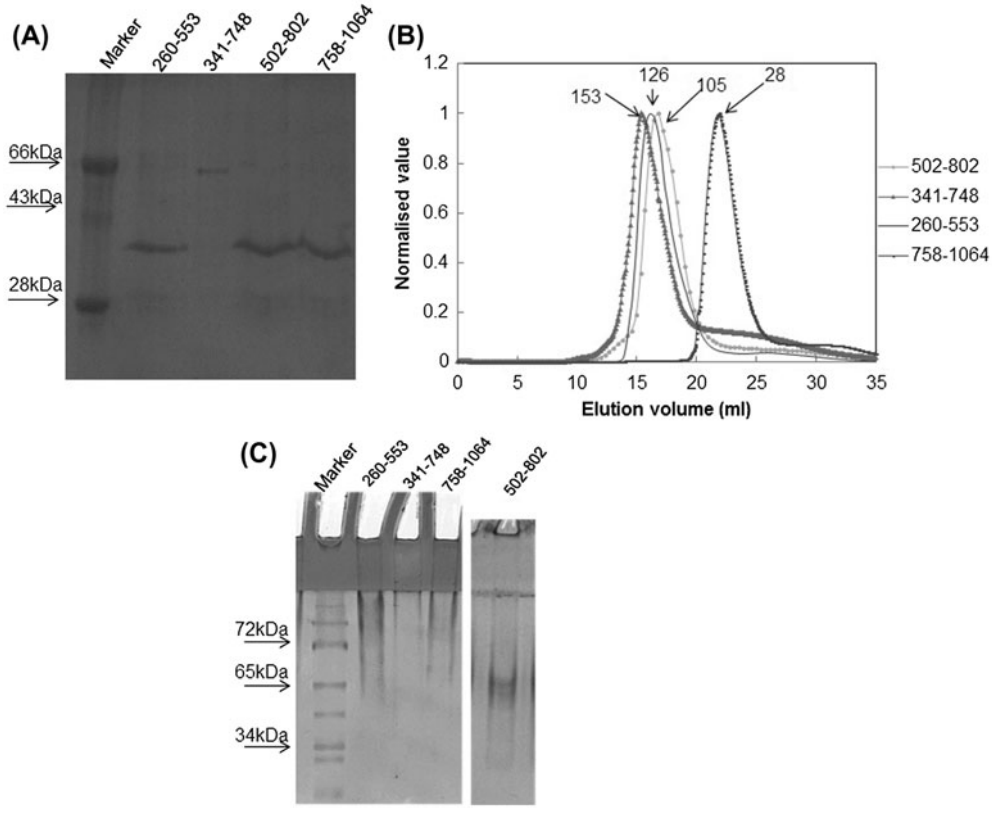


Figure 3. Oligomeric characterisation. (A) FPLC purified protein loaded on 10% SDS gel, (B) Gel filtration profile for the central domains. Molecular weight estimated from standards, and (C) Native gel profile of the BRCA1 central domain constructs.

however showed presence of a 60 kDa band in all the central regions constructs. To remove the contaminant washing with high concentration of sodium chloride, Triton X-100, ATP, magnesium chloride and β-mercaptoethanol have been performed (Figure 2). The final purified protein was obtained using gel permeation chromatography. Mass spectrometry based peptide mass finger printing was performed to confirm, different purified BRCA1 CRC proteins (Supplementary Figures 3–6).

2.3. Oligomeric characterization

Different central regions of BRCA1 were FPLC purified using superdex 200 column (Figure 3(A)). The retention volumes in FPLC chromatogram for the central regions of BRCA1 indicated them to form higher order oligomers except for CRC-4. The plot of log10 (molecular weight) vs. (elution volume) was used to estimate the apparent size of the eluted central regions of BRCA1. These apparent sizes of central regions of BRCA1 are labeled in (Figure 3(B)). The partition coefficient Kav, for different central region was calculated from the elution volume (Dulac et al., 2005; Wolfram, Morris, & Taylor, 2010). Stoke's radius (R_S) for different central region was interpolated from a plot of partition coefficient (Kav) vs. Stoke's radii (Blake-Hodek, Cassimeris & Huffaker 2010; Portier 1975; Sá-Moura et al. 2013). The dependency of log (R_S) vs. log (M) curves to different conformation of protein is described by straight line equation. Here, the straight line equation for BRCA1 CRC is $((-0.559) + (0.7167)\cdot\log(M))$. This equation matches with the protein which is natively unfolded indicating high hydrodynamic radius of BRCA1 CRC (Uversky, 2002b).

