Structural and Functional Basis of Fanconi Anemia (FANCI-FANCD2) Pathway: Studies of Protein-Protein Interactions required for DNA crosslink repair

By

Mohammad Quadir Siddiqui
[LIFE09201104013]

Tata Memorial Centre

Mumbai

A thesis submitted to the

Board of Studies in Life Sciences

In partial fulfillment of requirements

For the Degree of

DOCTOR OF PHILOSOPHY

of

HOMI BHABHA NATIONAL INSTITUTE

December, 2017

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 Mr. Mohammad Quadir Siddiqui, entitled "Structural and Functional Basis of Fanconi Anemia (FANCI-FANCD2) Pathway: Studies of Protein-Protein Interactions required for DNA crosslink repair" 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: 07/12/2017

Guide/Convener – Dr. Ashok K Varma Date: 07/12/2017

External Examiner – Dr. Petety V. Balaji Date: 7 Dec 2017

Member – Dr. Hari S. Misra Date: 7/12/17

Member – Dr. Sanjeev Waghmare Date: 7/12/17

Invitee – Dr. Prabhakar Dongre Date: 7-12-2017

Invitee – Dr. Ruchi Anand Date: 7-12-2017

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

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

Date: 07/12/2017
Place: Navi Mumbai

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.

M. Quadir

Navi Mumbai — Mohd. Quadir Siddiqui

Date: 15-12-17

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.

Navi Mumbai

Date: 15-12-17

M. Quadir

(Mohd. Quadir Siddiqui)

List of Publications arising from the thesis

Journal

1. "Structural and biophysical properties of h-FANCI ARM repeat protein", **M. Q. Siddiqui**, R.K. Choudhary, P. Thapa, N. Kulkarni, Y.S. Rajpurohit, H.S. Misra, N. Gadewal, S. Kumar, S.K. Hasan, A.K. Varma, **Journal of Biomolecular Structure and Dynamics,** 2016; p. 1-11

2. "Studies of Protein-Protein Interactions in Fanconi Anemia Pathway to Unravel the DNA Interstrand Crosslink Repair Mechanism", **M. Q. Siddiqui**, Y. S. Rajpurohit, P. Thapa, G. K. Maurya, Kuheli Banerjee, Pragnya Panda, M Ali Khan, S K Hasan, N Gadewal, H S Misra, A.K. Varma, **International Journal of Biological Macromolecules,** 2017; p. 1338-1344

Chapters in books and lectures notes NA

Conferences

International:

1. **Poster presentation** entitled "**Protein-Protein Interactions of Fanconi anemia proteins-FANCI, FANCD2 and BRCA2**" M. Q. Siddiqui, Y S Rajpurohit, H S Misra, A. K. Varma, 24th Congress -**International Union of Crystallography (IUCR)**, Hyderabad, India, August 2017

2. **Poster presentation** entitled "Structural evaluation of mutations identified on Cue domain of FANCD2" M. Q. Siddiqui, A. K. Varma, **American Association for Cancer Research (AACR), Cancer Research** 77(13 Supplement):4546-4546 · Washington, USA, July 2017

3. **Poster presentation** entitled "Biophysical Characterization of FANCD2 and Leu231Arg Mutant" M. Q. Siddiqui, A. K. Varma, **European Biophysical Societies' Association (EBSA)**, L809, 10th EBSA International Conference held in Dresden, Germany, 2015

National:

1. **Oral presentation** entitled "Structure, function and dynamics studies of Armadillo Repeat Domain of FANCI" in National Research Scholar's Meet in Life Sciences, at ACTREC, 19th-20th December 2016

2. **Oral presentation** entitled "Biophysical characterization FANCI ARM repeat" in Western region crystallography Conference at IIT Bombay, 2015

3. **Poster presentation** entitled "Structure, function and dynamics studies of FANCD2 CUE domain and Ser222Ala, Leu231Arg mutants" in Advances in Enzymology: Implications in Health, Disease, and Therapeutics " Indo-US Conference", February 2017, P26,

4. **Poster presentation** entitled "Molecular mechanism of Fanconi Anemia Complementation Group I (FANCI) ARM Repeat Protein to understand DNA intercrosslink repair" in International Conference "A Conference of New Ideas in Cancer – Challenging Dogmas" **European Journal of Cancer** 54 (2016) S1–S72 P165-C011 at NCPA**,** 26th to 28th February 2016 in Mumbai

5. **Poster presentation** entitled "Cloning, Purification & characterization of FANCI ARM repeat" Indian Biophysical Society 2015, **Journal of Proteins and Proteomics,** p. JPP91, National Conference at Jamia Millia Islamia, New Delhi

6. **Poster presentation** entitled "Mapping & Characterization of Armadillo Repeat Domain of FANCI" in "Indo-French Seminar" 19th-20th November 2015

Others

1. "Structural basis to stabilize the domain motion of BARD1-ARD BRCT by CstF50", R.K. Choudhary, **M.Q. Siddiqui**, P.S. Thapa, N. Gadewal, S.K. Nachimuthu, A.K. Varma, **Scientific Reports**, 2017; p. 3849

2. "Mislocalization of BRCA1-complex due to ABRAXAS Arg361Gln mutation", Vikrant, R. Kumar, **Q. Siddiqui**, N. Singh, S.K. Waghmare, A.K. Varma, **Journal of Biomolecular Structure and Dynamics,** 2015; p. 1291-301

3. "Multimodal approach to explore the pathogenicity of BARD1, ARG 658 CYS, and ILE 738 VAL mutants", R.K. Choudhary, Vikrant, **Q.M. Siddiqui**, P.S. Thapa, S. Raikundalia, N. Gadewal, N.S. Kumar, M.V. Hosur, A.K. Varma, **Journal of Biomolecular Structure and Dynamics,** 2015; p. 1533-44

4."New orally active DNA minor groove binding small molecule CT-1 acts against breast cancer by targeting tumor DNA damage leading to p53-dependent apoptosis", K.S. Saini, Hamidullah, R. Ashraf, D. Mandalapu, S. Das, **M.Q. Siddiqui**, S. Dwivedi, J. Sarkar, V.L. Sharma, R. Konwar, **Molecular Carcinogenesis**, 2017; p. 1266-1280

5. "A coumarin-chalcone hybrid instigates DNA damage by minor groove binding and stabilizes p53 through post translational modifications", R. Ashraf, Hamidullah, M. Hasanain, P. Pandey, M. Maheshwari, L.R. Singh, **M.Q. Siddiqui**, R. Konwar, K.V. Sashidhara, J. Sarkar, **Scientific Reports** 2017; p. 45287

Manuscript communicated/under revision

1. "Structural and functional evaluation of pathogenic mutations of FANCD2 CUE domain", **M. Q. Siddiqui,** P. Thapa, N. Kulkarni, R.K. Choudhary, A.K. Varma

M. Quadir

Date 15-12-17 Signature of Student

Dedicated to my parents

ACKNOWLEDGEMENTS

I take this opportunity to offer my sincere thanks to my mentor Dr. Ashok K. Varma for his guidance, incessant encouragement, and prudent suggestions. Without him, it would have been impossible to make this journey successful.

I am grateful to Prof. S.V. Chiplunkar (Director- ACTREC), Dr. R. Sarin (Former Director ACTREC), Dr. S. Zingde (Former Deputy Director ACTREC) for providing research facility and excellent infrastructure. I am fortunate to have Dr. Vinay Kumar, Dr. Hari Misra, Dr. Prabhakar Dongre, Dr. Ruchi Anand and Dr. Sanjeev Waghmare as my doctoral committee members. Their critical analysis and helpful suggestions have contributed significantly to the work. I am extremely thankful to Dr. M.V. Hosur, Dr. Syed K Hasan, Mr. Nikhil Gadewal, Er. Uday Dandekar, common instrument facility, Macromolecular Crystallography-X-Ray Diffraction Facility, MALDI-TOF Facility (Mr. Shashi, Ms. Savita), BTIS, Library, IT, administration and accounts department of ACTREC for their constant help and support.

I am thankful to all my lab members Mrs. Ulka Sawant, Dr. Dilip Badgujar, Dr. Vikrant, Dr. Lumbini, Dr. Bhanu Prakash, Dr. Rajan, Pankaj, Suchita, Lipi, Mudassar, Kuheli, Neha, Nishita, Jaspreet, Priyancka, Khushboo, Ankita and all my present lab members. I am thankful to Dr. Shafqat Ali Khan, Akash Dubey, Batch 2011 (Rahul, Moquitul, Sajad, Pawan, Rasika, Prasanna, Sonali, Pratik, Ekjot, Amit) and student's community of ACTREC who helped me in innumerable ways during the entire period.

I am thankful to Dr. H S Misra, Dr. Y S Rajpurohit, Mr. Ganesh Maurya, Dr. Anubrata Das, Mr. Sunil and Mr. Kamal of MGS Lab, BARC, for their encouragement and support.

I owe a lot more than I can express in words to my father, mother, brother and other family members for all their sacrifices and affectionate blessings. I would reserve my heartiest thanks for my family members Dr. Abdul Rasheed, Mrs. Shaheen Siddiqui, Jainab, Zeenat Qayyum, Qamar, Qasim, Fatima, Juggu and Mr. Rizwan for being my pillar of strength throughout my journey. I am thankful to my friends Dr. Hamidullah, Mr. Khursheed, Kashifa, Mr. Sushil, Dr. Tauqeer, and Mr. Ubaid for their blessings.

Last but not the least, I would like to thank my mother who has provided me love, strength, motivation and guiding lessons of life. I don't know where life will take me as I grow older, but one thing I know is that I will never be able to be a person better than her. I cherish precious memories of her and find strength in knowing that even though she is no more, she still lives in my heart forever.

Though all cannot be mentioned, but none is forgotten.

CONTENTS

Homi Bhabha National Institute

SYNOPSIS OF Ph.D. THESIS

1. **Name of the Student:** Mohd. Quadir Siddiqui
2. **Name of the Constituent Institution:** Tata Memorial Centre, ACTREC
3. **Enrolment No. and Date of Enrolment:** LIFE09201104013, 10th August 2011
4. **Title of the Thesis:** Structural and Functional Basis of Fanconi Anemia(FANCI-FANCD2) Pathway: Studies of Protein-Protein Interactions required for DNA crosslink repair
5. **Board of Studies**: Life Science

SYNOPSIS

Fanconi anemia (FA) is a cancer predisposition syndrome, which exhibits hallmark features of radial chromosome and sensitivity towards crosslinking agents [1]. FA arises due to inherited germ-line and somatic mutations in the different genes present in FA pathways. Fanconi Anemia Complementation Group- I (FANCI) and Fanconi Anemia Complementation Group -D2 (FANCD2) are the proteins present in FA pathway that facilitates the DNA inter crosslink (ICL)

damage repair by recruiting to DNA damage site. Recent reports suggest that proteins present in FA and BRCA pathway together form a complex to repair the DNA ICL damage [2]. FANCI, FANCD2, and BRCA2 are the important proteins that bridge the FA and BRCA pathway [2]. These proteins have functionally conserved domains that regulate the FA pathway. FANCI armadillo repeat (ARM repeat) domain encompassing the region from 985-1207 amino acids plays an important role in DNA ICL damage repair, and also perform functions associated with protein-protein interactions by forming the complex with FANCD2 protein [3]. FANCD2 protein contains coupling of ubiquitin conjugation to the endoplasmic reticulum (CUE) domain, which is required for the interactions with mono-ubiquitinated FANCI and known to regulate FA pathway [4].

In the present study, we have investigated the structural and functional basis of FANCI ARM repeat as well as binding studies with the dsDNA. Furthermore, structure and function of FANCD2 coupling of ubiquitin conjugation to the endoplasmic reticulum (CUE) domain, and protein-protein interactions with monoubiquitin have been evaluated using *in-vitro* and *in-silico* tools. Functionally relevant mutations reported in CUE domain, and their effect on the structure and functions of CUE domain has been explored comparatively. In addition, protein-protein interactions studies of CUE domain *wild-type* and mutant with monoubiquitin and BRCA2 was also explored. Protein-protein interactions (PPI) studies were carried out using *in-vivo* tools to understand the interplay between FANCI, FANCD2, and BRCA2, that would help in unraveling the molecular mechanism of DNA ICL damage repair. The results presented here, explore the effect of cancer predisposing point mutations in PPIs of FANCI ARM repeat, FANCD2 CUE domain, and BRCA2 C-terminal region using multimodal *in-vivo*, *in-vitro* and *in-silico* based approach.

This thesis encompasses seven chapters including a chapter on summary and overall conclusion at the end.

Chapter 1 provides an overview on structural and functional basis of Fanconi anemia (FANCI-FANCD2) pathway: protein-protein interactions required for DNA intercrosslink repair

Studies of Fanconi Anemia (FA) pathway which is one of the rare genetic disorder provides an opportunity to unravel the molecular mechanisms associated to DNA inter-crosslink (DNA ICL) repair [1]. Recently, it has been observed that DNA ICL pathway/FA pathway comprises eighteen complementation groups of proteins such as FANCA,B,C,D1/BRCA2,D2,E,F,G,I,J/BRIP1,L,M,N/PALB2,O/RAD51C,P/SLX4,Q/XPF,S/BRCA1 and T/UBE2T that specifically act in a cascade [5]. FA patients exhibit the diverse spectrum of clinical phenotypes and show high sensitivity to inter-crosslinking agents such as mitomycin C and di-epoxybutane [6-8]. Radial chromosomal aberration is one of the diagnostic hallmarks of FA cells [6]. Basic mechanism underlying the formation of the radial chromosome is the inter-crosslinking of the DNA strands. Inter-cross links (ICLs) occur due to covalent bond in DNA that has to be incised for a successful replication and transcription. Proteins present in FA pathway comprises a set of specific proteins that recruited to DNA damage sites to facilitate the DNA ICL repair [2]. In eukaryotes, the repair mechanism of intrastrand and interstrand crosslinks (ICL's) are not yet fully explored. FANCI (Fanconi anemia complementation group I) is known to be recruited at DNA damage sites to facilitate the DNA ICL repair [3]. FANCI protein also harbors armadillo repeat (ARM repeat) domain which is known to be involved in protein-protein and protein-DNA interactions [3, 9].

The FA proteins together with BRCA1 and BRCA2 function cooperatively in the FA/BRCA pathway [2]. The FA pathway is activated by exposure to DNA-damaging crosslinking agents such as mitomycin c and diepoxybutane during S-phase of the cell-cycle. This activated FA core complex which is composed of FANC A, B, C, E, F, G, L, M, proteins monoubiquitinate the FANCD2 and FANCI proteins [2]. Monoubiquitination of FANCD2 and FANCI targets

these proteins to chromatin-associated nuclear foci, where they facilitate the interactions with different key DNA repair proteins including BRCA1, FANCD1/BRCA2 and RAD51 [10]. FANCD2 protein is necessary for the recruitment of the FAN1 and SLX4/FANCP endonucleases to the DNA damage sites [11]. FANCD2 has a coupling of ubiquitin conjugation to endoplasmic reticulum domain (CUE domain) which encompasses the region from 1-254 amino acids at the N-terminus. It has been reported that CUE domain regulates the FA pathway by correcting the mitomycin c hypersensitivity and localizing the FANCD2 and FANCI proteins at chromatin site [4]. CUE domain is also responsible for the protein-protein interactions between FANCD2 and monoubiquitinated FANCI, and masking the target protein from poly-ubiquitination to regulate the protein homeostasis at the time of DNA damage repair [4]. FANCD2 N-terminus is also known to interact with BRCA2 C-terminal region (BRCA2 CTR) [12, 13]. In addition, BRCA2 proteins with C-terminal truncations were observed in cell lines derived from some FA patients [2]. Thus, it confirms the indispensable functional role of different domains of FANCI, FANCD2, and BRCA2 in DNA ICL repair/FA pathway.

Chapter 2 is a brief description of the different methodologies and techniques used to carry out research work reported in the thesis

The individual gene of interest was polymerase chain reaction (PCR) amplified using specifically designed primers having TEV-(ENLYFQG) protease site at N-terminus, and restriction enzyme site at both the end. The PCR amplified constructs were cloned in a different bacterial expression vector such as pET28a+, pGex4T-1. DH5-α and NOVABLUE cells were used for cloning purposes. *E.coli* Rosetta 2 (DE3) and BL 21 (DE3) strains were used to check the expression and solubility of the proteins. Soluble protein was passed through the pre-equilibrated glutathione or Ni-NTA affinity column, further washed by Tris-NaCl, additives (like imidazole) buffer to remove impurities bound to the resin. The fusion protein was treated with TEV protease to remove the affinity tag. The native protein was further purified using gel

filtration column (Superdex 75, GE Healthcare), mounted on AKTA –FPLC and BIORAD-FPLC. Protein secondary structure assessment and temperature dependent folding pattern were analyzed using a Circular-Dichroism (CD) polarimeter (Jasco J-810, Japan) with the wavelength ranges from λ= 200–240 nm. Fluorescence spectra were monitored using a Fluorescence spectrophotometer (Horiba, USA) at emission wavelength from λ= 300-450 nm, in order to understand the tertiary structure and melting temperature (Tm) of the protein. Limited proteolysis of purified protein was performed in different times 0, 10, 30, 60 and 120 minutes at 37°C using trypsin enzyme (Sigma-Aldrich), after that the reaction was terminated by adding 1 mM PMSF (Sigma-Aldrich). Protein samples were heated at 95°C before loading onto SDS-PAGE gel. Stable domain mapping was performed by peptide mass fingerprinting using mass spectrometry. Dynamic Light Scattering (Malvern and Wyatt DynaPro) was used to study the hydrodynamic radii and oligomeric behavior of the protein. Protein-protein interactions were analyzed using Surface Plasmon Resonance (BIAcore T200) to quantify the binding affinity and study the kinetics. Electrophoretic Mobility Shift Assay (EMSA) by autoradiography was performed to study the protein-DNA interactions. *In-vivo* protein-protein interactions were performed using Bacterial Two-Hybrid (BTH) system and beta-galactosidase assay. For *in-vitro* protein-protein interactions, GST pull-down assay was performed by co-expression strategy and using Snapcap column (Thermo-Fisher). *In-silico* Molecular Dynamics (MD) were performed using GROMACS 4.5.5 package [14]. Normal mode analysis (NMA), anisotropic network modeling (ANM) and Gaussian network modeling (GNM) were done by using R 3.2 package and ProDy (Protein Dynamics 1.7), respectively [15, 16]. Protein-protein and protein-DNA docking studies were performed by using HADDOCK [17] and HEX 8.0.0 [18].

Chapter 3 describes structural and biophysical properties of FANCI and FANCD2

Fanconi anemia complementation groups I (FANCI) protein facilitate DNA ICL (Inter-Crosslink) repair and plays a crucial role in maintaining genomic integrity [3]. FANCI protein comprises 1-1328 amino acids including armadillo (ARM) repeat between 985-1207 amino acids [3]. ARM repeats are functionally diverse and evolutionarily conserved domain that plays an important role in protein-protein and protein-DNA interactions [3, 19, 20]. Here, we are reporting the multimodal comprehensive *in-vitro*, *in-silico* and biophysical approach to understand the folding pattern of FANCI ARM repeat. Size exclusion chromatography, dynamic light scattering (DLS) and glutaraldehyde crosslinking studies suggest that FANCI ARM repeat exists as a monomer and in oligomeric forms. Circular-dichroism (CD) and fluorescence spectroscopy results indicate that protein has predominantly α-helices and well-folded tertiary structure. Temperature dependent CD, Fluorescence spectroscopy, and DLS studies suggest that protein unfolding and oligomer formation occurred beyond 30°C. The functionally stable domain of FANCI ARM repeat was examined using limited proteolysis and mass spectrometry. To our conclusion, FANCI ARM repeat has a compact domain at N-terminus and dynamic region at C-terminus. These findings are in consistent with previously reported results [21-23], stating that FANCI ARM repeat also has a stable and dynamic region which may acts as recognition and interacting module, respectively in cellular signaling.

Fanconi anemia complementation group D2 (FANCD2) protein is very important to understand the mechanism associated with DNA interstrand crosslink repair, cell-cycle checkpoint, and genome stability [24]. FANCD2 comprises ubiquitin binding domain at N-terminus, known as CUE domain [4] which plays an important role in DNA intercrosslink repair and protein stability [4]; however, its precise role at the structural level is not yet well characterized. Here, we have carried out a comprehensive interdisciplinary approach to elucidate the structure and function of FANCD2 CUE domain and cancer predisposing point mutations identified from a cohort of patients. DLS data from purified CUE domain proteins *wild-type*, Ser222Ala,

Leu231Arg mutants have revealed that *wild-type* protein is predominantly monomeric, having the homogenous population, whereas mutated counterparts Ser222Ala and Leu231Arg are devoid of homogeneity. Moreover, mass spectrometry data indicates the high intensity of monomeric species in *wild-type* compared to Ser222Ala and Leu231Arg mutants. Perturbation in α-helical contents in the mutants as compared to *wild-type* has been demonstrated by CD spectroscopy. In addition, thermodynamic stability calculated by CD and fluorescence-spectroscopy have revealed a significant decrease in melting temperature (Tm) of the mutant protein.

Chapter 4 describes the functional characterization of FANCI and FANCD2

It has been reported that FANCI C-terminus region is involved in DNA binding and binds to different forms of DNA such as double-stranded DNA, Holliday junction DNA, and Y-shaped DNA [25]. FANCI C-terminus contains ARM repeat and the crystal structure from mouse FANCI is available with ~78% homology [26], hence a model with a very good stereochemistry was built using *in-silico* homology modeling. Structural alignment of FANCI ARM repeat suggests the profound similarity with transcription factors. In general, transcription factors are known to have DNA binding elements such as helix-turn-helix (HTH) or helix-loop-helix [27-29]. Interestingly, sequence alignment from 20 master sets of different known HTH motifs [28, 29], revealed that FANCI ARM repeat has similarity with known HTH motif. Furthermore, it has conserved PHS (P= any charged amino acid mostly glutamate, H= hydrophobic amino acid and S= small amino acid residue) signature known to present in HTH motif [27]. Molecular dynamics and essential dynamics demonstrate that HTH-type motif present in ARM repeat at C-terminus has high dynamicity and anti-correlated motion. In addition, docking analysis has revealed that HTH-type motif binds to the major groove of DNA. The results presented in this study provide the first information on detection of the HTH-type motif, structural behavior, as well as dynamics of FANCI ARM repeat. Our findings offer

intriguing possibilities of ARM repeat function that would help to understand the molecular mechanism of DNA ICL repair.

CUE domain of FANCD2 interacts with monoubiquitin and known to regulate FA pathway [4]. We have performed protein-protein interactions between CUE domain *wild-type* and Leu231Arg mutant protein with monoubiquitin using surface plasmon resonance (SPR) to observe the binding affinity and kinetics. These data suggest a decrease in binding affinity towards monoubiquitin of mutant proteins. In addition, the chaperone-like assay has revealed the impaired chaperonin efficiency of mutants. Molecular modeling, simulation and docking results were in concordance with *in-vitro* findings. Interaction interfaces of *wild-type* and mutants were calculated, and found intact moderately conserved hydrophobic patch in *wild-type* and perturbed in Leu231Arg mutants.

Chapter 5 describes protein-protein interactions between FANCI and FANCD2

FANCI and FANCD2 interacts, and recruited to DNA damage sites during the DNA ICL damage repair [3]. FANCD2 binds to FANCI with the help of CUE domain and mask monoubiquitin, to rescue polyubiquitination that escapes the protein to degradation [3, 4]. However, little is known at domain level interactions between FANCI and FANCD2. Therefore, *in-vivo* bacterial two-hybrid assays followed by a beta-galactosidase assay as well as *in-silico* binding analysis were performed. To our conclusion, CUE domain of FANCD2 interacts with ARM repeat of FANCI and interestingly, the higher binding affinity between FANCI ARM repeat and FANCD2 were observed when N-terminus of FANCD2 ARM repeats were included.

Chapter 6 describes the interactions between FANCD2 and BRCA2

FA proteins are known to interact with BRCA1, FANCD1 (BRCA2), FANCN (PALB2) and FANCJ (BRIP1) [2]. These proteins function in an orchestrating fashion to repair the DNA ICL and prevent hypersensitivity to crosslinking agents and cellular transformation [30]. Recent reports suggest that FA/BRCA pathway coordinate important DNA repair processes such as nucleolytic incision, translesion DNA synthesis and homologous recombination [2]. It has been found that FA and BRCA pathway is bridged by protein-protein interactions (PPI) between FANCD2 and BRCA2 [2], however not much has been explored on the PPIs for the proteins present in FA and BRCA pathway. Here, we performed bacterial two-hybrid assays (BTH) followed by beta-galactosidase assay, between FANCD2 CUE and BRCA2 C-terminal region (BRCA2 CTR), and observed that these proteins have strong interactions. In addition, we also compared the interactions between Ser222Ala/Leu231Arg CUE domain mutants with BRCA2 CTR. To our finding, a decrease in interactions in both the mutant compare to *wild type* CUE domain was observed. *In-vitro* GST-pull down assay is also in full agreement with *in-vivo* BTH data. Furthermore, *in-silico* protein-protein docking analysis was performed to support the *in-vivo* and *in-vitro* findings. Structural analysis at the binding interface revealed that the CUE *wild-type* has the more number of hydrogen bonds (20 hydrogen bonds) compared to Ser222Ala (7 hydrogen bonds) and Leu231Arg (12 hydrogen bonds) mutants with BRCA2 CTR.

Chapter 7 Summary and future prospects

This chapter concludes that ARM repeat is predominantly composed of alpha-helices. We have analyzed oligomeric propensity, secondary and tertiary structure stability of FANCI ARM repeat. FANCI ARM repeat has an HTH-type motif and DNA binding property. However, it would be very interesting to do mutational studies of the conserved residues present at the interface of FANCI ARM repeat HTH motif and DNA, to analyze the quantitative binding

affinity, and ultimately co-crystallization with DNA to unravel the binding interface at the atomic level.

This study also presents the structural and biophysical properties of FANCD2 CUE domain *wild-type* to compare the folding pattern of DNA ICL repair mutant (Leu231Arg). It has been observed that mutant was comparatively less stable. Moreover, there was inefficient binding due to structural alterations in Leu231Arg mutant with a monoubiquitin. It has also been observed that the hydrophobic patches (interface) of CUE domain and monoubiquitin are perturbed in Leu231Arg mutant, and mutant is not able to form efficient interactions. The interdisciplinary approaches reported here clearly explain the role of Leu231Arg mutation on the CUE domain stability, overall conformation, chaperonin capacity and binding affinity with a monoubiquitin.

It has been observed that FANCI ARM repeat interacts with BRCA2 CTR and FANCD2 CUE domain. FANCD2 CUE domain showed strong interactions with BRCA2 CTR. However, the Ser222Ala and Leu231Arg mutants were having perturbed interactions with BRCA2 CTR. These results suggest that domain and regions present in FANCI, FANCD2, and BRCA2 are playing important role in PPIs, and mutations cause the failure of functions associated to PPIs between these proteins.

Comparative *in-vivo, in-vitro, in-silico* and biophysical studies of FANCI, FANCD2, BRCA2 and mutant proteins reported in this work will add to the existing knowledge of the molecular mechanism of DNA ICL repair. However, it would be interesting to explore further, the structure and function using high-resolution macromolecular crystallography and structural biology approach.

References

1. Cohen, M.M., et al., *The identification of fanconi anemia genotypes by clastogenic stress.* Am J Hum Genet, 1982. **34**(5): p. 794-810.
2. D'Andrea, A.D. and M. Grompe, *The Fanconi anaemia/BRCA pathway.* Nat Rev Cancer, 2003. **3**(1): p. 23-34.
3. Smogorzewska, A., et al., *Identification of the FANCI protein, a monoubiquitinated FANCD2 paralog required for DNA repair.* Cell, 2007. **129**(2): p. 289-301.
4. Rego, M.A., et al., *Regulation of the Fanconi anemia pathway by a CUE ubiquitin-binding domain in the FANCD2 protein.* Blood, 2012. **120**(10): p. 2109-17.
5. Castella, M., et al., *FANCI Regulates Recruitment of the FA Core Complex at Sites of DNA Damage Independently of FANCD2.* PLoS Genet, 2015. **11**(10): p. e1005563.
6. Alter, B.P., et al., *Cancer in Fanconi anemia.* Blood, 2003. **101**(5): p. 2072.
7. Rosenberg, P.S., B.P. Alter, and W. Ebell, *Cancer risks in Fanconi anemia: findings from the German Fanconi Anemia Registry.* Haematologica, 2008. **93**(4): p. 511-7.
8. Rosenberg, P.S., M.H. Greene, and B.P. Alter, *Cancer incidence in persons with Fanconi anemia.* Blood, 2003. **101**(3): p. 822-6.
9. Mitra, P., et al., *Evidence for the Nucleo-Apical Shuttling of a Beta-Catenin Like Plasmodium falciparum Armadillo Repeat Containing Protein.* PLoS One, 2016. **11**(2): p. e0148446.
10. Ameziane, N., et al., *Lack of large genomic deletions in BRIP1, PALB2, and FANCD2 genes in BRCA1/2 negative familial breast cancer.* Breast Cancer Res Treat, 2009. **118**(3): p. 651-3.
11. Airik, R., et al., *A FANCD2/FANCI-Associated Nuclease 1-Knockout Model Develops Karyomegalic Interstitial Nephritis.* J Am Soc Nephrol, 2016. **27**(12): p. 3552-3559.
12. Hussain, S., et al., *Direct interaction of FANCD2 with BRCA2 in DNA damage response pathways.* Hum Mol Genet, 2004. **13**(12): p. 1241-8.
13. Hussain, S., et al., *Direct interaction of the Fanconi anaemia protein FANCG with BRCA2/FANCD1.* Hum Mol Genet, 2003. **12**(19): p. 2503-10.
14. Hess, B., et al., *GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation.* J Chem Theory Comput, 2008. **4**(3): p. 435-47.
15. Bahar, I., et al., *Global dynamics of proteins: bridging between structure and function.* Annu Rev Biophys, 2010. **39**: p. 23-42.
16. Bakan, A., L.M. Meireles, and I. Bahar, *ProDy: protein dynamics inferred from theory and experiments.* Bioinformatics, 2011. **27**(11): p. 1575-7.
17. van Dijk, M., et al., *Information-driven protein-DNA docking using HADDOCK: it is a matter of flexibility.* Nucleic Acids Res, 2006. **34**(11): p. 3317-25.
18. Macindoe, G., et al., *HexServer: an FFT-based protein docking server powered by graphics processors.* Nucleic Acids Res, 2010. **38**(Web Server issue): p. W445-9.
19. Tewari, R., et al., *Armadillo-repeat protein functions: questions for little creatures.* Trends Cell Biol, 2010. **20**(8): p. 470-81.
20. Tsytlonok, M., et al., *Complex energy landscape of a giant repeat protein.* Structure, 2013. **21**(11): p. 1954-65.
21. Coates, J., *Armadillo repeat proteins: beyond the animal kingdom.* Trends Cell Biol, 2003. **13**(9): p. 463-471.
22. Crist, R.C., et al., *The armadillo repeat domain of Apc suppresses intestinal tumorigenesis.* Mamm Genome, 2010. **21**(9-10): p. 450-7.
23. Drechsel, G., et al., *C-terminal armadillo repeats are essential and sufficient for association of the plant U-box armadillo E3 ubiquitin ligase SAUL1 with the plasma membrane.* J Exp Bot, 2011. **62**(2): p. 775-85.
24. Andreassen, P.R., A.D. D'Andrea, and T. Taniguchi, *ATR couples FANCD2 monoubiquitination to the DNA-damage response.* Genes Dev, 2004. **18**(16): p. 1958-63.

25. Longerich, S., et al., *Regulation of FANCD2 and FANCI monoubiquitination by their interaction and by DNA.* Nucleic Acids Res, 2014. **42**(9): p. 5657-70.
26. Joo, W., et al., *Structure of the FANCI-FANCD2 complex: insights into the Fanconi anemia DNA repair pathway.* Science, 2011. **333**(6040): p. 312-6.
27. Aravind, L., et al., *The many faces of the helix-turn-helix domain: transcription regulation and beyond.* FEMS Microbiol Rev, 2005. **29**(2): p. 231-62.
28. Brennan, R.G., *The winged-helix DNA-binding motif: another helix-turn-helix takeoff.* Cell, 1993. **74**(5): p. 773-6.
29. Brennan, R.G. and B.W. Matthews, *The helix-turn-helix DNA binding motif.* J Biol Chem, 1989. **264**(4): p. 1903-6.
30. Andreassen, P.R. and K. Ren, *Fanconi anemia proteins, DNA interstrand crosslink repair pathways, and cancer therapy.* Curr Cancer Drug Targets, 2009. **9**(1): p. 101-17.
31. Kuntz, I.D., et al., *A geometric approach to macromolecule-ligand interactions.* Journal of molecular biology, 1982. **161**(2): p. 269-288.
32. SchuÈttelkopf, A.W. and D.M. Van Aalten, *PRODRG: a tool for high-throughput crystallography of protein–ligand complexes.* Acta Crystallographica Section D: Biological Crystallography, 2004. **60**(8): p. 1355-1363.
33. Jones, G., et al., *Development and validation of a genetic algorithm for flexible docking.* Journal of molecular biology, 1997. **267**(3): p. 727-748.
34. Longerich, S., et al., *FANCI binds branched DNA and is monoubiquitinated by UBE2T-FANCL.* J Biol Chem, 2009. **284**(35): p. 23182-6.
35. van Zundert, G.C., et al., *The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes.* J Mol Biol, 2016. **428**(4): p. 720-5.

Publications in Refereed Journal:

a. Published:

1. "Structural and biophysical properties of h-FANCI ARM repeat protein" **(Mohd. Quadir Siddiqui** *et al.,**Journal of Biomolecular Structure and Dynamics,*** 2016)

2. **"**Mislocalization of BRCA1-complex due to ABRAXAS Arg361Gln mutation" (Vikrant,Kumar R, **Siddiqui Q**, *et al*, ***Journal of Biomolecular Structure and Dynamics,*** 2015)

3 "Multimodal approach to explore the pathogenicity of BARD1, ARG 658 CYS, and ILE 738 VAL mutants" (Choudhary RK, Vikrant, **Siddiqui QM**, *et al, **Journal of Biomolecular Structure and Dynamics,*** 2015)

b. Accepted: NA

c. In Communication/Preparation:

1. "Structural and functional evaluation of pathogenic mutations of FANCD2 CUE domain, **(Mohd. Quadir Siddiqui** *et al.,*) (Manuscript Communicated)

2. "Studies of Protein-Protein Interactions in Fanconi Anemia Pathway to Unravel the DNA Interstrand Crosslink Repair Mechanism, (**Mohd. Quadir Siddiqui** *et al.,*) (Manuscript under revision)

d. Other publications:

1. "New orally active DNA minor groove binding small molecule CT-1 acts against breast cancer by targeting tumor DNA damage leading to p53-dependent apoptosis" (Karan Singh Saini, Hamidullah, Raghib Ashraf, Dhanaraju Mandalapu, Sharmistha Das**, Mohd Quadir Siddiqui**, *et al*., ***Molecular Carcinogenesis***, November 2016)

e. Conference abstracts:

International:

1. Poster presentation entitled "Biophysical Characterization of FANCD2 and Leu231Arg Mutant" 10th EBSA International Conference held in Dresden, Germany, 2015

National:

1. Poster presentation entitled "Structure, function and dynamics studies of FANCD2 CUE domain and Ser222Ala, Leu231Arg mutants" in Advances in Enzymology: Implications in Health, Disease, and Therapeutics "Indo-US Conference", February 2017, held at ACTREC

2. Oral presentation entitled "Structure, function and dynamics studies of Armadillo Repeat Domain of FANCI" in National Research Scholar's Meet in Life Sciences, at ACTREC, 19th-20th December 2016 at ACTREC

3. Poster presentation entitled "Molecular mechanism of Fanconi Anaemia Complementation Group I (FANCI) ARM Repeat Protein to understand DNA intercrosslink repair" in International Conference "A Conference of New Ideas in Cancer – Challenging Dogmas" at NCPA**,** 26th to 28th February 2016 in Mumbai

4. **Oral presentation** entitled "Biophysical characterization FANCI ARM repeat" in Western region crystallography Conference held at IIT Bombay, 2015
5. **Poster presentation** entitled "Cloning, Purification & characterization of FANCI ARM repeat" IBS 2015 National Conference held at Jamia Millia Islamia, New Delhi
6. **Poster presentation** entitled "Mapping & Characterization of Armadillo Repeat Domain of FANCI" in "Indo-French Seminar" 19th-20th November 2015 at ACTREC

Doctoral Committee:

S. No.	Name	Designation	Signature	Date
1.	Dr. Vinay Kumar-BARC	**Chairperson**		11/4/17
2.	Dr. Ashok Varma	**Guide & Convener**		11/4/17
3.	Dr. H. S. Misra-BARC	**Member**		11/4/17
4.	Dr. Sanjeev Waghmare	**Member**		11/4/17
5.	Dr. Ruchi Anand-IITB	**Invitee**		11/4/17
6.	Dr. Prabhakar Dongre – Mumbai University	**Invitee**		11-4-17

Forwarded through:

S. Chiplunkar 17/4/17

Dr. S.V. Chiplunkar
Director, ACTREC
Chairperson, Academic & Training Programme
ACTREC

Dr. S. V. Chiplunkar
Director
Advanced Centre for Treatment, Research & Education in Cancer (ACTREC)
Tata Memorial Centre
Kharghar, Navi Mumbai - 410 210.

To

Dean-Academic
Dr. K. Sharma,
Director, Academics,
T.M.C.