However one is not sure if the higher size is due to oligomerisation or due to unstructured protein. This point was further probed by native gel electrophoresis. In this gel each protein gave a smeared pattern with no bands in the gel suggesting the absence of discrete olgomeric species (Figure 3(C)). Hence, looking at the lower retention volume on the size exclusion column and smear pattern of native gel, it can be concluded that the central region has an intrinsically disordered structure.

2.4. Secondary structural analysis

Far UV CD spectra of CRC-1 to CRC-4 are shown in (Figure 4(A)). It may be seen that these proteins show negative ellipticity at λ_{200} nm. It has been reported that random coils or intrinsically-disordered proteins exhibit a negative ellipticity near λ_{200} nm and ~zero ellipticity around λ_{218} nm (Schweers, Schonbrunn-Hanebeck, Marx, & Mandelkow, 1994; Weinreb et al., 1996). Central regions of BRCA1 thus show CD spectra characteristic of random coil structures.

Further to confirm for absence of secondary structural characteristics, temperature induced changes in ellipticity was monitored at $\lambda = 222$ nm. A slight increase in negative ellipiticty was observed at around 70 °C (Figure 4(B)). This effect may be attributed to increased strength of the hydrophobic interaction at higher temperatures, which may act as a driving force for folding (Millett, Doniach, & Plaxco, 2002; Uversky, 2002a, 2009). This slight increase in negative ellipticity due to increase in temperature is defined as "turned out response" as witnessed with other proteins like α-synuclein (Uversky et al., 2001), caldesmon (Permyakov, Millett, Doniach, Permyakov, & Uversky, 2003), and phosphodiesterase γ-subunit (Uversky et al., 2002).

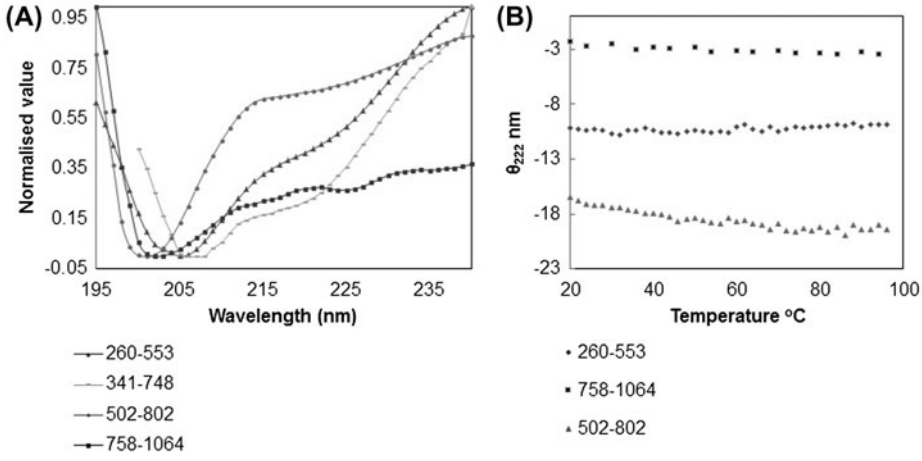


Figure 4. Circular dichroism spectroscopy. (A) Far UV CD scans of all the central region of BRCA1 and (B) Plot of ellipticity at θ_{222} nm vs. Temperature.

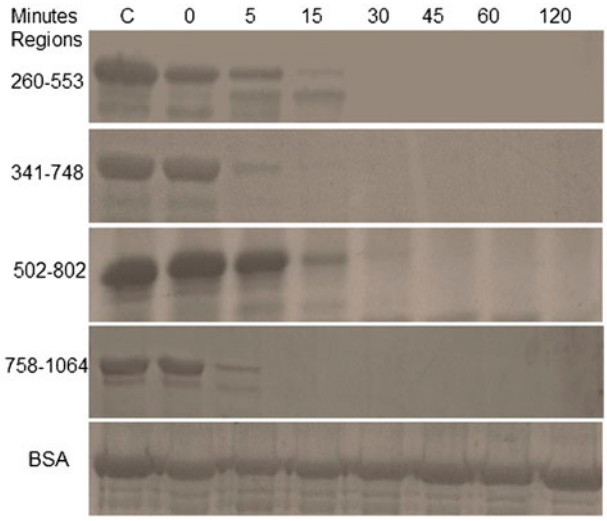


Figure 5. Protease sensitivity assay Central region purified protein treated with trypsin at concentration of 1:4000. All the different domains were incubated at 37 °C for different time points of 0, 5, 15, 30, 45, 60, 120 min, C is the untreated protein.