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

ABBREVIATIONS

ARM	: Armadillo repeat
ATM	: Ataxia Telangiectasia Mutated
ATP	: Adenosine Triphosphate
ATR	: Ataxia Telangiectasia and RAD3 related
BLAST	: Basic Local Alignment Search Tool
β-ME	: β-mercaptoethanol
BRCA1	: Breast Cancer susceptibility gene1
BRCA2	: Breast Cancer susceptibility gene2
bp	: base pairs
BSA	: Bovine serum albumin
BTH	: Bacterial two hybrid
Chk1	: Checkpoint kinase 1
Chk2	: Checkpoint kinase 2
CD	: Circular Dichroism
CUE	: Coupling of ubiquitin conjugation to endoplasmic reticulum
DLS	: Dynamic Light Scattering
DNA	: Deoxyribonucleic Acid
dNTPs	: De-Oxy Nucleoside Tri-Phosphate
EB	: Elution Buffer
EDTA	: Ethylene diamine tetra acetic acid
FANCI	: Fanconi anemia complementation group I
FANCD2	: Fanconi anemia complementation group D2
FFT	: Fast Fourier Transform
HR	: Homologous Recombination
FPLC	: Fast protein liquid chromatography
ICL	: DNA Intercrosslink

IMAC	: Immobilized Metal ion Affinity Chromatography
IPTG	: Isopropyl-β-D-thiogalactoside
kb	: kilo base pairs
K_d	: Dissociation constant
kDa	: kilo Dalton
LB	: Luria-Bertani
MALDI-TOF	: Matrix Assisted Laser Desorption Ionization Time of Flight
MBP	: Maltose binding protein
MS	: Mass Spectrometry
NHEJ	: Non-Homologous End Joining
NES	: Nuclear Export Signal
NLS	: Nuclear Localization Signal
NMA	: Normal mode analysis
NMR	: Nuclear Magnetic Resonance
MRE11	: Meiotic Recombination 11 homolog
MRN	: Mre11-Rad50-Nbs1 complex
MDS	: Molecular Dynamics Simulation
NTA	: Nitrilo Triacetic Acid
OD_{340}	: Optical Density at 340 nm
OD_{600}	: Optical Density at 600 nm
ORF	: Open Reading Frame
PAGE	: Poly-Acrylamide Gel Electrophoresis
PCA	: Principal Component Analysis
PCR	: Polymerase Chain Reaction
PDB	: Protein Data Bank
pI	: Isoelectric point
PPIs	: Protein-Protein Interactions

RMSD	: Root Mean Square Deviation
RMSF	: Root Mean Square Fluctuation
SEC	: Size Exclusion Chromatography
TD	: Thermal denaturation
TEV	: Tobacco Etch Virus
TFA	: Trifluoro Acetic acid
T_m	: Melting Temperature
UV	: Ultra Violet
Ub	: Ubiquitin

List of Tables

List of Figures

	codons in correct reading frame, (different colors show the nucleotides; Adenine=Green, Guanine=Black, Cytosine=Blue and Thymine=Red) **(C)** FANCI ARM Repeat purified protein on SDS-PAGE gel and **(D)** Peptide mass fingerprinting using mass spectrometry profile for the identification of the FANCI ARM Repeat protein (grey color shows the matching of captured peptides)
Figure 3.2	**Figure 3.2: (A)** Size exclusion chromatography of FANCI ARM repeat, elution peak 1, 2 and 3 showing oligomer, dimer and monomer **(B)** Molecular weight determination with standard proteins **(C)** Whole molecular weight determination by mass spectrometry **(D)** Time-dependent glutaraldehyde crosslinking SDS-PAGE gel profile showing different species of monomer, dimer and oligomer **(E)** DLS profile showing multimeric population of protein, each dot was showing the single DLS measurement and ≥7 scans were used for the distribution plot
Figure 3.3	**(A)** Far –UV, CD spectra of FANCI-ARM repeat, indicating the α-helical nature of protein, **(B)** Thermal denaturation profile by CD showing protein unfolds at 50°C **(C)** Quantification of α- helices of CD data using K2D2 server
Figure 3.4	**Figure 3.4: (A)** Chemical denaturation profile of fluorescence spectroscopy at 0M, 4M and 8M urea **(B)** Thermal denaturation profile using Fluorescence spectroscopy showing steep decrease in intensity beyond 30°C and **(C)** Temperature dependent DLS profile showing oligomer formation beyond 30°C
Figure 3.5	**(A)** Limited proteolysis profile of SDS-PAGE gel**, (B)** Disordered prediction by using PrDos showing disordered region in protein, **(C)** Expasy peptide cutter with trypsin, **(D)** Modelled structure of FANCI ARM repeat protein showing the stable region in yellow and dynamic region in green, **(E)** Full length tryptic peptide mass fingerprint and **(F)** stable region profile of tryptic peptide mass fingerprint of MALDI-TOF-TOF
Figure 3.6	**(A)** DALI structural alignment shows similarity with transcription factors and elongation factors proteins
Figure 3.7	**(A)** Modelled structure of FANCI ARM repeat protein and **(B)** validation by Ramachandran plot and **(C)** Ramachandran plot statistics
Figure 3.8	**(A)** Modelled structure of FANCI ARM repeat protein**, (B)** Multiple sequence alignment of FANCI ARM repeat of different mammalian species showing the conserved PHS signature and **(C)** MUSCLE alignment of eighteen different known HTH motif present in "Master Sets"
Figure 3.9	Molecular dynamics simulation profile of FANCI ARM repeat **(A)** RMSD profile & **(B)** Radius of gyration (Rg) profile of 80 nanoseconds showing structural transitions up to 60 ns, **(C)** RMSF profile showing large fluctuation of C-terminus and **(D)** Solvent accessible surface area (SASA) profile of h-FANCI ARM repeat protein showing high SASA value at C-terminus
Figure 3.10	**(A)** Cross-correlation diagonal matrix of h-FANCI ARM repeat, **(B)** RMSF sausage profile of FANCI ARM repeat, **(C)** Scree plot of N-

	terminus (1-100) and C-terminus (100-223) and **(D**) PCA profile of Eigenvector 1 and 2 of N-terminus (1-100) and C-terminus (100-223)
Figure 3.2.1	**(A)** Restriction enzyme digestion of FANCD2 CUE Domain cloned in pGEX4T-1 vector showing insert release of right size **(B)** DNA Sequencing chromatogram of FANCD2 CUE Domain cloned in pGEX4T-1 exhibiting TEV site and FANCD2 CUE Domain (different colors show the nucleotides; Adenine=Green, Guanine=Black, Cytosine=Blue and Thymine=Red) **(C)** FANCD2 CUE domain purified protein, Lane: M= Protein Marker, 1= FANCD2 CUE *wild type*, 2= FANCD2 CUE Ser222Ala and 3= FANCD2 CUE Leu231Arg **(D)** Peptide mass fingerprinting profile for the identification of FANCD2 CUE domain using mass spectrometry (grey color shows the matching of captured peptides)
Figure 3.2.2	Dynamic light scattering results for **(A)** *wild-type* CUE, **(B)** Ser222Ala and **(C)** Leu231Arg exhibiting population of proteins, **(D)** Mass spectrometry (intensity vs molecular weight) comparative profile of *wild type* and mutants
Figure 3.2.3	**Structure and stability profile: (A)** Far –UV, CD spectra of *wild-type* CUE and Leu231Arg mutant indicating the loss of α-helical nature in mutant proteins, **(B)** Secondary structure stability profile of *wild type* CUE and Leu231Arg mutant
Figure 3.2.4	**Structure and stability profile: (A)** Fluorescence spectroscopy profile for tertiary structure assessment and folding of *wild-type* CUE and Leu231Arg mutant **(B)** Tertiary structure stability profile for *wild-type* CUE and Leu231Arg mutant
Figure 3.2.5	**Molecular dynamics simulation profile of *wild-type* CUE domain and Leu231Arg mutant: (A)** Comparative RMSD data shows *wild type* CUE is able to attain stable conformation unlike mutants **(B)** Comparative RMSF data suggests higher residual fluctuation in mutants
Figure 3.2.6	**DSSP profile: (A)** *wild-type* CUE; secondary structure transition was stable as alpha helices were consistent throughout the simulation at the mutation site **(B)** Leu231Arg; mixed population of coil and beta bend was observed, and devoid of stable conformation at the mutation site
Figure 3.2.7	**Figure 3.2.7;** Cross-correlation profile of **(A)** FANCD2 CUE domain *wild-type* and **(B)** Leu231Arg mutant; **(C)** Molecular model FANCD2 CUE domain *wild-type* and **(D)** Leu231Arg mutant, showing the mutant residue in α-helices
Figure 3.2.8	**Molecular dynamics simulation profile of *wild-type* CUE domain and Leu231Arg mutant: (A)** Scree plot showing comparably higher eigenvalues of in Leu231Arg mutant and **(B)** PCA graph suggest large tertiary structure transitions predominant in mutant in contrast to *wild-type* CUE domain
Figure 4.1	**(A)** DNA binding analysis of FANCI ARM repeat with DNA by autoradiography (C=Control untreated probe, 1, 2, 3, & 4 with increasing concentration of protein 4.6µM, 9.29µM, 17.14µM and 34.29µM respectively) **(B)** Graph plot of bound fraction of protein to DNA, calculated Kd value showing profound binding activity of protein with DNA

Figure 4.2	**(A)** DSSP profile of FANCI ARM repeat protein, **(B)** Superimposed structures of different time points (10 ns to 40 ns) of trajectories showing large structural rearrangement in HTH-type motif **(C)** Docking profile of FANCI ARM repeat with DNA dodecamer shows major groove binding mode of putative HTH-type motif interaction analysis profile of docked pose indicates the conserve residues present at binding and **(D)** Intermolecular interface (E) Docking results from HADDOCK
Figure 4.3	Secondary structure intermolecular interaction analysis profile of docked pose indicates the conserve residues present at binding interface (marked with red dots)
Figure 4.4	**SPR sensorgram of *wild-type* CUE and Leu231Arg mutant with monoubiquitin: (A)** Binding profile of mono-ubiquitin with *wild-type* CUE (arrow shows specificity is higher because of low dissociation constant) **(B)** Binding profile of mono-ubiquitin with Leu231Arg (arrow shows specificity is lower because of high dissociation constant)
Figure 4.5	**(A)** Representative Comparative binding profile of *wild-type* CUE and Leu231Arg mutant, shows *wild-type* CUE has the higher affinity compare to the Leu231Arg mutant, while Leu231Arg showing lesser affinity and anomalous behavior in binding
Figure 4.6	**Ligplot analysis of docked complex: (A)** CUE *wild-type* with monoubiquitin **(B)** Leu231Arg with monoubiquitin; Red dot circle shows the point mutation position in α-helices (C) HADDOCK docking results of Cue *wild-type* and monoubiquitin (D) HADDOCK docking results of Cue Leu231Arg mutant and monoubiquitin
Figure 4.7	**Docked complex structures of CUE domain and monoubiquitin complex: (A)** CUE *wild-type* with monoubiquitin **(B)** Leu231Arg with monoubiquitin
Figure 4.8	**PDB*SUM* analysis of docked complex: (A)** CUE *wild-type* with mono-ubiquitin **(B)** Leu231Arg with monoubiquitin
Figure 4.9	**Comparative profile for chaperone-like assay: (A)** *wild-type* CUE and **(B)** Leu231Arg mutant
Figure 5.1	STRING analysis of FANCI, FANCD2 and BRCA2 proteins
Figure 5.2	*In-vivo* protein-protein interaction studies of FANCI ARM Repeat, FANCD2 CUE Domain and BRCA2 C-Terminal Region (CTR), 1= FANCI ARM Repeat+ FANCD2 (1-472), 2= FANCI ARM Repeat+ FANCD2 CUE Domain, 3= FANCI ARM Repeat + BRCA2 CTR. 4= Cells co-expressing pUT18 and pKNT25 tag empty vectors were used as negative control, and 5= FtsA-T18 and FtsZ-T25 protein was used as positive control, β-Galactosidase activity in Miller units are shown here as mean±SD ($n \geq 2$)
Figure 5.3	Protein-protein interaction interface of FANCI ARM with BRCA2 C-terminal region and FANCD2 CUE domain

Figure 6.1	**(A & B)** *In-vivo* protein-protein interaction studies of FANCD2 CUE Domain with BRCA2 C-Terminal region, 1= FANCD2 CUE domain +BRCA2 CTR, 2= Negative control (pUT18 and pKNT25 tag empty vectors), and 3= FtsA-T18 and FtsZ-T25 protein was used as positive control (**B**) β-Galactosidase activity in Miller units are shown here as mean±SD (n≥2)
Figure 6.2	*In-vivo* protein-protein interaction studies of FANCD2 CUE Domain *wild-type*, Ser222Ala, Leu231Arg mutants with BRCA2 C-Terminal region, 1= FANCD2 CUE Leu231Arg+BRCA2 CTR, 2= FANCD2 CUE Ser222Ala+ BRCA2 CTR, 3= Negative control (pUT18 and pKNT25 tag empty vectors), and 4= FANCD2 CUE *wild-type*+ BRCA2 CTR (as positive control). Cells co-expressing pUT18 and pKNT25 tag empty vectors were used as negative control, and FANCD2 CUE Domain *wild-type* pUT18 and BRCA2-CTR-pKNT25 protein was used as positive control, β-Galactosidase activity in Miller units are shown here as mean±SD (n≥2) and the significance differences were analyzed using Student's t-test, and P values, obtained at 95 % confidence intervals, are shown as (*) for <0.05, (**) for <0.01 and (***) for <0.001.
Figure 6.3	**(A)** Schematic presentation of FANCD2 CUE domain and BRCA2 CTR **(B)** *In-vitro* protein-protein interaction studies of FANCD2 CUE Domain, the supernatant of co-expressed pellet of FANCD2 CUE and BRCA2 CTR were incubated with GST beads and analyzed on SDS-PAGE by using Snapcap column, M= Protein Marker, 1= GST-resin+ GST protein+ BRCA2 CTR, 2= GST-resin+ BRCA2 CTR, 3= GST-resin+ GST-Fusion FD2 CUE *wild-type*+BRCA2 CTR, 4= GST-resin+ GST-Fusion FD2 CUE Ser222Ala+BRCA2 CTR, 5= GST-resin+ GST-Fusion FD2 CUE Leu231Arg+BRCA2 CTR
Figure 6.4	*In-silico* protein-protein interaction studies of FANCD2 CUE Domain, FANCD2 CUE Mutants and BRCA2 C-Terminal region, A= FANCD2 CUE *wild-type*+ BRCA2 CTR, B= FANCD2 CUE Ser222Ala+BRCA2 CTR and C= FANCD2 CUE Leu231Arg + BRCA2 CTR

Chapter 1

Introduction
&
Review
of Literature

1.1 Fanconi anemia, DNA repair pathways and Cancer

Fanconi Anaemia (FA) is a genomic instability syndrome clinically characterized by congenital abnormalities, bone marrow failure and predisposition to malignancy [1]. It was first diagnosed by the Swiss paediatrician Guido Fanconi in the year 1927 [6, 31]. Cancer predisposition syndromes are closely related to the accumulation of DNA damage, and repair abnormalities [32-34]. The advancement in interdisciplinary areas such as genomics, proteomics, and metabolomics clearly suggest that the deregulation of DNA damage repair and cell -cycle checkpoint signaling are the main etiological factors for the most human cancers.

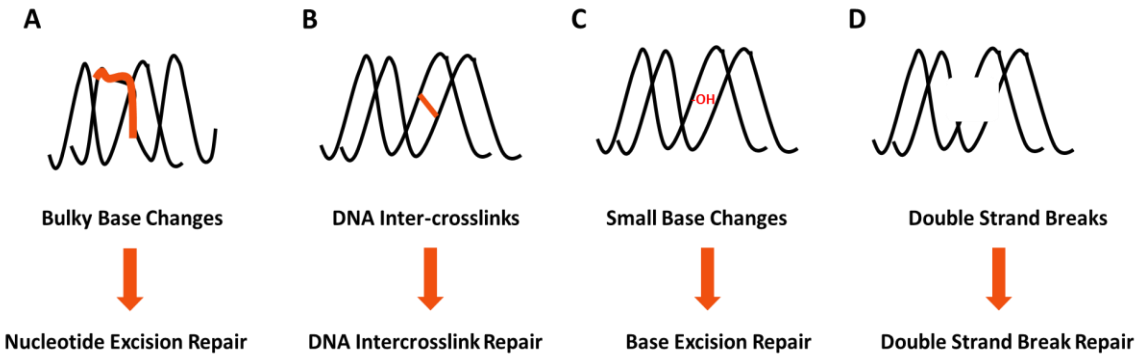


Figure 1.1:- Different types of DNA damage and repair mechanism; **(A)** Nucleotide Excision Repair **(B)** DNA Intercrosslink Repair **(C)** Base Excision Repair and **(D)** Double Strand Break Repair

However, the molecular mechanism of associated to different DNA-repair **(Figure 1.1)** and cell-cycle checkpoint pathways have still not been explored. Furthermore, the cells try to repair both endogenous and exogenous causes of DNA intercrosslink (ICL) damage [2]. The inherited or acquired deficiencies in conserved genomic blueprint system contribute ominously in the commencement of cancer predisposition [6, 31]. Fanconi

Anemia patients are affected by solid tumour and haematological malignancy at its early stage [6, 31]. Fanconi Anemia cells exhibit hypersensitivity to DNA interstrand crosslinking agents such as mitomycin c, diepoxy butane and cisplatin etc. Recent findings suggest that different proteins associated with Fanconi Anemia pathway play a pivotal role in DNA damage response especially in interstrand crosslink (ICL) repair mechanism [2, 6, 31, 35, 36].

Table 1.1 List of different proteins present in Fanconi anemia pathway with their chromosomal locations and functions

FA Proteins	Mol. Wt	Chromosomal Location	Functions	Functional Domains
FANCA	163	16q24.3	FANC Core	
FANCB	95	Xp22.31	FANC Core	
FANCC	63	9q22.3	FANC Core	
FANCD1 /BRCA2	380	13q12.13	Homologous Recombination	
FANCD2	162	3p25.3	DNA damages signalling Replication forks stabilization Transcription factor	Cue Domain, ARM Repeat
FANCE	60	6p21.22	FANC Core	
FANCF	42	11p15	FANC Core	
FANCI	140	15q25.16	DNA damages signalling Replication forks stabilization	ARM Repeat
FANCJ /BRIP1	140	17q22-q24	Helicase	
FANCL	43	2p16.1	FANC Core	
FANCM	250	14q21.3	Anchor the FANC Core Chromatin translocase	
FANCN /PALB2	130	16p12.1	Homologous Recombination	
FANCO /RAD51C	42	17q25.1	Homologous Recombination	
FANCP /SLX4	200	16p13.3	Endonuclease Transcription factor	
FANCQ /XPF	101	16p13.12	Endonuclease	
FANCS /BRCA1	207	17q21.31	E3 Ubiquitin ligase Homologous recombination	
FANCT/UBE2T	23	1q32.1	E2 conjugating Enzyme	
FANCR/RAD51	37	15q15.1	Homologous Recombination	
FANCU/XRCC2	32	7q36.1	Homologous Recombination	

Fanconi Anemia which is a highly heterogeneous genetic disorder comprising of 18 complementation groups with discrete genes FANC (A, B, C, D1/BRCA2, D2, E, F, G, I, J/BRIP1, L, M, N/PALB2, O/RAD51C, P/SLX4, Q/XPF, S/BRCA1 and T/UBE2T) [2, 5, 30, 31, 36, 37]. Since it is a rare human disorder, studies of Fanconi anemia pathway could

reveal many proteins and their interactive networks that preserve genomic integrity [6, 38]. The defect in DNA repair pathway specific for ICL's and molecular mechanism for hypersensitivity of Fanconi Anemia cells to DNA crosslinking agents are not well studied. The 18 Fanconi Anaemia proteins (FA proteins) have been reported to form different complexes and act in a common pathway to repair the intercrosslink [2, 5].

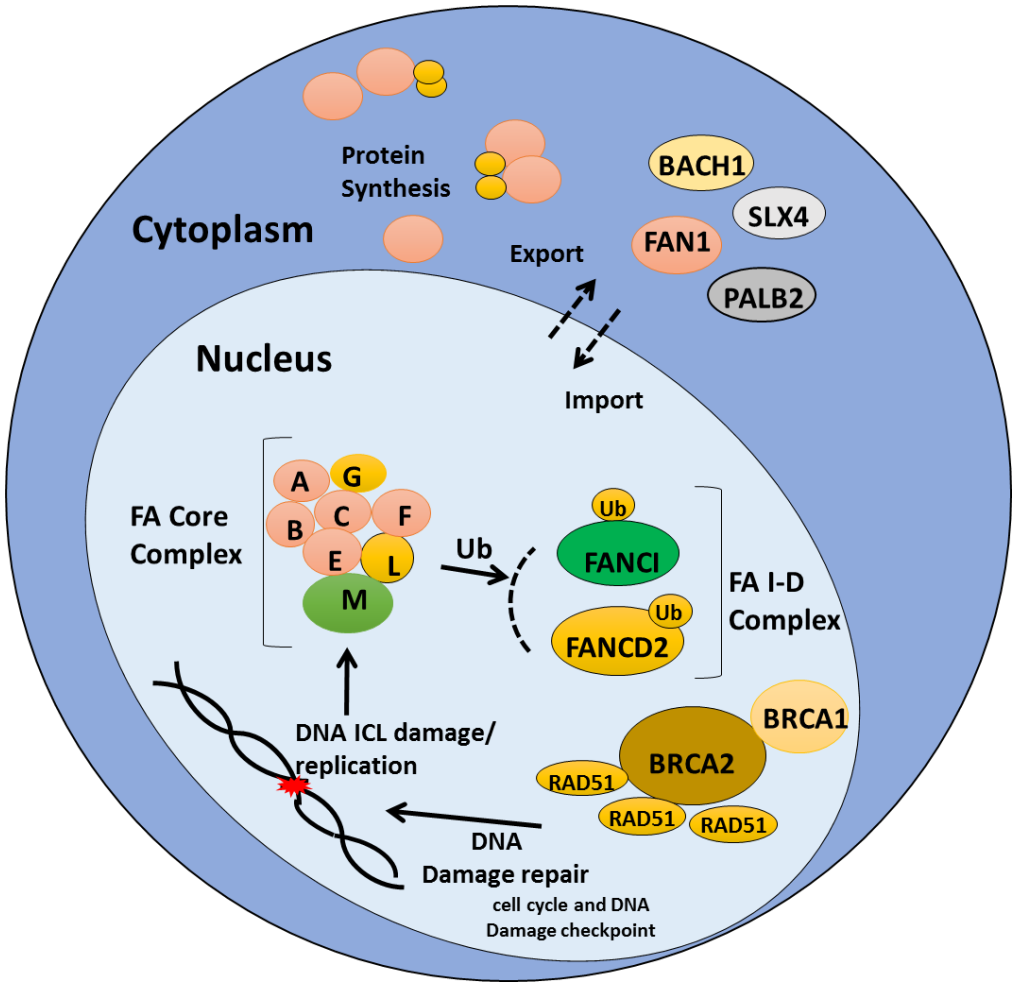


Figure 1.2:- DNA ICL repair pathway exhibiting DNA damage response specifically to FA core complex and FANCI-FANCD2 complex including BRCA proteins

The Fanconi Anemia (FA) proteins are categorized into three groups. Group first constitute eight FA proteins (A, B, C, E, F, L, G and M) having E3-ubiquitin ligase activity [2]. Furthermore, FANCL is the catalytic subunit and acts as a ubiquitin ligase. The second group comprises FANCI and FANCD2 proteins that forms FANCI-FANCD2 (I-D) complex. Proteins present in Fanconi Anemia core complex monoubiquitylates FANCI-FANCD2 on its conserved lysine residue. These monoubiquitinated FANCI and FANCD2 are recruited at DNA damage site for repair processes **(Figure 1.2**). The third group comprises of BRCA1 and BRCA2 that helps in DNA ICL repair by recruiting Fanconi anemia nuclease 1 (FAN1) and other FA-associated proteins which bridges the FA-BRCA pathway [2, 5, 31].

The Fanconi Anemia core complex is a downstream target of Ataxia telangiectasia Rad-3 related (ATR) checkpoint kinase which is activated by either DNA damage response or replication stress [24, 30, 39]. Proteins encompassing of Fanconi Anemia complexes are phosphorylated by ATR checkpoint kinase to get activated for DNA ICL repair [2, 3, 24, 40, 41]. These proteins act as signal transducers to promote DNA repair through homologous recombination and translesion synthesis pathways [3].

1.2 Fanconi Anemia and Breast Cancer Susceptibility

The Fanconi anemia proteins repair the DNA inter-crosslink by interacting to proteins involved in DNA double-strand break repair mechanisms such as BRCA1, BRCA2 (FANCD1), FANCN (PALB2) and FANCJ (BRIP1). **BR**east **CA**ncer susceptibility gene 1 (BRCA1) and **BR**east **CA**ncer susceptibility gene 2 (BRCA2) have been involved in DNA double-strand break repair through homologous recombination [2, 3, 42]. The DNA repair pathway in BRCA2 deficient cells leads to error prone single-strand annealing [43]. Furthermore, DNA repair processes such as nucleolytic incision, translesion DNA

synthesis, and homologous recombination have been coordinated by FA/BRCA pathway [44].

BRCA1 and BRCA2 are the tumor suppressor proteins, and mutations in any part of these genes are linked to breast and ovarian cancer [45-48]**.** Breast cancer neoplasms are closely associated with DNA damage repair and cell-cycle checkpoints defects [43, 45, 47, 49]. Different proteins expressed in homologous recombination repair pathway are the cellular binding partner of BRCA1 and BRCA2 [2]. BRCA1 and BRCA1-associated RING domain protein 1 (BARD1) are a ring finger protein which has ubiquitin ligase activity and together they can monoubiquitinate the FANCD2 [50]. BARD1 has an important role in domain stabilization with CstF50 complex [51]. Member of BRCA1-A complex proteins such as Coiled-coil domain-containing protein 98 (CCDC 98/ABRAXAS) and RAP 80 also interact with FANCD2 [52]. Furthermore, BRCA1 and BRCA2 are inevitably involved in various cellular processes like transcriptional regulation, cell-cycle checkpoint control, and cellular response to DNA damage repair processes [2, 42, 43].

Besides great achievement in the field of cancer biology, there are still different unanswered questions such as how the impairment in DNA repair pathway leads to developmental abnormalities, bone marrow failure, and predisposition to cancer. Furthermore, the mechanism underlying the activation of Fanconi Anemia pathway for DNA cross-link damage but not to other forms of DNA damage remains a serious question for study. The cellular, biochemical and molecular studies of Fanconi Anemia patients can unravel the molecular mechanisms for different proteins associated to Fanconi Anemia pathway in order to improve the diagnosis, early detection, and treatment of cancer.

1.3 Activation of Fanconi Anemia Pathway

Eukaryotic DNA is uniformly packed in chromatin and get exposed at the time of replication [53]. It has been reported that circular dsDNA is sufficient to activate FANCD2 monoubiquitylation in oocyte extracts of *Xenopus laevis* [54]. Interestingly, double strand break (DSBs), an intermediate during DNA ICL damage repair is not required for the activation of the Fanconi Anemia pathway proteins [54-58]. The DNA structures that are spawned during replication, such as ssDNA and Y-shaped DNA are not known for the activation of FANCD2 monoubiquitylation [25, 58-62].

Fanconi Anemia genes individually identified by linkage analysis, positional cloning and genome wide analysis are also known to function in the FA pathway for genomic maintenance [1, 3, 63-67]. The findings of different Fanconi Anemia genes i.e. FANCA, FANCC, FANCF, FANCG, FANCI, and FANCD2, have led to the identification of the DNA-repair molecules involved in the FA-BRCA network [1, 5, 44, 68, 69]. FANCI and FANCD2 protein expression increases when cells are exposed to various DNA ICL damaging agents [3, 44, 70].

Fanconi Anemia core complex is required for FANCI and FANCD2 monoubiquitination, however, if the cells are lacking any protein of the Fanconi Anemia core complex, the monoubiquitinated form of FANCI and FANCD2 will not be present [3, 5, 71, 72]. The proteins which connect the FA-BRCA pathway are FANCD1 (BRCA2), FANCN (Partner and Localizer of BRCA2, PALB2) and FANCJ (BRCA1-Interacting Protein BRIP1, or BRCA1-Associated C-terminal Helicase 1, BACH1) [2, 12, 13, 34, 39, 42, 73-76]. The cells those are defective in any of the members of third group proteins such as BRCA1, BRCA2 and BACH1 have normal levels of FANCI and FANCD2 monoubiquitination. This suggest that the corresponding proteins function either downstream of the ID complex or in a parallel pathway [2, 3, 69, 74, 77].

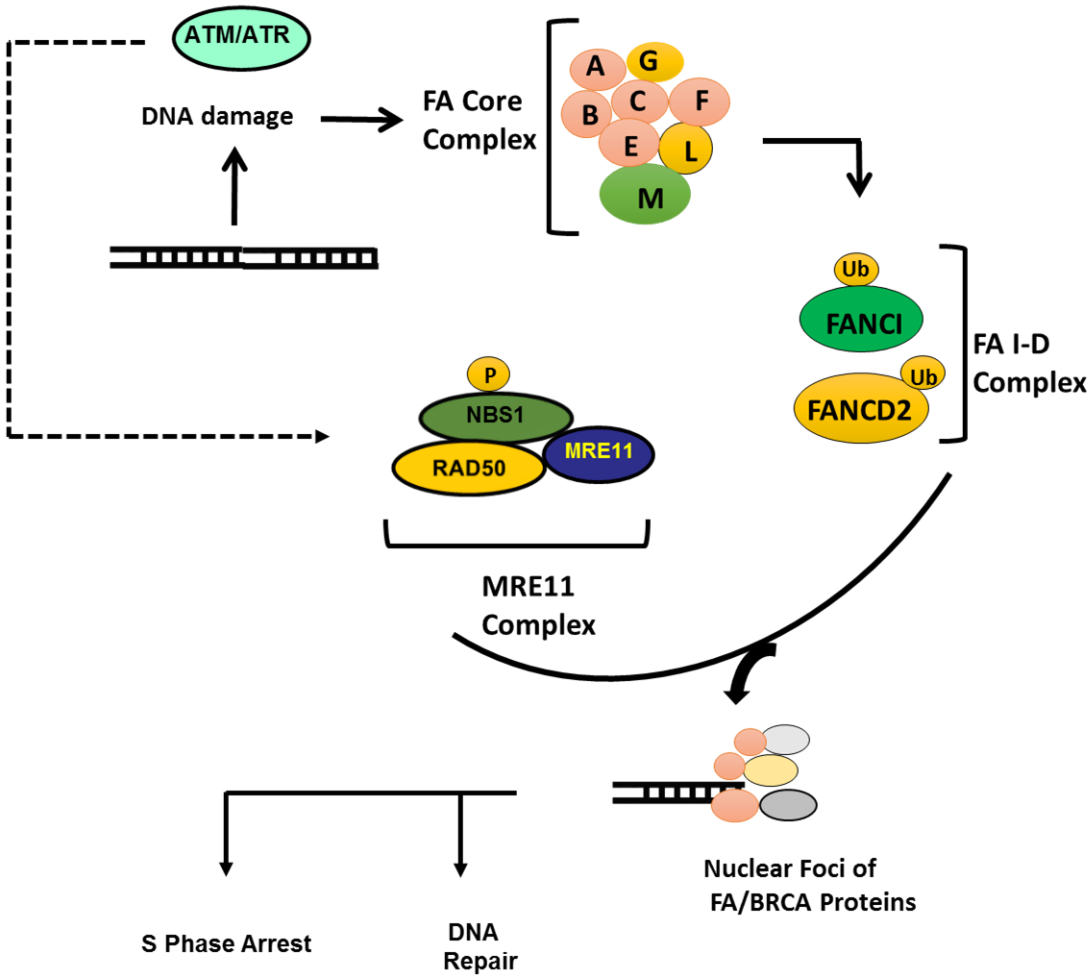


Figure 1.3:- Fanconi anemia pathway: DNA ICL repair pathway sensing DNA damage response by ATM/ATR kinases that phosphorylates specifically FA core complex, FA I-D and MRE11 Complex, plays an important role in S phase arrest and DNA repair

The homologous recombination repair mechanism for DNA double-stranded ended forks and replication stalled forks are different [78-80]. The adjacent to the break of the replication complex may facilitate recombination with the correct template sequences [3, 5, 44]. Furthermore, the MRE11–RAD50–NIBRIN (NBS1; known as the MRN) complex, is required for the end resection of replication forks which is further regulated by RAD51 to avoid the extensive resection [2, 3, 5, 44, 68] (**Figure 1.3**). The recruitment of homologous recombination repair and DNA ICL proteins that function specifically in DNA

intercrosslink repair mechanism suggest that these proteins are recognizing the structural differences in the DNA intermediates during the DNA damage repair process [5, 26, 58, 70, 75, 81-84].

1.4 Fanconi Anaemia; FANCI-FANCD2 complex

FANCI and FANCD2 are paralogues that share miniscule sequence homology. It has been reported that FANCI and FANCD2 are co-ordinately regulated to work as an ID complex. This ID complex is also recruited to DNA damage sites and perform independent functions [2, 3, 85, 86] (**Figure 1.4**). However, FANCI-FANCD2 co-immunoprecipitate with each other from cell lysates and both proteins are monoubiquitinated in response to DNA damage or replication [2, 3]. This monoubiquitylation of FANCI and FANCD2 depend on the protein present in Fanconi Anemia core complex, and their de-ubiquitination is catalyzed by the ubiquitin-specific protease 1 (USP1) [2, 3]. Once the cells are exposed to DNA ICL damaging agents, the sensor kinases ataxia telangiectasia RAD-3 related (ATR) or ataxia telangiectasia mutated (ATM) kinases phosphorylate the FANCI and FANCD2 proteins [2, 3, 87, 88]. Thereafter, they get localized into the chromatin at the sites of ICL DNA damage to form nuclear foci. Both FANCI and FANCD2, depends on the presence of its binding partner for stability and ubiquitination [2, 3]. Hence, the I-D complex is activated by DNA damage induced specific phosphorylation and monoubiquitination [2, 3, 5, 44]. The post-translational modifications such as phosphorylation and monoubiquitination serve as a stabilizer for FANCI and FANCD2 heterodimeric complex [3, 5, 44].

Recent results and available crystal structure of FANCI and FANCD2 complex suggest that both the protein have an ability to bind single as well as double-stranded DNA [26]. In addition, they also have an ability to bind single-stranded as well as double stranded

DNA independently [26, 85]. This unravel the facts that I-D complex can recognize DNA structures resulting from DNA ICL damage. Hence, FANCI-FANCD2 (I-D) complex facilitates as a protecting agent to these DNA structures and also provides specificity for initial incision of ICL damage [3, 26, 85].

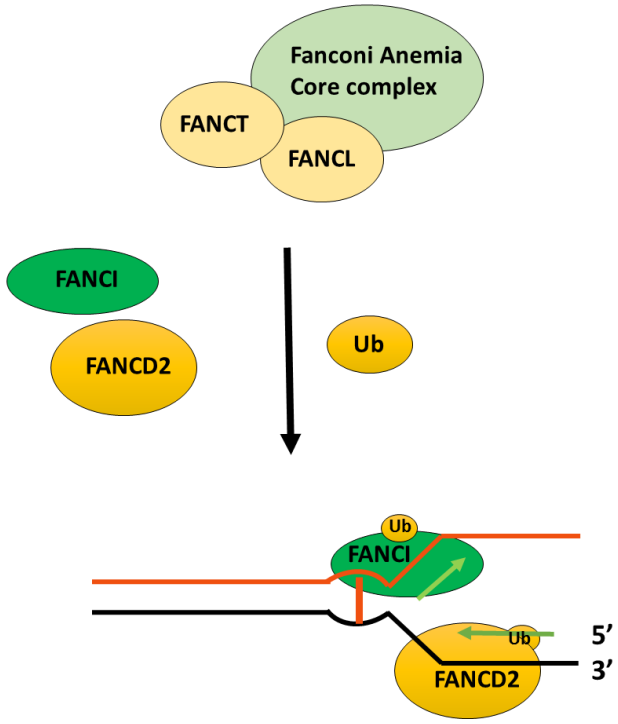


Figure 1.4: Schematic diagram for Intercrosslink DNA damage and response network; Role of FANCI-FANCD2 complex in Fanconi Anemia pathway

FANCI-FANCD2 (I-D) complex perform different functions independently of the FA pathway [2, 3, 26, 44, 85]. The clinical phenotypes in Fanconi Anemia patients with FANCD2 mutations are severe than those who are defective in protein present in Fanconi Anemia core complex [3, 89, 90]. Most of the identified FANCD2 mutations are hypomorphic which suggest the importance of FANCD2 in embryonic development [90, 91]. The *fancd2* knockout mice show phenotypes such as small eyes, perinatal lethality, and epithelial cancers. However, it has not been observed in the mice containing disrupted

genes encoding Fanconi Anemia core complex [92-94]. The majority of FA core complex genes are not present in many eukaryotic species that do contain orthologues of FANCD2 and FANCI [3, 95].

It has been reported that the FANCI-FANCD2 (I-D) complex is important for Fanconi Anemia pathway [1-3, 5, 12, 26, 40, 55, 59, 60, 62, 68, 69, 77, 83, 86, 93, 96-117]. The FANCD2 point mutations are causing severe malformations and manifestations of haematological malignancies [90, 91]. In addition, Acute myeloid leukemia (AML) patients with FA disease have a high risk of developing solid tumor compared to other cancer population [6-8, 36, 90, 91, 118-130].

1.5 Molecular basis of FA pathway

The understanding of Fanconi anemia at the molecular level is continuously evolving, as new members are being discovered. In context to genomic maintenance, Fanconi anemia proteins function to stabilize the replication forks and regulate the cytokinesis [131]. Fanconi anemia proteins together with BRCA proteins are the master regulators for genomic maintenance and integrity [25, 54, 131-134] **(Figure 1.5).** Consistent with the recent reports, Fanconi anemia deficient mice and Fanconi anemia patients are sensitive to aldehyde compounds which suggest that Fanconi anemia pathway has evolved intracellularly to avert with the endogenous cytotoxicity [56, 135-138].

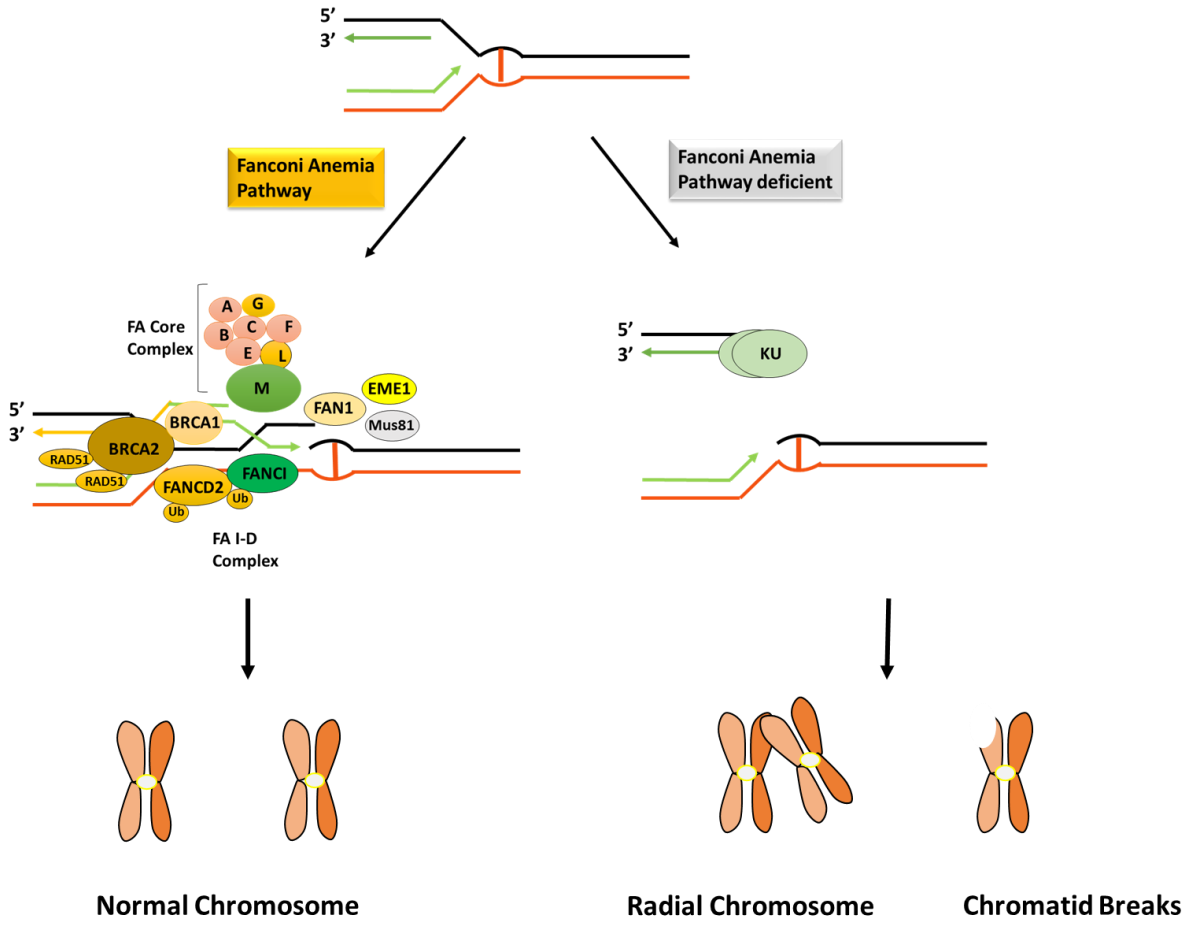


Figure 1.5: Schematic diagram of Fanconi Anemia pathway proteins during repair process

The DNA ICL damage arises from endogenous crosslinking agents such as acetaldehyde, malondialdehyde and nitrous acid [56, 135-138]. Furthermore, exogenous sources of DNA inter crosslink are clinically relevant as it employed during chemotherapy to the cancer patients [96, 138-143].

Fanconi anemia pathway undertakes the DNA ICL damage repair as follows, (1) DNA Lesion recognition (2) DNA Lesion incision (3) DNA Lesion bypass and (4) DNA Lesion repair [2, 5, 11, 44, 144-146].

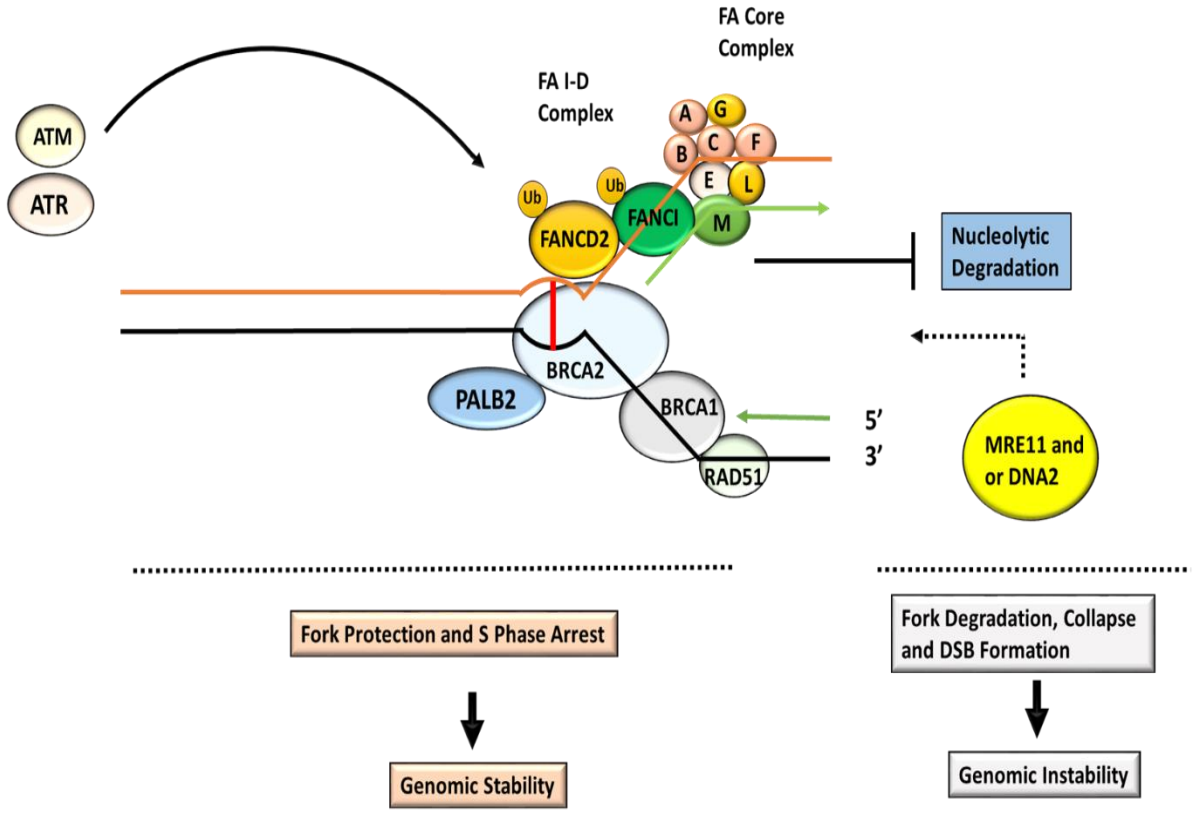


Figure 1.6: Schematic diagram of Intercrosslink DNA damage response network FANCI-FANCD2 complex. BRCA and RAD51 proteins that facilitate the lesion recognition, lesion incision and homologous recombination repair in DNA ICL repair process of Fanconi Anemia pathway

(1) **DNA Lesion recognition**- Once DNA Lesion is sensed by ATM-ATR kinases, the FANCM protein gets phosphorylated and recruited to DNA damage sites which in turn localize FA core complex proteins and forms nuclear foci during S and G2 phase of the cell-cycle [3, 5, 85]. The Fanconi anemia-associated proteins (MHF1/FAAP 24), histone fold protein (FAAP 16/CENPS) and MHF histone-fold complex subunit 2 (MHF2/FAAP 10/CENPX) helps during the DNA intercrosslink repair [3, 5, 26, 44]. These Fanconi anemia associated proteins together with ATR and BRCA1 (FANCS)

are required for the sensing, recognition, and recruitment to DNA ICL sites (**Figure 1.6**). The Fanconi anemia core complex comprises namely FANCA, FANCB, FANCC, FANCE, FANCF, FANCG, FANCL, FANCM, FANCT/UBE2T, FAAP100, MHF1, MHF2, FAAP 20 and FAAP 24 proteins. These proteins have multiple functions and are recruited to DNA ICL sites [2, 3, 5, 26, 44]. They also act as a large ubiquitin ligase complex and ubiquitinates the FANCI and FANCD2 proteins. The FANCI and FANCD2 monoubiquitination process are interdependent and reversible [3, 147, 148]. Furthermore, the phosphorylation of FANCI and FANCD2 are responsible for the heterodimer dissociation and their independent functions at the chromatin sites [3, 5, 85].

(2) **DNA Lesion incision**- Recent study suggest that converging replication forks activate the DNA ICL repair [54]. The replication fork initially paused from the DNA ICL sites 20 to 40 base pairs away to avoid the steric clashes levied by a replicative helicase complex (CMG complex) [59]. FANCI and FANCD2 control the nucleolytic incision of DNA ICL at converged replication forks, a process known as 'unhooking' [149, 150]. Monoubiquitinated FANCD2 recruits the nuclease scaffold SLX4 (FANCP), ERCC4 (FANQ) and Fanconi anemia nuclease 1 (FAN 1). ERCC 4 and ERCC 1 heterodimer has a crucial role in unhooking, therefore FAN 1 orient the DNA and unwinds the 5' flap to promote the DNA incision [101, 151]. Furthermore, FAN 1 is required to resume the replication of stalled forks [88].

(3) **DNA Lesion bypass-** After unhooking of the DNA ICL from the damage sites, translesion synthesis polymerase (TLS polymerase) is recruited with the help of ubiquitinated PCNA and Fanconi anemia core complex [152, 153]. TLS polymerase performed lesion bypass by adding nucleotides opposite to the DNA ICL and extend the nascent DNA strand [154, 155]. Extension of nascent DNA strand by TLS

polymerase complex which is composed of REV 1 and REV3-REV7 and have a role in the restoration of one intact DNA duplex, which is used as a template during homologous recombination processes [156-158].

(4) **DNA Lesion repair**- After the nucleolytic incision step, homologous recombination (HR) process complete the DNA ICL repair [54]. In homologous recombination, third group of FA proteins plays a major role in FA-BRCA pathway [2]. BRCA2 and Partner and Localizer of BRCA2 (PALB2) are known to perform the HR with RAD51, and also form the nucleofilaments of single stranded DNA to catalyze the strand invasion of homologous template for successful homologous recombination [159, 160]. In addition, BRCA1 interacting protein C-terminal helicase 1 (BRIP1/FANCJ) acts as a multifaceted repair protein because it perform BRCA1 independent DNA repair function not identical to canonical homologous recombination [161]. Furthermore, FANCJ and BRCA1 interactions are tightly regulated by post-translational modification with specific phosphorylation at Ser990 position in FANCJ to carry out proper translesion synthesis and homologous recombination [162]. Fanconi anemia mediated DNA ICL damage repair involves the processes like unhooking of DNA ICL lesion, nascent ssDNA formation, strand invasion, homologous recombination, resolution, and ligation which gives rise to an intact double-stranded DNA [2, 44, 163].