2.5. *Protease sensitivity*

Limited proteolysis is often used to explore disordered regions in proteins. This technique relies on extended conformation of protease target site, enhanced backbone flexibility and local unfolding of proteolytic site. Proteins with disordered structure are sensitive to degradation at very low protease concentration (Bouvier & Stafford, 2000; Iakoucheva et al., 2001). To check the sensitivity of this central region of BRCA1, purified proteins were subjected to trypsin degradation. The ratio of trypsin to protein used was 1:4000. All the purified central regions were degraded within 15 min of incubation at 37 °C, unlike bovine serum albumin which, under identical conditions, was not degraded even after 2 h of incubation (Figure 5). This sensitivity of protein to such a low concentration of protease indicates BRCA1 central region have a disordered structure accessible for protease degradation.

3. Discussion

The aim of this study was to determine the crystal structure of the central DNA binding region of BRCA1. Initial preliminary study was done to determine the structural elements presents in these different construct. Bioinformatics analysis of sequences of BRCA1 central regions carried out using primary and meta disorder servers, PONDR and Composition profiler classified them as intrinsically unstructured proteins based on their mean-hydrophobicity, mean charge at pH 7, presence of flexible loops, predicted secondary structure and missing coordinates in crystal structures.

The central regions of BRCA1 co-purified along with another protein of approximate molecular weight 60 kDa. This contaminant was identified as a bacterial chaperon, because removal of that protein required extensive wash of affinity resin with wash buffer containing Triton X-100, NaCl and ATP with $MgCl_2$. Co-purification of chaperons, especially Hsp70 is reported to assist in solubilising the aggregation-prone, intrinsic disorder proteins (Kyratsous, Silverstein, DeLong, & Panagiotidis, 2009). Hsp70 is known to bind extended form of protein in order to protect it from proteases and help in folding (Hendrick & Hartl, 1995). This co-purification of chaperons & purification with $MgCl_2$ and ATP suggests that the central regions of BRCA1 have an extended form which is stabilised by bacterial chaperons.

All the four proteins were found to be extremely prone to degradation during purification, supporting their unstructured nature. Limited proteolysis assay also showed high sensitivity of the four constructs to trypsin digestion. However, there appears to be a difference in the rates at which these four proteins are hydrolyzed by trypsin. The CRC-1 and CRC-3, which are involved in binding DNA, appear to be more stable compared CRC-2 and CRC-4. This observation is interesting in view of the fact that known DNA-binding modes of proteins involve protein secondary structures. Molecular simulation studies have suggested that unfolding of more than 10 residues of target is necessary to fit target residue's backbone into trypsin's active site (Hubbard, Eisenmenger, & Thornton, 1994). This sensitivity to protease observed here suggests extended conformation of target leading to unstructured conformation.

4. Materials and methods

All the enzymes used for cloning were procured from NEB (US). Oligonucleotides were purchased from Sigma Aldrich (US). Most of the other chemicals used were from Hi Media, Qualigens and Sigma-Aldrich. Bacterial culture media, IPTG and ampicillin were from Hi Media (India).

4.1. In-silico *analysis*

In-silico analysis was performed to evaluate the sequence and structure composition for disorder prediction Composition profiler analysis was performed to detect the differences in BRCA1 CRC against the PDB Select 25 reference database (Vacic, Uversky, Dunker, & Lonardi, 2007; Vucetic et al., 2005). PONDR analysis was also performed to check the positioning of BRCA1 CRC residues in the hydropathy plot (Romero, Obradovic, Kissinger, Villafranca, & Dunker, 1997). Structural

BRCA1 domains	F→ Forward, and R→Reverse Primers
BRCA1 (260–553)	BRCA1_F
CRC-1	5′-GTCGGATCCGAGAACCTGTACTTTCAGGGTAAGTATCAGGGTAGTTCTGTT-3′
	BRCA1_R
	5′-GTCCTCGAGCTATTAATGACCACTATTAGTAATATT-3′
BRCA1(341–748)	BRCA1_F
CRC-2	5′-GTCGGATCCGAGAACCTGTACTTTCAGGGTGATCTGAATGCTGATCCCCTG-3′
	BRCA1_R
	5′-GTCGAATTCCTATTATTTGGGGTCTTCAGCATTATT-3′
BRCA1(502–802)	BRCA1_F
CRC-3	5′-GTCGGATCCGAGAACCTGTACTTTCAGGGTTTAAAGCGTAAAAGGAGACCT-3′
	BRCA1_R
	5′-GTCGAATTCCTATTACACACATTTATTTGGTTCTGT-3′
BRCA1(758–1064)	BRCA1_F
CRC-4	5′-GTCGGATCCGAGAACCTGTACTTTCAGGGTCAAACTGAAAGATCTGTAGAG-3′
	BRCA1_R
	5′-GTCGAATTCCTATTAACTGGAACCTATTTCATTAA-3′

disorder in the four constructs of BRCA1 was also analyzed by submitting the amino acid sequence to following different online servers: Primary servers like Cspritz (Walsh et al., 2011), Anchor (Dosztanyi et al., 2009), and Spine D (Zhang et al., 2012), the two meta servers MFDp (Mizianty et al., 2010) and MetaDisorder (Kozlowski & Bujnicki, 2012) was used. The prediction algorithms of these servers take into account sequence complexity, intramolecular interactions, solvent accessibility, torsion angle fluctuation and secondary structure.