1.6 Functional domains of FANCI

Fanconi anemia complementation group I (FANCI) protein is one of the FA pathway protein recruited to the DNA ICL damage site to perform the repair function [3, 115, 156]. FANCI plays a key role in FANCD2 monoubiquitination, and also get monoubiquitinated at conserved lysine at 531 position which further triggers the activation of FANCI-

FANCD2 (I-D) Complex [41, 104, 164]. The monoubiquitinated FANCI is recruited to DNA damage site and forms the DNA repair foci [3, 5, 85]. FANCI protein is a leucine rich peptide (14.80% leucine residues) having miniscule homology with other reported proteins [3, 20, 115]. FANCI protein comprises of 1328 amino acids and has armadillo repeat (ARM repeat), EDGE motif at C-terminus [3] (**Figure 1.7)**. The ARM repeat forms α-α superhelix fold which is involved in the protein-protein and protein-DNA interactions [9, 19, 165-168]. The mutant FANCI has a dominant negative effect on FANCD2 monoubiquitination and DNA damage-induced FANCD2 foci formation [3, 115]. Therefore, the FANCI Lys523Arg mutant is not able to participate in DNA damage foci formation [3, 115]. The Lys525Arg mutant chicken DT40 cells have an inability to restore cellular resistance to the DNA cross-linking agent mitomycin C [169]. It also affects the FANCD2 chromatin recruitment, suggesting that FANCI monoubiquitination plays an important role in the chicken cells [3, 169]. However, additional studies needed to conclude the significance of FANCI monoubiquitination in the activation of the Fanconi anemia DNA repair pathway in mammalian and non-mammalian cells.

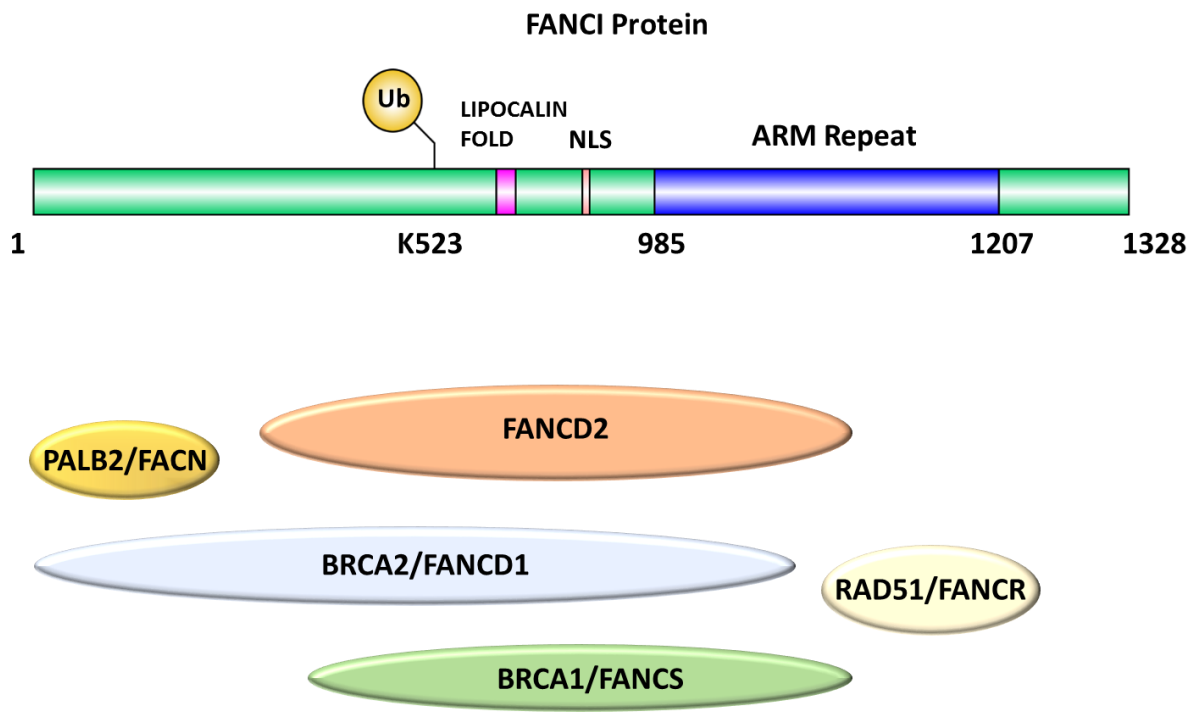


Figure 1.7: Domain organization of FANCI protein; ARM repeat at C-terminus and conserved Lysine at 523 amino acid position. FANCI protein interacts with specifically FANCD2 which further binds BRCA2, PALB2, RAD51, BRCA1 to form nuclear foci at the DNA damage site

The interesting observation regarding the phosphorylation status of FANCI is decisive for FANCD2 monoubiquitination, and activation of I-D complex [3, 169]. In the DT40 cell line, phosphomimetic mutations of ATR phosphorylates consensus sequences (S/T)Q containing motifs. A conserved cluster at the Lys-525 monoubiquitination acceptor site of FANCI act as a monoubiquitination switch for FANCD2 and FANCI-FANCD2 localization at chromatin DNA damage site [3, 86, 169]. However, the FANCI phosphorylation is not enough to promote localization of FANCD2/FANCI without other DNA repair proteins such as RAD51 and FA pathway proteins [2, 3, 5, 85, 86, 169].

After DNA damage response monoubiquitinated FANCI localizes to DNA ICL damage at chromatin site and facilitate the repair process with FANCD2 [2, 3, 85]. The C- terminal region of FANCI which harbours ARM repeat binds double-stranded DNA [3, 25, 83]. Interestingly it was found that the C-terminal region encompassing the 985-1328 amino acids has the similar DNA binding ability with similar DNA substrate preference [83]. Disease causing mutations are also found in the C-terminal region of the FANCI [2, 3, 169].

1.7 Functional domains of FANCD2

FANCD2 protein has multifaceted role such as DNA repair, cell cycle checkpoint control and cytokinesis [56, 58-61, 87, 99, 103-105, 114, 116, 117, 170-176]. Fanconi anemia core complex monoubiquitinates FANCD2 at conserved lysine 561 position and is recruited to DNA ICL damage site [3, 41]. FANCD2 interacts with other FA proteins mainly FANCI, BRCA2 and RAD 51 involve in DNA ICL repair [173]. Recent reports suggest that FANCD2 monoubiquitination is necessary for the recruitment of the Fanconi anemia-associated nuclease (FAN1) and SLX4/FANCP endonucleases to DNA damage sites [11, 69, 101, 109, 149, 177]. FANCD2 and FANCI play a critical role in interstrand crosslink repair mechanism [3], however, FANCD2 also has an important role in meiosis [178].

FANCD2 protein comprises of 1451 amino acids including coupling of ubiquitin conjugation to endoplasmic reticulum domain (CUE domain) and armadillo repeat (ARM repeat) [67, 179] **(Figure 1.8**). FANCD2 mono-ubiquitination occurs in response to DNA damage during the S-phase of the cell-cycle [173]. The monoubiquitinated FANCD2 is targeted to nuclear foci containing proteins such as BRCA1, BRCA2, and RAD51 [3, 173] and involved in DNA-damage signaling and homologous recombination repair

mechanism. The FANCD2 deficient DT40 cells are inefficient in homologous recombination-mediated DNA ICL repair, DNA double-strand break (DSB) repair and trans-lesion synthesis [71, 110, 113, 173, 180]. FANCD2 has multifunctional role in addition to have functions in the maintenance of genomic stability [3, 71, 85, 108, 109, 111-114, 117, 139, 149, 172, 173, 177, 180-187].

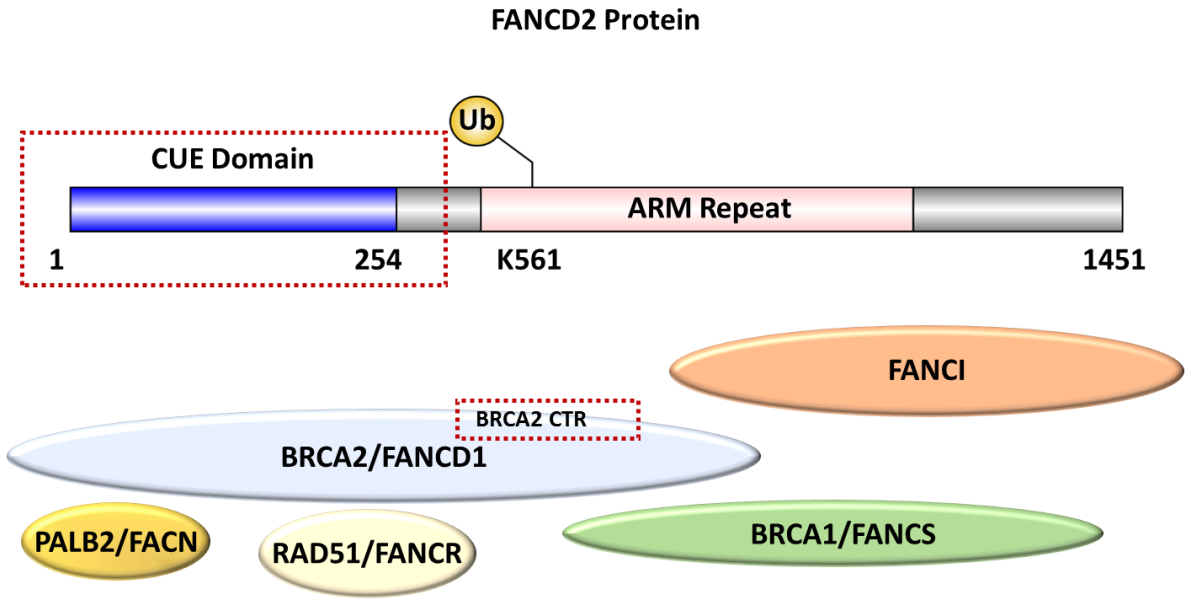


Figure 1.8: Domain organization of FANCD2 protein; the CUE domain at N-terminus, a conserved Lysine at 561 amino acid position and ARM Repeat. FANCD2 N-Terminus known to interact with BRCA2 C-terminal region (CTR) (in red dots).

Furthermore, *fancd2* knockout mice are severely affected from microphthalmia, early epithelial cancers and perinatal lethality, but it remains unclear whether the murine FANCD2 phenotype is severe than the corresponding murine knockouts of the FA genes [92, 94, 188]. It has been reported that in *Caenorhabditis elegans*, *fancd2* is required for

survival after DNA damage [97, 189]. *fancd2* deficient *zebrafish* embryos show severe developmental defects and increased apoptosis [190]. These reports clearly indicate the importance of *fancd2* function in the vertebrate developmental pathway, other than DNA ICL repair. In *Drosophila*, the *fancd2* knock-down causes pupal lethality [90, 191, 192]. Whereas in humans, it has been reported that complementation group FANCD2 accounts for <1% to 3% of Fanconi anemia mutations [90]. FANCD2 CUE domain encompasses the region from 1-254 amino acids at N-terminus of FANCD2 regulates the FA pathway by correcting the mitomycin C hypersensitivity and localizing the FANCD2 and FANCI proteins at DNA damage site [3, 4, 181]. FANCD2 CUE domain has an important role in the FANCI-FANCD2 complex formation and escaping the FANCI protein from polyubiquitination, thereby maintaining protein homeostasis (proteostasis) at the time of DNA ICL repair.

1.8 Domain organization and functional association of BRCA2

BRCA2 is one of Fanconi anemia protein identified as an FANCD1, implicated in the DNA inter-crosslink (DNA ICL) repair pathway failure [193-195]. Germline mutations in the *brca2* gene confer an elevated lifetime risk of developing breast and ovarian cancer [45, 47, 193, 195-197]. BRCA2 is composed of 3418 amino acids, a large protein that bridges the FA and BRCA pathway, and also indicate some common pathway of DNA double strand break and ICL repair [198]. The tumour suppressor function relates to a role of BRCA2 protein in homologous recombination [45]. Tumorigenesis is linked with loss of heterozygosity of the non-mutated allele. The gene encoding BRCA2 was identified by linkage analysis and positional mapping on the basis of the inheritance of chromosomal markers from a large cohort of affected families [2, 34, 45, 76, 159].

Fanconi anemia proteins interact with cellular partners like BRCA1, BRCA2 (FANCD1), FANCN (PALB2) and FANCJ (BRIP1) [2, 12, 13, 74, 76, 123, 178, 196]. FA/BRCA proteins function in a coordinated fashion to repair the DNA ICL and avert cellular transformation. FANCI, FANCD2, and BRCA2 act together to perform crucial functional role in DNA ICL repair. BRCA2 protein is specifically involved in DNA ICL repair, double-strand break repair through homologous recombination. The *brca*2 deficient cells cause error prone single-strand annealing pathway [159, 199].

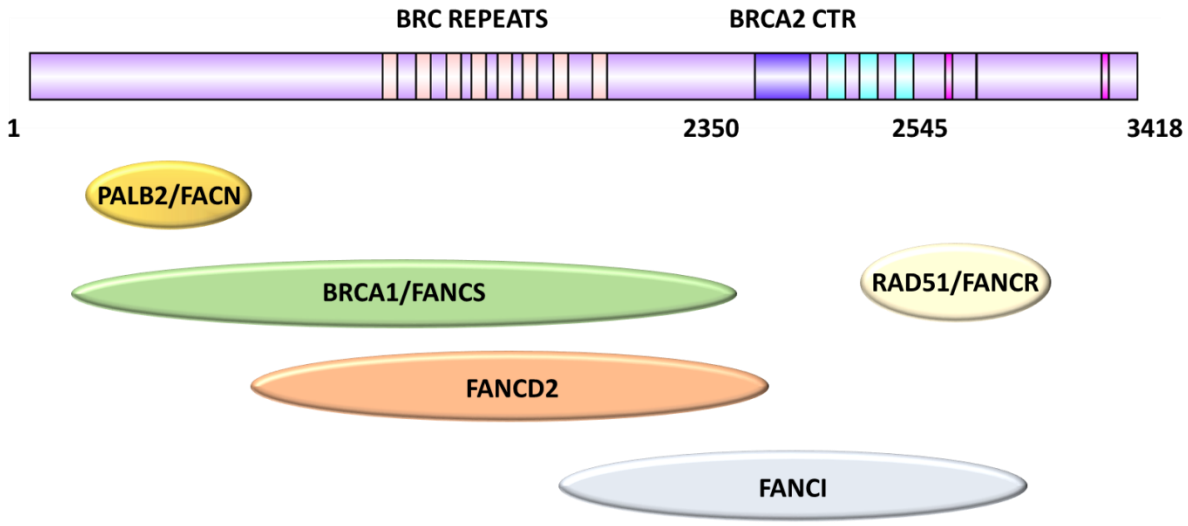


Figure 1.9: BRCA2 protein and its interacting partners

BRCA2 interacts with RAD51 which forms filaments on single-stranded DNA (ssDNA) to perform DNA strand invasion and homologous recombination [2, 12, 159, 196]. It has been reported that human BRCA2 binds six RAD51 molecules and promotes RAD51 binding to ssDNA coated by replication protein A (RPA) [196] (**Figure 1.9**). The reported crystal structures (PDB ID; 1IYJ and 1MIU) of 800-residue from the C-terminal domain of BRCA2 unravel the molecular insights of the protein-DNA binding at atomic level [200-

202]. The C-terminal region of BRCA2 is the most conserved region across metazoan, plant and fungal orthologs, and also known as BLAT domain (BRCA2-motif in *Leishmania, Arabidopsis, and Trypanosoma* domain) [12, 13]. Deleted in split hand /split foot syndrome (DSS1), which is a small acidic protein has also been recognized as a BRCA2 interacting partner in a yeast two-hybrid system [203]. Structural insights obtained from protein structure of the mouse and rat contain bound DSS1, and ssDNA [203, 204]. Five domains are structurally characterized, four of them are globular and packed well in a linear array [202, 204].

1.9 Ubiquitination, de-ubiquitination and its role in DNA ICL repair

Ubiquitin is a 7.6 kDa protein composed of seven lysine residues Lys 63, Lys 48, Lys 33, Lys 29, Lys 27, Lys 11 and Lys 6 with specific linkage during chain assembly [205-207]. Most of the proteins undergo degradation through proteasome targeted by Lys48-linked chains [208]. However, Lys63-linked chain present non-proteolytic signals involved in protein trafficking [209].

Ubiquitination is a systematic process which involves at least three classes of enzymes; a ubiquitin-activating enzyme E1, a ubiquitin-conjugating enzyme E2, and a ubiquitin ligase E3. In FA pathway FANCL acts as a ubiquitin ligase.

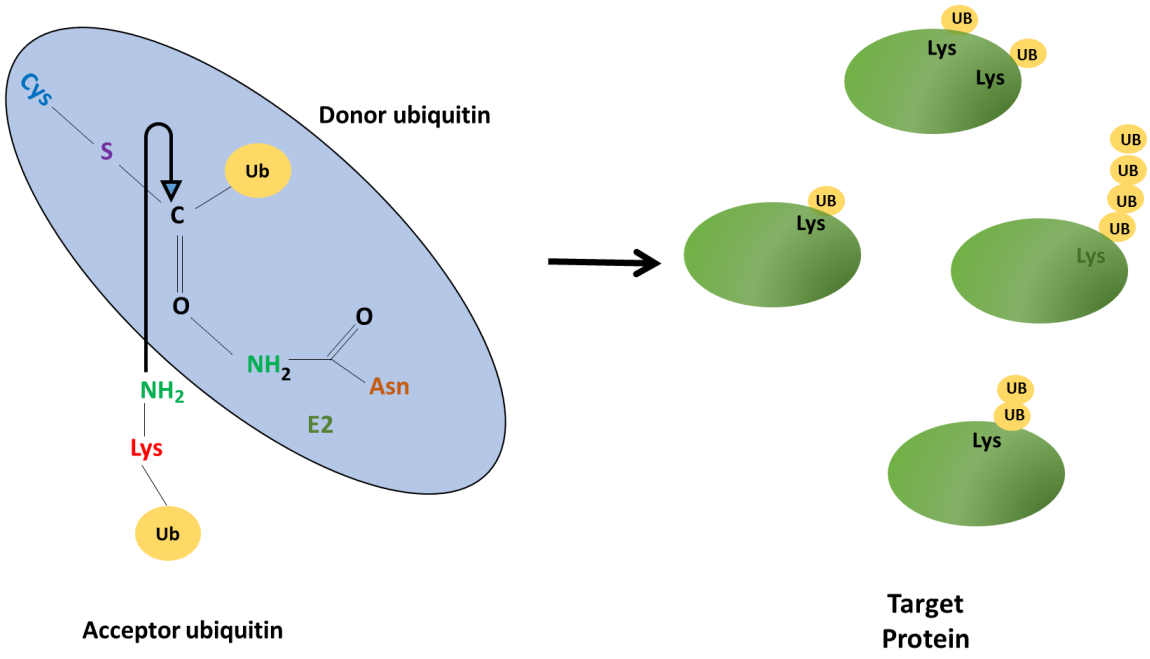


Figure 1.10: The model illustrates the E2-Catalytic mechanism. Isopeptide bond formation during ubiquitin chain synthesis, and different Lys residues with specific linkage during mono, di and polyubiquitination.

At the final stage of ubiquitination, a ubiquitin molecule forms covalent bond to a specific Lys residue of the target substrate protein by an isopeptide bond [205, 207, 209] **(Figure 1.10**).

Monoubiquitinated substrate Lys residue serves as a target for E3 to assemble next ubiquitin molecule and thereby forming polyubiquitin chain, and in some cases, monoubiquitin is masked by ubiquitin-binding domain to maintain proteostasis until DNA repair is not completed [3, 210, 211].

E1 forms a thioester linkage with cysteine located at active site and the C-terminus of ubiquitin through ATP molecule [205, 209]. The ubiquitin molecule is then moved to the active site cysteine of E2, and from E2 to cysteine residue of HECT domain of the E3 prior to the substrate attachment. Eukaryotic organisms produce many types of E3 to ensure the

specificity of ubiquitination. Most of E3s have RING domain and a structurally related Ubox that bring a substrate and a charged E2 together, thereby activating the E2 to ligate ubiquitin molecule [206, 207]. Ubiquitin usually attaches to the ε-amino group of Lys in substrates through isopeptide bond. Transfer of a single ubiquitin can occur to one (monoubiquitylation) or two (diubiquitylation) or multiple (multi monoubiquitylation) sites to recruit binding partners during DNA repair, protein-protein interactions, protein localizations and in cell signaling activities [205-207, 209].

1.10 Conclusion

Fanconi Anaemia regarded as genetic instability syndrome in which cells are highly sensitive to DNA crosslinking agents such as mitomycin c, diepoxy butane and cisplatin etc. [6, 31]. The identification and characterization of 19 Fanconi Anemia genes (One more is recently identified) have revealed a DNA-damage response network that comprises novel signaling cascades with phosphorylation and ubiquitination [3, 5]. Fanconi anemia proteins in association with other protein complexes performs various DNA-processing activities that facilitate the repair mechanism [5, 30, 44, 212]. FA proteins also crosstalk with other DNA repair proteins that intriguingly suggests their importance beyond DNA damage response and repair. Fanconi anemia proteins are also involved in acute myeloid leukemia (AML) and BRCA pathway to function in cell-cycle transition and chromosome segregation [41, 130, 178, 213]. Fanconi Anemia pathway could explore some important key target that might be useful for therapeutic application of Fanconi Anemia disease in near future.

Chapter 2

Materials

And

Methods

2.1 Materials

Molecular biology grade reagents were used to perform the experiments. Autoclaved, double distilled, 0.44 µM filtered Milli-Q water (Millipore, USA) was used to prepare the buffers for protein purification, protein-protein interactions, protein-DNA interactions and other spectroscopic studies. Buffer solutions and purified proteins were either centrifuged at 13000 rpm or filtered through 0.44-micron filter (Millipore, USA) and degassed prior to performing experiment. DNA polymerase and restriction enzymes for gene- cloning were purchased from Fermentas (https://www.thermofisher.com) and New England Biolabs (https://www.neb.com/). All other chemicals and consumables were procured from Sigma-Aldrich, Promega, Roche, GE Healthcare and Boston Biochem, or unless otherwise specified.

2.1.1 Materials and Instruments

(A) Bacterial medium, chemicals and materials; LB medium (Himedia, India, Cat. No. M575), Petri plates 90 mm (diameter) x 15 mm (Himedia, India, Cat. No. PW008), β-D-Isopropyl thiogalactopyranoside (IPTG) (MP Biomedicals, USA, Cat. No. 102101), Chemicals, salts, buffers, precipitants, detergents and organic solvents were procured from Sigma (USA), Merck (Germany), Fluka (Germany), GE Healthcare (Illinois, USA), Himedia (India), MP Biomedicals (USA) and SRL (India)

(B) Molecular biology grade antibiotics; Ampicillin (Himedia, India, Cat. No. CMS645), Kanamycin (Himedia, India, Cat. No. TC136), Chloramphenicol (Himedia, India, Cat. No. TC204)

(C) Protease inhibitor cocktail (Sigma, USA Cat. No. P2714), snake skin dialysis bag (Thermo Fisher, USA, Cat. No. 68035)

(D) Enzymes, Trypsin Singles (Sigma, USA Cat. No. T7575)

(E) Bacterial plasmid DNA isolation kits; Mini-prep kit (Qiagen, Germany, Cat. No. 27106), Plasmid DNA Maxi prep kit (Qiagen, Germany, Cat. No. 12163), Polymerase Chain Reaction extraction kit (Qiagen, Germany, Cat. No. 28106) and Agarose Gel DNA Extraction kit (Qiagen, Germany, Cat. No. 28706)

(F) Chromatographic affinity resins; Nickel- nitrilotriacetic acid (Ni-NTA) (Qiagen, Germany, Cat. No. 32169), Glutathione-S-Transferase (GST) resins (Novagen, Germany, Cat. No.70541) and Amylose (New England Biolabs, USA, Cat. No. E8021L)

(G) Pre-packed size exclusion chromatography columns; Superdex-200 and Superdex-75 GE Healthcare (Sweden)

(H) Crystallization screens; Hampton Research (USA) and Qiagen (Germany)

(I) Monoubiquitin (Boston Biochem, USA)

(J) Surface Plasmon Resonance; Sensor Chip, Carboxymethylated Dextrans-CM5, GE Healthcare (Sweden)

(K) Instruments; Centrifuge (Beckman Coulter, USA), Shaker incubator (Infors HT, Switzerland), Surface plasmon resonance-BIACore T200 (GE Healthcare, Sweden), Circular Dichroism polarimeter (JASCO 815, USA), Fluorescence spectrophotometer (Horiba, USA), Dynamic light scattering (Wyatt DynaPro, CA, USA and Malvern, Worcestershire, United Kingdom), GelDoc, Nanodrop (GE), cell sonicator (Branson, CT, USA) and MALDI-TOF-TOF (Bruker Daltonics, Germany)

(L) Super Computational Grid "UTKARSH" BARC and DBT-BTIS facility- ACTREC were used for molecular dynamics simulation and analysis by utilizing GROMACS 4.5.5 package

2.2 Methods

2.2.1 Gene Cloning

Full-length cDNA gene construct cloned in different mammalian or prokaryotic systems were used as a template for cloning full-length or specific gene region. Polymerase chain reaction (PCR) was performed to amplify the desired region. Most of the forward primers for PCR reaction have TEV protease site between the gene and fusion TAG region and reverse primers were designed with the two stop codons. Following steps were performed for gene cloning experiments.

2.2.1.1 Primer Designing and expression vector

Primers were designed carefully by checking the presence of restriction enzyme (RE) sites of the DNA insert by NEB CUTTER [214]. Furthermore, RE sites were used as per the vector map provided for specific vector, ensuring that DNA insert will be in correct reading frame with correct orientation between promoter and terminator site. Complementarity and thermodynamics parameters were validated using Primer-X software (www.bioinformatics.org/primerx/). Primer Length; 18-25 (base pairs) with GC clamp, GC content; 40-60%, melting temperature (Tm); 50-65°C parameters were taken into consideration. Bacterial expression vectors were used such as pGEX-KT, pGEX4T-1, and pET28a+ for cloning and expression of respective proteins. For the bacterial two-hybrid system pUT18, pUT18C, pKT and pKNT vector backbone were used. Details of the vector constructs, primers, and bacterial strains were described in **Table 5.1, 5.2 & 6.1 in chapter 5 and 6**, respectively.

2.2.1.2 PCR amplification

PCR amplification of the desired region was performed using the following parameters*, however, some modifications were made in the protocol as per the DNA insert size, particularly in annealing temperature and extension time. PCR amplification of desired genes was performed using gene specific primers, size and quality of the amplicons were further evaluated using the 1% agarose gel by electrophoresis with specific DNA ladder.

The following protocols were used for the PCR amplification-

*

1. Denaturation: 95°C for 5 minutes
2. Annealing: 55-65°C for 45 seconds
3. Extension: 72°C for 0.5 kb/minutes
4. Repeat cycles from step 2 to 4 for 32 cycles
5. Final extension: 72°C for 10 minutes
6. Storage: 4°C

With the help of DNA markers, the desired amplicon of respective sizes was gel excised. The gel portion containing amplicon of interest was purified using commercially available gel extraction kit (Qiagen, Germany). The purified PCR product was subjected to restriction enzyme digestion. The digested gene of interest was further purified using Enzyme clean up- kit (Qiagen, Germany) to remove enzymes.

2.2.1.3 Restriction digestion

The PCR amplified gene of interests and expression vectors were digested with the same restriction enzymes (RE), as per the choice of selected restriction site in the PCR amplicons and desired vectors. The digestion protocol is described below.

1. 1-2 µg (20µl) of DNA amplicons (PCR amplified gene of interest) was taken in the microfuge tube (1.7 ml tube).
2. As per volume of the reaction mixture, 10X digestion buffer was added.
3. Restriction enzymes (RE); 1µl each (10-unit) were added to the reaction mixture.
4. Final reaction mixture volume was supplemented with DNAase and protease free double distilled (dd) water.
5. The RE mixed reaction mixture was kept at 37°C for 2-3 hour in Water bath.
6. RE digested PCR amplicons were purified using enzyme clean-up kit (Qiagen).

The digested PCR amplicons were then ligated into double digested vector using following protocol-

2.2.1.4 Ligation

Double digested gene sticky or blunt end insert and bacterial expression vector were ligated either using T4 DNA ligase or Quick ligase.

1. Double digested vector and insert DNA were taken in such a way that they are in 1:3 molar ratio (1 vector and 3 inserts).
2. Ligation buffer (2x) was added in the reaction mixture.
3. 1 µl of T4 DNA ligase or Quick ligase was added in the reaction mixture.

4. The ligation mixture was adjusted to the final volume using nuclease free double distilled water and further incubated at room temperature for 30 minutes.

   This ligated product was used for transformation in a bacterial host such as *E.coli* DH5α strain or *E.coli* NOVABLUE strain.

2.2.1.5 Transformation

Uptake of exogenous genetic material in prokaryotes from the external environment or by means of extracellular stimulus such as temperature is called transformation. In prokaryotic transformation, DNA fragment, a vector with or without foreign gene are taken up by the living cells under *in-vitro* condition. However, cells have to be made competent enough by chemical treatment to uptake the extracellular naked DNA inside the cells.

Competency is defined as biological nature of bacterial cells to uptake DNA easily than in the normal conditions. The treatment of bacterial cells with buffers supplemented with divalent cations such as Mg^{2+} or Ca^{2+}, that make bacterial cells of the desired competency. Under the appropriate condition, the transformation efficacy of most genes would be about 1 in every 10^3 cells [215].

2.2.1.6 Bacterial expression host cells

The *E. coli* BL21 (DE3) and *E. coli* Rosetta (2DE3) are commonly used bacterial strain for heterologous protein expression [216], which conveniently expresses foreign proteins using T7 promoter. The (DE3) stands for the host that is a lysogen of lambda-prophase (DE3) and carries T7 RNA polymerase gene under the control of a lacUV5 promoter, which can be induced by a non-hydrolysable analogue of lactose such as Isopropyl β-D-1-thiogalactopyranoside (IPTG) [217]. This strain is also devoid of expressing Lon and OmpT proteases which make them advantageous and an excellent host for foreign protein expression.

Since codon usage of eukaryotic and prokaryotic proteins are different, therefore the expression of eukaryotic proteins in the bacterial system often gets complicated. To overcome such problems, Rosetta 2(DE3) which is a derivative of BL21 is specifically designed to enhance the expression of foreign eukaryotic proteins having rare codons. The Rosetta strain has an extra set of tRNAs for the rare codons such as AGA, CUA, AUA, AGG, CCC, GGA and CGG on a compatible chloramphenicol-resistant plasmid; called pRARE. [218]. This strain offers a "universal" translation system, and their tRNA genes are under the control of their own promoters. These bacterial strains are the pLacI and pLysS derivatives and the rare tRNA genes are present on the plasmids that carry the T7 lysozyme and *lac* repressor genes.

2.2.1.7 Competent cells preparation protocol

1. The desired competent cells stored at -80 ˚C were streaked on Luria Broth agar plate, and allowed to grow over night at 37˚C.
2. The inoculum was prepared by inoculating single colony into 5 ml LB media and allowed it to grow at 37˚C for 12 hours.
3. The freshly prepared inoculum was added into 500 ml LB media with/without antibiotics as per the host competent cells, and incubated at 37˚C with constant shaking condition at 220 rpm, till the O.D at λ=600 nm reaches between 0.5-0.6.
4. The growth was synchronised by transferring the culture at low temperature (preferably on ice) for 5 minutes.
5. Bacterial cells were harvested in the collection Sorvall tubes by centrifuging at 5000 rpm for 15 minutes at 4˚C.

6. The supernatant was discarded and harvested cells were kept on ice immediately in a sterile condition.
7. The cell pellet was re-suspended in 100 ml of pre-chilled Transfer Buffer 1 (TB buffer 1) without any bubbles and cell clumps.
8. Re-suspended cells were transferred to pre-chilled autoclaved sorvall tubes and further centrifuged at 5000 rpm for 8 minutes.
9. The supernatant was discarded and centrifuged cell pellet was re-suspended in 10 ml of pre-chilled transfer buffer 2 (TB2).
10. Further aliquots of 100 µl in 1.7 ml microfuge tubes were made.
11. This aliquot was snap frozen in liquid nitrogen and stored at -80°C for further use.

Composition of Transfer Buffer 1 (TB 1): 300 mM Potassium acetate, 50 mM Magnesium chloride, 100 mM Potassium chloride, 10 mM Calcium chloride**,** 15% glycerol

Composition of Transfer Buffer 2 (TB 2): 10 mM MOPS, 10 mM Potassium chloride, 75 mM Calcium chloride, 15% glycerol

Prepared both the buffers TB1 and TB2 freshly with autoclaved ddH2O and stored at 4°C.

The competent cells were tested by streaking on LB agar plates with or without antibiotics. Result: with antibiotic; No colony, without antibiotic; colony grown

2.2.1.8 Protocol of DNA transformation in bacterial cell

1. The competent cells brought from -80°C were allowed to thaw on ice.
2. 50-100 ng/µl of cloned DNA was added to the tube containing 100 µl of competent cells.
3. The DNA with cells was incubated on ice for 45 minutes.

4. A brief heat shock was given for <2 minutes at 42°C and immediately brought back on to the ice.
5. Autoclaved LB medium of 800 µl was added in sterilized conditions.
6. Cells were incubated at 37°C for ~45 minutes.
7. The 100 ul of grown cells were spread the on LB agar plate containing the selective antibiotic with sterile L-shaped glass rod/ autoclaved plastic L-shaped rod.
8. This Petri plate was placed in inverted position at 37°C for overnight in the incubator to get well-isolated colonies.

2.3 Protein Expression and Purifications

The advancements in the genetic engineering and recombinant DNA technology have empowered us for over-expression of the protein of interest using a bacterial host system. Protein overexpression in a strong and controlled promoter can efficiently convert bacterial cells into a micro-protein production factory with high yield. The over expression of a non-bacterial protein may possess a threat to host system by intoxicating them. Therefore, the desired gene of interest must be in under very stringent and strong promoter control system. The expression vector generally available with strong inducible promoters that can sense the presence of inducer in the system e.g. Lac operon induced by addition of IPTG (a non-hydrolysable inducing agent and analogue of allolactose). When the inducer is absent in the growth medium, lac promoter gets repressed by the Lac repressor protein (LacI). Induction of the Lac promoter is mediated by the addition of either Lactose or its non-hydrolysable analogue isopropyl-β-D-thioglactopyranoside (IPTG) to the medium. These

substances prevent the Lac repressor from binding to the Lac operator, as a result enabling over-expression of the genes.

In *E. coli* BL21 (DE3), the lacUV5 promoter controls the expression of the T7 polymerase. BL21 (DE3) carries the LacI gene in its chromosome and applies a negative control over the expression of the T7 polymerase. Subsequently, the addition of inducer releases LacI inhibitory effect that turned on the expression of T7 polymerase. Moreover, the addition of IPTG to the culture induces the expression of T7 polymerase which further over-expresses the desired protein.

Requirements:

Sterile LB broth medium containing the desired antibiotic such as Ampicillin (100 mg/ml) or Kanamycin (50 mg/ml), 1M IPTG, micropipette with sterile tips, sterile 10 ml pipette, table top centrifuge etc.

Protocol:

1. **Inoculation:** A single isolated healthy colony from freshly transformed agar plate was selected to inoculate in 100 ml LB broth containing desired antibiotic as per expression vector. The pre-inoculum was incubated in shaker incubator at 220 rpm at 37°C overnight (12-16 hrs).
2. **Culture Dilution:** 10 ml of pre-inoculum was further diluted into the 1000 ml culture with (1:100 ratios) of autoclaved LB broth containing an appropriate concentration of antibiotic. Diluted culture was then allowed to grow in the shaker incubator at 37°C until all the cells reached mid-log phase i.e. A_{600} between 0.6-0.8.

3. **Protein Induction:** The cells growth was synchronized by cooling the culture. The culture was induced by adding an optimized concentration of IPTG and incubated on a shaker incubator at optimal temperature, at 22°C overnight.
4. **Cell Harvesting:** The induced culture was centrifuged for 15 minutes at 6000 rpm in a pre-cooled centrifuge.
5. **Cell Storage:** The cell pellet obtained after harvesting was stored at -80°C till further use.

2.3.1 Protein Purification

Protein purification is a chronological way to take out protein of interest from the cellular protein pool. Homogenous and highly purified protein is a prerequisite for structure determination, functional and interactions analysis such as protein-ligand, protein-DNA, and RNA. Extraction of soluble protein from bacterial extract is usually the most difficult part of protein purification. Protein purification steps depend on protein size, folding pattern, binding affinity to its respective tag etc. Chromatography is one of the common techniques used for separation and fractionation of desirable protein molecule from the mixture of complex cellular protein pool. Affinity chromatography is widely used for the protein purification because it purifies protein of interest in a high purity. Further, size exclusion generally performed to get an idea of oligomeric behaviour, and the homogenous population of proteins. The selection of stationary phase and mobile phase of the chromatographic system depends on the biophysical properties of the proteins. Proteins are generally purified using different chromatographic techniques such as affinity chromatography, size exclusion chromatography, ion exchange chromatography and hydrophobic interaction chromatography. Purified proteins are generally monitored using

UV detector present in Fast Protein Liquid Chromatography (FPLC) as they traverse through the column by their absorbance at λ=280 nm and at λ=215.

Protocol for Protein Purification

1. **Cell re-suspension**: The induced bacterial cells were re-suspended in the desired volume of buffer (Buffer A), complemented with protease inhibitor (200 µl of 200 mM PMSF and 20 µl of protease inhibitor cocktail).
2. **Sonication**: Resuspended cells were disrupted by using cell-sonicator (BRANSON Sonifier). This process involves the high-frequency sound waves that disrupt the bacterial cell wall. In sonication process, the sound energy of an ultrasonic bath or probe was used in a temperature controlled manner to disrupt the cell wall thereby exuding out of cellular contents. Sonication conditions were 70-pulse rate, 70 power with 1 minutes of duty cycle intermittently in an ice bucket.
3. **Fractionation by Centrifugation:** These disrupted cells were centrifuged at 18000 rpm for 45 minutes in 4°C to remove unwanted cell debris. The Clear supernatant was collected in a fresh tube for further purification.
4. **Equilibration of GST/Ni-NTA affinity resins**: Prior to binding, equilibration of GST/Ni-NTA affinity resins were performed with the desired buffer. Initially, five column volume of water was passed to remove traces of ethanol as the beads were stored in 20% ethanol. Further, equilibrated with 5 column volumes of buffer.
5. **Protein Binding**: The soluble protein was prudently transferred into the column containing affinity resin and allowed to pass by gravity flow. The flow through from the column was collected to check the protein fraction unbound to the resin.

6. **Column Washing**: To remove non-specific bound proteins from the affinity matrix, washing with ten column volumes of wash buffer were performed. To detect the binding efficiency, ~20 µl of protein bound on resin was taken and loaded onto the SDS-PAGE gel.

7. **TEV Protease Cleavage**: To remove the affinity tag like GST/His tag from the fusion protein, 200 µl (20 units) of 6XHis-TEV protease in wash buffer was treated with fusion protein for 3 hours. After cleavage, ~20 µl of resin was taken out to analyse on a SDS-PAGE gel to check the cleavage efficiency. The TEV protease is highly site-specific cysteine protease found in Tobacco Etch Virus (TEV). The recognition sequence for the TEV enzyme is Glu-Asn-Leu-Tyr-Phe-Gln-(Gly/Ser), and cleavage occurs between the Gln and Gly/Ser residues.

8. **Size exclusion chromatography:** 2 ml or 1 ml concentrated protein (as per the availability of FPLC loop size) was injected onto the AKTA-FPLC/BIORAD-FPLC pre-equilibrated with respective FPLC buffer. Size exclusion column as per the size of the protein (superdex-75 or superdex 200) was used. Eluted protein fractions were collected by monitoring the absorbance at λ=280 nm and λ=215 nm wavelength.

9. **Sample Loading on SDS-PAGE**: 12 µl of FPLC fractions were loaded onto desired percentage of SDS-PAGE gel, and stained with Coomassie brilliant blue dye. Furthermore, destained to visualize the protein bands under gel documentation system.

10. **Concentrate the protein**: The purified protein fractions of the expected size as observed on the SDS-PAGE gel were concentrated using desired Centricon filter (Millipore, USA).

2.3.2 Regeneration of GST resin

Composition of Wash Buffer: Buffer A, 5% glycerol, pH-8.0

Buffer A: 300 mM NaCl, 50 mM Tris (pH 8.5)

Buffer 1: 0.5 M NaCl, 0.1 M Tris (pH 8.5)

Buffer 2: 0.1 M Na-acetate, 0.5 M NaCl (pH 4.5)

Protocol:-

1. Freshly prepared 20 mM reduced glutathione in pre-chilled buffer 50 mM Tris pH 8.0, was passed twice through the column containing GST resin in order to remove the traces of bound GST and GST-fusion protein.
2. Five column volume of autoclaved double distilled water was passed.
3. Further, 1 column volume wash with washing buffer (300 mM NaCl, 50 mM Tris, 5% glycerol, pH-8.0) followed by 2 column volumes wash with Buffer A+1% Triton-X 100.
4. Passed two column volumes of 70% ethanol followed by washing with buffer-1 and buffer-2 alternatively.
5. Finally, passed single column volume of wash buffer followed by distilled water.
6. The resin was stored in 20% ethanol at 4^{o}C for further use.

2.3.3 Regeneration of Ni-NTA resin

Regeneration buffer: 6 M GuHCl, 0.2 M acetic acid

Protocol:

1. 750 mM imidazole prepared in Buffer A (pH 8.0) was passed, from the column containing Ni-NTA resin followed by two column washes with regeneration buffer.

2. The column was washed with double distilled water to remove the traces of GuHcl and further three column volumes of 2% SDS was passed subsequently.

3. Washing with a single column volume of 25%, 50%, 75%, and 100% ethanol. Further, repeated the washing steps in reverse order of ethanol concentration.

4. Double distilled water of one column volume was passed through the column, followed by five column washes with 100 mM EDTA (pH-8.0) to strip the chelated Ni^{2+} ions.

5. Two column volumes of 100 mM $NiSO_4$ were passed to re-charge Ni-NTA resins followed by three column volumes of water to remove unbound Ni^{2+}.

6. The resin was stored in 20% ethanol at 4^oC for further use.

2.4 Protein Estimation

It is important to estimate the accurate protein concentration. Using protein sequences, theoretical extinction coefficient was determined. Bradford (Expedon) assay and Nanodrop (Thermo Scientific, USA) was used to measure the concentration of the purified proteins.

1. A gradient concentration of BSA from 0.1, 0.5, 1.0, 2.5, 5, 10 mg/ml were prepared as a standard reference and the absorbance was recorded at λ=595 nm using split beam UV spectrophotometer (Shimadzu, Japan).

2. The concentration of the protein was obtained with the absorbance at λ=280 nm using Nanodrop spectrophotometer (Thermo Scientific, USA).

2.4.1 Mass Spectrometry

The mass spectrometry is an important tool to study proteomics and identify protein/DNA sequences, mass, post-translational modifications, and other biophysical nature of the

proteins [219]. The sequence of the short stretches of polypeptides can be determined using a technique called tandem *MS or MS/MS,* also known as peptide mass fingerprinting. Mass spectrometry technique requires only miniscule amount (~1-10 ng) of protein for study. Mass/Charge ratio of the analyte can be used to determine the mass (M) of the analyte with very high resolution and accuracy almost in attomole range [219].

A solution containing the protein to be identified was digested with specific protease or chemical reagent to get the mixture of oligopeptides. The specific peptide mixture is then subjected into the device for peptide mass fingerprinting [219].

Types of Mass Spectrometry: Mass spectrometry of following types-

1. Electron Impact (EI): This is useful for a low mass range of molecules up to 1000 Daltons.
2. Chemical Ionization (CI): It is a method for the generation of the molecular ion peak $[M+H]^+$ for molecules of up to 1000 Daltons.
3. Electrospray Ionization (ESI): It is a powerful method for the production of manifold charged species and can be engaged for the molecules of up to 200,000 Daltons.
4. Matrix Assisted Laser Desorption (MALDI): It is a good method for the generation of multiple charged species and can be used for the identification and characterization of the molecules of up to 500,000 Daltons [219].