4.2. *Gene-cloning of different regions of BRCA1*

The following four regions of BRCA1 were cloned in different bacterial expression vectors: BRCA1 (260–553) in pGEX-4T, BRCA1 (502–802) and BRCA1 (758–1064) in pGEX-kT (generous gift from John A Ladias, BIDMC, USA), and BRCA1 (341–748) in pET41 a + vector from cDNA of *BRCA1* (generous gift from Prof. Richard Baer, Columbia University, USA). Primers used to amplify the different constructs are given in table below. The protein fragments encoded by these constructs are described hereafter for convenience as CRC-1, CRC-2, CRC-3 and CRC-4.

All the clones were PCR amplified and ligated in the corresponding vectors. Each clone was sequenced using 3500 Genetic Analyser (Applied Bio systems).

4.3. *Protein expression and purification*

For protein expression, CRC-1 and CRC-4 were successfully transformed in *E. coli* bacterial strain BL21 (DE3), while *CRC-2* and *CRC-3* were transformed in *E. coli* Rosetta2 (DE3) strain. A single transformed colony was inoculated in Luria Bertani broth with appropriate antibiotics. Bacterial cultures were grown at 37 °C until $O.D_{600}$ reached between 0.6 and 0.8. CRC-2 and CRC-3 were induced with 0.4 mM IPTG and incubated for 16 h at 18 °C whereas CRC-1 and CRC-4 were induced with 0.4 mM IPTG and incubated for 3 h at 37 °C, after which the cells were harvested. Harvested bacterial pellet of CRC-1, CRC-2 and CRC-4 were resuspended in lysis buffer I containing 50 mM Tris pH 7.5, 300 mM NaCl, 1 mM EDTA, 5% Glycerol, 0.1% Triton X 100, 20 mM β-mercaptoethanol and protease inhibitor. Lysis buffer II containing 20 mM phosphate pH 6.5, 500 mM NaCl, 1 mM EDTA, 2% Glycerol, 0.1% Triton X 100, 20 mM β-mercaptoethanol and protease inhibitor were used for purification of CRC-3. Resuspended cells were lysed by sonication, for three times with 1 min duty cycle at 60 pulse rate. Cell debris was removed by centrifugation for 40 min at 36,000 g and 4 °C. The supernatant was passed through the pre-equilibrated glutathione resin for affinity purification. The non-specifically bound proteins for CRC-2 and CRC-4 were removed by washing the resin with wash buffer containing lysis buffer I with 1 M NaCl, 0.5% Triton X 100, 20 mM $MgCl_2$ and 5 mM ATP (Rial & Ceccarelli, 2002). CRC-3 resin was washed by adding lysis buffer II with 1 M NaCl and protease inhibitor in phosphate-lysis buffer, while CRC-1 bound resin was washed by adding 1 M NaCl, 1% Triton X-100 and protease inhibitor in lysis buffer I. To get protein in native form, on-beads cleavage using TEV protease was performed. The concentrated protein was further purified by gel filtration using AKTA-FPLC, and Superdex-200 column.

4.4. *Size exclusion chromatography*

BRCA1 CRC was purified to homogeneity using AKTA explorer system (GE Healthcare). Concentrated protein was passed through Superdex 200 column. Protein standards (BSA, lysozyme, and carbonic anhydrase) were eluted from the same column under similar conditions to prepare the standard plot for stokes radii estimation. The sample was chromatographed at a flow rate of

0.5 ml/min, and 1 ml fractions were collected. Standard plot of (−log Kav)*1/2 vs. stokes radii was used to estimate stokes radii of BRCA1 CRC.

Stokes radii for BRCA1 CRC was determined from elution parameter Kav = $(V_e - V_0)/(V_t - V_0)$. V_e is elution volume, V_t is total volume and V_0 is void volume of column (Dulac et al., 2005; Wolfram et al., 2010).

4.5. Mass spectrometry

Mass spectrometry (MALDI-TOF) was performed to confirm the identity of different central domains of BRCA1. Affinity chromatography purified protein loaded onto SDS-PAGE gel, and size exclusion chromatography purified protein was used for, In gel and In solution trypsin digestion respectively (Shevchenko, Wilm, Vorm, & Mann, 1996). This digested peptides were mixed with sinapinic acid matrix and spotted onto the MALDI plate. Peptide mass fingerprinting were performed using MALDI-TOF-MS spectroscopy (Ultraflex Bruker Daltonics system). MS spectrum was analysed in Flex analysis 3.0 and Bio-tools software (Bruker Daltonik, Germany). The MS peaks were further analysed for searching against SwissProt/NCBI nr database using MASCOT search engine (Perkins, Pappin, Creasy, & Cottrell, 1999). The score shown is with value of $p \leq 0.05$.

4.6. Limited proteolysis assay

Limited proteolysis of CRC-1, CRC-2, CRC-3 and CRC-4 with trypsin was carried out to probe the stability of the purified protein. Protein of concentration of 8–12 µM and 3 µl of trypsin at a concentration of 20 ng/µl were used for the assay. The ratio of trypsin to protein used was 1:4000. Protein was incubated with trypsin at 37 °C, and 60 µl aliquot of the reaction mixture was taken at time points of 0, 5, 15, 30, 45, 60 and 120 min. The reaction was stopped by adding 1 µl of 1 mM PMSF. Proteolysis digests were analyzed on 10% SDS/PAGE.

4.7. Circular dichroism

Far UV CD scan for the BRCA1 central region CRC-1, CRC-2, CRC-3 and CRC-4 were recorded using CD spectrometer (Jasco J-815). All the measurements were baseline corrected. The scan was recorded at 20 °C for the λ range of 250–195 nm, with a scanning speed of 20 nm/min and 3 accumulations .Thermal stability studies for BRCA1 CRC constructs were carried out in the temperature range of 20–95 °C, and over a wavelength range of 240–195 nm. All the samples were incubated for 200 s, and scans were recorded at a data pitch of 5 °C. Plot of θ_{222} nm as a function of temperature was analyzed to check for secondary structural changes.

4.8. Native gel electrophoresis

To check the folding pattern of proteins and its homogeneity, 5 µM of purified protein sample was loaded on acidic (CRC-1, CRC-2 and CRC-4) and basic (CRC-3) native polyacrylamide gel. Protein concentration was estimated from molar extinction coefficient. For basic native polyacrylamide gel electrophoresis, 7% polyacrylamide gel was prepared in potassium acetate buffer. β-alanine acetic acid buffer at pH 4.3 was used for electrophoresis. Methyl green was used as a tracking dye. For acidic gel, 10% polyacrylamide gel was prepared in Tris-Cl buffer. Tris glycine buffer was used for electrophoresis.

Abbreviations

HBOC	Hereditary breast and ovarian cancer syndrome
BRCA1	Breast Cancer associated gene 1
MFDp	Multilayered Fusion-based Disorder predictor
CRC	Central Region Construct
MALDI-TOF	Matrix-assisted laser desorption/ionization time-of-flight mass spectrometer

Supplementary material

The supplementary material for this paper is available online at http://dx.doi.10.1080/07391102.2014.1000973.

Acknowledgements

The authors would like to acknowledge Mahamaya Biswal, Neha Behare and Hafiza Mahadik for their help in cloning and purification. The authors would like to acknowledge the Mass Spectrometry, BTIS facility of TMC-ACTREC.

Funding

Funding for this study was supported by DBT (BT/513/NE/TBP/2013), Seed in Air grant from TMC. MVH thanks DAE, for the award of Raja Ramanna fellowship.

References

Bau, D. T., Mau, Y. C., & Shen, C. Y. (2006). The role of BRCA1 in non-homologous end-joining. *Cancer Letters, 240*, 1–8.

Blake-Hodek, K. A., Cassimeris, L., & Huffaker, T. C. (2010). Regulation of microtubule dynamics by Bim1 and Bik1, the budding yeast members of the EB1 and CLIP-170 families of plus-end tracking proteins. *Molecular Biology of the Cell, 21*, 2013–2023.

Bouvier, M., & Stafford, W. F. (2000). Probing the three-dimensional structure of human calreticulin. *Biochemistry, 39*, 14950–14959.

Cantor, S. B., Bell, D. W., Ganesan, S., Kass, E. M., Drapkin, R., Grossman, S., … Livingston, D. M. (2001). BACH1, a novel helicase-like protein, interacts directly with BRCA1 and contributes to its DNA repair function. *Cell, 105*, 149–160.

Choi, J. D., Park, M. A., & Lee, J. S. (2012). Suppression and recovery of BRCA1-mediated transcription by HP1γ via modulation of promoter occupancy. *Nucleic Acids Research, 40*, 11321–11338.

Cissell, K. A., Shrestha, S., Purdie, J., Kroodsma, D., & Deo, S. K. (2008). Molecular biosensing system based on intrinsically disordered proteins. *Analytical and Bioanalytical Chemistry, 391*, 1721–1729.

Dosztanyi, Z., Meszaros, B., & Simon, I. (2009). ANCHOR: Web server for predicting protein binding regions in disordered proteins. *Bioinformatics, 25*, 2745–2746.