Mass spectrometry sample preparation protocol

1. Cut down the protein band from the SDS-PAGE gel into small pieces and incubate into fresh destainer by keeping in vortex conditions overnight to remove the stain.
2. Repeated washing was provided (3 changes for 10 minutes each) to the gel pieces with ddH2O to remove any staining content from the gel pieces.

3. The excised gel pieces were further treated with 50 mM NH_4HCO_3 + 50% acetonitrile (1:1) cocktail to provide buffering and de-staining. It was kept for 15 minutes at 25˚C.
4. The excess liquid was discarded and the gel pieces were submerged with sufficient amount of 100% acetonitrile.
5. Excess acetonitrile was removed, and 50 mM NH_4HCO_3 was added to rehydrate the gel pieces.
6. An equal volume of 100% acetonitrile was added subsequently for 5 minutes to shrink the gel pieces.
7. The gel particles were dried out in vacuum spin-dry (for 10-25 minutes).

Reduction and Alkylation

1. The gel pieces were swelled in freshly prepared buffer 10 mM DTT, 50 mM NH_4HCO_3 (300 µl/tube) for reduction reaction.
2. It was incubated at 56˚C for 45-55 minutes. The tube was brought to room temperature once the incubation period was over.
3. The additional fluid was removed and substituted quickly the identical volume of freshly prepared 55 mM iodoacetamide (IAA) in 50 mM NH_4HCO_3 for alkylation.
4. It was incubated for 30 minutes more in dark environment.
5. IAA solution was removed and washed the gel particles with the mixture of 50mM NH_4HCO_3 and 50% acetonitrile (1:1).
6. Excess of 100% acetonitrile (about 200 µl) was added to cover the gel pieces.
7. Excess acetonitrile was removed, once the gel pieces were shrunk.
8. Gel particles were dried out in speed vacuum (5-10 minutes).

In -Gel Digestion

1. The required amount of freshly prepared enzyme solution (Trypsin + NH_4HCO_3) was added to the gel particles. The working concentration of trypsin enzyme was in the ratio of 1:100 or 1:10.
2. The tubes containing gel pieces were incubated at 37°C overnight in an upright position (up to 16 hours).

Extraction of Peptides

Composition of peptide extraction buffer: 100% acetonitrile + 5% TFA

1. 30 µl of extraction buffer was added three times after vortexing intermittently.
2. Centrifuge the extract from the sample for 1-2 minutes at 1000 rpm.
3. The sample (45-50 minutes) was dried using vacuum drier.
4. The vacuum dried samples were stored at -20°C till further use.

2.4.2 Circular Dichroism

Circular Dichroism (CD) employs an optical phenomenon, which arises due to the differential absorption of left and right circularly polarized light by chiral molecules [220, 221]. The only asymmetrical molecule can display the differential absorption activity when exposed to circularly polarized light. Biological molecules like amino acids which are chiral in nature except glycine can display differential absorption of right circular polarized light (R-CPL) and left circularly polarized light (L-CPL). In addition, proteins having a different composition of a secondary structural element like α- helices, β-strand or random coil has a unique signature peak in CD spectrum [220, 221]. The far-UV region ranges

from λ=250-190 nm, CD spectroscopy of proteins reveals secondary structural elements and their quantification of the protein. However, near-UV region ranges from λ=350-250nm. CD spectroscopy also predicts tertiary structure of the protein with specific peaks. Proteins showing the characteristic absorption minima at λ=208 nm and λ=222 nm indicate predominant α-helical structure, whereas a minimum at λ=218 nm is a characteristic content of β-strands or sheet [220, 221]. The CD signal is often affected by the composition of the buffer in which the protein is present. Therefore, it is important that spectra of buffer components may also require additional monitoring at lower wavelengths. For example, CD spectrum at λ= 208, 218, 222 and 228 nm wavelength allows the acquisition of higher signal to noise ratio for the secondary structural changes at different urea concentration. The higher concentrations of denaturant may incorporate noise during data acquisition with an increase of HT voltage.

CD Polarimeter; Jasco J-815 spectropolarimeter, inbuilt with a Peltier temperature controller.

Protocol:-

1. CD cuvette was washed with 0.1% t-Octylphenoxypolyethoxyethanol (Triton-X) and subsequently with 100% ethanol and air-dried.
2. 5-10 µM of protein solution in the buffer 10 mM HEPES/10 mM Tris, pH 7.5, 50 mM NaCl was scanned using wavelength ranges from of λ =200-250 nm at 20°C.
3. The mean residual ellipticity was calculated using the protein concentration and molecular mass.

2.4.3 Fluorescence Spectroscopy

Each biological molecule has a lower energy state and attains higher energy level when excited from the external source. Electronic level comprises of closely spaced energy bands, which exist at different vibrational modes. Each electronic state is divided into multiple sublevels demonstrating the vibrational modes of the molecule. The energies of the vibrational levels are further separated by 100 cm^{-1}. Fluorescence intensity measurement allows the determination of the intrinsic fluorophores and their absorptions in the local environment. The fluorescence emission spectrum of a well-folded protein is a combination of the fluorescence emission of distinct aromatic residues. Aromatic amino acids like phenylalanine (Phe), tyrosine (Tyr) and tryptophan (Trp) have intrinsic fluorescence properties due to the presence of conjugated aromatic ring system. Emission from Tyrosine and Tryptophan are generally used because their quantum yields are very high and provide a fluorescence signal. The major involvement in intrinsic fluorescence emissions of a folded protein is due to tryptophan residue, with compliment from tyrosine and phenylalanine. Disulphide bonds also have considerable absorption in the same wavelength range. Emission spectra were collected from λ=310 to 450 nm wavelength, following excitation at λ= 280 nm or at λ= 295 nm wavelength. Temperature and chemical denaturants were used to monitor the folding pattern. In both, we collect emission spectrum for the native and unfolded protein using a single wavelength to observe the structural changes.

Spectrofluorimeter: JOBIN YVON Horiba Fluorolog 3

Protocol

1. The tryptophan, tyrosine and phenylalanine microenvironments were observed using Fluorolog 3 (Horiba, USA) at an excitation wavelength of λ=280 nm or λ=295 nm at the temperature range of 10°C to 60°C.

2. Fluorescence emission scans were recorded from λ=310-450 nm wavelength.

There are two widely used methods to monitor the protein folding and melting temperature of the desired protein.

1) Thermal Denaturation

A protein can attain different conformations in different time points, however each conformation is closely related to the initial one. Intramolecular interactions like hydrogen bonding, ionic interactions, van der Waal interactions, and hydrophobic interactions play a very important role in protein folding and dynamics. The kinetics and thermodynamic parameters of protein folding can be explored by utilizing intrinsic probes in fluorescence and CD spectroscopy.

Proteins found in complexes generally have an internal features such as exposed hydrophobic patch, which leads to its aggregation or precipitation. For thermal denaturation, CD spectra of ~10 µM protein at λ=200-250 nm wavelengths in a temperature dependent manner from the range of 10-80°C have been collected. Furthermore, fraction unfolded (*unF*) were calculated at each temperature interval and plotted to determine the melting temperature and folding pattern transition.

In fluorescence spectroscopy tyrosine, phenylalanine and tryptophan microenvironment can be monitored at an excitation wavelength of λ=280 nm and λ=295 at 10°C, respectively. Emission spectra were collected from λ=310-450 nm wavelength. 10-µM *wild-type* and mutant proteins were denatured in the different temperature range from 10-80°C or 20-60°C with the emission spectra from λ=310-450 nm wavelength.

Data analysis and curve fitting

Data analysis and curve fitting of CD and Fluorescence spectra were performed using Origin software (Origin Lab, Northampton, MA) for the two state and three state transition model. Thermodynamic parameters were calculated considering native (N) to intermediate (I) and intermediate to unfolded (U) transition [222].

$$N \overset{K_1}{\longleftrightarrow} I \overset{K_2}{\longleftrightarrow} U \quad (1)$$

Here, K_1 and K_2 represent the equilibrium constants for the two reactions of intermediate fractions and unfolded fractions, respectively.

2.4.4 Time-dependent Proteolysis

Time-dependent proteolysis or limited proteolysis was performed to analyze the structural features of proteins. Specific proteolytic enzyme probes such as trypsin and chymotrypsin were used to locate the sites of a compact and disordered polypeptide chain. Furthermore, to get a clear picture of domain boundary in high-resolution mass spectrometry or protein sequencing has been performed. The target proteins would exist in the folded, partially folded or partially unfolded state in different conditions, such as low/high temperature, pH, and ionic strength. It can also be useful to design the stable protein fragments that can fold independently.

Protocol for limited proteolysis of FANCI ARM Repeat

1. Equal concentration of FANCI ARM Repeat (1 mg/ml) was incubated with trypsin independently in two different set of experiment, keeping the final concentration of trypsin 10 ρg/μl.

2. The reaction mixture was incubated for the different period 1, 5, 10, 15, 30, 60 and 180 minutes at 37°C with trypsin enzyme.
3. The reaction was terminated by adding 1mM PMSF protease inhibitor.
4. Samples were mixed by adding an equal volume of laemmli buffer and analyzed on the SDS-PAGE gel.
5. In conclusion, the band corresponding to 14 kDa was considered as a stable fragment.
6. The different domain of interest was identified by peptide mass fingerprinting, mass spectrometry and Mascot analysis using Biotool software (Bruker Daltonics, USA) at a resolution of attomole range [223].

2.4.5 Molecular Modeling

Characterization of protein structure and its associated function is an important area in the structure biology. Protein structure unravels the functional domains present in the molecule at the atomic level resolution. X-ray crystallography, Cryo-electron microscopy, and NMR spectroscopy are the important experimental methods to determine protein structure [224-226]. Structure determination of protein using X-ray crystallography or NMR spectroscopy requires a large amount of pure and homogenous protein, that the limitation of these techniques. However, recent advances in computational and bioinformatics resources has empowered us to model the protein structure. Currently, *ab initio* prediction and homology modeling are commonly used methods to model the three-dimensional structure of protein [224-226].

Ab initio or *de novo* prediction is established on the fundamental physical and chemical properties of atoms and molecules present in amino acid residues. However, *Ab-initio* or

de novo method is limited by computational power and accuracy. Nevertheless, homology modeling of protein structure depends mainly on available structures of a similar sequence in the Protein Data Bank (PDB) or in other structural databases. Homology modeling is based on the principle of similar amino acid sequences or the amino acids having similar properties which are evolutionarily interrelated and fold into analogous three-dimensional structure.

Two widely used approach for protein modeling are described here:-

1. Homology Modeling

Knowledge-based structural datasets combine the data to information about the structure of polypeptides in a sequential manner, such as homology modeling [227, 228]. When the amino acids sequence similarity is high with homologous protein and stereochemical structure is also known, the model building of an unknown structure by sequence comparison can be carried out with reasonable success [229, 230]. Upon the examination of related proteins that have different amino acid sequences but similar tertiary fold indicates conserved folding pattern [230, 231]. For desired protein sequences, the homology modeling procedure requires investigation of homologous sequences with crystal structures, and alignment of the query sequences to the target template structure.

The homology modeling generally comprises of four steps:

1. Begin with the available PDB structure of known sequences.

2. Assemble fragments with the substructures or missing loops if present, from known homologous structures.

3. Carry out the investigation about the limited structural changes from a known adjoining protein structure.

4. Optimize the structure geometry by energy minimization.

Homology modeling relies on the three-dimensional structures between the structure of the protein available with similar sequences [230]. Therefore, the precise identification of homologous sequences is critical to generating the correct structural models. Usually, identification of closed homolog occurs through the comparison of protein sequence profiles with Hidden Markov Models (HMMs). Additional, use of innovative methods of sequence assessment, such as sequences to profiles alignment position-specific iterative-basic local alignment search tool (PSI-BLAST), profiles to profiles (FFAS), or HMMs to HMMs, helps significantly in homolog identification [225, 226].

Amino acid sequence searches are performed from pairwise sequence alignment using BLAST [232]. The reliability of BLAST and PSI- BLAST is often compromised when the output is in the “twilight zone” (<30% identity with query sequence). Further, the profile based identification method allows the identification of distantly related proteins. It has been reported in the SCOP database, PSI-BLAST validates with twice the accuracy of BLAST, whereas FFAS search enhances the additional accuracy by 20% in PSI-BLAST. These augmentations in the search specificity ultimately led to the better alignment of the query sequence with the experimental structures [225, 226].

Once a set of homologues structure is carefully chosen to use as templates, alignment of the query sequences to the template structures can often be followed by adjustments to optimize placement of insertion and deletions within the range of secondary structure elements. In addition, building the large loops with low structural homology or no homology can be performed by searching structural libraries [233]. Loops longer than six

or seven residues increases challenges because of a large number of possible random conformations [234]. Modern homology modeling techniques have achieved the structure predictions in an accuracy that it can even be successfully used in site-directed mutagenesis, virtual screening, and in drug designing applications.

2. *Ab initio* Prediction of Protein Structure

Energy-based calculations of the theoretical models based on amino acids composition, energy parameters and energy minimization are involved in *ab initio* prediction [235]. The prediction of the three-dimensional structure from the sequences of amino acids in the absence of template structure is a challenging task [235]. An alternate option is energy based structure prediction, which generally relies on energy minimization and molecular dynamics. It is challenging because it generates a large number of possible multiple structures that have very close energy minima. For example, a protein composed of 100 amino acids can have $(2x5)^{100}$ different main chain conformations when only two torsion angles per residue are presumed to have five probable values per torsion angle. Furthermore, the assumption and calculation of other features of proteins like constant bond lengths, bond angles, and torsion angles for the side chains, which is really tedious to evaluate and predict all the possible conformations in order to find the global minima. Several constraints and approximations have to be introduced to simplify the problem associated with protein structure prediction. The total molecular energy E can be expressed as under the assumption of constant bond lengths, and torsion bond angles are as follows:

$$\mathbf{E= E}_{tor} \mathbf{+E}_{elec} \mathbf{+ E}_{vdw} \mathbf{+E}_{pse} \quad \textbf{(1)}$$

Where E_{tor}, $E_{elec,}$ E_{vdw}, and E_{pse} are torsion angle potential, electrostatic potential, van der Waals and pseudo entropic interaction energies, respectively.

Moreover, it is possible that an incorrectly predicted three-dimensional structure may have fewer stabilizing hydrogen bonds, electrostatic, and van der Waals interactions, larger solvent accessible surface and a greater fraction of hydrophobic side-chain atoms exposed to the solvent [236]. Classically, force fields that provide *ab initio* modeling conduct a conformational expedition under the direction of a pre-designed energy function. This method produces a number of possible conformations also known as structure decoys. Hence, successful *ab-initio* modeling depends on the following three factors:

(1) A precise energy function in which the native structure of a protein resembles the utmost thermodynamically stable state, related to decoy structures.

(2) An effective search method, which can precisely identify the low-energy states from different conformational search model.

(3) An assortment of native-like protein models from a pool of decoy structures.

One of the classical known methods for *ab-initio* modeling is pioneered by Bowie and Eisenberg [237], who generated protein models by collecting small fragments (mainly 9-mers) taken from the PDB library [238]. Based on a related idea, Baker and coworkers have designed ROSETTA software, which is immensely successful for the free energy modelling (FM) [237, 239]. In the recent extensions of ROSETTA, models were generated in a reduced form with conformations specified heavy backbone and Cβ atoms. Furthermore, the low-resolution decoys were subjected to all-atom refinement including van der Waals interactions, pairwise solvation free energy, and an orientation dependent hydrogen-bonding potential [240, 241]. However, analysis of conformational searches, multiple rounds of minimization through Monte Carlo approach also needed attention [242]. The success of the ROSETTA algorithm, as well as the restricted accessibility of its energy functions to other groups, different investigators started developing new energy

functions using the idea of ROSETTA, like the derivatives of ROSETTA are Simfold and Profesy [243, 244].

2.4.6 *In-silico* Docking

Protein-protein interactions are the basis of cellular signaling, communication, and functions. The interacting complex molecules generally adopt energetically and entropically best binding interface. In addition, determining the structures of the protein complexes by experimental methods X-ray and NMR solution is still challenging. Therefore, *in-silico* protein-protein docking can help to illustrate the binding interface. It is usually achieved by the following three steps:

(**i**) **Sampling**: A large number of randomized conformations of the three-dimensional protein structures to be docked is generated.

(**ii**) **Optimizing the complexes**: The docked complexes are optimized with respect to different energy terms, stereochemistry or other additional information by performing translational, rotational, and conformational parameters.

(**iii**) **Scoring:** The output structures are categorized to find the best solutions close to the native complex structure. This is attained by calculating appropriate scores that measure the quality of the complexes.

Generally, in molecular docking, one ligand molecule is brought into close proximity of target protein while investigating the energies of the interactions in different possible mutual orientations. In the complex structure, the ligand and protein molecules are bound in such a way that they adopt the energetically favorable conformation. There are majorly two kind of docking. The first treats a whole target molecule as a beginning point and

involves a search algorithm to inspect the energy contour for the ligand at the binding site. Furthermore, it observes optimal reasonable solutions for a specific scoring function. The search algorithms comprise geometrical complementary fit, simulated annealing, genetic algorithms and molecular simulation dynamics.

The second methodology of docking is based on placing one or several substructures/geometrical conformers of a ligand into the binding pocket of target molecule and then further orienting the rest of the molecule. In this method of docking, an initial knowledge of the binding site is usually required which helps in placing the ligand onto the binding site.

Beside this, knowledge of geometric restraints such as distances, angles, and dihedrals between bonded or non-bonded atoms may simplify the docking process. In the interatomic distance, the bond length can be restrained to remain constant during a simulation for representative complex structure. Such constraints also promise the confinement of the ligand molecule at the juxtaposition of the binding site of the receptor during energy minimization process. In an automated docking, the ligand is allowed to fit into the possible binding cleft of the target molecule. HADDOCK (High Ambiguity Driven biomolecular DOCKing) is established by EMBL which takes the advantage of experimental biochemical and biophysical data [17, 245]. HADDOCK uses an approach of Ambiguous Interaction Restraints (AIRs) to carry out the docking process. An AIR can be defined as an ambiguous distance between all residues shown to participate in the binding. Furthermore, HADDOCK also uses more quantitative restraints between pairs of nuclei from Nuclear Overhauser Effect (NOEs) and residual dipolar couplings (RDCs) [17, 245]. HADDOCK deal with a large variety of data and complexes in addition to protein-protein and protein-DNA docking. Docking by HADDOCK is performed by transmuting the data into ambiguous interaction restraints (AIRs). This describes a grid of between residues

likely to be involved in the binding mode without forcing it to any specific position or orientation of the components. HADDOCK also controls other NMR sources of knowledge such as residual dipolar couplings (RDCs), diffusion relaxation, and pseudo-contact shifts. Furthermore, experimental techniques such as small angle X-ray scattering (SAXS) and cross-linking data from mass spectrometry can be used for scoring and generating models of docked interfaces [17, 246-248].

Docking protocol of the HADDOCK

Docking using HADDOCK can be completed in three major successive stages:

1. **Randomization of orientations and rigid body energy minimization**

The receptor and ligand are positioned at a geometrical separation of 150 Å from each other in space and are allowed to randomly rotate around their center of mass. After the translation and rotation, ligand and receptor are docked by rigid body energy minimization. The best decoys are selected based on the energy minima.

2. Annealing

Two proteins present as rigid bodies with optimized orientations. However, their side chains at the interfaces are allowed to move in. Furthermore, chains and backbone at the interfaces are permitted to move in within the conformational rearrangements. The best top scoring models 200 out of 1000 from the 1st stage are selected using a simulated annealing (SA) method in torsion angle space procedure.

3. Final refinement in Cartesian space with explicit solvent

The simulated annealing models are subjected to a molecular dynamics simulation in an explicit solvent shell. In the final phase, they are solvated in 8Å shell of TIP3P water molecules [249, 250]. In the TIP3P model, each atoms are assigned a point charge and the

oxygen atom gets the Lennard-Jones parameters [251]. A large number of energetically improved structures poses the problem to select the better model. However, comparative analysis of micromolecular properties, such as steric surface complementarity, electrostatic interactions, hydrogen bonding, knowledge based pair-potentials, desolvation energies and van der Waals interactions supports to find better-scored model [247, 252-260].

2.4.7 Molecular Dynamics Simulation (MD simulation)

Computer-based MD simulations act as a connection between microscopic length and time scales to develop the prospect of complication to split 'solvable' from 'unsolvable'. MD simulation determines the motion of atoms in a multi-dimensional potential energy landscape which is resultant of the one electronic state of classical mechanics or quantum mechanics. MD simulation imposes Newtonian mechanics on a particle system to solve a set of interactions in the time evolution. MD simulation comprises of the mathematical and conventional calculations of motion.

$$\mathbf{F}_i = m_i \frac{\mathrm{d}^2 \mathbf{r}_i(t)}{\mathrm{d}t^2} \qquad m_i \ddot{\boldsymbol{r}}_i = \boldsymbol{f}_i \qquad \boldsymbol{f}_i = -\frac{\partial}{\partial \boldsymbol{r}_i} \mathcal{U} \qquad \textbf{(1)}$$

$$\mathbf{r}_i(t + \Delta t) \cong 2\mathbf{r}_i(t) - \mathbf{r}_i(t - \Delta t) + \frac{\mathbf{F}_i(t)}{m_i} \Delta t^2$$

In order to calculate the forces f_i acting on the atoms, originated from a potential energy U (r^i), where r_i = (r1; r2; ….. r_i) represents the entire set of 3N atomic coordinates. The position and velocity vector describes the time evolution of the system in phase space, which is defined in the MD trajectory. The objective of the numerical integration of

Newton's equations of motion is to reveal an expression that describes positions r_i ($t_1\Delta t$) at time $t_1\Delta t$ in terms of the previously known positions at time t. Verlet algorithm is generally used in MD simulations because of its simplicity and consistency [261].

The physical and chemical properties in a functional form of a system are defined by empirical potentials, defined by the atomic force field. The permissible restraints are chosen such that the empirical potential signifies a good fit to the appropriate sections of the *ab-initio* Born–Oppenheimer surface [262]. The atomic force field model defines physical systems as collections of atoms packed together by inter and intra-atomic forces.

A characteristic force field, used in the simulations of biological molecules, where the potential energy *V* is calculated from bonded and non-bonded interactions can be written as follows-

$$V = E_{bonded} + E_{non-bonded} = \\ = \sum_{bonds} k_b(r - r_0)^2 + \sum_{angles} k_\theta(\theta - \theta_0)^2 + \sum_{torsions} \frac{V_n}{2}[1 + \cos(n\tau - \phi)] \\ + \sum_i \sum_{j>i} 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}^{12}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}^{12}}\right)^{6}\right] + \sum_i \sum_j \frac{q_i q_j}{4\pi\varepsilon_0\varepsilon_{rel} r_{ij}} \quad \textbf{(4)}$$

where $r_{ij} = r_i - r_j$, kb is the bond stretching constant, r_0 equilibrium bond distance, K_θ is the bond angle constant, θ_0 is the equilibrium bond angle, τ is the torsion angle, ϕ is the phase angle, V_n is the torsional barrier, σ is the van der Waals diameter, q is the charge of each atom, and ε is dielectric constant. The last two non-bonded are Lennard-Jones potential and Coulomb interaction. Biological macromolecules force fields are generally implicated with long-range electrostatic and dispersive interactions.

An aqueous medium is an illustrative environment for biological macromolecules and has to be implicated in the realistic simulations. The treatment of long-range forces is

associated with the choice of boundary conditions applied to a particle system to deal with its size and surface effects. The two common techniques are developed on first periodic boundary conditions and on the Ewald method for lattice summations. Particle mesh Ewald method or fast multiple methods allow efficient computation of the long-range interactions and do not account for basic cut-off approximation, in which contributions of sites separated by a distance larger than a certain cut-off are not considered.

Docking Protocol for FANCI ARM Repeat, FANCD2 CUE Domain, and BRCA2 CTR

1. Protein structure of FANCI ARM Repeat and FANCD2 CUE Domain was modeled using Robetta server (http://robetta.bakerlab.org/), BRCA CTR was modeled using FALCON server (http://protein.ict.ac.cn/FALCON/).
2. The good quality model was validated and selected based on overall stereochemistry, Ramachandran plot, "**SAVES**" (Metaserver for analyzing and validating protein structures, (http://nihserver.mbi.ucla.edu/SAVES/) and SWISSMODEL Workspace (https://swissmodel.expasy.org/).
3. SAVES and SWISSMODEL Workspace (http://swissmodel.expasy.org/workspace/) comprises programs like, Procheck. What_check, Errat, Verify_3D, Anolea, DFire, QMEAN, Gromos, DSSP, Promotif and Prove. Modeled structure of FANCI ARM Repeat was simulated for 80 ns and FANCD2 CUE Domain for 50 ns using GROMACS.
4. Geometrical docking was carried out using HADDOCK server (http://milou.science.uu.nl/services/HADDOCK2.2/).

5. PDBsum was performed to analyze the interactions of the docked cluster.
6. DALI server (http://ekhidna.biocenter.helsinki.fi/dali_server/start) is an important tool that predicts the best structure homologs of protein. It has been reported that structurally similar proteins more likely play a similar role in a particular biological system.

Protocol for FANCD2 CUE Domain

Validated protein structure model was used for molecular dynamics simulation.

1. Mutants were generated from available structure in protein data bank or independently modeled by Robetta server [263].
2. Protein was simulated for 50 ns using GROMACS 4.5.5 [14, 264-266].
3. Protein-protein docking was carried out using HADDOCK server [246, 247].
4. PCA analysis and clustering were performed on the *wild-type* and mutant trajectory using R 3.2 and Prody software [16, 267].
5. Interaction analysis was performed using Ligplot (http://www.ebi.ac.uk/thornton-srv/software/LIGPLOT/) to observe changes in the molecular environment [268].

2.4.8 Principal Component Analysis (PCA)

Inside the cell, proteins are dynamic in nature. Protein dynamics have a crucial role in understanding mode of functioning at the molecular level. Molecular dynamics simulations of the biological macromolecules can mimic the cellular conditions, but exact meaningful information particularly biological relevant motion from large diverse data set can be derived from PCA [269, 270]. A characteristic MD trajectory comprises the datasets of time-evolution of the coordinates in the system being investigated. Commonly, the MD

time steps are of 1 femtosecond while the simulation time fluctuate from a few to tens of nanoseconds. So, a single resultant trajectory after MD simulation can generate a huge amount of data. For the *N*-atom system, the input dataset for PCA can be built as a trajectory matrix on that each column comprises a Cartesian coordinate for a given atom at output time step (***x** (t)*). Before performing the PCA, it is necessary to remove any net translational and rotational motion of the system by aligning the coordinate data to an input structure (reference structure) to get the appropriate trajectory matrix (*X*). The standardized trajectory data is then used to generate a covariance matrix (*C*) and elements of the matrix can be defined as below

$$C_{ij}=<(X_i-<X_i>)(X_j-<X_j>) \quad \textbf{(1)}$$

Where <···> signifies an average done over all the time steps of the trajectory. The following step comprises of diagonalization of the 3*N*x3*N* covariance matrix that can be accomplished by eigenvector decomposition method

$$C=T\Lambda T^T \quad \textbf{(2)}$$

Where T is a matrix of column eigenvectors and Λ is a diagonal matrix comprising the corresponding eigenvalues. The equation 2 leads to the transformation of the actual trajectory matrix in a new orthonormal basis set composed of the eigenvectors. The eigenvalues are suggestive of the mean squared displacements of atoms along with the corresponding eigenvector. There will be 3*N* eigenvalues if the number of arrangements (*M*) is more than 3*N*. But if $M<3N$ the number of eigenvalues will decrease to *M*.

The simplest mode of visualizing these results involves categorization of eigenvectors in a descending order with respect to their eigenvalues. The plot of eigenvalues against the respective eigenvector index can then be plotted which is also known as 'scree plot' [271]. Usually, a scree plot can indicate that only first few eigenvectors possess large eigenvalues

and increasing eigenvector shows lower eigenvalues. As a result, most of the large biological atomic motion in the data is illustrated by first few modes. It is often imperative to hypothesize that the motions along with these 'essential eigenmodes' explained the most significant global information which shows predominantly the dynamics of the macromolecular systems. Visualization of the components of the individual eigenvector is helpful to estimate the landscape of the eigenmodes and atomic fluctuation. After the identification of a subset of essential eigenmodes, further, the analysis of specifying each mode can be assumed by projecting the native trajectory along with a specified eigenvector. The respective projection matrix (P) can be acquired as

$$P=XT$$

The projection matrix yields excitation amplitude of a given eigenvector in a time evolution. The column vectors in P ($\mathbf{p}(t)$) are termed as the 'principal components'. To explore the motion along any specified eigenvector, the column vector from P multiplied by the respective eigenvector T yields a compact trajectory. The first 20 modes which occupy high eigenvalue position thus encompasses the essential information of the macromolecular dynamics.

2.4.9 Normal Mode Analysis (NMA)

The main goal in the protein dynamics is to explain slow large-amplitude motions. These motions, characteristically indicate conformational fluctuations crucial for the protein functions. Normal mode analysis (NMA) of the protein provide valuable information concerning to the mechanism of slow large-amplitude motions in multi-domain protein assembly in a time-independent manner [272, 273]. Normal mode analysis is a fast technique to compute the vibrational modes and protein flexibility.

In NMA the modeled atoms are considered as point masses associated with springs that specify the interatomic force fields. Each interatomic bonds in NMA are considered as springs connecting each atom (known as a node) to all other neighboring nodes. However, only the inter-connected atom within a limited distance is accounted. NMA in Cartesian coordinate space usually involves three calculation steps:

(1) Conformational potential energy; minimization of the molecule as a function of the atomic Cartesian coordinates;

(2) "Hessian" matrix calculation; the matrix of second derivatives of the potential energy with respect to the mass-weighted atomic coordinates.

(3) The diagonalization of the Hessian matrix; the final step that produces eigenvalues and eigenvectors known as normal modes.

These three steps require high-end computational power. Furthermore, the energy minimization and diagonalization are computationally challenging, and it comprises the diagonalization of a $3N \times 3N$ matrix, where N is the number of atoms in the particle system. It is generally considered that at a potential energy minimum, the potential energy function V can be prolonged in a Taylor series [274] in terms of the mass-weighted coordinates [275]. This can be written for Cartesian coordinate (q_i) as $qi=\sqrt{mi}\Delta x_i$, where Δx_i is the displacement of the $_i$th coordinate from the energy minimum and m_i is that mass of the respective atom. If the extension is completed at the quadratic level, then the linear term is zero at an energy minimum; the potential energy function for the particle system can be written as follows-

$$V = \frac{1}{2} \sum_{i,j=1}^{3N} \left. \frac{\partial^2 V}{\partial q_i \partial q_j} \right|_0 q_i q_j \quad \textbf{(1)}$$

The second derivatives in Eq. 1 can be transcribed in a matrix "Hessian," F. Eigenvalues and eigenvectors (equivalent to diagonalization) can be determined as follows-

$$Fwj = \omega^2 j w j \quad \textbf{(2)}$$

Where wj is the jth eigenvector and $\omega^2 j$ will be jth eigenvalue. There are 3N such eigenvector equations so each eigenvector classifies a normal mode coordinate using following equation-

$$Q_j = \sum_{i=1}^{3N} w_{ij}\, q_i . \quad \textbf{(3)}$$

Further, the calculation of low-frequency normal modes and the directions of large-amplitude fluctuations in molecular dynamics simulations designate a close similarity [270, 276]. The close directional coincidence of the first few normal modes and the first principal component acquired from molecular dynamics simulation can be calculated using equation (3) [276].

2.5.0 Protein Oligomerization

Protein oligomerization can be distinguished and characterized by different analytical techniques. These techniques are analytical ultracentrifugation, native PAGE, chemical crosslinking assays, size exclusion chromatography, dynamic and static light scattering

techniques, small angle X-ray scattering, NMR spectroscopy, electron microscopy and mass spectrometry.

2.5.1 Glutaraldehyde Cross-Linking

Protein-Protein Interactions can be either transient or stable, depend on their role in different biological functions [277]. The chemical cross-linking assay helps to find both transient and stable interactions involved between two proteins [278]. In chemical cross-linking, covalent bonds form between different bi-functional reagents such as glutaraldehyde, that contains reactive aldehyde group. The reactive aldehyde group reacts with functional groups present in the protein side chains, such as primary $–NH_2$ and sulfhydryls of amino acid residues. Moreover, the cross-linking experiment can also explore the interacting interface areas.

Protocol for FANCI ARM Repeat glutaraldehyde cross-linking

1. FPLC purified proteins at the concentration of 1 mg/ml were prepared.
2. Reaction mixtures with 15 μg of protein in 10 mM HEPES buffer (pH 7.5) were incubated with a freshly prepared glutaraldehyde (final concentration 0.1%) at 37°C in a time-dependent manner [279].
3. This reaction was terminated by adding 5 μl of 1 M Tris-HCl, pH 8.0. The cross-linked product was mixed with equal amount of Laemmli buffer and analyzed over 10% or 12% SDS-PAGE.

2.5.2 Dynamic Light Scattering

Dynamic Light Scattering (DLS) also known as Photon Correlation Spectroscopy measures the time-dependent fluctuations of the scattering intensity from particles undergoing random Brownian motion [280]. Furthermore, Dynamic light scattering is used to differentiate small oligomers and dimer from the monomer. This technique is further complemented with size exclusion chromatography and chemical crosslinking assay. In addition, the DLS is capable of measuring several other biophysical parameters such as molecular weight, hydrodynamic radius, correlation coefficient and translational diffusion constant etc. [280].

Dynamic Light Scattering Protocol for FANCI ARM Repeat

1. Molecular size measurement was performed using MALVERN Zetasizer, Brookhaven 90 plus particle size analyzer (Brookhaven, NY) and also with DynaPro Nanostar (Wyatt Technology, USA).

1. Purified 1 mg/ml protein was scanned at an interval of 5 minutes for 15 minutes and the effective diameter of each measurement was recorded.

2. For temperature dependent DLS; 1 mg/ml protein was scanned at an interval of 5 minutes for 15 minutes on temperature ranging from 25°C to 45°C at 5°C interval. The effective diameter of each measurement was considered and plotted with percent intensity versus diameter (nm) values.

Dynamic Light Scattering Protocol for FANCD2 CUE Domain and mutants

1. Molecular size measurement was performed using DynaPro Nanostar (Wyatt Technology, USA).

2. 1 mg/ml *wild-type* and mutant proteins were scanned at an interval of 5 minutes for three times and the effective diameter of each measurement was recorded. However, for frequency distribution graph percent intensity and radius (nm) were taken into consideration.

2.5.3 Crystallization

Three-dimensional structural analysis of a protein at the atomic level is very important to understand the protein structure and function. Furthermore, it helps in designing the small molecule inhibitors or peptides to modify its function as a potential drug target. During crystallization, purified protein forms the nucleation from an aqueous solution and tends to crystallize. In order to achieve crystallization, purified protein undergoes slow precipitation from an aqueous solution. The systematic arrangement of protein molecules in the form of lattice are held together by non-covalent interactions, forming the smallest unit known as "unit cells". Protein crystallography is a commonly used technique to determine of the three-dimensional structure using X-ray diffraction method. To get a good quality crystal, a highly purified protein with homogeneous nature is required with high concentration. Furthermore, a good crystal quality, shape, and size are also important to get a better diffraction pattern especially in the case of protein crystals.

Furthermore, in order to produce a well-shaped crystal, appropriate uniformity with purity of the protein is required at optimal pH conditions. However, to overcome the difficulty on the size of protein crystals, the microseeding and macroseeding methods can be employed to improve the quality of the protein crystals. For crystallization, protein should be present in appropriate buffer system and precipitants [281].

Many factors affect the crystallisation process of the proteins, such as protein purity, homogeneity, concentration, temperature, pH, ionic strength, etc. The process of crystallization is commonly divided into two steps:

1) Nucleation process
2) Crystal growth.

Nucleation and growth can occur in the supersaturated regions as plotted in phase diagrams (**Figure 2.1**). The diagram mainly includes three regions; unsaturated region, saturated region and supersaturated region. The crystal grows in the saturated or supersaturated region while nucleation starts in the supersaturated region. The major focus of crystallization process is to obtain well-ordered diffraction quality crystals. The best method to get the protein crystals is the systematic treatment of the protein solution in a diverse range of buffers. These buffers have different combinations of precipitants, such as salts, polyethylene glycols, etc. For better crystallization conditions, the protein precipitant mixture should reach to the nucleation phase progressively. Most commonly used precipitants in crystallization are salts of sodium, ammonium, and potassium ($(NH_4)_2SO_4$, NaCl, KH_2PO_4), organic polymers like polyethylene glycols (PEG) and alcohols (methanol, ethanol, propanol, acetonitrile). Salt generally preserve the ionic environment of the solution although organic polymers reduce the protein solubility by reducing the dielectric constant of the buffer. The solubility of a protein is lowest at its isoelectric point (pI) because it carries a net zero charge at particular pH. Therefore, buffer also plays an important role in crystallization process.

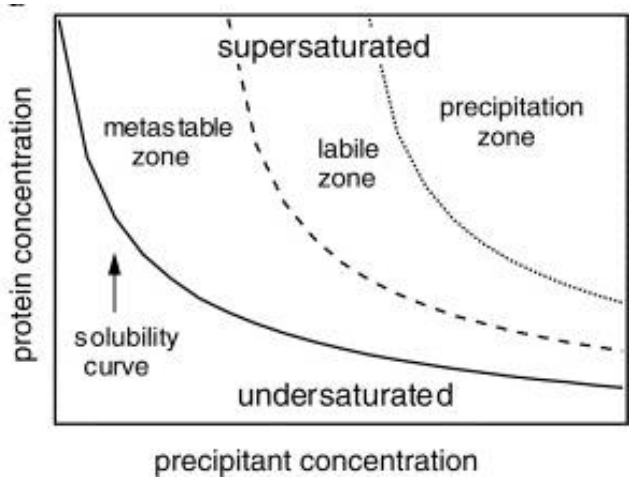


Figure 2.1: The phase diagram of crystallization.

2.5.1 Crystallization Methods: Various methods have been used to grow the crystals of functionally important proteins. Frequently used methods are vapor diffusion, micro-batch, dialysis, and liquid–liquid diffusion technique.

1) **Vapor Diffusion method**: Vapour diffusion method is further divided into hanging drop and sitting drop method.

In both the methods, a drop of purified protein is mixed well with the precipitant. It further allows equilibrating with reservoir solution comprising the same precipitant. At the beginning, protein and precipitant concentration in the droplet is different, but as the system equilibrate to start the diffusion from higher concentration to lower concentration which in turn leads to nucleation. Nucleation is an optimal level for the protein crystallization. Vapor diffusion is a simple but powerful technique as it consumes less amount of protein, and many crystallization conditions can be explored with a limited volume of protein.

2) **Micro-batch**: The micro batch method of crystallization is applied to get the better crystals from the problem of microcrystals. In the micro batch system, low-density paraffin oil (~0.87 mg/ml) which floats on the surface of the protein-precipitant mixture and moderates the rate of evaporation can be used [282]. In this process, a mixture of paraffin and silicone oil in combination can also be used. It is an excellent technique for filtering the already known crystallization conditions to optimize crystal growth [283].

3) **Dialysis**: The dialysis process is based on the slow diffusion of inorganic molecules through a semi-permeable membrane, which leads to the supersaturation of protein solution. The dialysis bag is kept in a chamber containing the pool solution with a precipitant and as system tries to maintain equilibration between precipitant and solute particles by permitting selective passage of water and precipitant the crystal grow better [284].

4) **Liquid–liquid diffusion method**: In this method, protein and precipitant are in a direct contact with the dense solution at the bottom. A concentration gradient is made by the diffusion of protein and precipitant, and thus crystallization may occur at appropriate protein and precipitant concentration.

2.5.4 Domain organization of the genes cloned in expression constructs

FANCI ARM Repeat, BRCA2 CTR, and FANCD2 CUE Domain were cloned in pET28a+ vector and pGEX4T-1 vector, respectively. Respective protein expression was checked using bacterial system Rosetta 2(DE3)/ BL21 (2DE3) strain.

Site-directed-mutagenesis with mismatch primers for respective mutations was performed over the *wild-type* template. All the reported mutations were confirmed by DNA

sequencing. *Wild-type* and mutant constructs were transformed into bacterial expression host, *E. coli* Rosetta 2(DE3)/ BL21 (2DE3) strain, for protein expression.

2.5.5 Bacterial Two-Hybrid Assay

Bacterial two-hybrid (BTH), is based on "Bacterial Adenylate Cyclase-based Two-Hybrid" system, to detect and characterize protein-protein interactions *in vivo* [285, 286]. Comparison to yeast two-hybrid system, BTH offers many advantages of working with *Escherichia coli*. BTH requires techniques such as PCR, plasmid preparations, transformation and blue-white colony screening [285]. The BTH system has been established by the group of Dr. D. Ladant at the Pasteur Institute. The basis of this technique relies on the interaction-mediated reconstitution of the adenylate cyclase activity in *E. coli*. The catalytic domain of adenylate cyclase (CyaA) from *Bordetella pertussis* comprises of two complementary catalytic fragments, T25 and T18. However, they are not able to restore adenylate cyclase activity when physically separated [285, 286] (**Figure 2.2**).

These two fragments become active when they are fused to interact polypeptides such as X and Y, and the heterodimerization of these hybrid proteins, in turn, leads to functional complementation between T25 and T18 fragments.

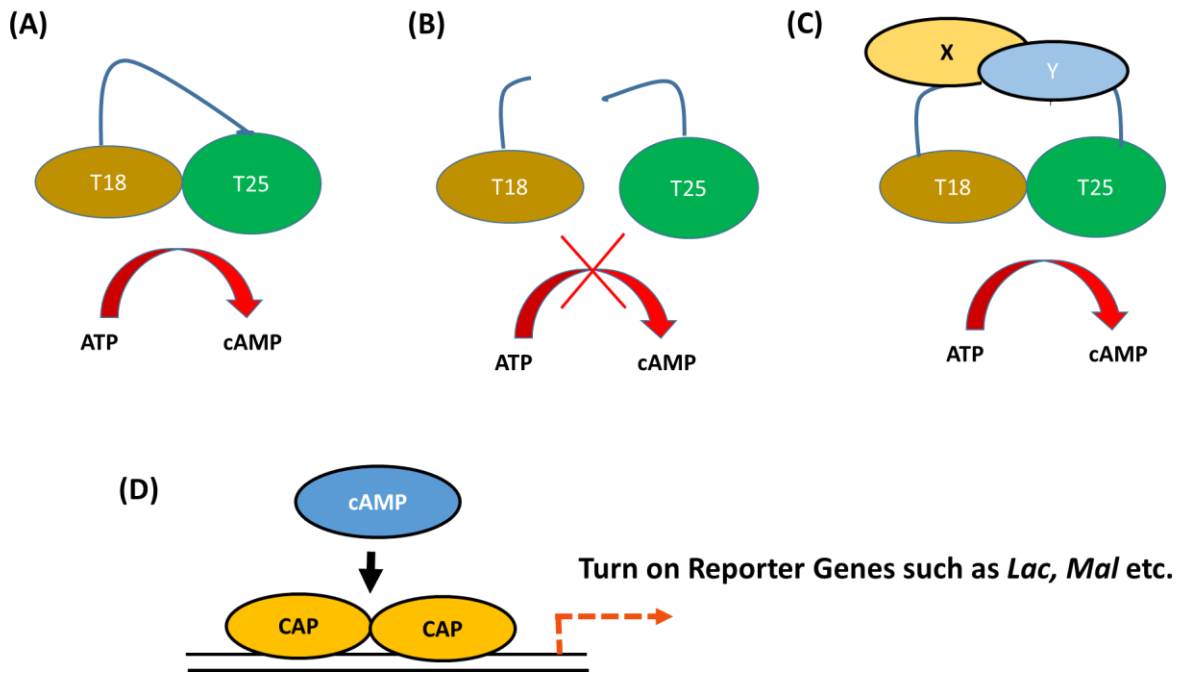


Figure 2.2: Principle of Bacterial Two-Hybrid Assay **(A)** T18 and T25 fragments interacts and forms cAMP, **(B)** T18 and T25 fragments not able to interact and does not form cAMP, and **(C)** T18 and T25 fragments interact through interacting partner X and Y proteins and restore the catalytic functions and forms cAMP **(D)** cAMP binds with CAP and turn on the expression of reporter genes

After complementation, the cAMP synthesis starts that binds to the catabolite activator protein (CAP). The cAMP/CAP complex is a pleiotropic regulator for gene transcription in *E. coli* that turns on the expression of several genes including genes of the *lac* and *mal* operons involved in lactose and maltose catabolism. Hence, bacteria become capable of utilizing lactose or maltose as the unique carbon source that can be easily detected on indicator media such as X-Gal.

Protocol for Bacterial Two-Hybrid Assay

In vivo interactions of different proteins were monitored using bacterial two-hybrid system (BACTH) as described in following:-

(1) FANCI (985-1207) aa ARM repeat, FANCD2 (1-254) aa CUE Domain, and BRCA2 (2350-2545) aa C Terminal Region, were cloned in bacterial two-hybrid plasmids such as pUT18, pUT18C, pKT25 and pKNT25. A series of recombinant plasmids were generated in different combinations as given in Tables of chapter 5 and 6.

(2) These constructs were co-transformed in specialized *E. coli* BTH101 host in different combinations, to detect possible interactions between co-expressing recombinant proteins with T18 and T25 fragments that can be detected by spotting recombinant cells on LB agar plates supplemented with appropriate antibiotics.

(3) Cells were grown overnight in triplicates, 5 μL of cell culture were spotted on LB agar plate containing 5' bromo 4 chloro-3-indolyl-β-D-galactopyranoside (X-Gal) (40 μg / mL), IPTG (0.5 mM) in the presence of antibiotics as required. Spotted LB-X-Gal-IPTG plates were incubated at 30°C overnight and the appearance of blue-white colored colonies was analyzed.

2.5.6 Beta-galactosidase Assay

(1) An aliquot of the cell culture from was grown in LB broth overnight with 0.5 mM IPTG in appropriate antibiotics. The β-galactosidase activity was calculated from liquid cultures.

(2) Further, the cultures were diluted 1:4 into LB broth medium and optical density at λ=600 nm were monitored. 100 μL of cultures were mixed with 400 μL Z- buffer containing 60 mM Na_2HPO_4, 40 mM NaH_2PO_4, 10 mM KCl, 10 mM $MgSO_4$, 50 mM β-mercaptoethanol pH 7.0.

(3) Further 0.01 % SDS and 20 μL chloroform was added to permeabilize the cells. Enzyme activity was measured in triplicates with 50 μL of clear supernatant using 0.4 % O-

nitrophenyl-*β*-D-galactoside (ONPG) as a substrate. The β-galactosidase activity was calculated and plotted in Miller units by using a spectrophotometer.

2.5.7 Chaperone-like Assay

Chaperone-like Assay [287] was performed by using lysozyme as protein substrate in following steps-

(1) Lysozyme (0.3 mg/ml) in phosphate buffer (pH 7.40) was pre-incubated for 5 min at 37°C either alone or in the presence of same quantities desired proteins in 96 well plate.

(2) The reaction was started by adding DTT up to the final concentration of 20 mM and aggregation of lysozyme was monitored in the optical density of 340 nm wavelength at different time points.

2.5.8 Surface Plasmon Resonance

Surface Plasmon Resonance (SPR) used for oscillation of electrons in a solid or liquid field by incident radiation [288]. The resonance condition occurs when the natural frequency of surface electrons oscillating against the restoring force of positive nuclei that coincide with the frequency of light photons. SPR measures adsorption of material on the inert substances such as gold surface or onto the metal nanoparticles. The basic principle utilizes the change in the reflective index as a result of absorption (interaction) of molecules on the immobilized ligand [289].

SPR experiments were performed using BiacoreT200 Surface Plasmon Resonance (GE Healthcare). SPR was performed in the following steps:-

(1) **Immobilization**: Immobilization of mono-ubiquitin (Boston Biochem, USA) was performed on carboxymethyl (CM5) sensor chip using amine coupling protocol (as described in amine coupling kit).

(2) **Immobilization level detection**: Immobilization level detection of 2500 RU was selected for immobilization on the sensor chip. HBSEP+ buffer used as running buffer and glycine-HCl pH 2.50 used for regeneration (30-second pulse in 30µl/min flow rate) with stabilization period of 60 seconds.

(3) **The flow of analytes on the Immobilized surface:** Purified proteins used as analytes with concentrations from 20, 40, 60, 80,100, 200, 400, 600 and 800 nM, with a 30 µl/min flow rate, contact time of 120 seconds and a dissociation time of 60 seconds. Data analysis was done using Biacore T200 evaluation software.

2.5.9 Pulldown Assay for protein-protein interactions

Pull down assay [290] of the desired proteins were performed in following steps:

(1) FANCD2 (1-254) CUE Domain, FANCD2 Ser222Ala CUE domain mutant, FANCD2 Leu231Arg CUE domain mutant and BRCA2 (2350-2545) C-terminal region, cloned in pGEX4T-1 and pET28a+ vectors.

(2) Co-transformation of desired constructs in respective combination was done for recombinant proteins expression with BRCA2 (2350-2545) C-terminal region in chemically competent *E. coli* Rosetta 2(DE3) cells.

(3) The overnight grown a culture of *E. coli* Rosetta 2(DE3) harbouring different constructs combinations were diluted 1:100 in fresh LB broth and incubated at 37°C, using shaking incubator at 200 rpm till OD (λ= 600 nm) reaches 0.60–0.80.

(4) The culture was then induced by the addition of 0.40 mM IPTG (Isopropyl β-D thiogalactoside) for 16 hours at 21°C.

(5) Further, cells were lysed in buffer containing 300 mM NaCl, 50 mM Tris, 0.1% triton-X, 5mM beta-mercaptoethanol, pH 8.00, protease inhibitor cocktail (Sigma) and sonicated. The cell lysate was centrifuged at 15,000 rpm for 40 min at 4°C.

(6) Clear cell supernatant was allowed to pass from GST resin, on Snapcap column (Thermo-Fisher). Bound protein was eluted using the buffer containing 20 mM glutathione, 100 mM NaCl, 25 mM Tris, pH 8.50, followed by three washing in buffer (50 mM NaCl, 10 mM Tris, pH 8.5).

(7) Eluted samples from Snapcap column (Thermo-Fisher) were mixed with laemelli loading buffer, and heated at 95°C for 10 minutes before loading onto 12% SDS-PAGE gel. The gel was stained and analyzed using Coomassie brilliant blue.

Chapter 3

Structural and Biophysical Properties of FANCI and FANCD2

3.1 Structural and Biophysical Properties of FANCI

3.1.1 Introduction:

To preserve the fidelity of genetic information, cells evolved in a complex, dynamic and tightly regulated network called DNA damage response (DDR) [291]. In genotoxic stress, DDR coordinates different cellular processes like cell cycle regulation, chromatin remodeling, DNA-repair and transcriptions [291]. The DDR signaling cascade involves the sensing of the DNA damage, assembly, and recruitment of many DDR sensors, mediators and effectors molecules at the sites of damage [292]. Intercrosslink DNA strand elicits the activation of ATM and ATR kinases which phosphorylate different protein specifically proteins present in Fanconi anemia (FA) pathway [3].

Germline inactivation in any one of these genes causes Fanconi anaemia [293]. Fanconi anemia complementation groups I (FANCI) protein, one of the FA pathway protein facilitates DNA ICL (Inter-Crosslink) repair and plays a crucial role in maintaining the genomic integrity [3]. FANCI protein comprises 1328 amino acids including armadillo (ARM) repeat between 985-1207 amino acids [3]. ARM repeats are functionally diverse and evolutionarily conserved domain that plays a key role in protein-protein and protein-DNA interactions [9, 294]. ARM repeats are composed of tandem copies of the degenerate protein sequences that form conserved three-dimensional structures [295]. ARM repeat was first discovered in segment polarity gene of Drosophila and afterwards in other proteins like junctional plaque protein plakoglobin [3, 296, 297], tumour suppressor adenomatous polyposis coli (APC) protein [22, 167] , nucleocytoplasmic transport factor protein importin, Fanconi anemia complementation group I (FANCI) and Fanconi anemia complementation group D2 (FANCD2) protein [3, 296]. The role of ARM repeat in tumour

suppression is well documented [167]. ARM repeat proteins are generally composed of the compact and dynamic region which performs as molecular recognition component and interacting module [20].

Here, we are reporting the multimodal comprehensive *in-vitro, in-silico* and biophysical approach to understand the structural and biophysical properties of FANCI ARM repeat protein.

3.1.2 Materials and methods:

3.1.2.1 Gene cloning, protein expression, and purification

FANCI ARM repeat region (985-1207) aa was PCR (Thermocycler, Biorad) amplified using full-length FANCI cDNA as template with forward primer 5'-GTCGGATCCGAGAACCTGTACTTTCAGGGTCTAGTCACGGTTCTTACCAG-3' and reverse primer 5'-GTCCTCGAGCTATTAGGGGGTCAGATGAGAACCAG-3'. The forward primer was designed with the TEV protease site (ENLYFQG) for native protein purification, having restriction BamH1/Xho1 sites. The ARM repeat region from the gene of interest was amplified using following PCR condition: initial denaturation 95 °C for 5 minutes, denaturation 95 °C for 45 seconds, annealing 58°C for 45 seconds, extension 72°C at 0.50 kb/min, final extension 72°C for 10 minutes for 32 cycles. PCR product of amplified ARM repeat was subcloned into the pET28a vector (Novagen). Positive clones were selected by colony screening, restriction digestion and finally confirmed by DNA sequencing. DNA sequencing was 100% matching with the gene of interest.

FANCI ARM repeat cloned in pET28a+ vector was induced in the *E. coli* Rosetta (2DE3) cells (Novagen) at $O.D_{600}$ between 0.60-0.80 by adding 0.40mM IPTG and kept at

22^0C overnight. Protein was purified in buffer containing 300mM NaCl, 50 mM Tris (pH 8.50), 0.10% Triton-X, 5mM beta-mercaptoethanol. The 6xHis-tag fusion protein was purified by affinity chromatography (Ni-NTA beads, Qiagen).

3.1.2.2 Protein characterization

The concentration of purified protein was determined by NanoDrop 1000 (Thermo Scientific, USA) and Bradford method [298]. The absorbance of protein sample(s) was recorded after blank correction and concentration was determined by considering extinction coefficient. Bradford assay using spectrophotometer was performed using the absorbance of different concentrations of BSA as standard. The concentration of protein of interest was determined by a standard curve.

Size exclusion chromatography was performed using Superdex-75 pre-packed (AKTA explorer, GE Healthcare) FPLC column as described in **Chapter 2**. Prior to injecting the protein, the FPLC column was pre-equilibrated with 300mM NaCl, 50mM Tris, 5mM β-Mercaptoethanol, pH 8.50. Standard proteins of known molecular weight were used to calculate void volume and the total volume of the column. The experiment were repeated twice and averaged for elution volume calculation.

3.1.2.3 Chemical crosslinking assay

Purified protein was incubated in 0.10% glutaraldehyde for 0, 2.5, 5, 10, 15, 30, and 60 min. This reaction was terminated in a time-dependent manner by adding 5 µl of 1 M Tris pH-8.50. Untreated purified protein sample was taken as control. Crosslinked product was mixed with equal amount of Laemmli buffer to analyze on 12% SDS-PAGE gel.

3.1.2.4 Limited proteolysis and mass spectrometry

FPLC purified 1 mg/ml protein was treated with trypsin in time-dependent manner till the final concentration of 10 pg/µl and the untreated protein was taken as control. The reaction mixture was incubated at 37 °C (trypsin) for different time period 1, 5, 10, 15, 30, 60, and 180 min. Reactions were terminated by adding 2 µl of 200mM PMSF (Sigma-Aldrich). Reaction samples were heated with laemmli buffer at 95°C for 5 minutes to analyze on SDS-PAGE by Coomassie staining. The band corresponding to 14 kDa was considered as a stable fragment and was further subjected to trypsin digestion followed by mass spectrometry (MALDI TOF-TOF Ultraflex- II from Bruker Daltonics, Germany). Digested peptides were analysed with high sensitivity at attomole range using peptide mass fingerprinting. The domain of interest was identified by Mascot analysis with Biotool software (Bruker Daltonics).

3.1.2.5 Dynamic light scattering

FANCI ARM repeat protein at a concentration of 1 mg/ml was filtered 0.22µm filter and degassed at 4°C prior to DLS measurements. Malvern zeta-sizer (https://www.malvern.com/en/products/product-range/zetasizer-range/zetasizer-nano-range) was used to study the oligomeric characteristic of protein with different temperatures. Wyatt DynaPro nanostar (http://www.wyatt.com/products/instruments/dynapro-nanostar-dynamic-light-scattering-detector.html) was used for DLS experiment.

3.1.2.6 Circular -Dichroism

Secondary structure characterization was performed using CD polarimeter (Jasco J-815, Japan) in the far UV range at 20°C/25°C at the wavelength ranges between of λ=260-190nm. An average of seven spectra were taken for final representation in mean residual ellipticity. Further, averaged spectra were used for secondary structure quantification by K2D3 server [299] . 10 µM protein was taken to record the spectra from 10°C to 50°C to understand the thermal denaturation. For chemical unfolding experiments, 10 µM protein was treated with 0 M, 4 M and 8 M urea to record the emission spectra were recorded as described earlier in **Chapter 2**.

3.1.2.7 Fluorescence spectroscopy

Micro-environment of tryptophan residues (intrinsic fluorophore) at the hydrophobic core of the protein was monitored using fluorescence spectrophotometer (Horiba, Japan) at the excitation wavelength of λ=295nm. The emission spectra of 10 µM were recorded from λ=310nm to λ=450nm to evaluate thermal and chemical unfolding behavior of protein.

.

3.1.2.8 Model building, molecular dynamics, principle component analysis and normal mode analysis

FANCI amino acids from 985 to 1207 was retrieved from UniProtKB [300] (Uniprot ID Q9NVI1) and submitted to Robetta server [263] for 3D model building [301]. Stereochemical refinement of Ramachandran outliers in the model was performed by ModLoop server [302]. Further model was validated for stereochemistry using SAVES server (https://services.mbi.ucla.edu/SAVES/) which includes PROCHECK, ProSA (Protein Structure analysis) and PROVE plot [303]. The validated model was used for the

molecular dynamics simulation using GROMACS 4.5.5 [266] with the implementation of OPLS-AA/L force field [14, 304]. The system (solute configurations) was solvated using TIP3P water model in a cubic box with periodic boundary conditions. Furthermore, counter-ions were added to neutralize the system. The systems were first energy minimized using steepest descent algorithm with a tolerance of 1000 kJ/mol/nm. Electrostatic interactions were calculated using Particle Mesh Ewald (PME) summation [265] with 1 nm cut-offs. Coulombic interactions and Vander Waal interactions were calculated at a distance cut-off of 1.40 nm [305]. System equilibration was performed by applying positional restraints on the structure using constant number (N), volume (V) and temperature (T) [NVT] followed by a constant number (N), pressure (P) and temperature (T) [NPT] ensemble for 100 picoseconds each. The temperature of 300K was coupled using Berendsen thermostat with pressure of 1 bar, in Parrinello-Rahman algorithm [306]. The equilibrated systems were subjected to 80 nanoseconds of the production run with the time-step integration of 2 femtoseconds. The trajectories were saved at every 2 picoseconds and analyzed using Gromacs 4.5.5. Root mean square deviation (RMSD), root mean square fluctuations (RMSF), the radius of gyration (Rg), hydrogen bonds, solvent-accessible surface area (SASA), Dictionary of the secondary structure of the protein (DSSP) and the volume were calculated [307]. Furthermore, cross-correlation analysis for principle component analysis (PCA) was performed [308]. Eigenvector and Eigenvalues were calculated after diagonalizing the covariance matrix. Trace of covariance matrix was calculated by adding all the eigenvalues. The eigenvalue was plotted for each eigenvector which suggests that 90% of the ARM repeat protein dynamics is exhibited by the first two eigenvectors 1 and 2 and 7.50% by eigenvector 2 and 3. Therefore, to understand the dynamics we have projected the eigenvector 2 on 1 [270, 271, 309]. Normal mode analysis

(NMA), Anisotropic network modelling (ANM) and Gaussian network modelling (GNM) were performed using R 3.2 package and ProDy (Protein Dynamics 1.7) [310, 311].

3.1.3 Results and discussion:

3.1.3.1 Cloning, expression, purification and identification of FANCI ARM repeat

Potential clones were selected by colony screening and restriction digestion with the BamH1 and Xho1 which showed the insert release of appropriate size (**Figure 3.1**). DNA sequencing results confirmed the presence of ligated gene of interest in the vector with desired reading frame of the codon sequences.

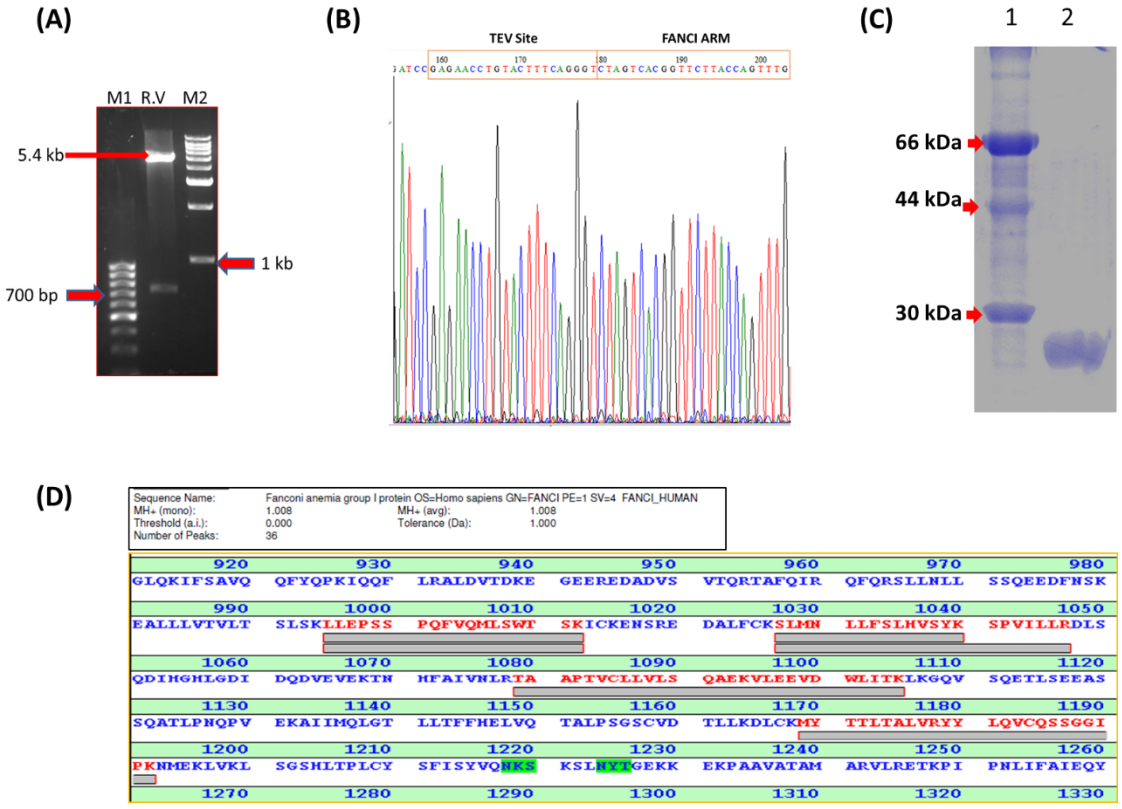


Figure 3.1: (A) Restriction enzyme digestion of FANCI ARM Repeat cloned in pET28a+ vector **(B)** DNA Sequencing chromatogram of FANCI ARM repeat cloned in pET28a+, exhibiting TEV site and FANCI ARM Repeat codons in correct reading frame, (different

colors show the nucleotides; Adenine=Green, Guanine=Black, Cytosine=Blue and Thymine=Red) **(C)** FANCI ARM Repeat purified protein on SDS-PAGE gel and **(D)** Peptide mass fingerprinting using mass spectrometry profile for the identification of the FANCI ARM Repeat protein (grey color shows the matching of captured peptides)

The protein was over-expressed and purified as mentioned in Material and Methods (**Section 2.2, Chapter 2**). The purified protein was subjected to mass spectrometry for the confirmation of the presence of FANCI ARM repeat sequences.

3.1.3.2 Oligomeric behavior of FANCI ARM repeat

In order to understand the oligomeric behavior of the ARM repeat of FANCI protein, size exclusion chromatography, glutaraldehyde crosslinking and dynamic light scattering studies were performed.

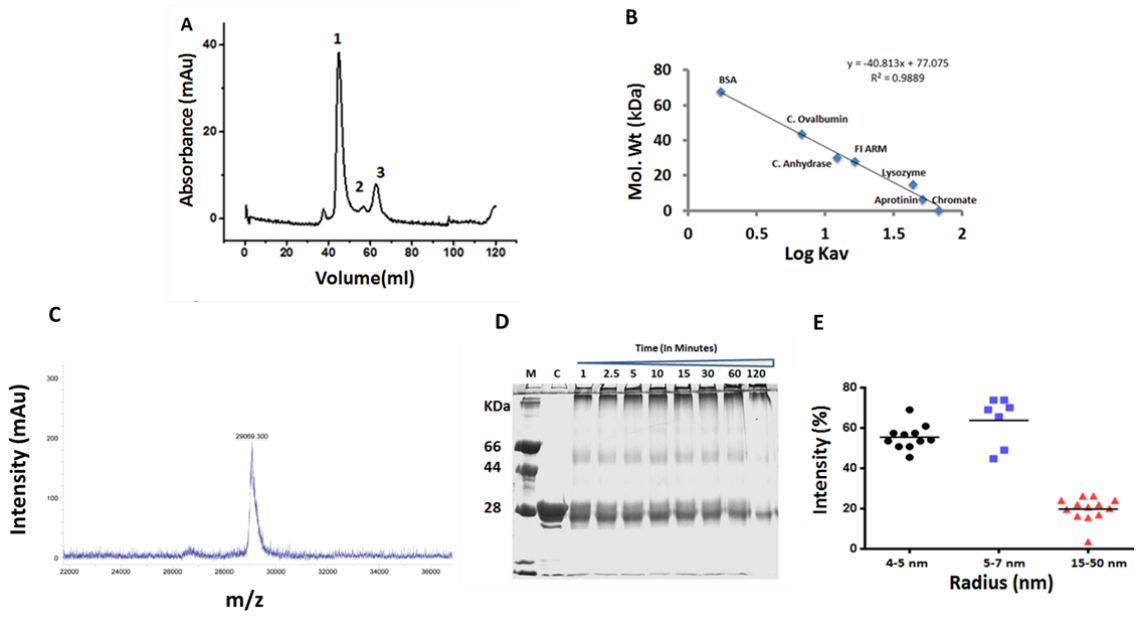


Figure 3.2: (A) Size exclusion chromatography of FANCI ARM repeat, elution peak 1, 2 and 3 showing oligomer, dimer and monomer **(B)** Molecular weight determination with standard proteins **(C)** Whole molecular weight determination by mass spectrometry **(D)**

Time-dependent glutaraldehyde crosslinking SDS-PAGE gel profile showing different species of monomer, dimer and oligomer **(E)** DLS profile showing multimeric population of protein, each dot was showing the single DLS measurement and ≥7 scans were used for the distribution plot

Table 3.1 [84]

Molecular weight estimation of purified protein

		Experimentally derived mol. wt. (kDa)	
Theoretical mol. wt.(kDa)[a]	Ve/Vo[b]	Size exclusion chromatography	Mass spectrometry
28.98	1.40±0.01	26.6±1	29.08±0.05

Note: Ve/Vo: Elution volume/Void volume ratio in gel filtration chromatography (superdex 75 16/60)
[a]Determined from Protparam, Expasy.
[b]Determined from standards chromate,aprotinin,lysozyme,carbonic anhydrase, ovalbumin, albumin, ferritin, dextran

Affinity-purified protein was subjected to size exclusion chromatography (SEC) using Superdex 75 column. It has been found that protein elutes at different column volumes of 62.7±0.25 ml, 56.91±0.16 ml and 44.9±0.56 ml suggesting monomeric, dimeric and oligomeric behavior respectively (**Figure 3.2 A**). Further to confirm oligomeric property of ARM repeat, time-dependent glutaraldehyde cross-linking and DLS studies (**Figure 3.2 C and D**) were performed. In crosslinking, dimeric and oligomeric species were predominant at the size of 58 kDa and >66 kDa respectively (**Figure 3.2 C)**. In DLS protein shows different hydrodynamic radii which suggest its polydispersive characteristics (**Figure 3.2 D**). In conclusion, these results suggest that FANCI ARM protein exists in monomeric and multimeric forms at a physiological temperature that might be due to the presence of some structurally distorted region. The molecular weight of protein was determined by mass spectrometry and SEC **(Table 3.1)** [84].

3.1.3.3 Secondary structure characterization of FANCI ARM repeat

The secondary structure of FANCI ARM repeat protein was characterized by far-UV CD spectroscopy using FPLC purified protein. CD spectra of ARM repeat show elliptical minima (in millidegree ellipticity) at λ= 218 nm and 222 nm indicating the predominantly presence of α- helices (Fig 2 A), which has sustained resemblance with the crystal structure (PDB ID; 3S4W) and the modelled structure (**Figure 3.5 D**) [84].

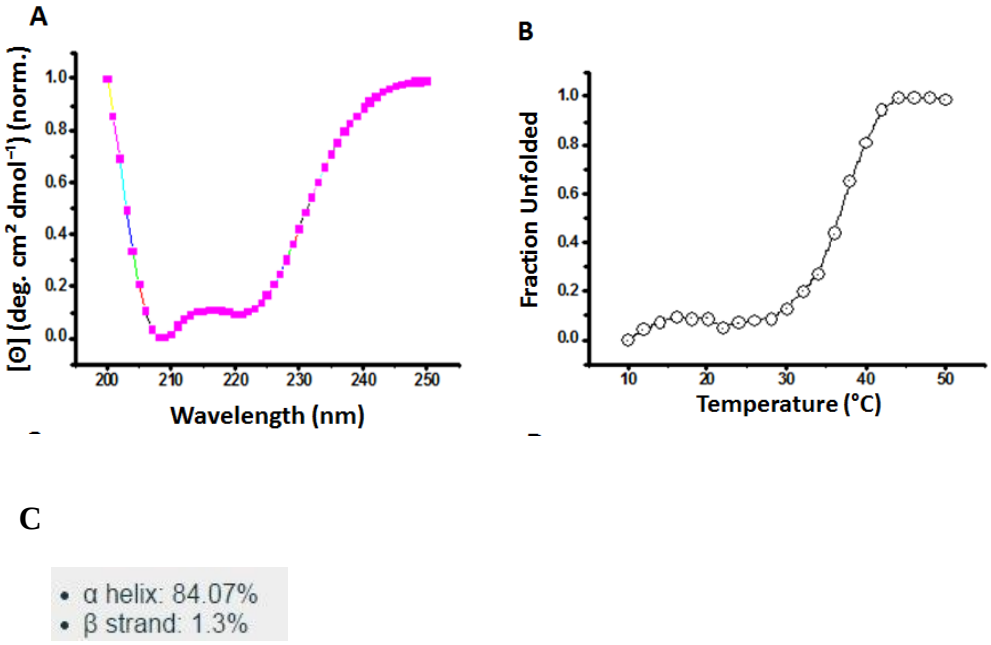


Figure 3.3: (A) Far –UV, CD spectra of FANCI-ARM repeat, indicating the α- helical nature of protein, **(B)** Thermal denaturation profile by CD showing protein unfolds at 50°C **(C)** Quantification of α- helices of CD data using K2D2 server

To study the thermodynamic stability, temperature dependent unfolding of the protein was performed from range of 10°C to 50°C with the interval of 2°C, and fraction unfolded was calculated at λ=222 nm. Furthermore, the melting temperature (T_m) was calculated by fitting into a two-state unfolding pathway [312] (**Figure 3.3**) (**Table 3.1**). These results suggest that protein predominantly composed of α-helices and have T_m of ~36°C [84].

3.1.3.4 FANCI ARM protein tertiary structure conformation

Folding behavior of the protein varies with protein type and tertiary structure conformation. It often becomes complicated because of different folding pattern and sensitivity of fluorophore environment [313]. To investigate the overall folding and compactness of ARM repeat, fluorescence spectroscopy was carried out with two intrinsic fluorophores tryptophan at 25 and 118 positions present in ARM repeat (numbering positions of fluorophores were considered in ARM repeat (985-1207 aa) from 1 to 222 amino acids). To get an idea about the tertiary structure conformation, we have recorded the fluorescence scan of both folded as well as unfolded (protein treated with 8M urea) FANCI ARM repeat protein. It was found that the fluorescence emission maxima of the folded protein are at λ= ~333 nm whereas unfolded protein (protein in 8M Urea) shows the emission maxima at λ= ~345 nm. However, protein treated with 4M urea shows a decrease in fluorescence intensity (**Figure 3.4**) suggesting the presence of intermediates such as molten globule or oligomeric species which has sustained resemblance with FPLC and DLS data [84].

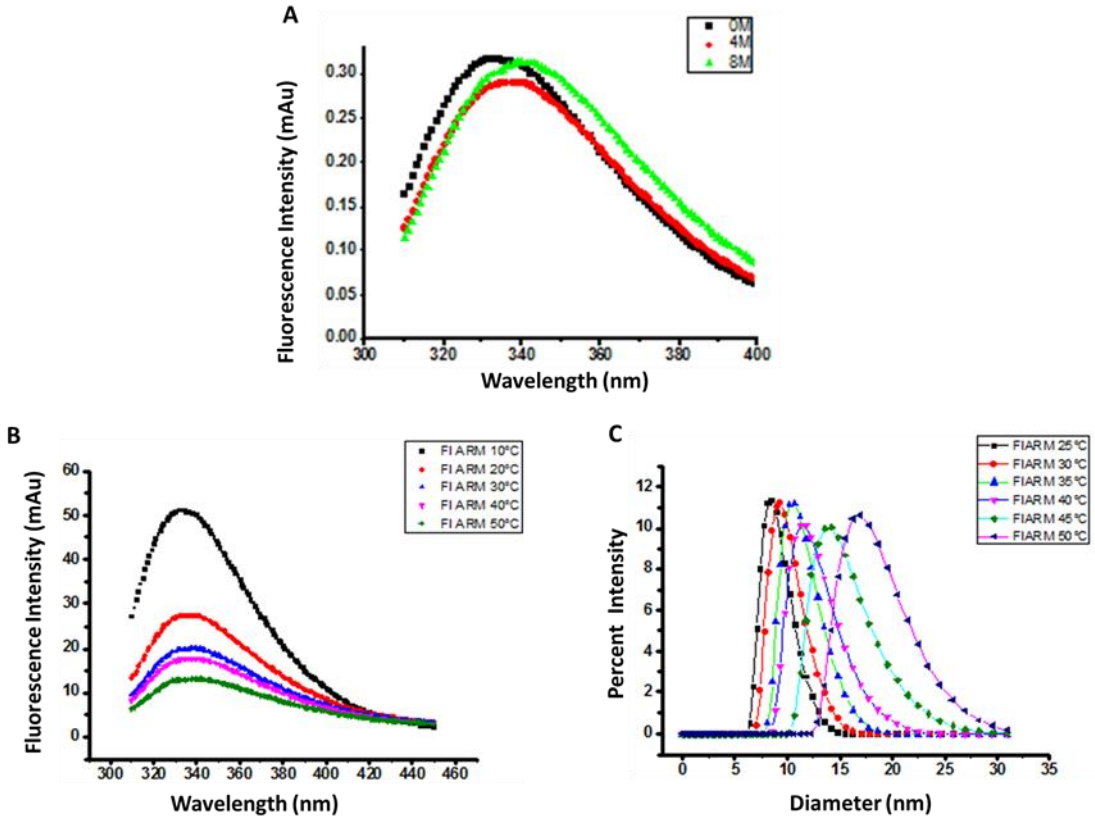


Figure 3.4: (A) Chemical denaturation profile of fluorescence spectroscopy at 0M, 4M and 8M urea **(B)** Thermal denaturation profile using Fluorescence spectroscopy showing steep decrease in intensity beyond 30°C and **(C)** Temperature dependent DLS profile showing oligomer formation beyond 30°C

Temperature (°C)	Red Shift (nm)
10°C	333 nm
20°C	338 nm
30°C	339 nm
40°C	339 nm
50°C	341 nm

Interestingly, temperature dependent unfolding of the FANCI ARM repeat from 10°C to 50°C, a decrease in fluorescence intensity and characteristic red shift of the emission maxima observed at 30°C, suggests the misfolded molten globule protein fractions are present predominantly beyond 30°C (**Figure 3.4**). Further to support this observation, temperature dependent DLS studies (**Figure 3.4**) were performed and found that protein hydrodynamic radii have increased beyond 25°C. In temperature dependent DLS, CD and

Fluorescence spectroscopy, protein shows high propensity of polydispersity and perturbed secondary and tertiary structure conformation beyond 20°C.

3.1.3.5 Peptide mass fingerprinting and *in-silico* analysis of stable region of FANCI ARM repeat

In general, the compact globular domain of protein shows the resistivity towards protease digestion that helps to determine the stability and dynamic conformation of the proteins [314]. To identify the compact region of ARM repeat peptide mass fingerprinting using MALDI-TOF-TOF was performed for ARM repeat as well as the region which withstands with proteolysis against trypsin digestion. It has been found that at the N-terminus about 95 amino acids of ARM repeat is forming compact region and shows prominent resistivity towards trypsin digestion (**Figure 3.5**). To further validate the compactness of this region the *in-silico* prediction for ARM repeat disorderness using PrDOS [315] were performed. PrDOS results suggest that at N-terminus about hundred amino acids are forming ordered structure, which is in agreement with peptide mass fingerprinting (**Figure 3.5**) [84].

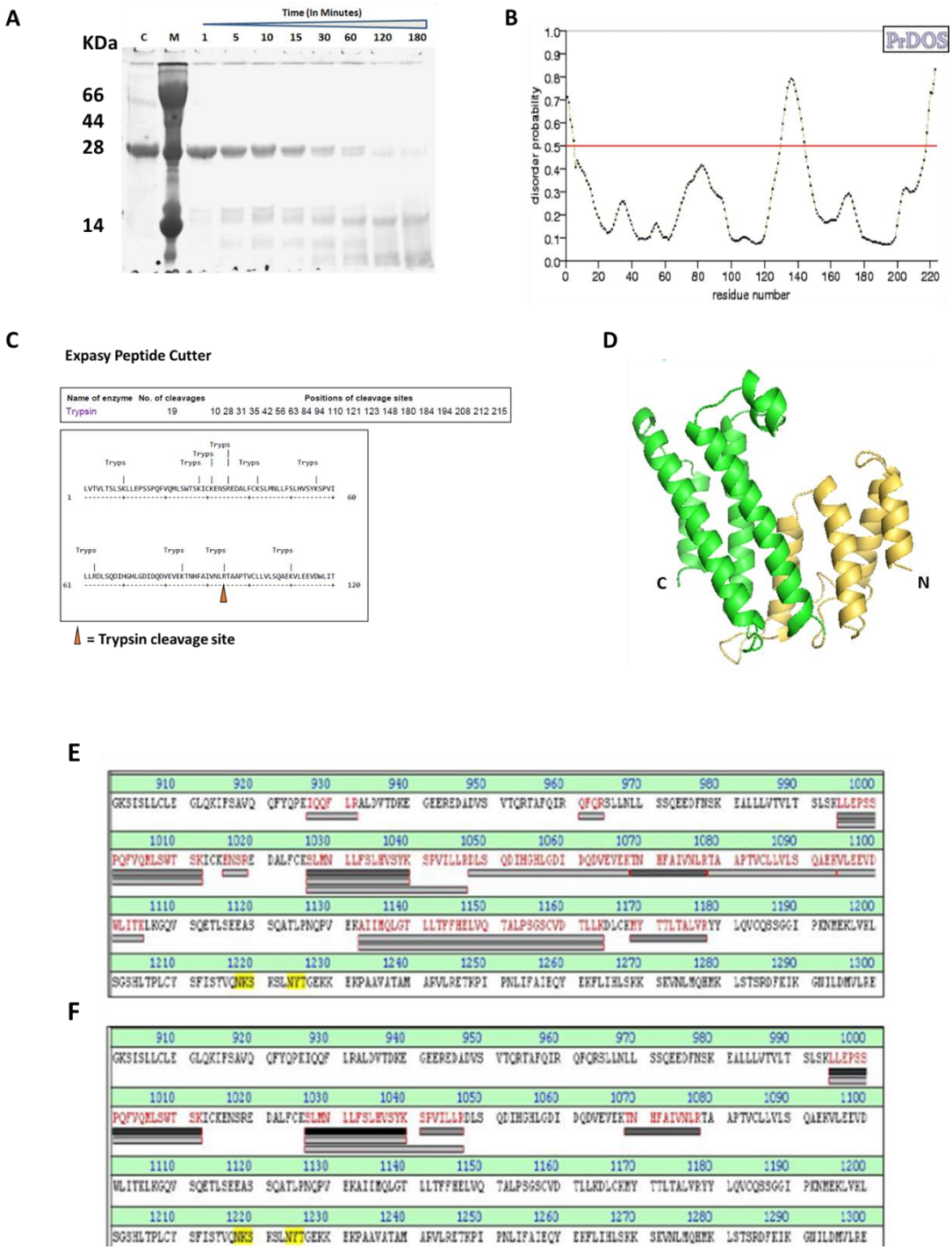


Figure 3.5: (A) Limited proteolysis profile of SDS-PAGE gel **(B)** Disordered prediction by using PrDos showing disordered region in protein **(C)** Expasy peptide cutter with trypsin **(D)** Modelled structure of FANCI ARM repeat protein showing the stable region in yellow and dynamic region in green **(E)** Full length tryptic peptide mass fingerprint and **(F)** stable region profile of tryptic peptide mass fingerprint of MALDI-TOF-TOF

To rule out the possibility of miscleavage of the tryptic digested product, we have used *in-silico* trypsin digestion prediction Expasy peptide cutter [316] and observed a trypsin site, that with the peptide mass fingerprinting results (**Figure 3.5 C**).

Cumulative results from limited proteolysis, mass fingerprinting concludes that ARM repeat has a stable region at its N-terminus of about 100 amino acids (**Figure 3.5 A B C E F**). From modeled structure, we have speculated stable region at N-terminus and dynamic region at C-terminus of FANCI (**Figure 3.5 D**) [84].

3.1.3.6 Observation of HTH- type structural motif in FANCI ARM repeat

It has been reported that C-terminal of ARM repeat has tumor suppressor activity whereas FANCI C-terminal region binds to DNA [22, 83]. To understand the structural property, molecular modelling of FANCI ARM repeat has been carried out with acceptable stereochemistry (**Figure 3.7**). The DALI [317] structural alignment exhibited that ARM repeats of FANCI is structurally conserved and shows significant similarity with elongation factors and DNA-binding transcription factors (**Figure 3.6**).

Since the primary sequences of amino acids dictate protein folds and functions, therefore we have performed sequence alignment with known helix-turn-helix (HTH) structural motif "master sets" [29] and found similarity with dynamic part of twenty amino acids of FANCI ARM repeat (**Figure 3.8)** [84].

A

DALI Structural Alignment

No:	Chain	Z	rmsd	lali	nres	%id	PDB	Description
1:	3s4w-A	27.2	0.7	208	1206	83	PDB	MOLECULE: FANCONI ANEMIA GROUP I PROTEIN HOMOLOG;
2:	3s51-A	24.2	0.8	194	1134	83	PDB	MOLECULE: FANCONI ANEMIA GROUP I PROTEIN HOMOLOG;
3:	1y1u-C	9.2	3.1	131	544	8	PDB	MOLECULE: SIGNAL TRANSDUCER AND ACTIVATOR OF TRANSCRIPTION
4:	1y1u-B	9.2	3.1	131	544	8	PDB	MOLECULE: SIGNAL TRANSDUCER AND ACTIVATOR OF TRANSCRIPTION
5:	1y1u-A	9.1	3.2	132	544	8	PDB	MOLECULE: SIGNAL TRANSDUCER AND ACTIVATOR OF TRANSCRIPTION
6:	3s4w-B	8.9	4.3	171	1146	9	PDB	MOLECULE: FANCONI ANEMIA GROUP I PROTEIN HOMOLOG;
7:	5cqs-A	8.2	4.7	123	325	8	PDB	MOLECULE: ELONGATOR COMPLEX PROTEIN 1;
8:	5cqs-B	8.2	7.3	128	323	7	PDB	MOLECULE: ELONGATOR COMPLEX PROTEIN 1;
9:	4pju-A	8.1	5.2	156	896	7	PDB	MOLECULE: COHESIN SUBUNIT SA-2;
10:	3cwg-A	8.1	3.3	129	501	7	PDB	MOLECULE: SIGNAL TRANSDUCER AND ACTIVATOR OF TRANSCRIPTION
11:	2jqq-A	8.0	2.6	111	154	9	PDB	MOLECULE: CONSERVED OLIGOMERIC GOLGI COMPLEX SUBUNIT 2;
12:	5cqs-D	8.0	7.2	127	324	7	PDB	MOLECULE: ELONGATOR COMPLEX PROTEIN 1;
13:	1bg1-A	8.0	3.2	128	559	7	PDB	MOLECULE: PROTEIN (TRANSCRIPTION FACTOR STAT3B);
14:	3cwg-B	7.9	3.4	130	507	9	PDB	MOLECULE: SIGNAL TRANSDUCER AND ACTIVATOR OF TRANSCRIPTION
15:	1yvl-B	7.9	3.4	128	653	6	PDB	MOLECULE: SIGNAL TRANSDUCER AND ACTIVATOR OF TRANSCRIPTION

Figure 3.6: (A) DALI structural alignment shows similarity with transcription factors and elongation factors proteins

The well conserved “PHS” signature (where P is the charged residue mostly glutamate, H is any hydrophobic amino acid and S is small residue) in HTH motif [27] has also been identified in the amino acid sequences (**Figure 3.8**). These results suggest that the long suspecting DNA-binding HTH-type structural motif is present in ARM repeat of FANCI [84].

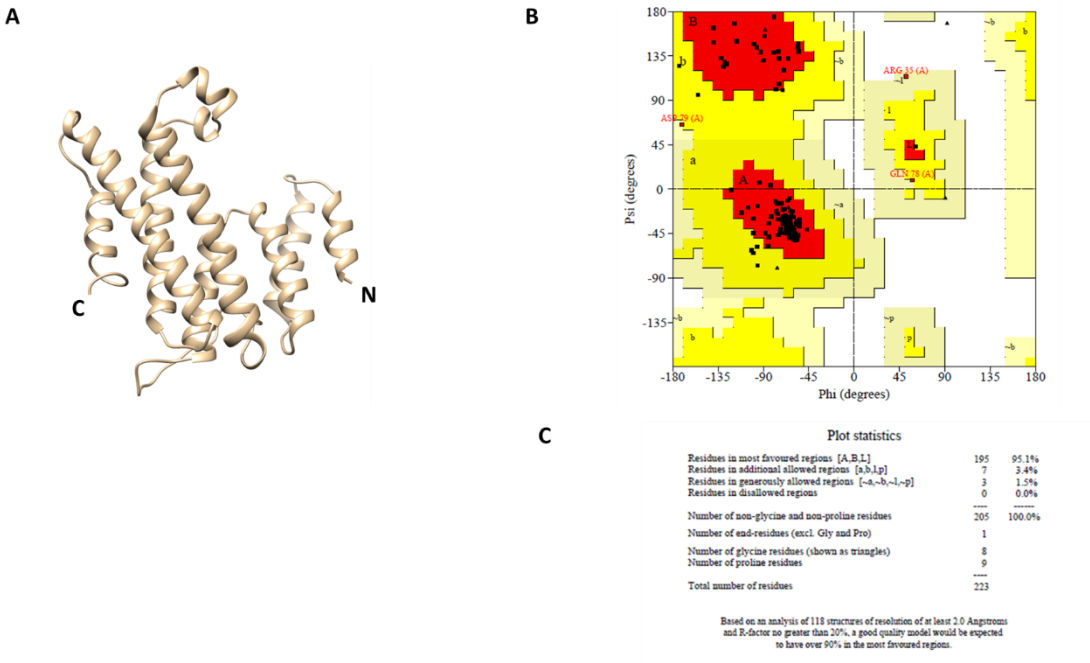


Figure 3.7: (A) Modelled structure of FANCI ARM repeat protein and **(B)** validation by Ramachandran plot and **(C)** Ramachandran plot statistics

3.1.3.7 Molecular dynamics simulation FANCI ARM repeat

To delineate the flexibility and dynamics of the ARM repeat domain and HTH motif, molecular dynamics simulation (MDS) studies for 80 nanoseconds using GROMACS [14, 264-266] were carried out in the protein as well as the folding pattern of HTH-type motif. MDS data were analyzed by plotting RMSD, RMSF, Rg, and SASA (**Figure 3.9**) [84].