Dulac, C., Michels, A. A., Fraldi, A., Bonnet, F., Nguyen, V. T., Napolitano, G., … Bensaude, O. (2005). Transcription-dependent association of multiple positive transcription elongation factor units to a HEXIM multimer. *Journal of Biological Chemistry, 280*, 30619–30629.

Dunker, A. K., Brown, C. J., & Obradovic, Z. (2002). Identification and functions of usefully disordered proteins. *Advances in Protein Chemistry, 62*, 25–49.

Dunker, A. K., Lawson, J. D., Brown, C. J., Williams, R. M., Romero, P., Oh, J. S., … Obradovic, Z. (2001). Intrinsically disordered protein. *Journal of Molecular Graphics and Modelling, 19*, 26–59.

Dunker, A. K., & Obradovic, Z. (2001). The protein trinity-linking function and disorder. *Nature Biotechnology, 19*, 805–806.

Gast, K., Damaschun, H., Eckert, K., Schulze-Forster, K., Maurer, H. R., Mueller-Frohne, M., … Damaschun, G. (1995). Prothymosin alpha: A biologically active protein with random coil conformation. *Biochemistry, 34*, 13211–13218.

Goldgar, D. E., Easton, D. F., Deffenbaugh, A. M., Monteiro, A. N., Tavtigian, S. V., & Couch, F. J. (2004). Integrated evaluation of DNA sequence variants of unknown clinical significance: Application to BRCA1 and BRCA2. *The American Journal of Human Genetics, 75*, 535–544.

Hendrick, J. P., & Hartl, F. U. (1995). The role of molecular chaperones in protein folding. *FASEB J, 9*, 1559–1569.

Hu, Y. F., Hao, Z. L., & Li, R. (1999). Chromatin remodeling and activation of chromosomal DNA replication by an acidic transcriptional activation domain from BRCA1. *Genes & Development, 13*, 637–642.

Hubbard, S. J., Eisenmenger, F., & Thornton, J. M. (1994). Modeling studies of the change in conformation required for cleavage of limited proteolytic sites. *Protein Science, 3*, 757–768.

Huth, J. R., Mountjoy, K., Perini, F., Bedows, E., & Ruddon, R. W. (1992). Domain-dependent protein folding is indicated by the intracellular kinetics of disulfide bond formation of human chorionic gonadotropin beta subunit. *Journal of Biological Chemistry, 267*, 21396–21403.

Iakoucheva, L. M., Kimzey, A. L., Masselon, C. D., Bruce, J. E., Garner, E. C., Brown, C. J., … Ackerman, E. J. (2001). Identification of intrinsic order and disorder in the DNA repair protein XPA. *Protein Science, 10*, 560–571.

Iakoucheva, L. M., Radivojac, P., Brown, C. J., O'Connor, T. R., Sikes, J. G., Obradovic, Z., & Dunker, A. K. (2004). The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Research, 32*, 1037–1049.

Kozlowski, L. P., & Bujnicki, J. M. (2012). MetaDisorder: A meta-server for the prediction of intrinsic disorder in proteins. *BMC Bioinformatics, 13*, 111. doi: 10.1186/1471-2105-13-111

Kyratsous, C. A., Silverstein, S. J., DeLong, C. R., & Panagiotidis, C. A. (2009). Chaperone-fusion expression plasmid vectors for improved solubility of recombinant proteins in *Escherichia coli*. *Gene, 440*, 9–15.

Mark, W. Y., Liao, J. C., Lu, Y., Ayed, A., Laister, R., Szymczyna, B., ... Arrowsmith, C. H. (2005). 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, 345*, 275–287.

Mészáros, B., Simon, I., & Dosztányi, Z. (2009). Prediction of protein binding regions in disordered proteins. *PLoS Computational Biology, 5*, e1000376.

Miki, Y., Swensen, J., Shattuck-Eidens, D., Futreal, P. A., Harshman, K., Tavtigian, S., ... Skolnick, M. H. (1994). A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science, 266*, 66–71.

Millett, I. S., Doniach, S., & Plaxco, K. W. (2002). Toward a taxonomy of the denatured state: Small angle scattering studies of unfolded proteins. *Advances in Protein Chemistry, 62*, 241–262.

Mizianty, M. J., Stach, W., Chen, K., Kedarisetti, K. D., Disfani, F. M., & Kurgan, L. (2010). Improved sequence-based prediction of disordered regions with multilayer fusion of multiple information sources. *Bioinformatics, 26*, i489–i496.

Moynahan, M. E., Chiu, J. W., Koller, B. H., & Jasin, M. (1999). BRCA1 controls homology-directed DNA repair. *Molecular Cell, 4*, 511–518.