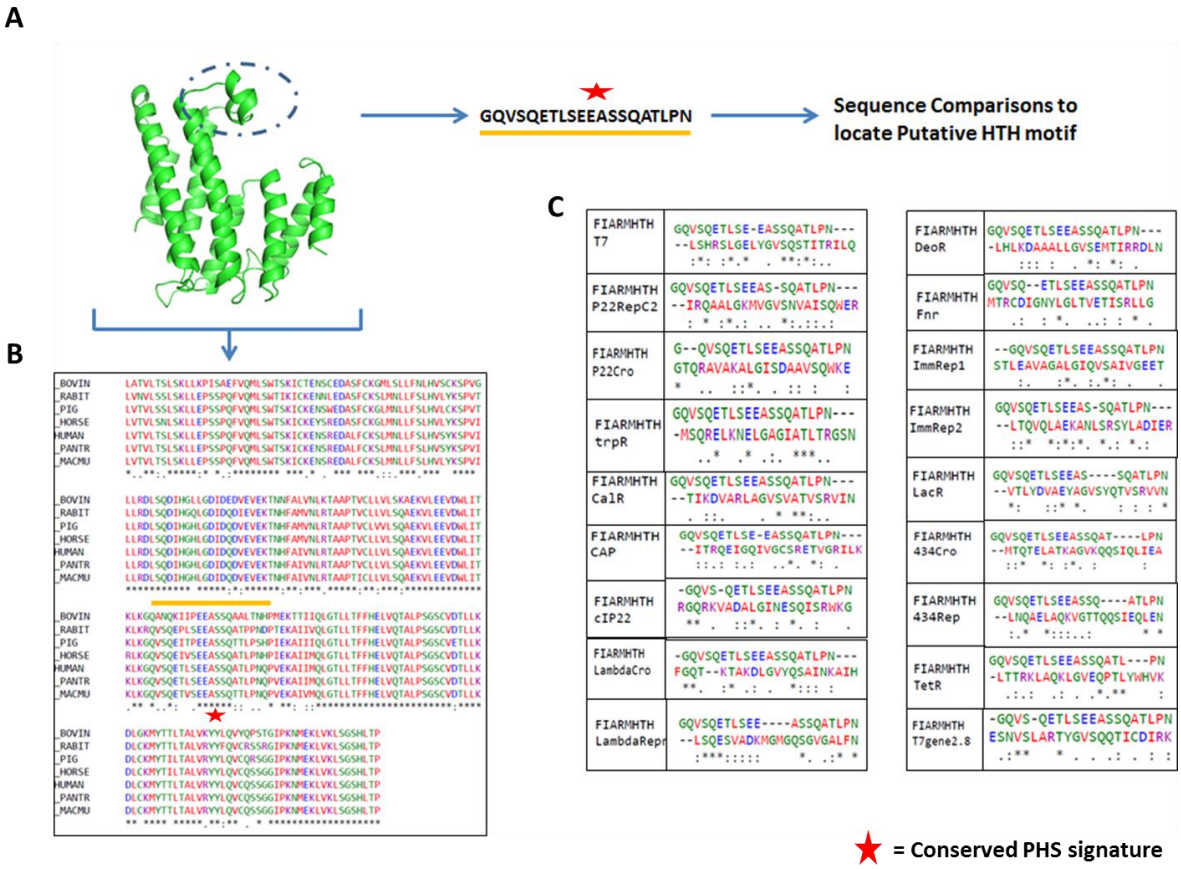


Figure 3.8: (A) Modelled structure of FANCI ARM repeat protein**, (B)** Multiple sequence alignment of FANCI ARM repeat of different mammalian species showing the conserved PHS signature and **(C)** MUSCLE alignment of eighteen different known HTH motif present in "Master Sets"

RMSD profile of ARM repeat shows dynamic behavior with different conformations which stabilizes after 60 ns. Rg fluctuation is another determinant of overall compactness and structural flexibility which suggests the predominant structural rearrangements during simulation. RMSF for C-alpha of ARM repeat domain residues indicates amplitude of fluctuation that unravels the dynamic residual regions [84].

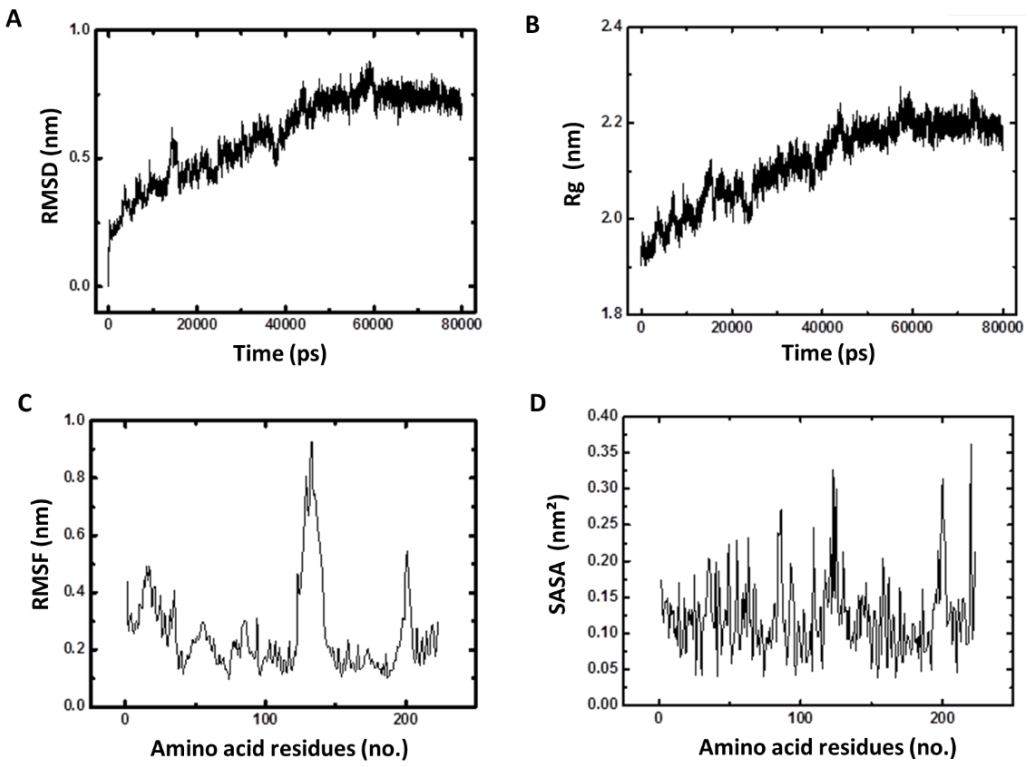


Figure 3.9: Molecular dynamics simulation profile of FANCI ARM repeat **(A)** RMSD profile & **(B)** Radius of gyration (Rg) profile of 80 nanoseconds showing structural transitions up to 60 ns, **(C)** RMSF profile showing large fluctuation of C-terminus and **(D)** Solvent accessible surface area (SASA) profile of h-FANCI ARM repeat protein showing high SASA value at C-terminus

RMSF predicts that C-terminus region of FANCI specifically HTH-type region comprising 124 – 143 amino acids is highly flexible (**Figure 3.9 C**). It is also evident with the projection of eigenvector 1 versus residual RMSF ARM repeat High values of SASA at C-terminus indicates that the HTH-type structural motif showing the high accessible surface area might act as interactions motif in ARM repeat (**Figure 3.9 D**).

Cross-correlation analysis and RMSF sausage plot suggest that HTH structural motif has anti-correlated motion with high fluctuation (**Figure 3.9 A and B**).

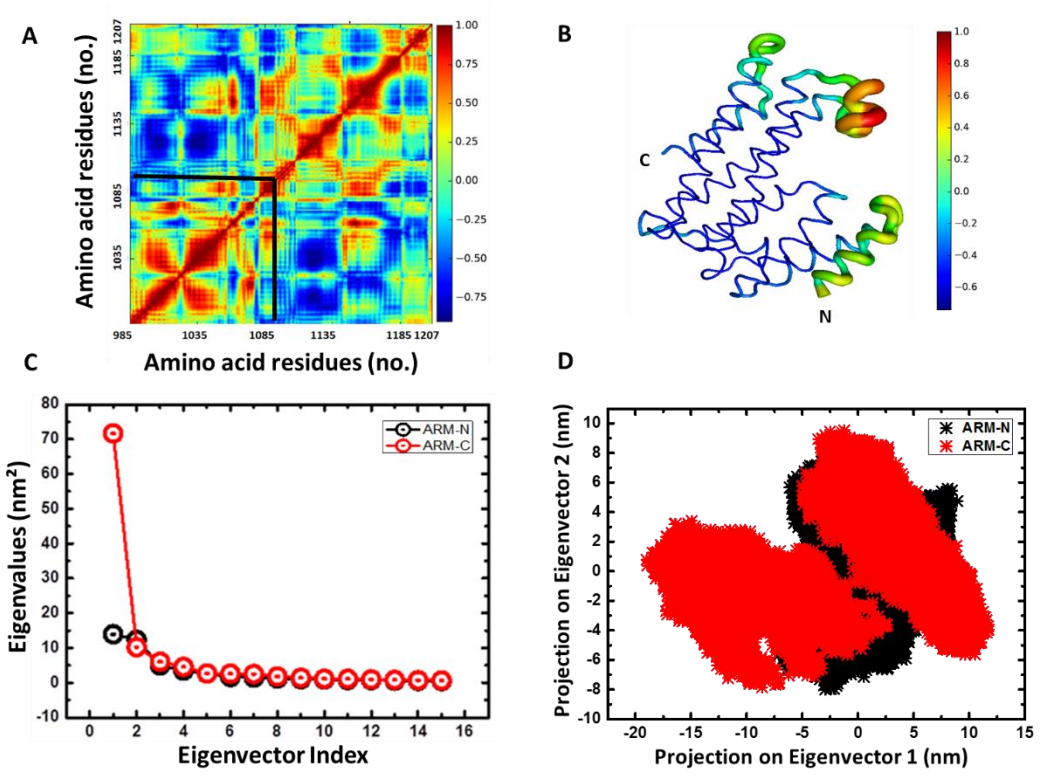


Figure 3.10: (A) Cross-correlation diagonal matrix of h-FANCI ARM repeat, **(B)** RMSF sausage profile of FANCI ARM repeat, **(C)** Scree plot of N-terminus (1-100) and C-terminus (100-223) and **(D**) PCA profile of Eigenvector 1 and 2 of N-terminus (1-100) and C-terminus (100-223)

Principal component analysis (PCA) was performed for the ARM repeats of N-terminus (1-100) and C-terminus (100-223) amino acids to understand the dynamics in essential subspaces. Scree plot revealed that C-terminus having higher eigenvalue than N-terminus that suggest large conformational motion of C-terminus domain (**Figure 3.10 C**) [84].

The trace of covariance matrix values calculated for N-terminus and C-terminus were 64.71 and 131.62 nm² respectively. Therefore, C-terminus comprising HTH-type motif is more dynamic than the N-terminus. Projection of eigenvector 2 on 1 indicates a large periodic tertiary structural transition in C-terminus than the N-terminus of ARM repeat protein (**Figure 3.10 D**). PCA results suggest that protein has large concerted motion due to the presence of HTH-type structural motif. Furthermore to demonstrate cross-

correlation and domain mobility, Gaussian network modelling and normal mode analysis were performed. It has been observed that HTH-type motif at C-terminus part has anti-correlated motion, more dynamicity than N-terminus, and are in opposite direction [84]. Essential Dynamics (ED) results are in agreement with MDS and suggest that structurally dynamic HTH structural motif might be stabilized by binding to DNA during ICL DNA repair [84].

3.1.3.7 Conclusions

FANCI ARM repeat predominantly composed of α-helices and has a Tm of about 36°C. It has HTH-like motif at the C- terminus which has higher dynamic nature than the other region of the protein [84]. Assessment of HTH region suggests that it has motif character. The presence of the good helix former amino acids [318] in putative HTH region also suggests that this dynamic region forms the α- helices. Furthermore, conserved PHS signature suggests HTH motif like character. Docking results (**Chapter 4 Figure 4.2 page no. 145**) indicate that FANCI ARM binds to the major groove of the DNA with conserved residues such as Gln, Ser, Glu and Thr [319, 320] that further corroborates the observation of HTH motif in FANCI ARM protein.

3.2 Structural and Biophysical Properties of FANCD2

3.2.1 Introduction

The FA proteins together with BRCA1 and BRCA2 function cooperatively in the FA/BRCA pathway to repair intercrosslink DNA damage [2]. Activated FA core complex assembles in the nucleus and monoubiquitinates the FANCD2 and FANCI proteins. This monoubiquitinated FANCD2 and FANCI proteins are recruited at DNA damage sites to interact with crucial DNA repair proteins including BRCA1, FANCD1/BRCA2 and RAD5 [3]. In addition, FANCD2 monoubiquitination is necessary for the recruitment of the Fanconi associated nuclease (FAN1) and SLX4/FANCP endonucleases to the DNA damage sites [177]. A Recent report suggested that FANCD2 has a coupling of ubiquitin conjugation to endoplasmic reticulum domain (CUE), an ubiquitin binding domain required for efficient DNA ICL repair [4]. FANCD2 CUE domain encompasses the region from (1-254) amino acids at N-terminus and regulates the FA pathway by correcting the mitomycin c hypersensitivity and localizing the FANCD2 and FANCI proteins at DNA damage site [4]. CUE domain is also responsible for the efficient interactions between FANCD2 and FANCI protein to regulate protein homeostasis at the time of DNA repair [4].

Ubiquitin-binding domains play a pivotal role in protein targeting, cell -cycle control, DNA replication, transcription and DNA repair processes [207, 321]. Due to diverse functions, different ubiquitin binding regions are characterized in proteins such as ubiquitin-associated domains (UBA domains), ubiquitin-interacting motifs (UIMs) and coupling of ubiquitin conjugation to endoplasmic reticulum domain (CUE domains). CUE domains have a role in ubiquitin binding as well as in facilitating intra-molecular monoubiquitination [322]. However, not much has been explored at the structure and function level especially in context to DNA ICL repair or FA pathway.

In the present study, we have focused on important missense mutation Leu231Arg found in CUE domain of FANCD2. The Leu231Arg mutant has been reported to be responsible for the deregulation of DNA ICL repair [90]. The structural and functional aspects of FANCD2 CUE domain *wild-type* and mutant have been explored using different *in-vitro, in-silico* and biophysical approach. It has been found that FANCD2 CUE domain *wild-type* is more stable than the Leu231Arg mutant. Furthermore, FANCD2 CUE domain has the chaperone-like activity which is impaired in the mutant. To our knowledge, this is the first multidisciplinary report which describes the comparison of DNA ICL repair mutant present in FANCD2 CUE domain using structural, dynamics and functional approach.

3.2.2 Materials and Methods

3.2.2.1 Gene cloning, site -directed -mutagenesis, protein expression, and purification

FANCD2 CUE domain comprising 1-254 amino acids was PCR amplified from full-length cDNA of FANCD2 (a generous gift from Markus Grompe, Oregon Health and Sciences University, USA) as a template using forward and reverse primers 5'-GTCGGATCCGAGAACCTGTACTTTCAGGGTCTAGTCACGGTTCTTACCAG-3' & 5'-GTCCTCGAGCTATTAGGGGGTCAGATGAGAACCAG-3', respectively. PCR product of amplified CUE domain was sub-cloned in the pGEX4T-1 vector (Novagen). Mutants Ser222Ala and Leu231Arg were generated by site- directed -mutagenesis using forward and reverse primer 5'-ATCCTAGGGGATGCCCAGCACGCTGAT-3' & 5'-ATCAGCGTGCTGGGCATCCCCTAGGAT-3' and 5'-GTGGGGAAAGAA**CGC**AGTGACCTACTG-3' & 5'-CAGTAGGTCACTGCGTTCTTTCCCCAC-3', respectively. CUE domain and mutants

cloned in pGEX4T-1 were expressed and purified as described in material method **Section 2.1 and 2.2**. Proteins were purified in buffer containing 300mM NaCl, 50 mM Tris, 0.10% Triton-X, 5mM beta-mercaptoethanol, pH 8.50. GST-fusion protein bound on affinity column (GST beads, Qiagen) was further treated with TEV-protease to get the protein in native condition. This native protein was further passed through AKTA FPLC gel filtration column Superdex 75 in buffer containing 300 mM NaCl, 50 mM Tris, 5 mM β-Mercaptoethanol, pH 8.50 to get highly purified protein.

3.2.2.2 Dynamic light scattering (DLS) and mass spectrometry

FPLC purified FANCD2 CUE domain and Ser222Ala, Leu231Arg mutants proteins in buffer (300mM NaCl, 50mM Tris, 5mM β-Mercaptoethanol, pH 8.50) at a concentration of 1 mg/ml were filtered using 0.22 µm pore size filter (Millipore, USA) and degassed at 4°C prior to DLS (Wyatt DynaPro nanostar) and mass spectrometry (MALDI-TOF-TOF) (Bruker Daltonics, Ultraflex II, Germany) experiments. These experiments were performed as described in **Chapter 2**.

3.2.2.3 Circular - Dichroism and Fluorescence Spectroscopy

Secondary structure characterization was performed using CD polarimeter (Jasco J-815, Japan) in the far -UV range (λ= 200nm -260nm). An average of three spectra was recorded for final representation in mean residual ellipticity. The secondary structure was quantitated by K2D3 server (http://cbdm-01.zdv.uni-mainz.de/~andrade/k2d3/). Thermal denaturation spectra were recorded from 10°C to 80°C with 2°C temperature interval and data fitting was done as described in **Chapter 2** [309].

Micro-environment of tryptophan (intrinsic fluorophore) at the hydrophobic core of the protein was monitored by fluorescence spectrophotometer (Horiba-Fluorolog, Japan) at the excitation wavelength of λ=295nm. The emission spectra were recorded from λ =310-450nm. For thermal unfolding experiments, spectra were recorded in temperature dependent manner using 10 µM protein as described earlier [49, 309].

3.2.2.4 Model building, molecular dynamics simulation, DSSP, normal mode analysis and principle component analysis

FANCD2 CUE domain (1-254) amino acids was retrieved from UniProtKB [300] (Uniprot ID; Q9BXW9), *wild-type* and mutant sequences were submitted to Robetta server (http://robetta.bakerlab.org/) for 3D model building. Stereo chemical refinement of Ramachandran outliers in the models was performed by ModLoop server [302]. Further, models were validated by Swiss model workspace encompassing the package of Anolea, DFire, QMEAN, Gromos, DSSP, Promotif and ProCheck (http://swissmodel.expasy.org/workspace/). Validated models were used for molecular dynamics simulation by GROMACS 4.5.5 [84, 309]. Molecular modelling and simulation were performed as described in **chapter 2**.

3.2.3 Results and Discussion

3.2.3.1 Gene cloning, site-directed mutagenesis, protein expression and purification

Selected potential clones when digested with the BamH1 and Xho1 restriction enzymes, showed the insert release of appropriate size (**Figure 3.2.1**).

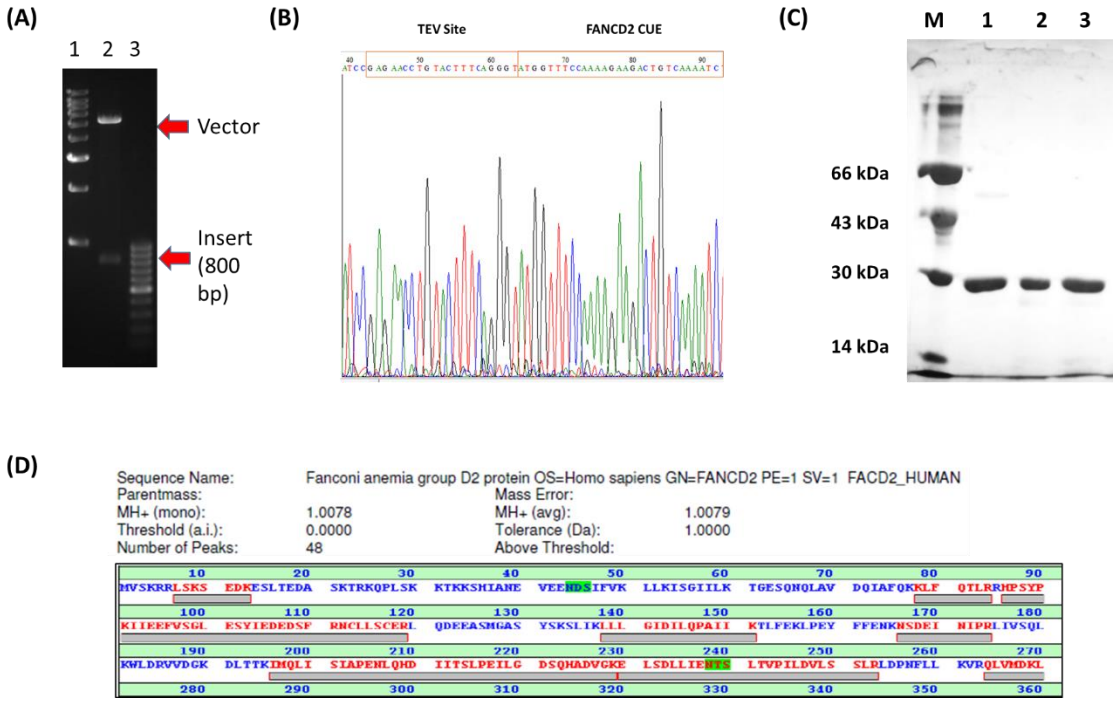


Figure 3.2.1: (A) Restriction enzyme digestion of FANCD2 CUE Domain cloned in pGEX4T-1 vector showing insert release of right size **(B)** DNA Sequencing chromatogram of FANCD2 CUE Domain cloned in pGEX4T-1 exhibiting TEV site and FANCD2 CUE Domain (different colors show the nucleotides; Adenine=Green, Guanine=Black, Cytosine=Blue and Thymine=Red) **(C)** FANCD2 CUE domain purified protein, Lane: M= Protein Marker, 1= FANCD2 CUE *wild type*, 2= FANCD2 CUE Ser222Ala and 3= FANCD2 CUE Leu231Arg **(D)** Peptide mass fingerprinting profile for the identification of FANCD2 CUE domain using mass spectrometry (grey color shows the matching of captured peptides)

DNA sequencing results confirmed the presence of ligated gene of interest in the vector with a correct reading frame of the codon sequence. FANCD2 CUE domain *wild-type*, Ser222Ala, and Leu231Arg mutants were FPLC purified and subjected to SDS-PAGE analysis for purity assessment **(Figure 3.2.1)**. Further, the identity of proteins was confirmed by peptide mass fingerprinting **(Figure 3.2.1)**.

3.2.3.2 Oligomeric behavior of FANCI CUE domain and Ser222Ala, Leu231Arg mutants

To understand the oligomeric behavior, effective radii and polydispersity of the proteins, Dynamic Light Scattering (DLS) were performed. It has been observed that FANCD2 CUE domain *wild-type* protein exists in predominantly two different hydrodynamic radii such as 2-3 nm and 30-90 nm with the intensity of >60% and ~40% respectively. However, Ser222Ala and Leu231Arg mutant proteins have completely different hydrodynamic radii pattern compared to *wild type*.

The hydrodynamic radii, for Ser222Ala mutant is: 3-5 nm and (~20% intensity), 110-190 nm (~60% intensity) and 800 nm (~10% intensity); and for Leu231Arg mutant is: 5-6 nm (~80% intensity), 55-80 nm (~10% intensity) **(Figure 3.2.2)**. The FANCD2 CUE domain *wild-type* shows lower hydrodynamic radii with higher intensity than the Ser222Ala and Leu231Arg mutants, suggesting the compact structure of *wild-type* protein. However, Ser222Ala and Leu231Arg mutations affect the overall packing of the protein. Further, mass spectrometry was performed to understand the monomeric intensity at a respective molecular weight of the FANCD2 CUE domain *wild-type* and mutant proteins.

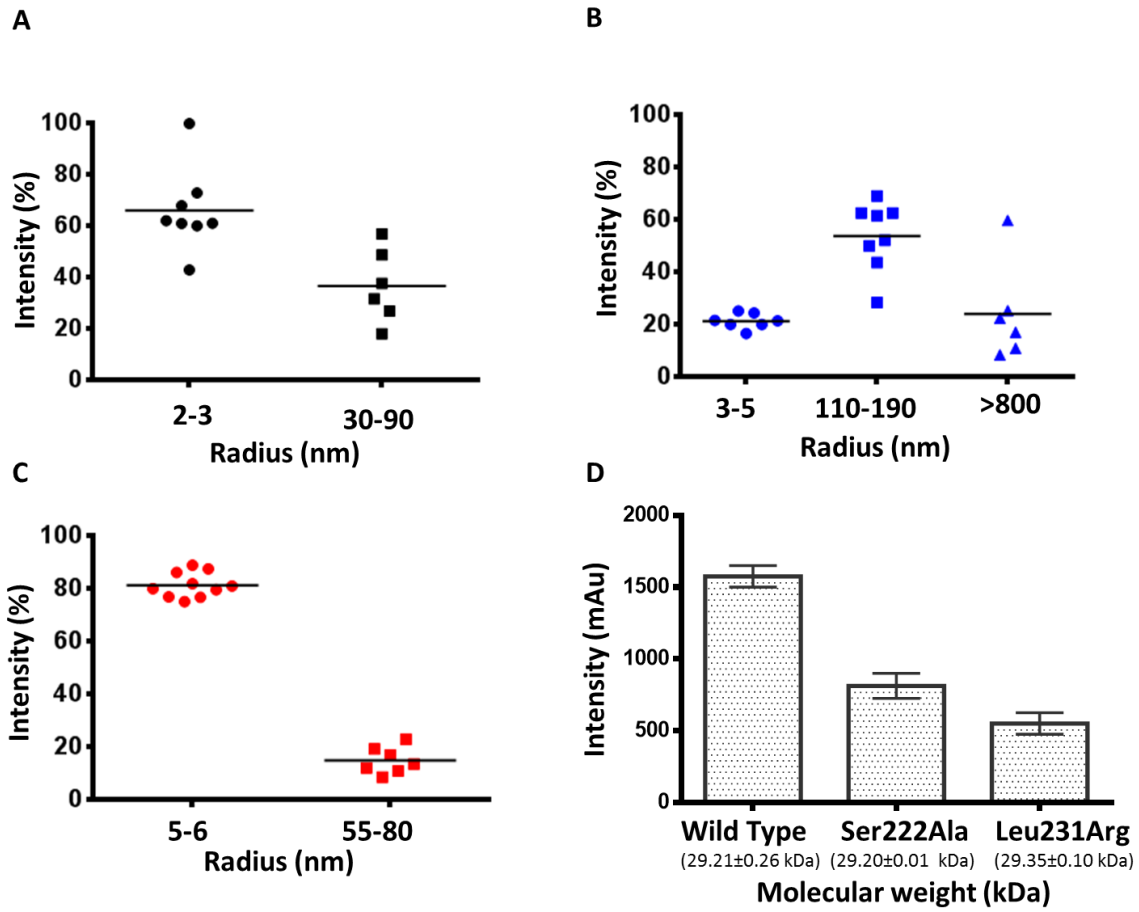


Figure 3.2.2; Dynamic light scattering results for **(A)** FANCD2 CUE domain *wild-type* **(B)** Ser222Ala and **(C)** Leu231Arg, each dot was showing the single scan of DLS measurements and ≥7 scans were plotted in distribution graph **(D)** Mass spectrometry (intensity vs molecular weight) comparative profile of *wild type* and mutants

The mass spectrometry results showed that the FANCD2 CUE domain *wild -type* protein has higher intensity at 29.12±0.26 kDa of monomeric size compared to Ser222Ala, Leu231Arg at 29.20±0.01 kDa and 29.35±0.10 kDa respectively. These results suggest that mutations affect the overall compactness and monomeric nature of the protein **(Figure 3.2.2)**.

3.2.3.3 Structure Characterization of FANCD2 CUE domain and Leu231Arg mutant

Circular-Dichroism (CD) spectroscopy was performed to estimate the secondary structure and protein stability. The Circular-Dichroism (CD) results revealed that the molar ellipticity of FANCD2 CUE domain *wild-type* and Leu231Arg mutant are different at λ=208 nm and λ=222 nm wavelength respectively, that suggests a change in the α-helical property of mutant protein **(Figure 3.2.3)**. Furthermore, the α-helical content for FANCD2 CUE domain *wild-type* and Leu231Arg mutant were found 70.97% and 62.30%, respectively, which indicates a reduction in α-helical content of Leu231Arg mutant protein.

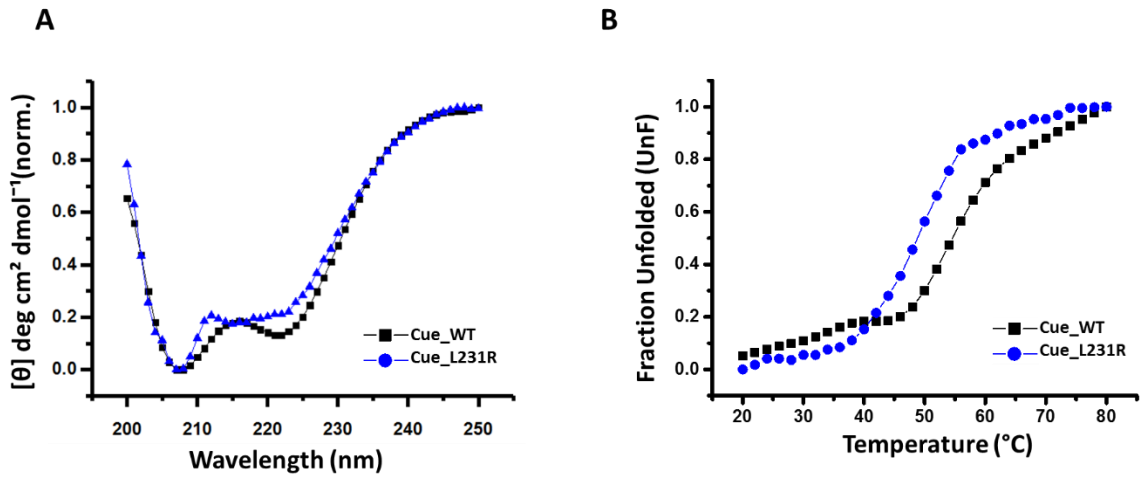


Figure 3.2.3; **Structure and stability profile: (A)** Far –UV, CD spectra of *wild-type* CUE and Leu231Arg mutant **(B)** Secondary structure stability profile of *wild type* CUE and Leu231Arg mutant

Thermodynamic stability of FANCD2 CUE domain *wild-type* and mutant protein were calculated using temperature dependent unfolding by circular dichroism spectroscopy. Melting temperature (Tm) of FANCD2 CUE domain *wild-type* was 54.69±1.60 °C as compared to Leu231Arg at 48.19±1.50 °C indicating that Leu231Arg mutant protein is less

stable than the *wild-type* (**Figure 3.2.3)**. However, due to poly-dispersive nature of Ser222Ala mutant, α- helical content and Tm could not be calculated.

Fluorescence spectroscopy of FANCD2 CUE domain *wild-type* and Leu231Arg mutant were performed to explore the tertiary structure conformations. A decrease in fluorescence intensity in the Leu231Arg mutant was observed compared to the FANCD2 CUE domain *wild-type* **(Figure 3.2.4)**.

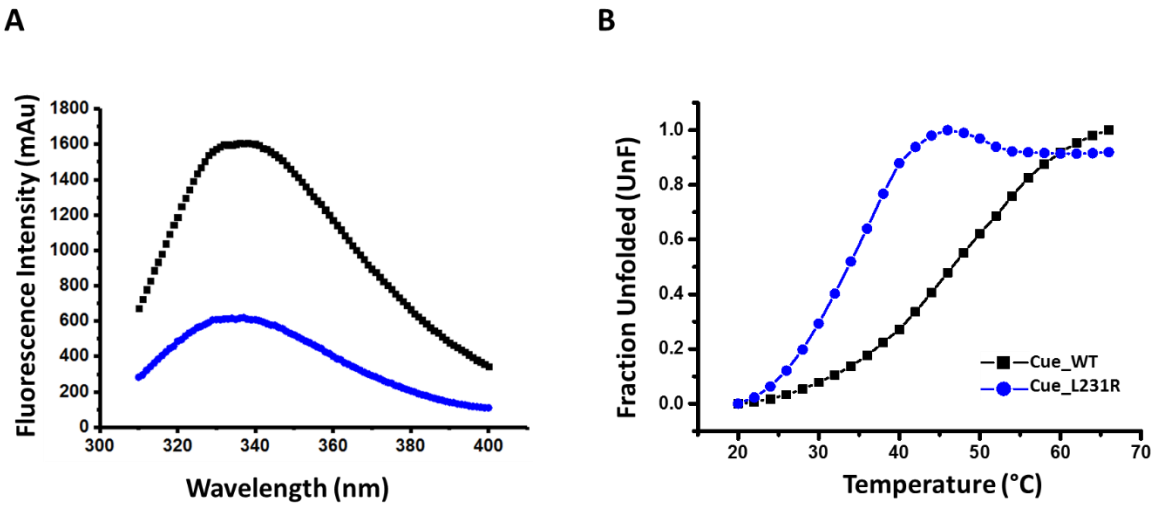


Figure 3.2.4; **Structure and stability profile: (A)** Fluorescence spectroscopy profile for tertiary structure assessment and folding of FANCD2 CUE domain *wild-type* and Leu231Arg mutant **(B)** Tertiary structure stability profile for FANCD2 CUE domain *wild-type* and Leu231Arg mutant

This loss in fluorescence intensity suggests that the intrinsic fluorophore tryptophan was exposed to solvent due to the perturbed tertiary structure. Furthermore, temperature dependent tertiary structure unfolding of FANCD2 CUE domain *wild-type* and Leu231Arg mutant protein were performed. The tertiary structure Tm for FANCD2 CUE domain *wild-type* and Leu231Arg mutant were 48.97±1.60 °C and 33.31±1.70 °C, respectively. The

missing pre-transition was observed in Leu231Arg sigmoidal curve suggesting that Leu231Arg mutant protein was partially unfolded **(Figure 3.2.4)**. These results are corroborating the DLS experiments in which an increase in hydrodynamic radii of Leu231Arg mutant was observed. Therefore, these findings demonstrate that mutant proteins have impairment in secondary and tertiary structure integrity as well as stability.

3.2.3.4 *In-silico* studies of FANCD2 CUE domain and Leu231Arg mutant

To analyze the effect of mutation at the molecular level, homology modelling and molecular simulation based approach were employed. To understand the structure and dynamics, molecular dynamics simulation (MDS) of 50 nanoseconds (ns) was performed for FANCD2 CUE domain and Leu231Arg mutant. The root means square deviation (RMSD) of trajectories revealed lesser deviation (0.20 to 0.40 nm) for FANCD2 CUE domain *wild-type* compared to Leu231Arg (0.20 to 0.80 nm) mutant **(Figure 3.2.5)**. In the root mean square fluctuation (RMSF) analysis, unlike FANCD2 CUE domain *wild-type*, the Leu231Arg revealed that an increase in flexibility from 0 to 50 and from 100 to 150 amino acid residues. However, a loss in flexibility was observed in Leu231Arg from 150 to 200 amino acid region in contrast to FANCD2 CUE domain *wild-type* **(Figure 3.2.5).**

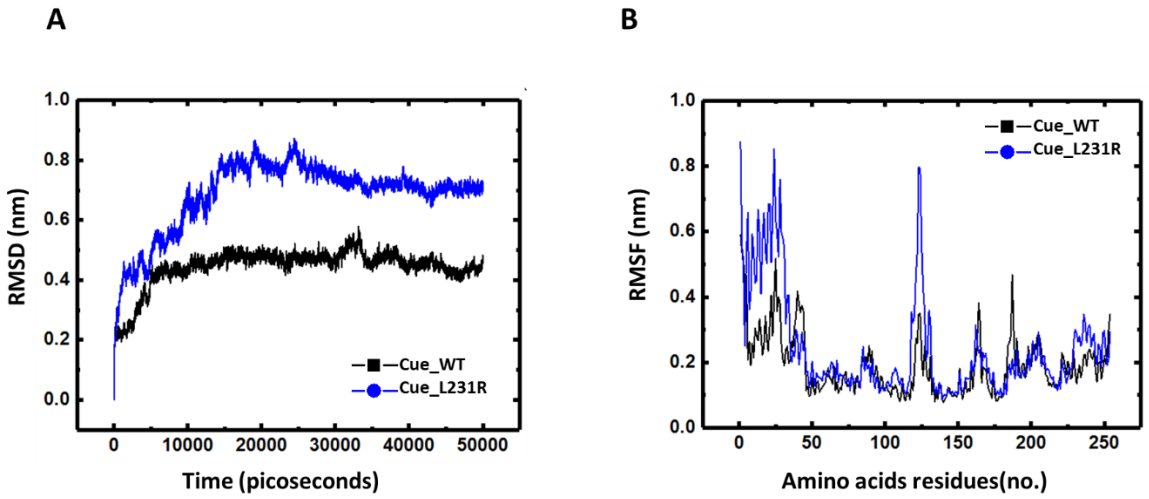


Figure 3.2.5- Molecular dynamics simulation profile of FANCD2 CUE domain *wild-type* and Leu231Arg mutant: (A) Comparative RMSD data shows FANCD2 CUE domain *wild-type* is able to attain stable conformation unlike mutant **(B)** Comparative RMSF data suggests higher residual fluctuation in mutant

Secondary structure dynamics of FANCD2 CUE domain *wild-type* and Leu231Arg mutant were characterized in a time-dependent manner using Dictionary of Secondary Structure of Protein (DSSP). It has been observed that large helix to β-bend and turn transition is predominant beyond 2 ns in Leu231Arg between 220 to 240 amino acids region which includes the point mutation site. However, FANCD2 CUE domain *wild-type* shows α-helices consistently from 0 ns to 50 ns, unlike Leu231Arg mutant **(Figure 3.2.6)**.

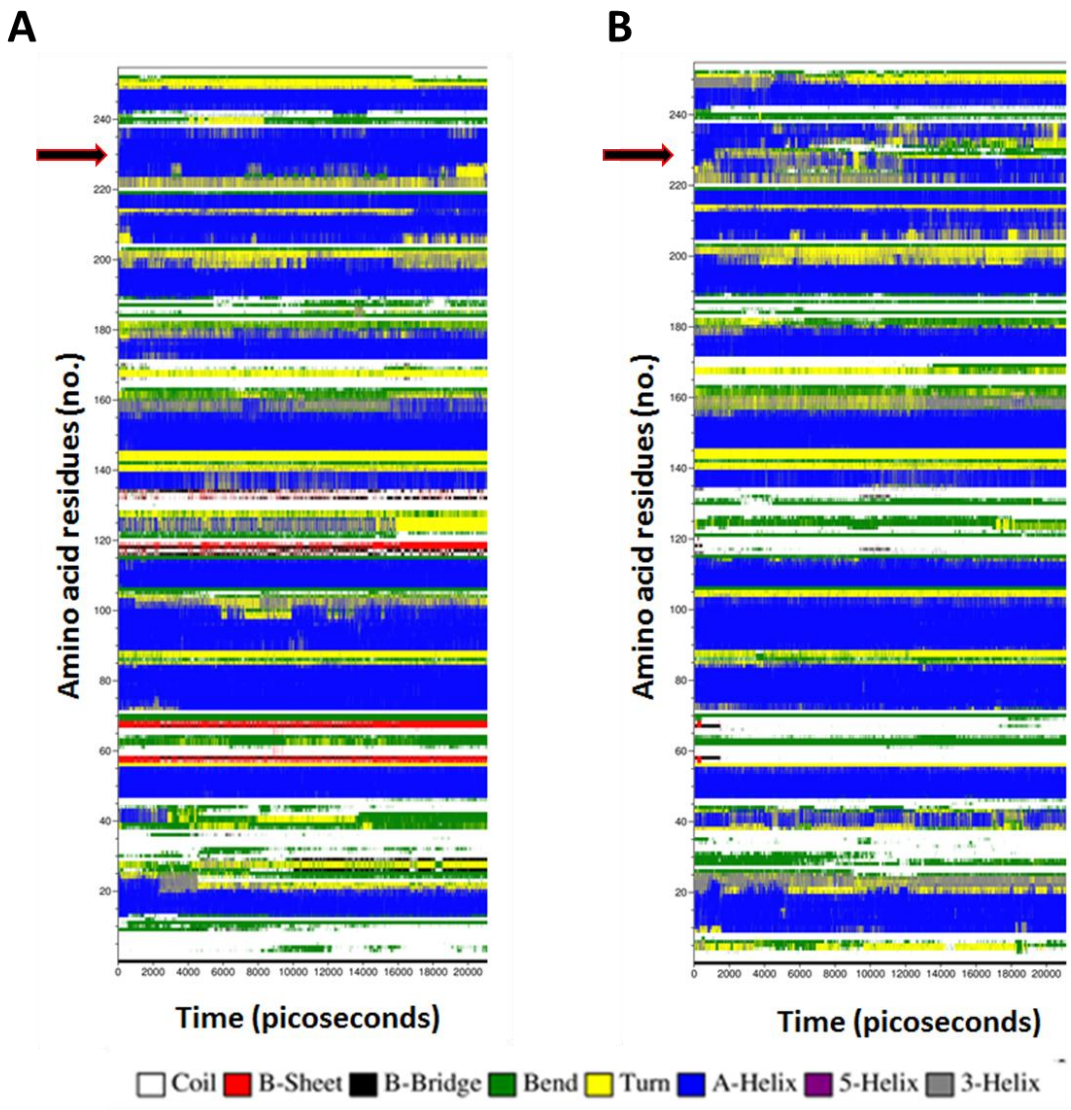


Figure 3.2.6; **DSSP profile: (A)** FANCD2 CUE domain *wild-type*; secondary structure transition was stable as α-helices are consistent throughout the simulation at the mutation site **(B)** Leu231Arg; mixed population of coil and beta bend was observed, marked with arrow

FANCD2 CUE domain *wild-type* shows α-helices consistently from 0 ns to 50 ns between 220 to 240 amino acid residues (aa), which further corroborate with the CD results. Furthermore, the correlated dynamics of Leu231Arg suggest an increase in anti-correlated motion in the amino acids region from 50 to 150 aa and 220 to 240 aa, in contrast to *wild-type* CUE **(Figure 3.2.7)**. In Leu231Arg mutant there is an increase in positive correlation and negative correlation from 50 to 175 amino acid region. Furthermore, increase in

positive correlation and negative correlation in Leu231Arg mutant from 200 to 250 amino acids region were observed (this region also include the mutation region Leu231Arg) **(Figure 3.2.7)**. These transformations reflect the differences in the internal dynamics of mutant Leu231Arg in comparison to FANCD2 CUE domain *wild-type* protein.