Narod, S. A., & Foulkes, W. D. (2004). BRCA1 and BRCA2: 1994 and beyond. *Nature Reviews Cancer, 4*, 665–676.

Ouchi, T., Lee, S. W., Ouchi, M., Aaronson, S. A., & Horvath, C. M. (2000). Collaboration of signal transducer and activator of transcription 1 (STAT1) and BRCA1 in differential regulation of IFN-gamma target genes. *Proceedings of the National Academy of Sciences of the United States of America, 97*, 5208–5213.

Paull, T. T., Cortez, D., Bowers, B., Elledge, S. J., & Gellert, M. (2001). Direct DNA binding by BRCA1. *Proceedings of the National Academy of Sciences of the United States of America, 98*, 6086–6091.

Perkins, D. N., Pappin, D. J., Creasy, D. M., & Cottrell, J. S. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis, 20*, 3551–3567.

Permyakov, S. E., Millett, I. S., Doniach, S., Permyakov, E. A., & Uversky, V. N. (2003). Natively unfolded C-terminal domain of caldesmon remains substantially unstructured after the effective binding to calmodulin. *Proteins: Structure, Function, and Genetics, 53*, 855–862.

Portier, C. (1975). Quaternary structure of polynucleotide phosphorylase from *Escherichia coli*: Evidence of a complex between two types of polypeptide chains. *European Journal of Biochemistry, 55*, 573–582.

Ptitsyn, O. B. (1995). Structures of folding intermediates. *Current Opinion in Structural Biology, 5*, 74–78.

Rial, D. V., & Ceccarelli, E. A. (2002). Removal of DnaK contamination during fusion protein purifications. *Protein Expression and Purification, 25*, 503–507.

Risch, H. A., McLaughlin, J. R., Cole, D. E., Rosen, B., Bradley, L., Fan, I., … Narod, S. A. (2006). Population BRCA1 and BRCA2 mutation frequencies and cancer penetrances: A kin-cohort study in Ontario, Canada. *Journal of the National Cancer Institute, 98*, 1694–1706.

Romero, P., Obradovic, Z., Kissinger, C., Villafranca, J. & Dunker, A. (1997). Identifying disordered regions in proteins from amino acid sequence. Neural Networks, 1997., International Conference. Houston, TX: IEEE.

Romero, P. R., Zaidi, S., Fang, Y. Y., Uversky, V. N., Radivojac, P., Oldfield, C. J., … Dunker, A. K. (2006). Alternative splicing in concert with protein intrinsic disorder enables increased functional diversity in multicellular organisms. *Proceedings of the National Academy of Sciences of the United States of America, 103*, 8390–8395.

Roy, R., Chun, J., & Powell, S. N. (2012). BRCA1 and BRCA2: Different roles in a common pathway of genome protection. *Nature Reviews Cancer, 12*, 68–78.

Rubin, S. C., Blackwood, M. A., Bandera, C., Behbakht, K., Benjamin, I., Rebbeck, T. R., & Boyd, J. (1998). BRCA1, BRCA2, and hereditary nonpolyposis colorectal cancer gene mutations in an unselected ovarian cancer population: Relationship to family history and implications for genetic testing. *American Journal of Obstetrics and Gynecology, 178*, 670–677.

Sá-Moura, B., Simões, A. M., Fraga, J., Fernandes, H., Abreu, I. A., Botelho, H. M., … Ramos, P. C. (2013). Biochemical and biophysical characterization of recombinant yeast proteasome maturation factor Ump1. *Computational and Structural Biotechnology Journal, 7*, e201304006. doi: 10.5936/csbj.201304006

Schweers, O., Schonbrunn-Hanebeck, E., Marx, A., & Mandelkow, E. (1994). Structural studies of tau protein and Alzheimer paired helical filaments show no evidence for beta-structure. *Journal of Biological Chemistry, 269*, 24290–24297.

Scully, R., Chen, J., Plug, A., Xiao, Y., Weaver, D., Feunteun, J., … Livingston, D. M. (1997). Association of BRCA1 with Rad51 in mitotic and meiotic cells. *Cell, 88*, 265–275.

Shevchenko, A., Wilm, M., Vorm, O., & Mann, M. (1996). Mass spectrometric sequencing of proteins from silver-stained polyacrylamide gels. *Analytical Chemistry, 68*, 850–858.

Tan, W., Zheng, L., Lee, W. H., & Boyer, T. G. (2004). Functional dissection of transcription factor ZBRK1 reveals zinc fingers with dual roles in DNA-binding and BRCA1-dependent transcriptional repression. *Journal of Biological Chemistry, 279*, 6576–6587.

Urbauer, J. L., Short, J. H., Dow, L. K., & Wand, A. J. (1995). Structural analysis of a novel interaction by calmodulin: High-affinity binding of a peptide in the absence of calcium. *Biochemistry, 34*, 8099–8109.