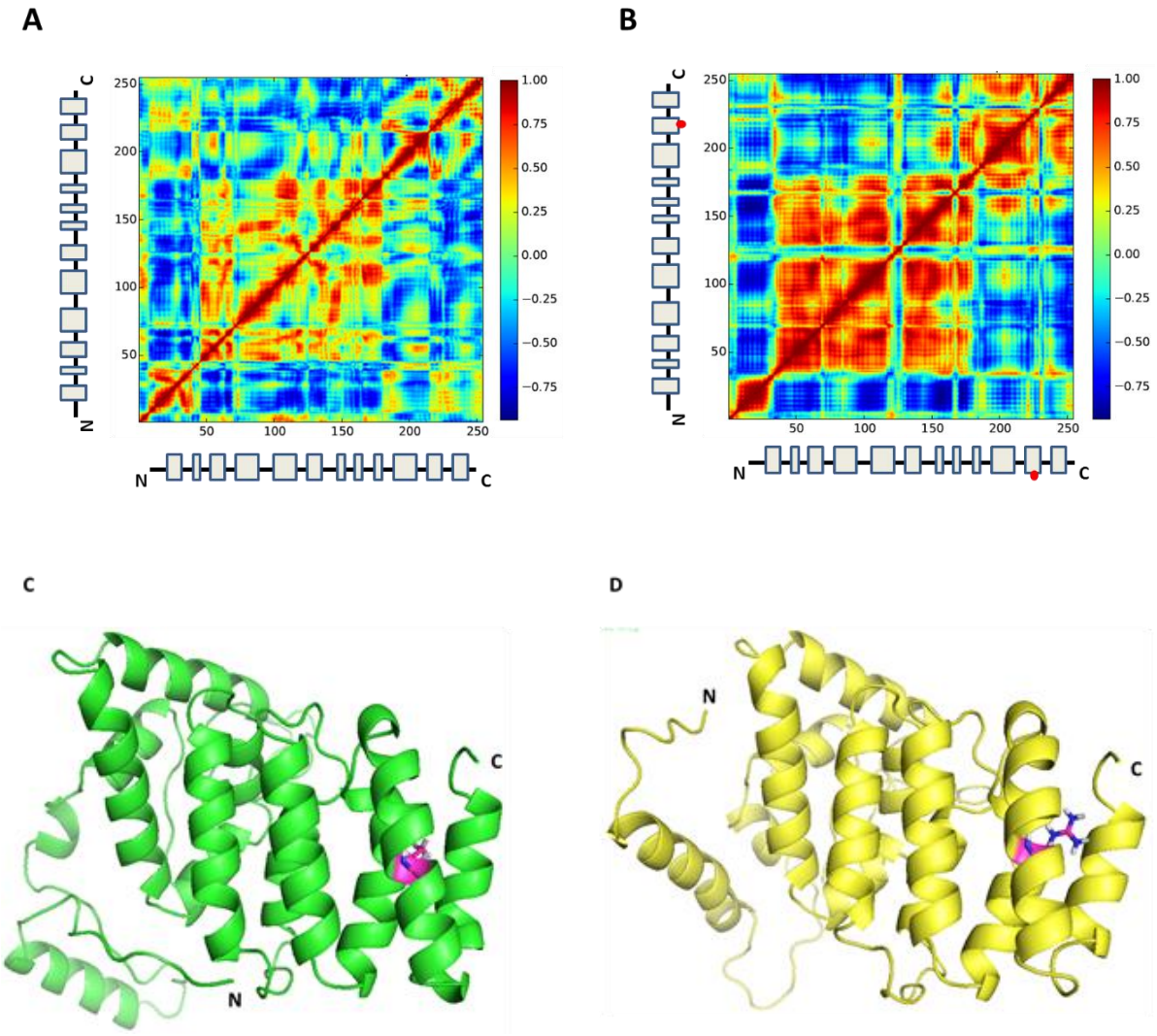


Figure 3.2.7; Cross-correlation profile of **(A)** FANCD2 CUE domain *wild-type* and **(B)** Leu231Arg mutant; **(C)** Molecular model FANCD2 CUE domain *wild-type* and **(D)** Leu231Arg mutant

Furthermore, cross-correlation and principal component analysis (PCA) were performed to understand periodic transitions in tertiary structure and dynamics in essential subspaces. Scree plot [323] revealed that FANCD2 CUE domain *wild-type* had lower eigenvalue (~50 nm²) than Leu231Arg (~225 nm²) **(Figure 3.2.8)**. The trace of covariance matrix values calculated for the FANCD2 CUE domain *wild-type* and Leu231Arg mutant are 8.82 and 19.57 nm² respectively. In PCA, the projection of eigenvector 2 on 1 indicates a large periodic tertiary structural transition in Leu231Arg mutant protein compared to FANCD2 CUE domain *wild-type* **(Figure 3.2.8)**.

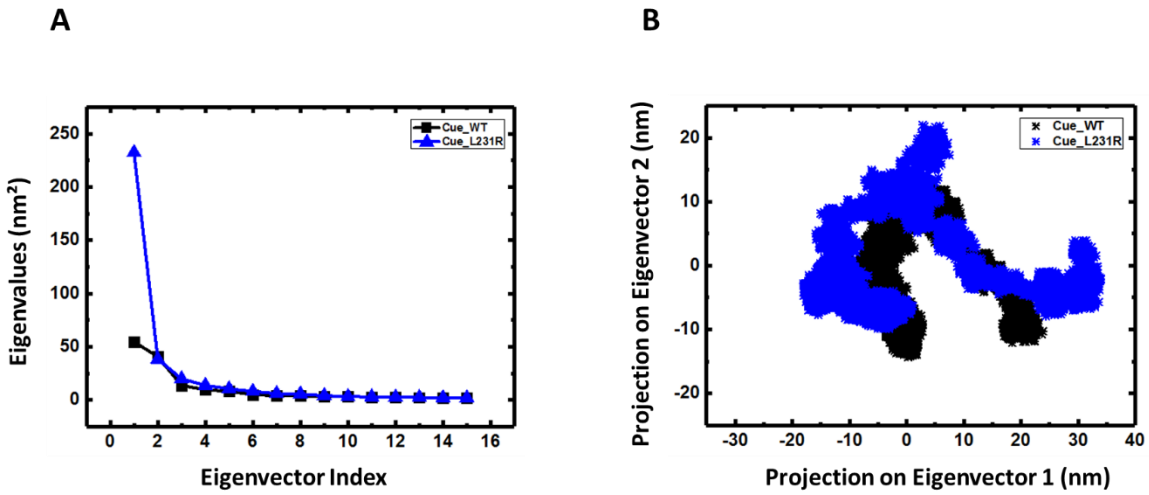


Figure 3.2.8; **Molecular dynamics simulation profile of FANCD2 CUE domain *wild-type* and Leu231Arg mutant: (A)** Scree plot showing comparably higher eigenvalues in Leu231Arg mutant compare to FANCD2 CUE domain *wild-type* and **(B)** PCA graph suggest large tertiary structure transitions predominant in mutant in contrast to FANCD2 CUE domain *wild-type*

This suggests that mutant protein has large concerted motion. Hence, to our conclusion FANCD2 CUE domain *wild-type* has maintained its structural integrity, unlike Leu231Arg mutant.

3.2.3.4 Conclusion

FANCD2 protein plays a key role in DNA ICL repair. However, detailed structure and associated function of the protein is largely unknown. Among FA patients, one missense mutation Leu231Arg is present in newly identified CUE domain, whereas Leu231 is in evolutionary conserved region. The Leu231Arg mutation is located in the α-helices of the FANCD2 CUE domain. This mutation affects the polydispersity, monomeric behavior, α-helical propensity, tertiary structure conformation and thermodynamic stability. Molecular modeling, simulation, and PCA data shows the perturbed behavior of the mutant protein. Two simulations, each of 50 ns, showing relative deviations of structural transitions in a time-dependent manner. The cumulative results obtained from molecular dynamics, normal mode analysis and principal component analysis studies suggest that Leu232Arg mutant was not able to attain stable structure conformation, compared to the FANCD2 CUE domain *wild type*.

Chapter 4

Functional Characterization of FANCI and FANCD2

4.1 Functional Characterization of FANCI

4.1.1 Introduction

Fanconi anemia pathway protein FANCI facilitates the repair of replication stalling DNA lesions [1, 5, 6]. FANCI protein is phosphorylated and monoubiquitinated during the ICL DNA damage response. After the posttranslational modification FANCI protein dissociates from its heterodimeric I-D complex state (FANCI-FANCD2) [3]. An overview of FANCI has been described in **Chapter 1**. In brief, ARM repeats are functionally diverse and evolutionarily conserved domain that plays a key role in protein-protein and protein-DNA interactions [3, 9, 294, 324, 325]. FANCI C-terminus region also has the binding ability to different forms of DNA such as double-stranded DNA, Holliday junction DNA, and Y-shaped DNA [83]. Here, we are reporting the functional aspects of FANCI ARM repeat by utilizing the techniques such as electrophoretic mobility shift assay, DSSP analysis, and protein-DNA docking. To our knowledge, the results presented here provide the first information on DNA binding studies of the FANCI ARM repeat. These findings may offer intriguing possibilities to unravel the ARM repeat function that would help to understand the molecular mechanism of DNA ICL repair.

4.1.2 Materials and Methods

4.1.2.1 DNA binding by electrophoretic mobility shift assay

DNA binding activity of FANCI ARM repeat protein was determined using electrophoretic mobility shift assay (EMSA) [326]. 82 long random sequence oligonucleotide was used as dsDNA substrate made by annealing it with its complementary strand (**Table 4.1**) [84].

Table 4.1 Radioactive labeled [^{32}P] oligonucleotide sequence used in EMSA study

1.	82F	5'GAATTCGGTGCGCATAATGTATATTATGTTAAATCATGTCCCTGCCCCAATATAAACCAAGCGTATGCAGTAAGCTTCGATC3'	EMSA
2.	82R	5'GATCGAAGCTTACTGCATACGCTTGGTTTATATTGGGGCAGGGACATGATTTAACATAATATACATTATGCGCACCGAATTC3'	EMSA

The dsDNA was labeled with [^{32}P] γATP using polynucleotide kinase and purified by G-25 column (GE Healthcare, Sweden). The 0.20 picomole of labeled probe (dsDNA) was incubated with increasing concentrations of FANCI ARM repeat protein in 10 µl of reaction mixture containing 35 mM Tris-HCL, pH 7.50, 2.50mM $MgCl_2$, 25mM KCl, 1mM ATP and 1mM DTT for 20 min at 37°C. Products of radio-labelled DNA and proteins were analyzed on 6% native polyacrylamide gel, and signals were recorded by autoradiography. DNA band intensity both in free form and bound to the protein was quantified using GelQuant software (http://biochemlabsolutions.com/GelQuantNET.html). The amount of DNA bound to the protein was divided by total DNA taken to obtain the fraction of bound DNA. The percent bound fractions of DNA were plotted against protein concentration using GraphPad Prism 5 (http://www.graphpad.com/scientific-software/prism/). The Kd for curve fitting of the individual plot was determined by the software designed on the principle of the least squares method using formula Y= Bmax *[X]/Kd+[X], where [X] is the protein concentration and Y is the bound fraction.

4.1.2.2 DSSP analysis

MD trajectory of 80 nanoseconds of the molecular dynamics simulation was subjected to DSSP analysis [307] in GROMACS 4.5.5 package [266] using do_dssp reads. DSSP computes the secondary structure in different time frame.

4.1.2.3 Docking studies

Molecular docking studies were performed using HADDOCK server [17, 245, 246]. The structure of the DNA dodecamer (PDB ID: 1BNA) was downloaded from the protein data bank (http://www.rcsb.org./pdb). Docking was performed as described in **Chapter 2.**

4.1.3 Results and Discussion

Considering the importance of FANCI in DNA inter-crosslink (DNA ICL) repair mechanism, the functional domain of FANCI, Armadillo (ARM) repeat protein was purified. The multidisciplinary *in-vitro* and *in-silico* approaches were carried out to understand the DNA binding ability of the FANCI ARM repeat. The FANCI protein is a leucine-rich protein (LRPs) which mediates protein-protein and protein-DNA interactions. LRPs are generally composed of ARM, HEAT, leucine- rich repeat and leucine zipper [20, 115]. FANCI harbours ARM repeat, EDGE motif and nuclear localization signal at C-terminus [3]. However, the molecular mechanism associated to DNA binding of FANCI C-terminus specifically ARM repeat is still not well understood. Longerich *et. al.* reported that C-terminus of FANCI have similar DNA binding activity as the full length of FANCI and binds preferentially to double stranded DNA [83]. Structural alignment of FANCI ARM repeat suggests the profound similarity with elongation factors and transcription factors [84]. In general, elongation factors and transcription factors are known to involve in direct DNA binding and regulating crucial processes like replication and transcription [327-329]. These factors contain DNA binding elements such as helix-turn-helix (HTH) or helix-loop-helix motif [319, 320, 330, 331]. Sequence alignment from 20 master sets of different known HTH motifs [29], have revealed that FANCI ARM repeat has sequence

similarity with these motifs. Furthermore, it has conserved PHS (P= any charged amino acid mostly glutamate, H= hydrophobic amino acid and S= small amino acid residue) signature known to present in HTH motif [27]. Molecular and essential dynamics demonstrate that HTH-type motif present in ARM repeat at C-terminus has high dynamicity and anti-correlated motion. In addition, *in-silico* docking analysis has revealed that HTH-type motif binds to the major groove of DNA [84].

In this study, we are reporting that ARM repeat has structurally dynamic region majorly composed of α-helices and binding to major groove of DNA suggest that FANCI ARM repeat harbors HTH structural motif. Molecular modelling, dynamics, and docking studies are in concordance with *in-vitro* results derived from limited proteolysis, sequence and structural alignment [317]. Importantly, the helix propensity [318] of the amino acids sequences of HTH motif also revealed the presence of helix former amino acids such as Ala, Glu, Gln, and Ser [84]. It is well known that the HTH structural motif generally binds to the major groove of DNA and conserved amino acids such as Glu, Gln and Ser are at the binding interface between HTH motif and DNA [27, 28, 330, 332].

4.1.3.1 FANCI ARM repeat binds to double strand DNA

DNA binding activity of FANCI ARM repeat was monitored with radiolabeled double-stranded DNA substrates using electrophoretic mobility shift assay (EMSA). The observed dissociation constant (Kd) value for FANCI ARM repeat was 3.96±1.71 µM (**Figure 4.1)**. The Kd value in the range of µM concentration suggests the greater affinity of FANCI ARM repeat to double strand DNA.

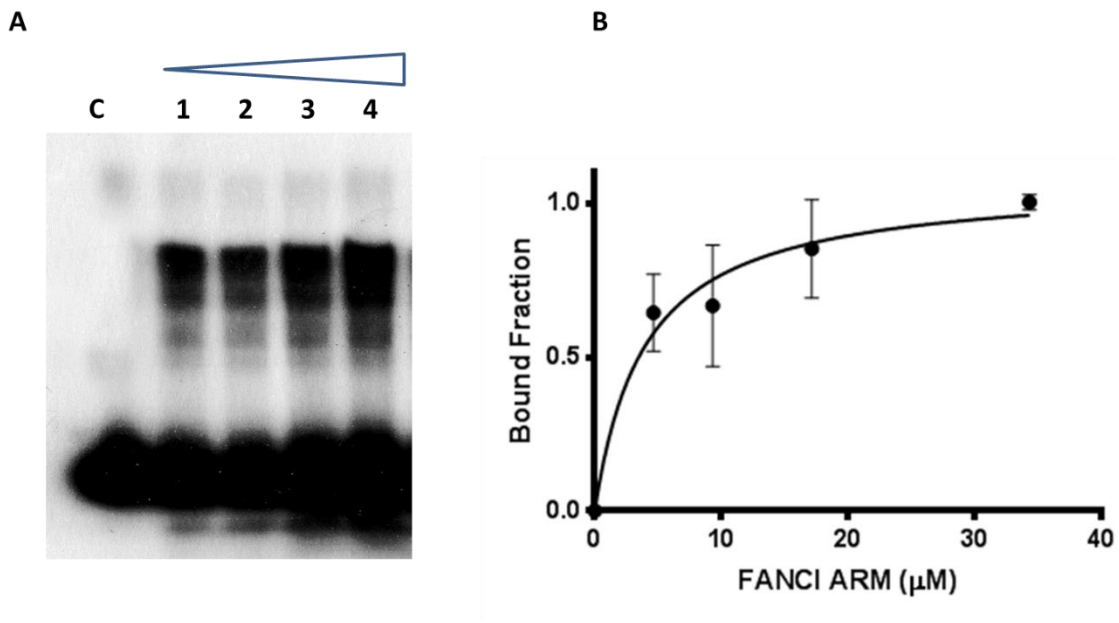
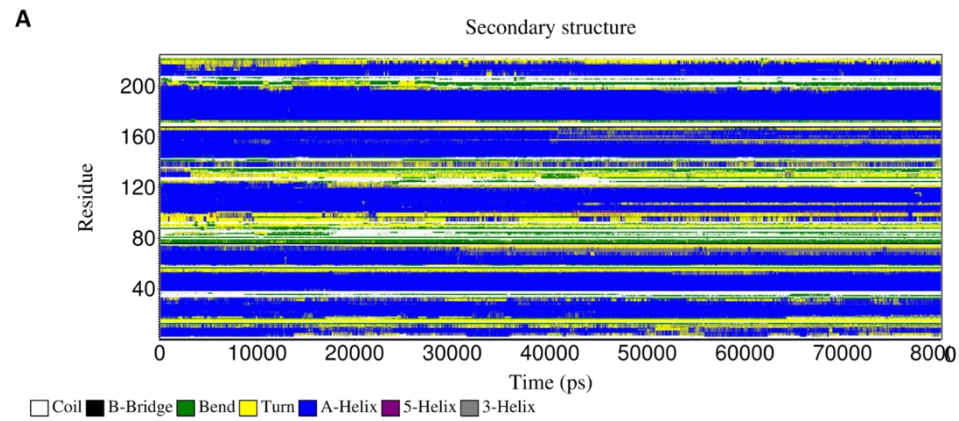
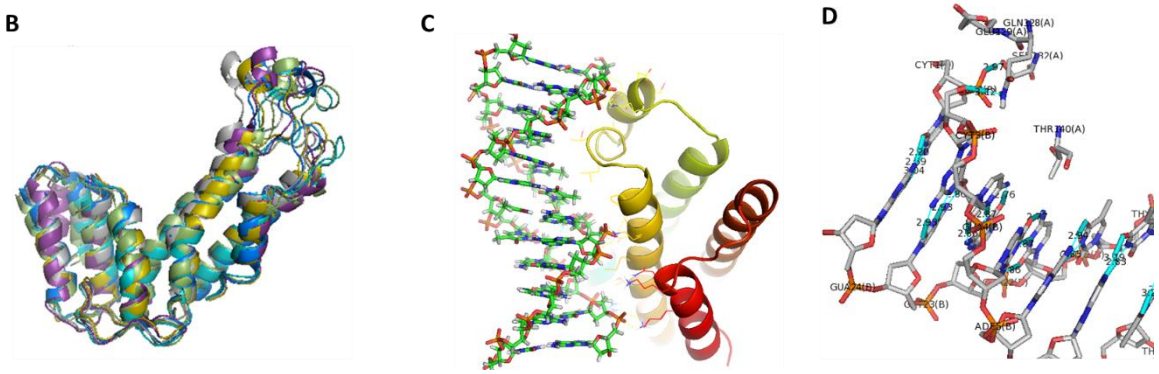


Figure 4.1: (A) DNA binding analysis of FANCI ARM repeat with DNA using autoradiography (C=Control; untreated probe; 1, 2, 3, & 4 with increasing concentration of protein 4.60μM, 9.29μM, 17.14μM and 34.29μM respectively) **(B)** Graph plot of bound fraction of protein to DNA

4.1.3.2 Characterization and docking analysis of HTH-structural motif of FANCI ARM repeat

Molecular dynamics studies of FANCI ARM repeat indicate that HTH structural motif is highly flexible (as described in **Chapter 3**). The secondary structure of ARM repeat was characterized using a dictionary of the secondary structure of the protein (DSSP). It has been observed that large helix turn to loop transition is predominant in the HTH motif which also corroborates well with the results obtained from structural alignment over the extracted structures from the trajectory of different time points [84].





E

Cluster 1

HADDOCK score	-52.1 +/- 5.9
Cluster size	85
RMSD from the overall lowest-energy structure	1.5 +/- 0.9
Van der Waals energy	-48.4 +/- 5.2

http://milou.science.uu.nl/enmr/serviceresults/HADDOCK/4737591987/run13/

4/16/2016 HADDOCK webserver

Electrostatic energy	-443.0 +/- 34.9
Desolvation energy	37.4 +/- 1.7
Restraints violation energy	475.6 +/- 52.02
Buried Surface Area	1720.8 +/- 126.6
Z-Score	-1.8

Figure 4.2 (A) DSSP profile of FANCI ARM repeat protein, **(B)** Superimposed structures at different time points (10 ns to 40 ns) of trajectories showing large structural rearrangement in HTH-type motif **(C)** Docking profile of FANCI ARM repeat with DNA dodecamer shows

major groove DNA binding mode of putative HTH-type motif interactions and **(D)** Intermolecular interface (E) Docking results from HADDOCK

DSSP analysis of the 80 ns simulation, suggest that only HTH structural region showing unfolding at 15 ns, and unable to form a stable structure in the due course of simulation (**Figure 4.2 A and B**).

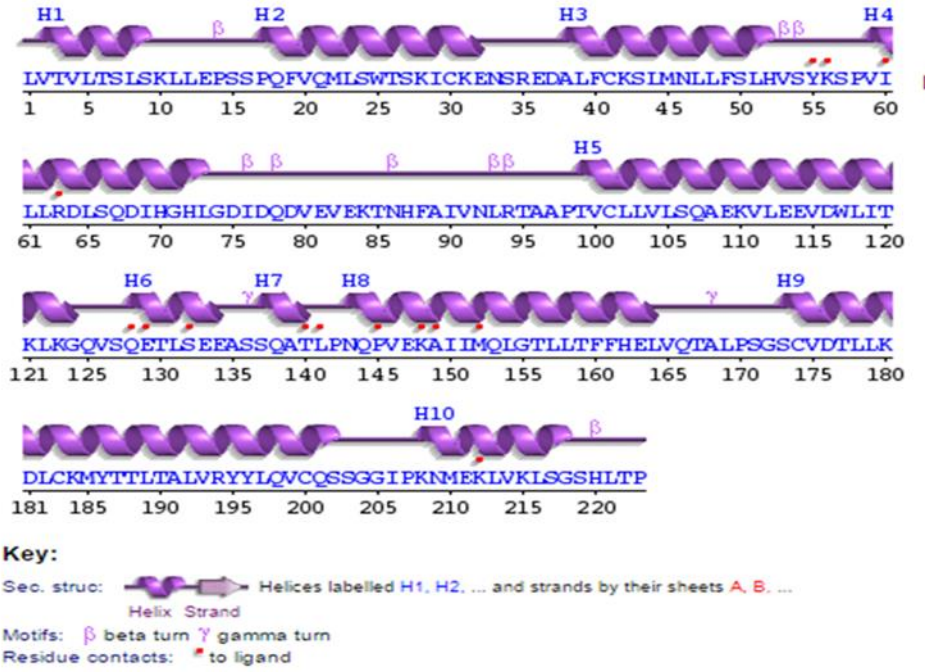


Figure 4.3; Secondary structure intermolecular interaction analysis profile of docked pose between FANCI ARM Repeat & DNA dodecamer (conserved residues marked with red dots) [84]

Thus, it suggests that HTH-type region has a motif character. It is well established that HTH motif binds specifically to the major groove of DNA [330]. To explore the binding of FANCI ARM repeat with DNA, we have performed docking analysis and found that HTH-type region binds to the major groove of DNA (**Figure 4.2 C**). It has also been observed that conserved amino acids such as 128 Gln, 129 Glu, 132 Ser and 140 Thr are

at the interactions interface between DNA and HTH-type motif of ARM repeat (**Figure 4.2 D and Figure 4.3**) [84].

4.1.4 Conclusion

To our conclusion, the FANCI ARM repeat has DNA binding ability and HTH structural motif with inherent structural flexibility. The structural flexible region has revealed the presence of helix former amino acids such as Ala, Glu, Gln and Ser [318], therefore the dynamic region forms α-helices. Furthermore, DSSP results also indicate that the structural flexible region has motif character. Major groove DNA binding of FANCI ARM Repeat with conserved interface instigate the presence of HTH motif as well as DNA binding ability. The multidisciplinary results reported here suggest that structural flexibility provides dynamic conformations to FANCI ARM repeat that binds to the major groove of DNA [84].

4.2 Functional Characterization of FANCD2 CUE Domain

4.2.1 Introduction

The Fanconi anemia proteins comprise the FA core complex, ID complex and group third BRCA proteins [2]. The activated FA- core complex assembles in the nucleus and monoubiquitinates the FANCD2 and FANCI proteins [3]. Monoubiquitinated FANCD2 and FANCI proteins are recruited at DNA damage sites to interact with crucial DNA repair proteins including BRCA1, FANCD1/BRCA2, and RAD51 [109, 173, 333]. A detailed overview of FANCD2 CUE domain has been described in **Chapter 1 and 3**. In brief, FANCD2 harbors CUE domain an ubiquitin binding domain known as Coupling of ubiquitin conjugation to endoplasmic reticulum domain (CUE domain) [334, 335]. CUE domain has a role in ubiquitin binding and also facilitate the intra-molecular mono-ubiquitination [336]. However a little is known at the structure and function level.

Recently, Leu231Arg mutation associated with deregulation DNA ICL repair was identified in the FANCD2 CUE domain region [90]. To evaluate the domain level association; the interaction studies were carried out using different *in-vitro, in-silico* and biophysical approach. It has been observed that FANCD2 CUE domain *wild-type* binds to the mono-ubiquitin efficiently compared to DNA ICL repair mutant Leu231Arg. In addition, we found that FANCD2 CUE domain has the chaperone-like activity and that is impaired in mutant protein. Furthermore, protein-protein docking was carried out to understand the residue located at interactions interface of FANCD2 CUE domain *wild-type* and Leu231Arg mutant with monoubiquitin. Here, the comparison of DNA ICL repair mutant with the FANCD2 CUE domain *wild-type* at the structure, dynamics, and function level have been evaluated.

4.2.2 Materials and methods

4.2.2.1 Surface Plasmon Resonance

SPR experiment was performed using BiacoreT200 (GE Healthcare). Mono-ubiquitin was immobilized (Boston Biochem, USA) on carboxymethyl (CM5) sensor chip using amine coupling protocol (http://proteins.gelifesciences.com/products-for-proteins/spr-systems/biacore-t200/). Immobilization level of 2500 Resonance Units was selected on the sensor chip (flow cell 2). HBSEP+ (0.10 M HEPES, 1.50 M NaCl, 0.03 M EDTA and 0.50% v/v Surfactant P20) buffer was used as running buffer and glycine-HCl pH 2.5 was used for regeneration (30-second pulse in 30µl/min flow rate) with stabilization period of 60 seconds. FPLC purified protein was used as analytes using the concentrations from 10, 20, 40, 60, 80, 100, 200, 400, 600 and 800 nM, with a 30 µl/min flow rate, contact time of 120 seconds and a dissociation time of 60 seconds. Data analysis was performed using Biacore T200 evaluation software (https://www.gelifesciences.co.jp/catalog/pdf/29050011aa.pdf).

4.2.2.2 Docking studies

Molecular docking study was performed using HADDOCK (http://haddock.science.uu.nl/services/HADDOCK2.2/). HADDOCK is an online tool to determine binding modes of protein-protein interactions [245]. Three-dimensional structural coordinates of monoubiquitin PDB file (PDB ID 5DK8) was retrieved from Protein data bank. Both the ligand and the receptor were formatted as per PDB format and different chain ID was assigned using PyMol (www.pymol.org) and PDB editor. Prior to docking, all hetero atoms were removed from protein and ligand. Best docked poses were selected for PDB-*SUM* (http://www.ebi.ac.uk/thornton-srv/databases/cgi

bin/pdbsum/GetPage.pl?pdbcode=index.html) and Ligplot analysis (www.ebi.ac.uk/thornton-srv/software/LigPlus/download.html). Docked poses interface was visualized using the PyMol molecular graphics software.

4.2.2.3 Chaperone-like assay

Chaperone-like activity was performed using lysozyme as protein substrate. Lysozyme of 0.30 mg/ml in phosphate buffer (pH 7.40) was pre-incubated for 5 min at 37°C alone and in presence of 0.30 mg/ml of FANCD2 CUE domain and Leu231Arg mutants in 96 well plate. The reaction started by addition of DTT up to the final concentration of 20 mM. Aggregation of lysozyme was monitored at the optical density of λ= 340 nm wavelength in different time points.

4.2.3 Results and Discussion

4.2.3.1 Binding affinity and kinetics of FANCD2 CUE domain and Leu231Arg mutant

Binding affinity and kinetics of FANCD2 CUE domain *wild-type* and Leu231Arg with monoubiquitin were performed by Surface Plasmon Resonance (SPR) (BiaCore T200, GE Healthcare). The decrease in binding of Leu231Arg mutant with monoubiquitin compared to FANCD2 CUE domain *wild-type* was observed. However, Resonance Units versus concentration plot (binding affinity) suggests that FANCD2 CUE domain *wild-type* protein has a stronger binding affinity with monoubiquitin than the Leu231Arg mutant **(Figure 4.3)**.

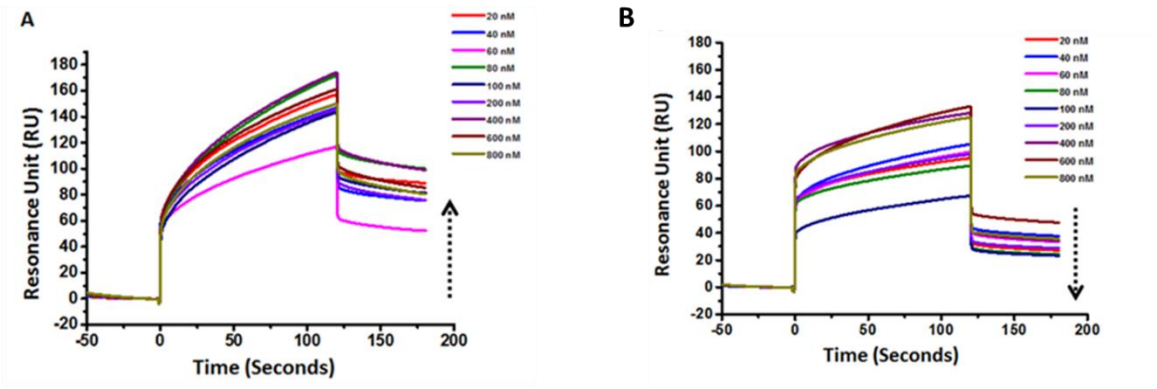


Figure 4.4; **SPR sensorgram of FANCD2 CUE domain *wild-type* and Leu231Arg mutant with mono-ubiquitin: (A)** Binding profile of mono-ubiquitin with FANCD2 CUE domain *wild-type* **(B)** Binding profile of mono-ubiquitin with Leu231Arg

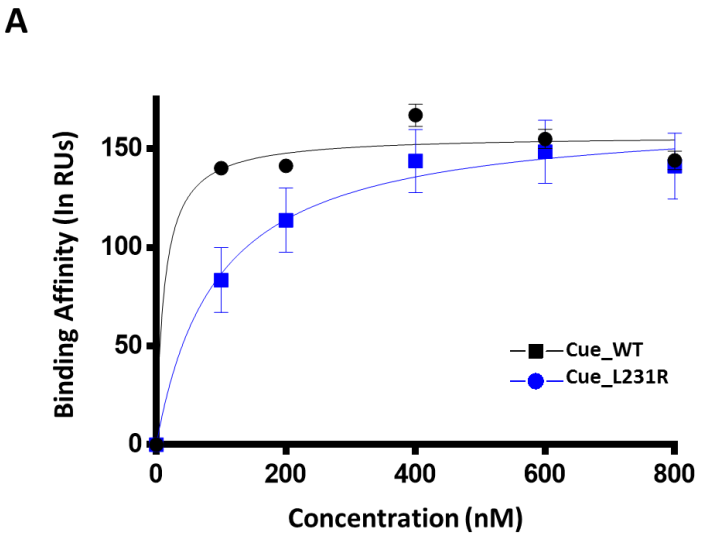


Figure 4.5; (A) Comparative binding profile of FANCD2 CUE domain *wild-type* and Leu231Arg mutant with monoubiquitin, shows FANCD2 CUE domain *wild-type* has the higher affinity compare to the Leu231Arg mutant

Further, to reveal the specificity of binding using SPR sensorgram, it has been found that FANCD2 CUE domain *wild-type* protein has specific binding curve unlike mutant **(Figure 4.3 A and B)**. In addition, we detected some anomalous behavior of Leu231Arg mutant

during binding; which was devoid of specificity. The observed binding affinity with monoubiquitin for FANCD2 CUE domain *wild-type* (KD: 3.41 ±1.8 x 10^{-1} nM) was significantly higher as compared to mutant Leu231Arg (KD; 0.92±1.5 µM). Association rate constant of *wild-type* CUE domain was found to be higher (Ka: 2.90±0.10 x 10^5 $M^{-1}s^{-1}$) than Leu231Arg mutant (Ka: 3.52±0.4 x 10^5 M^{-1} s^{-1}). Beside this, *wild-type* CUE domain showed lesser dissociation rate (kd: 9.78±8.510^{-2} x 10^{-6} s^{-1}) as compared to Leu231Arg (kd: 1.74±0.1 x 10^{-3} s^{-1}) mutant (**Figure 4.4 A**). Furthermore, *in-silico* docking results are also support the SPR results. Therefore, it can be concluded that *wild-type* CUE domain has higher binding affinity and specificity for the mono-ubiquitin, than the mutants. Binding affinity between mutant protein with monoubiquitin suggests that mutant might form unstable complex due to high dissociation rates and less binding affinity.

4.2.3.2 Structural orientation of FANCD2 CUE domain and Leu231Arg mutant

Reduction in the number of salt bridges, hydrogen bond and interface area in Leu231Arg mutants was observed compared to FANCD2 CUE domain *wild-type* protein at the binding interface (**Figure 4.5 A and B**). Furthermore, Leu231Arg mutants show the impairment in the interactions with monoubiquitin, and change in the folding pattern in *in-vitro* experiments.

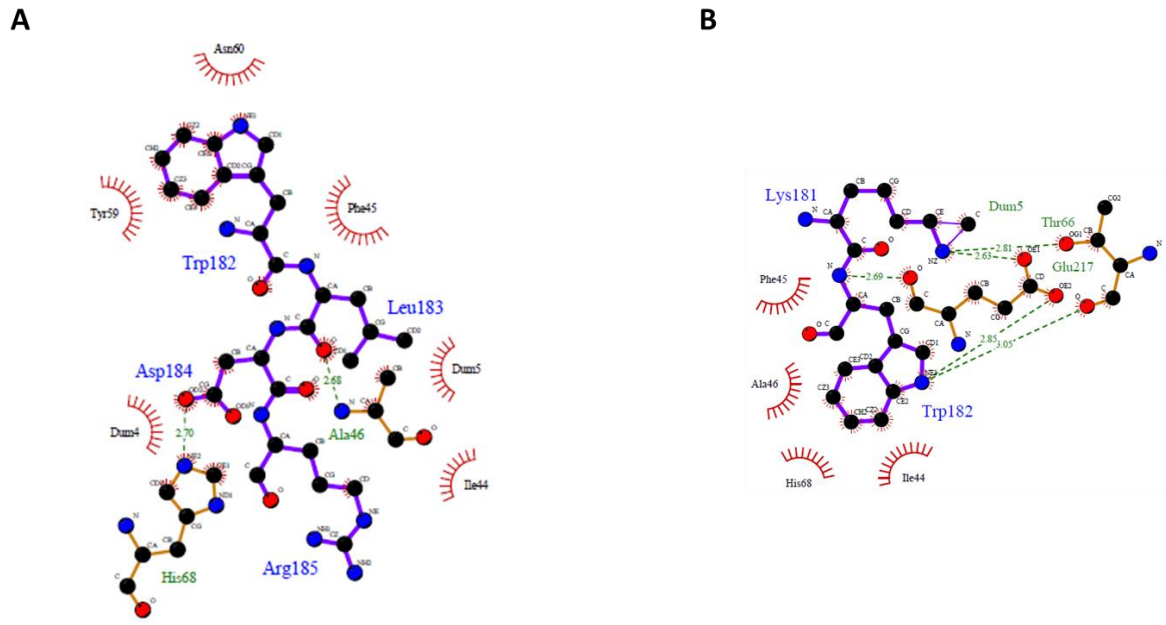
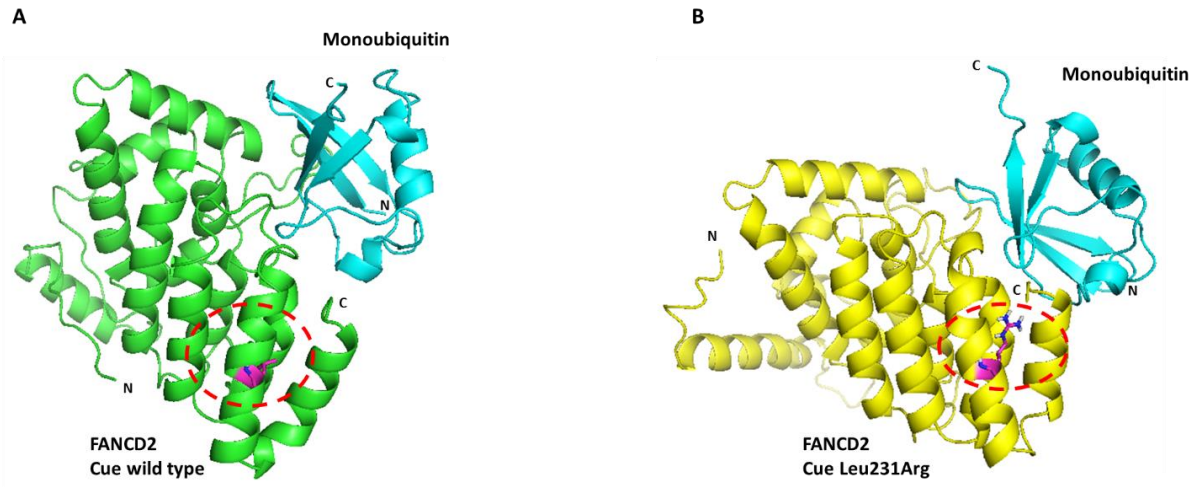


Figure 4.6; Ligplot analysis of docked complex: (A) Weak intermolecular interaction between FANCD2 CUE domain *wild-type* and monoubiquitin **(B)** Leu231Arg mutant and monoubiquitin,



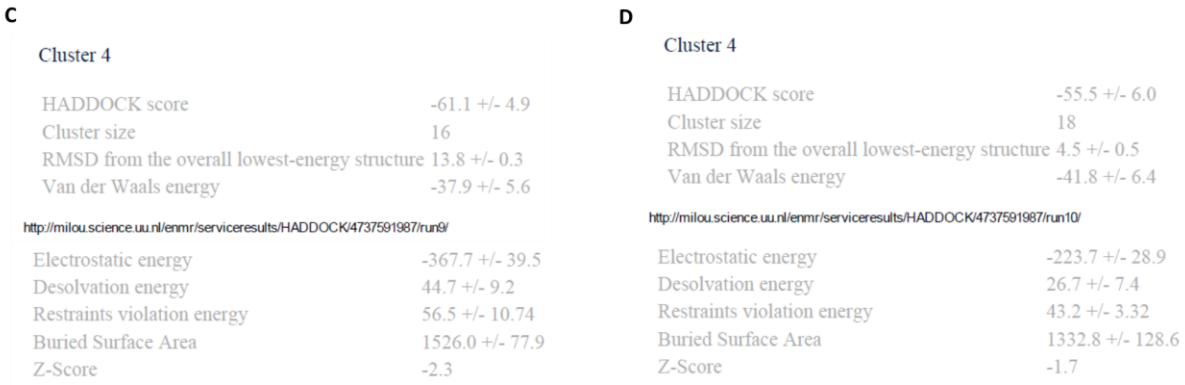


Figure 4.7; Docked complex structures of FANCD2 CUE domain and monoubiquitin: (A) FANCD2 CUE *wild-type* and monoubiquitin **(B)** Leu231Arg and monoubiquitin; Red dot circle shows the point mutation position in α-helices (C) HADDOCK docking results of Cue *wild-type* and monoubiquitin (D) HADDOCK docking results of Cue Leu231Arg mutant and monoubiquitin

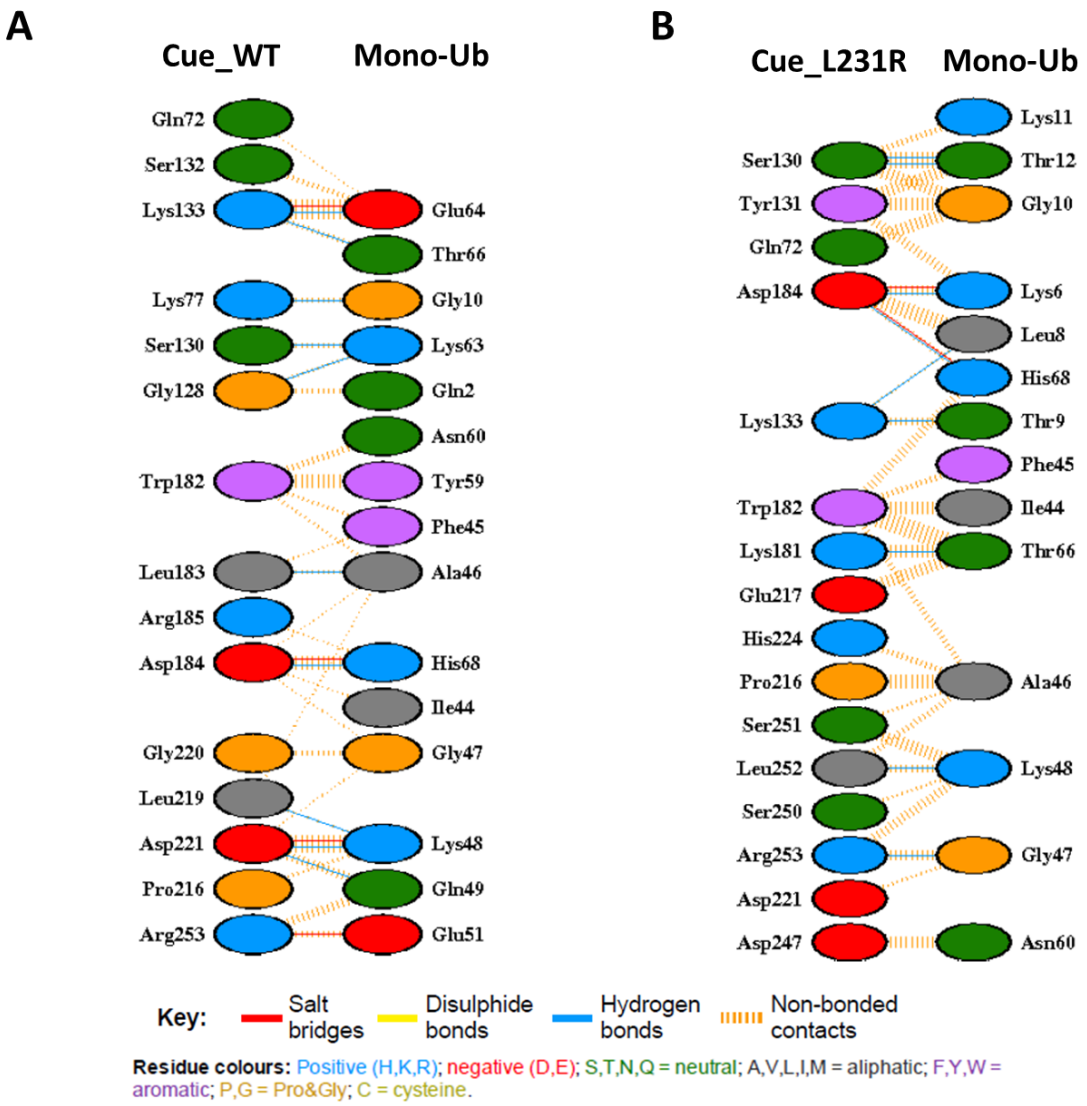


Figure 4.8; PDB*SUM* analysis of docked complex: (A) FANCD2 CUE *wild-type* with mono-ubiquitin **(B)** Leu231Arg with mono-ubiquitin

In-silico docking suggests that conserved hydrophobic patch of monoubiquitin intact at the binding interface of FANCD2 CUE *wild- type* and perturbed in Leu231Arg mutant.