Uversky, V. N. (2002a). Natively unfolded proteins: A point where biology waits for physics. *Protein Science, 11*, 739–756.

Uversky, V. N. (2002b). What does it mean to be natively unfolded? *European Journal of Biochemistry, 269*, 2–12.

Uversky, V. N. (2003). Protein folding revisited. A polypeptide chain at the folding-misfolding-nonfolding cross-roads: Which way to go? *Cellular and Molecular Life Sciences (CMLS), 60*, 1852–1871.

Uversky, V. N. (2009). Intrinsically disordered proteins and their environment: Effects of strong denaturants, temperature, pH, counter ions, membranes, binding partners, osmolytes, and macromolecular crowding. *The Protein Journal, 28*, 305–325.

Uversky, V. N. (2014). Proteins without unique 3D structures: Biotechnological applications of intrinsically unstable/disordered proteins. *Biotechnology Journal*.

Uversky, V. N., Gillespie, J. R., & Fink, A. L. (2000). Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins: Structure, Function, and Genetics, 41*, 415–427.

Uversky, V. N., Li, J., & Fink, A. L. (2001). Evidence for a partially folded intermediate in alpha-synuclein fibril formation. *Journal of Biological Chemistry, 276*, 10737–10744.

Uversky, V. N., Permyakov, S. E., Zagranichny, V. E., Rodionov, I. L., Fink, A. L., Cherskaya, A. M., … Permyakov, E. A. (2002). Effect of zinc and temperature on the conformation of the γ subunit of retinal phosphodiesterase: A natively unfolded protein. *Journal of Proteome Research, 1*, 149–159.

Vacic, V., Uversky, V. N., Dunker, A. K., & Lonardi, S. (2007). Composition profiler: A tool for discovery and visualization of amino acid composition differences. *BMC Bioinformatics, 8*, 211.

Vucetic, S., Obradovic, Z., Vacic, V., Radivojac, P., Peng, K., Iakoucheva, L. M., … Dunker, A. K. (2005). DisProt: A database of protein disorder. *Bioinformatics, 21*, 137–140.

Walsh, I., Martin, A. J., Di Domenico, T., Vullo, A., Pollastri, G., & Tosatto, S. C. (2011). CSpritz: Accurate prediction of protein disorder segments with annotation for homology, secondary structure and linear motifs. *Nucleic Acids Research, 39*, W190–W196.

Wang, Y., Cortez, D., Yazdi, P., Neff, N., Elledge, S. J., & Qin, J. (2000). BASC, a super complex of BRCA1-associated proteins involved in the recognition and repair of aberrant DNA structures. *Genes & Development, 14*, 927–939.

Wang, B., Matsuoka, S., Ballif, B. A., Zhang, D., Smogorzewska, A., Gygi, S. P., & Elledge, S. J. (2007). Abraxas and RAP80 form a BRCA1 protein complex required for the DNA damage response. *Science, 316*, 1194–1198.

Weinreb, P. H., Zhen, W., Poon, A. W., Conway, K. A., & Lansbury, P. T. (1996). NACP, a protein implicated in Alzheimer's disease and learning, is natively unfolded. *Biochemistry, 35*, 13709–13715.

Wolfram, F., Morris, E., & Taylor, C. W. (2010). Three-dimensional structure of recombinant type 1 inositol 1,4,5-trisphosphate receptor. *Biochemical Journal, 428*, 483–489.

Wright, P. E., & Dyson, H. J. (1999). Intrinsically unstructured proteins: Re-assessing the protein structure-function paradigm. *Journal of Molecular Biology, 293*, 321–331.

Xia, Y., Pao, G. M., Chen, H. W., Verma, I. M., & Hunter, T. (2003). Enhancement of BRCA1 E3 ubiquitin ligase activity through direct interaction with the BARD1 protein. *Journal of Biological Chemistry, 278*, 5255–5263.

Zhang, T., Faraggi, E., Xue, B., Dunker, A. K., Uversky, V. N., & Zhou, Y. (2012). SPINE-D: Accurate prediction of short and long disordered regions by a single neural-network based method. *Journal of Biomolecular Structure & Dynamics, 29*, 799–813.

Zhong, Q., Chen, C. F., Li, S., Chen, Y., Wang, C. C., Xiao, J., … Lee, W. H. (1999). Association of BRCA1 with the hRad50-hMre11-p95 complex and the DNA damage response. *Science, 285*, 747–750.

Zwahlen, C., Li, S. C., Kay, L. E., Pawson, T., & Forman-Kay, J. D. (2000). Multiple modes of peptide recognition by the PTB domain of the cell fate determinant Numb. *The EMBO Journal, 19*, 1505–1515.