Table 4.1 Interface statistics of FANCD2 CUE domain and Leu231Arg mutant with monoubiquitin

Interface statistics of monoubiquitin with Cue domain and mutants			
Monoubiquitin	**Salt bridges** (In number)	**Hydrogen bond** (In number)	**Interface area** (Å²)
Cue (*wild type*)	4	10	783,775 (Chain A, Chain B)
Cue (*Leu231Arg*)	2	9	736,737 (Chain A, Chain B)

Moreover, we have observed that FANCD2 Trp 182 non-covalently interacts with the monoubiquitin Phe 45, Ala 46 & Tyr 59. However, in Leu231Arg mutant Trp 182 forms non-covalent interactions with Ile 44, Phe 45 & Thr 66. Furthermore, there was a decrease in number of hydrogen bonds, salt bridges and interface area **(Table 4.1)** and **(Figure 4.6 A and B)**. Hence it can be concluded that FANCD2 CUE domain *wild-type* complex with monoubiquitin is stable compared to Leu231Arg mutant.

4.2.3.3 Chaperone-like assay of FANCD2 CUE domain and Leu231Arg mutant

FANCD2 protein has histone chaperone activity [333] which plays a central role during ICL repair by modifying in the chromatin structure. This allows the access for the DNA repair proteins required for the downstream steps of the DNA intercrosslink repair in Fanconi Anemia pathway [333].

In-vitro chaperone-like assay [337] was performed for FANCD2 CUE domain *wild-type* and Leu231Arg mutants to evaluate their ability to resist the aggregation of substrate model protein (Lysozyme). Reduction in disulfide bonds induces denaturation, unfolding, and aggregation of lysozyme accompanied by an increase in the optical density at λ= 340 nm. The addition of small quantity of FANCD2 CUE domain *wild-type* retarded the onset of lysozyme aggregation, however long incubation resulted in significant increase of light scattering.

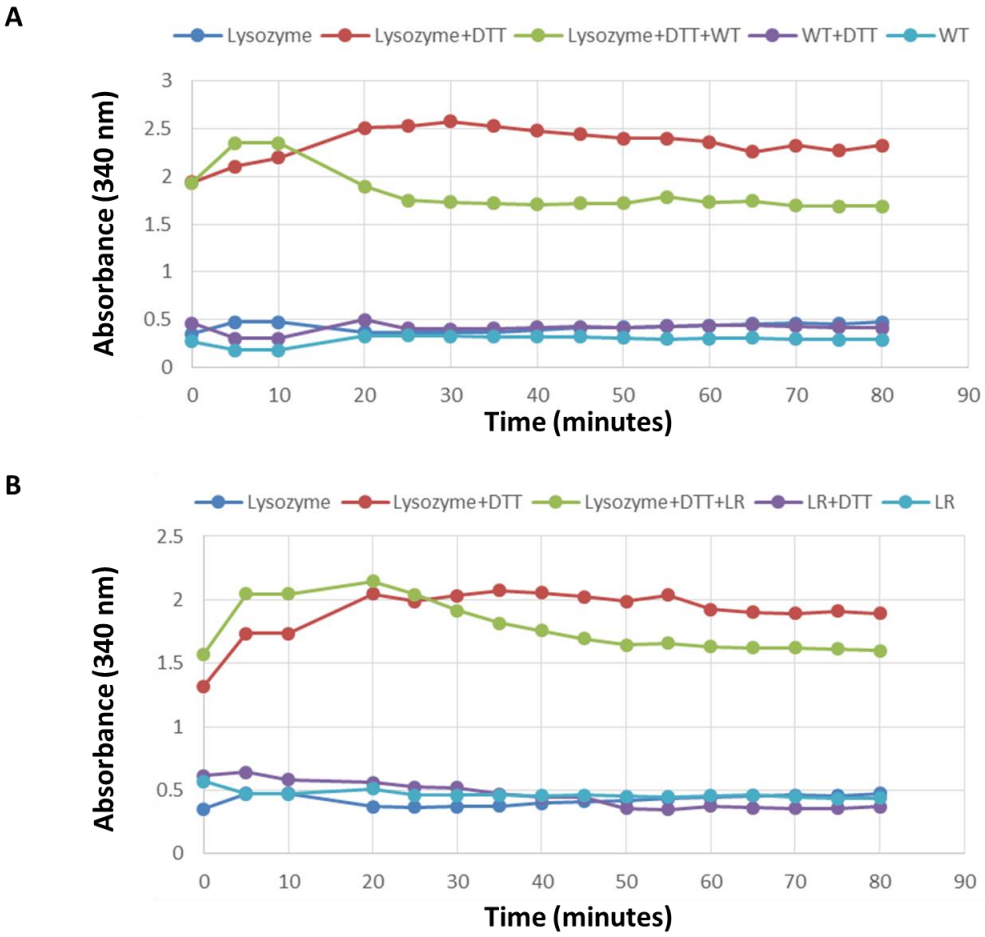


Figure 4.9; Comparative profile for chaperone-like assay: (**A**) FANCD2 *wild-type* CUE domain and (**B)** Leu231Arg mutant

We have observed that FANCD2 CUE domain *wild-type* protein is more effective in retardation or prevention of lysozyme aggregation than Leu231Arg mutant. These results suggest that at least in part, the FANCD2 CUE domain has the chaperone activity. However, further studies needed to validate the chaperonin activity of FANCD2 CUE domain.

4.2.3.4 Conclusion

The comparative studies of FANCD2 CUE domain and Leu231Arg mutant suggest that *wild-type* protein has higher affinity and specificity towards mono-ubiquitin than the Leu231Arg mutant. Furthermore, *in-silico* docking analysis also supports the results obtained from surface plasmon resonance. Chaperone-like assay of FANCD2 CUE domain *wild-type* and Leu231Arg mutant suggest that FANCD2 CUE domain *wild-type* protein has higher chaperonin activity compared to Leu231Arg mutant.

Chapter 5

Protein-Protein Interactions Between FANCI and FANCD2

5.1 Introduction

FANCI and FANCD2 (I-D) complex comprises the second group of proteins in the FA pathway and has an important role in DNA ICL damage repair [3]. Amino acids sequence alignment between FANCI and FANCD2 showed 13% identity and 20% similarity across the entire protein. However, less sequence similarity were observed in the FANCD2 and FANCI paralogs in other species including *A. thaliana* and *D. melanogaster* [3]. The striking amino acids conservation was found throughout species surrounded Lys523 in FANCI and Lys 561 in FANCD2, which is known to be monoubiquitinated [3, 41]. Knockdown of siRNAs to FANCI and FANCD2 revealed a reduced recombination from 78% to 47% compared to controls [3]. These reports suggest that FANCI is an important component in the HR-repair pathway [3, 85]. FANCI and FANCD2 in association with RAD51 repair the damaged DNA in the stalled replication fork and protect the nascent DNA. Furthermore, I-D complex facilitates the recruitment of FAN1 nucleases to incise the DNA lesion [3, 85, 109]. FANCD2 binds to monoubiquitinated FANCI with the help of CUE domain and mask mono-ubiquitin to rescue poly-ubiquitination that escapes the protein to degrade [179]. However, little is known at domain level interactions between FANCI and FANCD2. Therefore, we have performed comprehensive *in-vivo* (bacterial two-hybrid assay) followed by a beta-galactosidase assay as well as *in-silico* (protein-protein docking) binding analysis to understand PPIs at the domain level. To our conclusion, CUE domain of FANCD2 interacts with ARM repeat of FANCI. Interestingly, the higher binding affinity between FANCI ARM repeat and FANCD2 CUE domain were observed when FANCD2 has ARM repeats with it.

5.2 Materials and Methods

5.2.1 Materials

Detailed materials and methods have been described in **Chapter 2** and **Chapter 6**.

5.2.2 Bacterial strains, plasmids, and materials

Already described in **Chapter 2**. In brief, the *E. coli* BTH101 was used for co-expression of the respective interacting partners such as FANCI ARM Repeat and FANCD2 CUE domain to study *in-vivo* protein-protein interactions. The pET28a(+), pGEX4T-1, pUT18 and pKNT25 expression vector were maintained in *E. coli* DH5α/NOVABLUE strains **(Table 5.1).** Selectable antibiotic markers ampicillin (100 mg/ml), kanamycin (50 mg/ml) and chloramphenicol (34 mg/ml) were used as required. Recombinant *E. coli* harboring expression vectors and derivatives were grown in the presence of respective antibiotics.

Table 5.1 . Bacterial strains and plasmids used in this study

Bacterial strains	**Description**
E. coli BL21 (DE3)	B strain, *fhu*A2 (*lon*) *ompT* gal (λ*DE3*)(*dcm*) Δ*hsd*S *i21* Δ*nin5*
E. coli Novablue	*end* A1 *hsd* R17(rK12$^-$ mK12+) supE44 *thi*-1 *rec*A1 *gyr*A96 *rel*A1 *lac*F′ [*proA*$^+$ *B*$^+$ *lacI*^{q} ZΔ*M15*::Tn*10*] Tet^R
E. coli DH5α	*fhuA2 Δ(arg*F-*lac*Z*)U169 pho*A *gln*V44 *Φ80 Δ(lacZ)*M15 *gyrA96 rec*A1 *rel*A1 *end*A1
E. coli Rosetta 2(DE3)	F$^-$ *ompT hsdS*$_B$(r_B^- m_B^-) *gal dcm* (DE3) pRARE2 (Cam^R)
E. coli BTH101	F', cya-99, *ara*D139, *gal*E15, *gal*K16, *rps*L1 (*Str*R), *hsd*R2, *mcr*A1 *mcr*B1, *rel*A1
Plasmids	
pKNT25	Origin of replication from pSU40. N-terminal fusion with T25 adenylate cyclase fragment, 3.4 kb, Kanamycin resistance
pUT18	col*E1* origin of replication from pUC19. N-terminal fusion with T18 adenylate cyclase fragment, 3 kb, Ampicillin resistance
pET28a+	pBR322 origin, f1 origin, lac*I* coding sequence, T7 promoter, T7 transcription start, His-TAG, T7 terminator
pGEX4T-1	pBR322 origin, GST (N-term), tac promoter, size 4969 bp, Ampicillin resistance

5.2.3 Gene cloning and site-directed-mutagenesis

Construction of recombinant plasmids for the expression and purification of proteins were made as described in **Chapter 2**. FANCI (985-1207), and FANCD2 (1-254) amino acids region were PCR amplified using respective full-length human cDNA template and specific primers **(Table 5.2)**. PCR amplicons and vectors were digested by restriction enzyme (RE) such as *Bam*HI, *Eco*RI, Kpn1 & *Xho*I as required and sticky ends were ligated using ligation reaction. Further, recombinant clones were confirmed by double RE digestion and DNA sequencing. Details of bacterial strains, plasmids, and primers used for the study are listed in **Table 5.1** and **Table 5.2**.

Table 5.2. List of the primers used in this study

S.No.	Name of Primer	Primer Sequence (Forward & Reverse)	Vector
1	FI_ARM_BTH_F	5'GCGGATCCGGCAGCTCTAGTCACGGTTCTTACCAG 3'	pUT18
2	FI_ARM_BTH_R	5'CGGGGTACCGCGGCCGCGGGGGTCAGATGAGAACCAG 3'	pUT18
3	FD2_CUE_BTH_F	5'GCGGATCCGGCAGCTATGGTTTCCAAAAGAAGACTG 3'	pUT18
4	FD2_CUE_BTH_R	5'CCGGAATTCGCGGCCGCAAGTCGGAGGCTTGAAAG 3'	pUT18
5	BRCA2_BTH_F	5'GCGGATCCGGCAGCTACCGCACCTGGTCAAGAATTTCTG 3'	pKNT25
6	BRCA2_BTH_R	5'CCGGAATTCGCGGCCGCAACGCCATACGTATA 3'	pKNT25
7	FD2_CUE_F	5'GTCGGATCCGAGAACCTGTACTTTCAGGGTATGGTTTCCAAAAGAAGACTG 3'	pGEX4T-1
8	FD2_CUE_R	5'GTCCTCGAGCTATTAAAGTCGGAGGCTTGAAAGGAC 3'	pGEX4T-1
9	BRCA2_F	5'GTCGGATCCGAGAACCTGTACTTTCAGGGTACCGCACCTGGTCAAGAATTTCTG 3'	pET28a+
10	BRCA2_R	5'GTCGAATTCCTATTAAACGCCATACGTATACAGCTG 3'	pET28a+

5.2.4 *In vivo* protein-protein interaction studies

In vivo interactions between different proteins were monitored using bacterial two-hybrid system (BACTH) as described [285, 286, 338]. Recombinant plasmids FANCI (985-1207) and FANCD2 (1-254) amino acids considering full length from human cDNA template were cloned in bacterial two-hybrid plasmids pUT18 and pKNT25 and a series of recombinant plasmids were generated as listed in **Table 5.2**. Detailed methodology is

described in **Chapter 2**. In brief, the cells were grown overnight in triplicates, and 5 µL of it was spotted on LB agar plate containing 5' bromo 4 chloro-3-indolyl-β-D-galactopyranoside (X-Gal) (40 µg / mL), IPTG (0.50 mM) in the presence of antibiotics as required. Spotted LB-X-Gal-IPTG plates were further incubated at 30°C overnight and the appearance of blue-white colour colonies was analysed and β-galactosidase activity was measured from liquid cultures using 0.40 % O-nitrophenyl-*β*-D-galactoside (ONPG) as a substrate. For the first set of bacterial two-hybrid experiment Synergy H1 Hybrid Multimode Microplate Reader (Biotek, USA) was used, and for the second set BMG-Labtech Microplate reader (Germany) was used.

5.2.5 *In vivo* protein-protein interactions

The β-galactosidase activity was calculated and plotted in Miller units with P value using Graphpad Prism (www.graphpad.com/scientific-software/prism/). β-Galactosidase activity (units/mg protein) was calculated as mean±SD (n≥2). The significance differences were analyzed using Student's t-test, P values obtained at 95% confidence intervals are shown as (*) for <0.05, (**) for <0.01 and (***) for <0.001.

5.2.6 *In-silico* protein-protein interactions

Amino acids sequences of FANCI (985-1207) (Uniprot ID: Q9NVI1) and FANCD2 (1-254) (Uniprot ID Q9BXW9) were retrieved from UniProtKB (http://www.uniprot.org/) using respective accession number. Molecular modelling of FANCI (985-1207) & FANCD2 (1-254) were modelled as described earlier [49, 84, 309]. Ramachandran plot

was analyzed to check the stereochemistry of amino acids in allowed and generously allowed regions. Stereo-chemical details of the FANCI ARM repeat in Ramachandran plot are as follows: 95.10% core region, 3.40% allowed region, 1.50% generously allowed, 0.00% disallowed region, and for FANCD2 CUE domain is Ramachandran Plot: 91.10% core region, 8.90% allowed region, 0.00% generously allowed region, 0.00% disallowed region. HADDOCK was performed as described in **Chapter 2.**

5.3 Results and Discussion

FANCI and FANCD2 protein-protein interactions were observed using STRING analysis (https://string-db.org/) **(Figure 5.1)**. An interesting feature is that the FANCI and FANCD2 are paralogs having the ARM repeat, even though the length and sequences of the protein and ARM repeat is not identical [3].

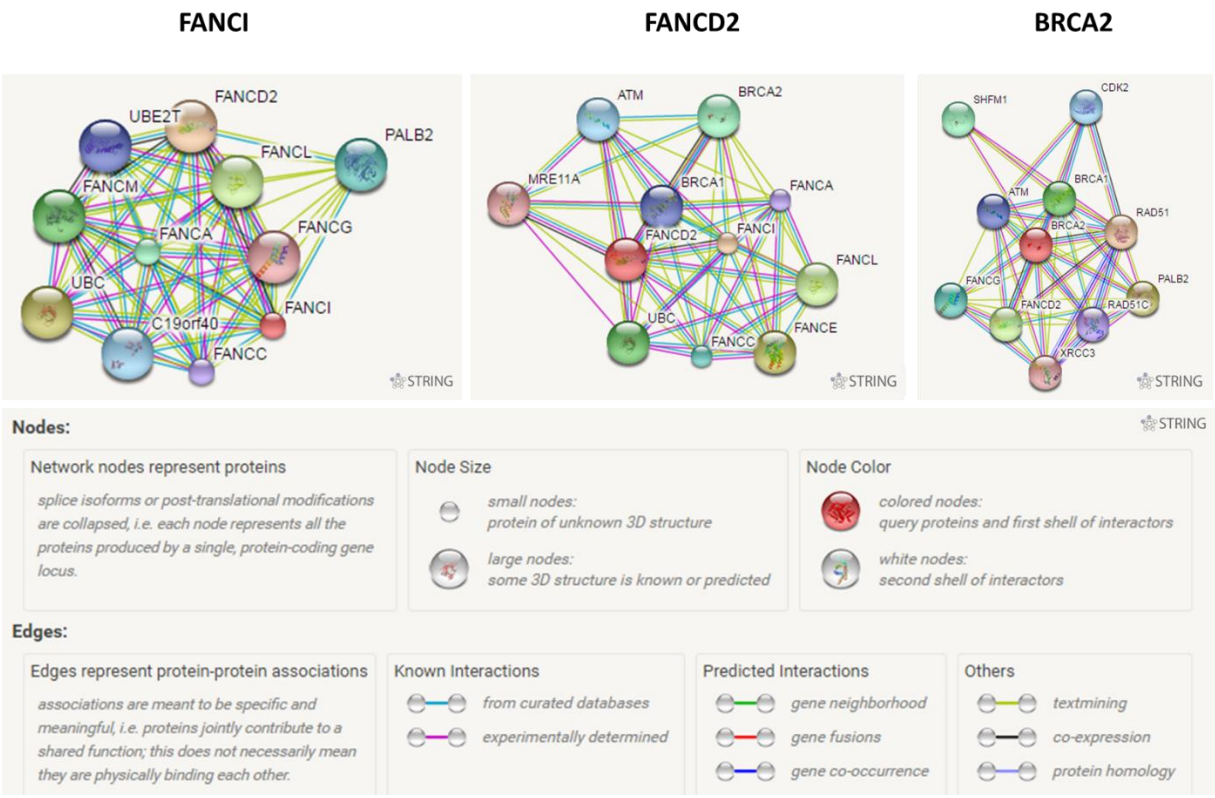


Figure 5.1; STRING analysis of FANCI, FANCD2 and BRCA2 proteins

5.3.1 *In-vivo* protein-protein interactions studies between FANCI ARM repeat with FANCD2 and BRCA2

Bacterial two-hybrid assays were performed to evaluate the interactions between FANCI and FANCD2. In the first set of experiment, It has been observed that FANCI ARM repeat shows interactions of 100.10±24.90 Miller Units with FANCD2 CUE domain in the β-Galactosidase assay (p=0.005) **(Figure 5.2)**. However, in the case of FANCI ARM and BRCA2 C-terminal region (BRCA2 CTR), the Miller Units was significantly higher (206.09±9.11, p <0.0001).

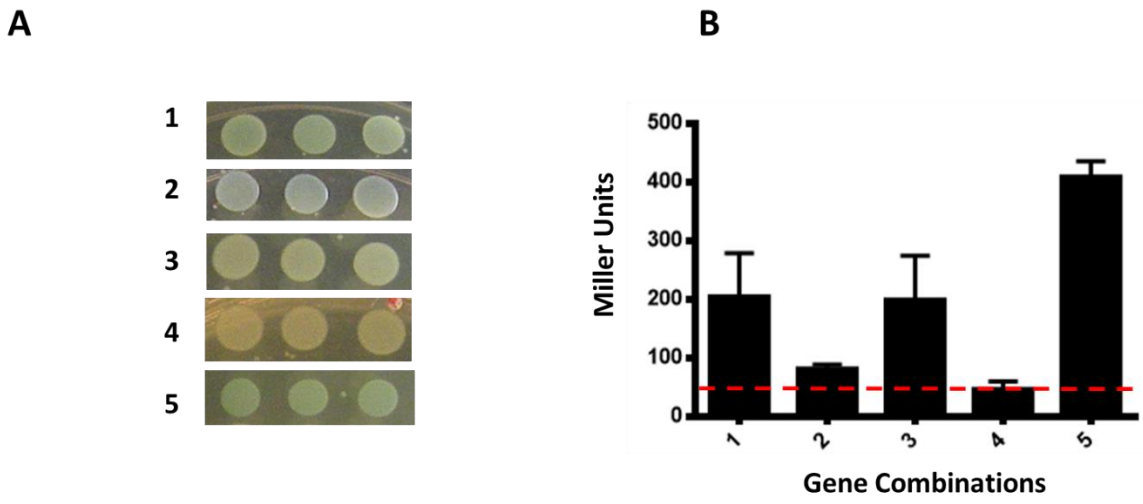


Figure 5.2; (A) *In-vivo* protein-protein interaction studies of FANCI ARM Repeat, FANCD2 CUE Domain and BRCA2 C- Terminal Region (CTR), 1= FANCI ARM Repeat and FANCD2 (1-472), 2= FANCI ARM Repeat and FANCD2 CUE Domain, 3= FANCI ARM Repeat and BRCA2 CTR. 4= Cells co-expressing pUT18 and pKNT25 tag empty vectors were used as negative control, and 5= FtsA-T18 and FtsZ-T25 protein was used as positive control, **(B)** β-Galactosidase activity in Miller units are shown here as mean±SD (n≥2)

Interestingly, we have observed that the interaction between FANCI ARM and FANCD2 (1-472) is higher than FANCI ARM and FANCD2 CUE domain. The data revealed that FANCI ARM interacts to the BRCA2 C-Terminal Region. FANCI ARM also interacts with the FANCD2 CUE and binds with higher binding affinity when a part of ARM repeat of FANCD2 was included. The experiments were performed independently twice for biological and technical duplicates.

5.3.2 *In-silico* protein-protein interactions of FANCI ARM repeat with FANCD2 and BRCA2

To further support the *in-vivo* results, *in-silico* protein-protein interactions were performed to analyze the binding interfaces **(Figure 5.3).** FANCI protein, particularly through ARM

repeat, play an important role in protein-protein interactions in FA-BRCA pathway [3, 22, 165, 166, 168]. It has been observed that FANCI ARM repeat interacts with BRCA2 C-terminal region. Interestingly, from the FANCI ARM repeat and BRCA2 CTR docked complex structure, Trp 117 and Thr 120 of FANCI ARM forms π-cation and hydrogen bonding interactions, with the Lys 64 of BRCA2 CTR respectively. Usually, protein-protein interaction hotspots are known to enrich with tryptophan, tyrosine, and arginine [339]. Furthermore, the presence of 13 hydrogen bonds with 6 salt bridges in FANCI ARM and BRCA CTR docked complex interface structure suggest that these regions might help in the interactions between FANCI and BRCA2 proteins. We also performed the docking studies between FANCI ARM and FANCD2 CUE domain and found that Trp 182 of the FANCD2 CUE domain forms the cation-π interaction and non-bonded interactions with Lys 215 and Leu 216 of FANCI ARM Repeat. In addition, interface statistics calculated from FANCI ARM and FANCD2 CUE docked complex structure indicate the presence of 8 hydrogen bonds and 5 salt bridges at the interaction interface.

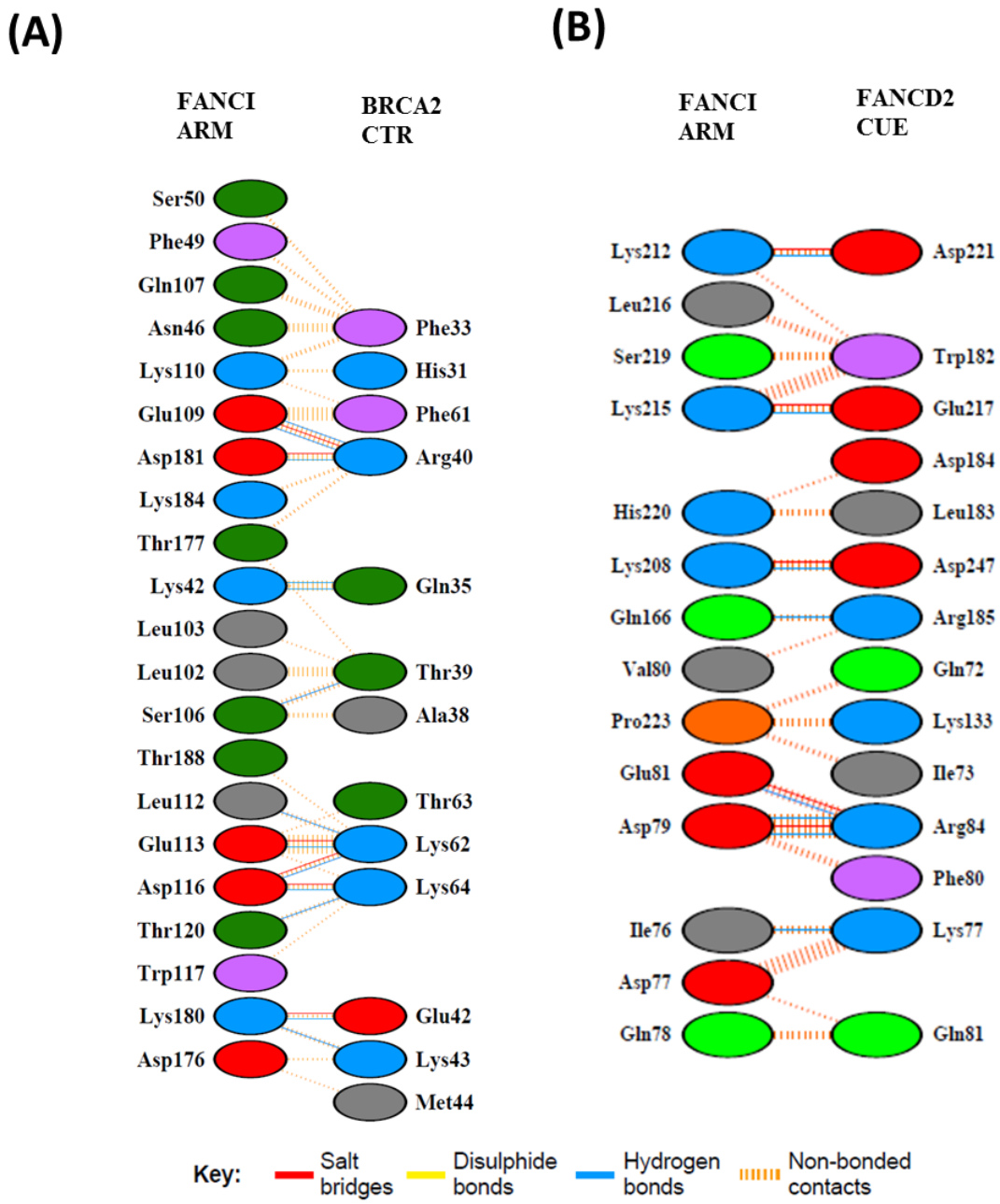


Figure 5.3; Weak intermolecular interactions between **(A)** FANCI ARM with BRCA2 CTR and **(B)** FANCI ARM with FANCD2 CUE domain

5.3.3 Conclusion

The interactions between full-length FANCI and FANCD2, and FANCD2 with BRCA2 protein have been reported by different groups [2, 4, 12, 42, 173], but how the protein-protein interactions associated with different domains in FA-BRCA pathway proteins have not been characterized. Here, it has been reported that FANCI ARM repeat interacts with FANCD2 CUE domain and BRCA2 C-terminal region. Furthermore, the interaction between FANCI ARM repeat and FANCD2 CUE domain increases by the addition of ARM repeat region of FANCD2. These interactions studies indicate that different functional domain present in FANCI, FANCD2, and BRCA2 proteins are playing an important role in PPIs.

Chapter 6

Protein-Protein Interaction Studies between FANCD2 and BRCA2

6.1 Introduction

Detailed introduction and functional properties of FANCD2 and BRCA2 have already been described in **Chapter 1, 3 and 4**. The FANCD2 protein is known to interact with BRCA2 protein (**Figure 5.1)**. DNA repair pathway in BRCA2 deficient cells leads to error prone single-strand annealing [12, 13, 82]. Furthermore, DNA repair processes such as nucleolytic incision, translesion DNA synthesis, and homologous recombination have been co-ordinated by FA/BRCA pathway [55, 59, 88, 101]. In FA pathway, the *fancd2* gene homologues is present in other eukaryotic species including *Drosophila melanogaster*, *Caenorhabditis elegans*, *Arabidopsis thaliana*, Fugu and zebrafish [2].

The C-terminus region of BRCA2 harbors BLAT (BRCA2-motif in *Leishmania, Arabidopsis, and Trypanosoma*) domain, which is highly conserved in divergent species [12, 13]. The BLAT domain of BRCA2 region encompassing the (2350-2545) amino acids interacts to FANCD2 domain comprising (1-1102) amino acids. However, FANCD2 lacking N-terminus residues from (1-325) amino acids failed to interact with BRCA2 [12]. In addition, BRCA2 proteins with C-terminal truncations were observed in cell lines derived from Fanconi Anemia patients [2]. Thus, confirming the importance of FANCD2 N-terminus and BRCA2 C-terminus for the protein-protein interactions (PPIs) in FA/BRCA pathway.

FA proteins particularly FANCI and FANCD2 form interdependent complexes required for chromatin association and DNA ICL repair [26, 55, 59, 101, 340]. FANCD2 harbors evolutionarily conserved ubiquitin binding domain, known as the coupling of ubiquitin conjugation to endoplasmic reticulum domain (CUE domain) [179]. This ubiquitin binding domain of FANCD2 is known for regulation of the FA pathway and reported to have a major role in DNA ICL repair processes [179]. Two biologically relevant mutations have

been identified in FANCD2 CUE domain, first cell-cycle checkpoint mutant Ser222Ala [341, 342] and second DNA ICL repair mutant Leu231Arg [90]. Considering the importance of mutations, and functional diversity of different domains present in FANCD2 and BRCA2, the protein-protein interactions between FANCD2 CUE domain (FD2 CUE) and BRCA2 C-terminal region (BRCA2 CTR) has been analysed. We also evaluated the interactions between FANCD2 CUE domain Ser222Ala and Leu231Arg mutants with BRCA2 C-terminal region using multi-disciplinary *in-vivo*, *in-vitro* and *in-silico* approach.

6.2 Materials and Methods

Materials and chemicals are described in **chapter 2**.

6.2.1 Bacterial strains, plasmids, recombinant constructs and materials

All recombinant techniques, expression, and purification of proteins have been performed as described **in Chapter 2, 5 and in Table 5.1.**

6.2.2 Cloning of different expression constructs and site-directed-mutagenesis

Construction of recombinant plasmids for the expression and purification of proteins were made as described in **Chapter 2 and 5**. FANCD2 (1-254) and BRCA2 (2350-2545) amino acids region were PCR amplified using respective full-length human cDNA template and specific primers listed in **Table 6.1**. PCR amplicons and vectors were digested by restriction enzyme (RE) such as *Bam*HI, *Eco*RI & *Xho*I, and ligated. Recombinant clones were confirmed by double RE digestion and DNA sequencing.

Table 6.1. List of the primers used in this study

S.No.	Name of Primer	Primer Sequence (Forward & Reverse)	Vector
1	FD2_CUE_BTH_F	5'GCGGATCCGGCAGCTATGGTTTCCAAAAGAAGACTG 3'	pUT18
2	FD2_CUE_BTH_R	5'CCGGAATTCGCGGCCGCAAGTCGGAGGCTTGAAAG 3'	pUT18
3	BRCA2_BTH_F	5'GCGGATCCGGCAGCTACCGCACCTGGTCAAGAATTTCTG 3'	pKNT25
4	BRCA2_BTH_R	5'CCGGAATTCGCGGCCGCAACGCCATACGTATA 3'	pKNT25
5	FD2_CUE_F	5'GTCGGATCCGAGAACCTGTACTTTCAGGGTATGGTTTCCAAAAGAAGACTG 3'	pGEX4T-1
6	FD2_CUE_R	5'GTCCTCGAGCTATTAAAGTCGGAGGCTTGAAAGGAC 3'	pGEX4T-1
7	FD2_CUE_SA_F	5'ATCCTAGGGGATGCCCAGCACGCTGAT 3''	pGEX4T-1, pUT18
8	FD2_CUE_SA_R	5'ATCAGCGTGCTGGGCATCCCCTAGGAT 3'	pGEX4T-1 pUT18
9	FD2_CUE_LR_F	5'GTGGGGAAAGAACGCAGTGACCTACTG 3'	pGEX4T-1, pUT18
10	FD2_CUE_LR_R	5'CAGTAGGTCACTGCGTTCTTTCCCCAC 3'	pGEX4T-1, pUT18
11	BRCA2_F	5'GTCGGATCCGAGAACCTGTACTTTCAGGGTACCGCACCTGGTCAAGAATTTCTG 3'	pET28a+
12	BRCA2_R	5'GTCGAATTCCTATTAAACGCCATACGTATACAGCTG 3'	pET28a+

6.2.3 *In vivo* protein-protein interactions

Recombinant plasmids FANCD2 (1-254) and BRCA2 (2350-2545) amino acids using full-length human cDNA template were cloned in bacterial two-hybrid plasmids pUT18 and pKNT25 and a series of recombinant plasmids were generated as given in **Table 6.1**. The detailed methodology has been described in **Chapter 5** of **section 5.2.4**. *In vivo* interactions of different proteins were monitored using bacterial two-hybrid system (BTH) [285, 286].

6.2.4 *In vitro* protein-protein interactions studies

FANCD2 (1-254), FANCD2 Ser222Ala, FANCD2 Leu231Arg and BRCA2 (2350-2545) cloned in pET28a+, pGEX4T-1 vectors [49, 84, 309, 338], and further co-transformed for recombinant proteins expression using chemically competent *E. coli* Rosetta 2(DE3) cells. The overnight grown culture containing different constructs were diluted to 1:100 in fresh LB broth and further incubated at 37°C using shaking incubator at 200 rpm till OD (λ= 600 nm) reaches 0.60–0.80. The culture was induced by the addition of 0.40 mM IPTG (Isopropyl β-D thiogalactoside) for 16 hours at 21°C. Cells were sonicated in the buffer

containing 300 mM NaCl, 50 mM Tris, 0.10% triton-X, 5mM beta-mercaptoethanol, pH 8.00, protease inhibitor cocktail (Sigma). The sonicated cells were centrifuged at 15,000 rpm for 40 min at 4°C. Clear cell supernatant was allowed to pass from Snapcap column (Thermo-Fisher) containing GST resin. GST bound protein was eluted in the buffer containing 20 mM glutathione, 100 mM NaCl, 25 nM Tris, pH 8.50. Eluted samples from Snapcap column (Thermo-Fisher) were mixed with laemelli buffer and heated at 95°C for 5 minutes before loading onto 12% SDS-PAGE gel. The gel was stained using Coomassie brilliant blue and further destained for the visualization of protein bands.

6.2.5 *In-silico* protein-protein interactions

Amino acids sequences of BRCA2 (2350-2545) (Accession number: P51587) and FANCD2 (1-254) (Accession number: Q9BXW9) were retrieved from UniProtKB (http://www.uniprot.org/). Molecular modelling of BRCA2 (2350-2545) was performed using FALCON (http://protein.ict.ac.cn/FALCON/), however, FANCD2 (1-254) were modelled using Robetta server [49, 84, 309]. The molecular models were validated as described in **Chapter 3, 4 and 5**. Ramachandran plot ensuring that all the amino acids are present in allowed and generously allowed regions. Ramachandran Plot for FANCD2 CUE domain: 91.10% core region, 8.90% allowed region, 0.00% generously allowed region, 0.00% disallowed region and for BRCA2 CTR has: 81.70% core region, 16.00% allowed region, 2.30% generously allowed, 0.00% disallowed region. Protein-protein docking using HADDOCK [245, 246] were performed as described in **Chapter 2, 3 and 4.**

6.3 Results and Discussion

Protein-protein interactions play very important role in DNA damage repair mechanism. In the context of DNA ICL repair, it is known that FANCD2 and BRCA2 interact particularly in FA/BRCA pathway to repair DNA ICL damage [2].

6.3.1 *In-vivo* protein-protein interactions studies of BRCA2 C-terminal region with FANCD2 CUE domain and Ser222Ala, Leu231Arg mutants

FANCD2 is known to interact with BRCA2, to bridge the FA/BRCA pathway. Bacterial two-hybrid (BTH) assays were performed to understand domain level interactions. Higher value of Miller Units 1349.27±289.65 between FANCD2 CUE domain and BRCA2 CTR ($p<0.0001$) was observed which suggests that FANCD2 CUE domain binds strongly with BRCA2 CTR (**Figure 6.1 A and B)**.

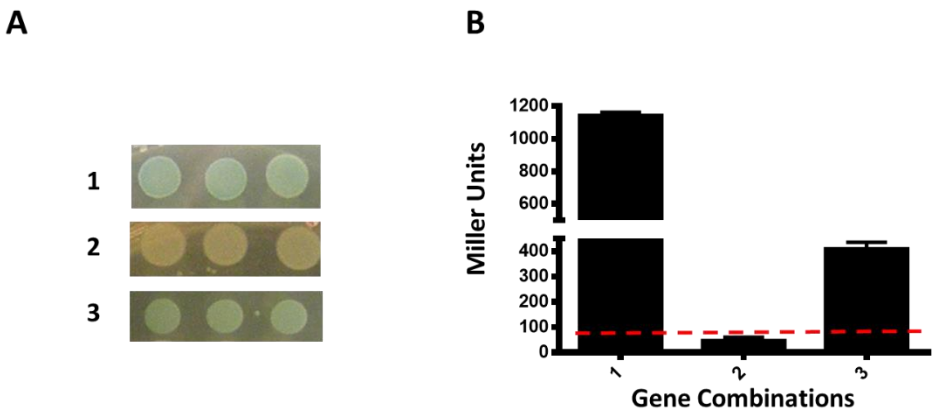


Figure 6.1; (A & B) *In-vivo* protein-protein interactions studies of FANCD2 CUE Domain with BRCA2 C- Terminal region, 1= FANCD2 CUE domain and BRCA2 CTR, 2= Negative control (pUT18 and pKNT25 tag empty vectors), and 3= FtsA-T18 and FtsZ-T25 protein was used as positive control (**B**) β-Galactosidase activity (units/mg protein) is shown here as mean±SD (n≥2)

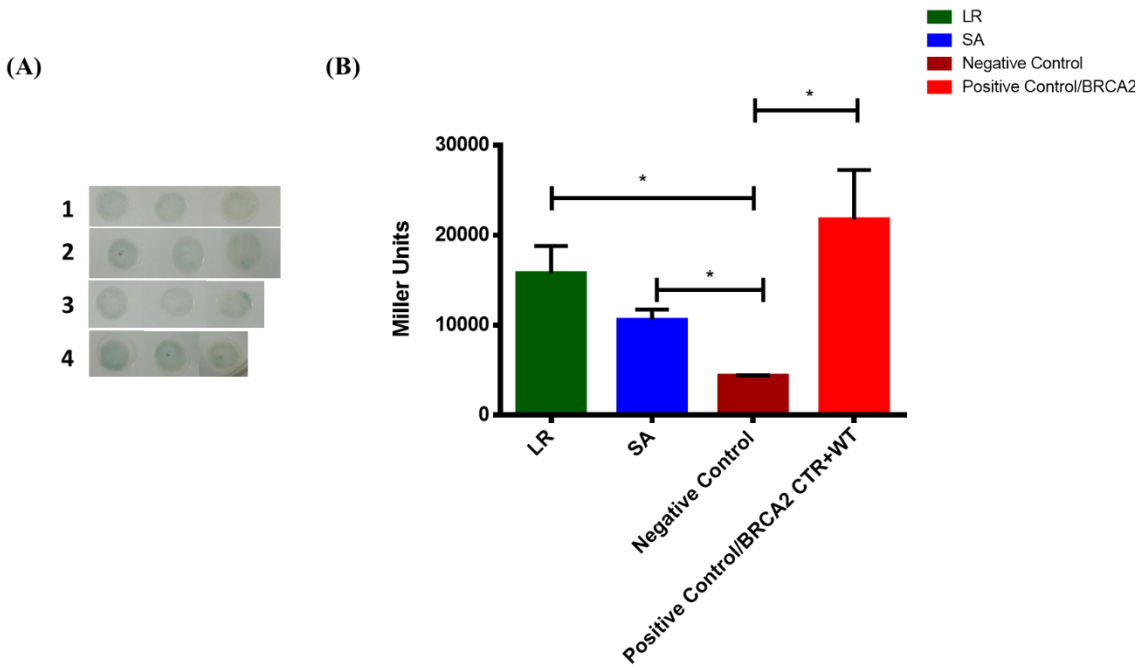


Figure 6.2; (A & B) *In-vivo* protein-protein interaction studies of FANCD2 CUE Domain *wild-type*, Ser222Ala, Leu231Arg mutants with BRCA2 C- Terminal region, 1= FANCD2 CUE Leu231Arg and BRCA2 CTR, 2= FANCD2 CUE Ser222Ala and BRCA2 CTR, 3= Negative control (pUT18 and pKNT25 tag empty vectors), and 4= FANCD2 CUE *wild-type* and BRCA2 CTR (as positive control). Cells co-expressing pUT18 and pKNT25 tag empty vectors were used as negative control, and FANCD2 CUE-pUT18 and BRCA2-CTR-pKNT25 protein was used as positive control, (**B**) b-Galactosidase activity calculated in Miller Units are shown here as mean±SD (n≥2) and the significance differences were analysed using Student's t-test, and P values obtained at 95 % confidence intervals are shown as (*) for <0.05, (**) for <0.01 and (***) for <0.001.

Furthermore, a decrease in Miller units of FANCD2 Ser222Ala (10609.54±1156.80 Miller Units, p=0.016), Leu231Arg mutants (15785.20±3028.03 Miller Units, p=0.033) with BRCA2 CTR compares to FANCD2 CUE domain *wild-type* (21788.06±5473.40 Miller Units, p= 0.046) were observed **(Figure 6.2 A and B)**. Furthermore, binding interaction has been impaired in Ser222Ala mutant compared to Leu231Arg mutant. These data revealed that point mutations mitigate the protein-protein interaction between FANCD2 CUE domain and BRCA2 CTR. Both the experiments were performed independently twice for biological and technical duplicates.

6.3.2 *In-vitro* protein-protein interaction between BRCA2 C-terminal region and FANCD2 CUE domain

BRCA2 C- terminal region (BRCA2 CTR) also binds with FANCG/XRCC [13]. Furthermore, BRCA2 and FANCD2 co-immuno-precipitates in the cellular extracts of Chinese hamster *wild-type* as well as in human cells [12]. Considering the importance of BRCA2 CTR and FANCD2 CUE, GST-pull down assays were performed to support the *in-vivo* protein-protein interactions.

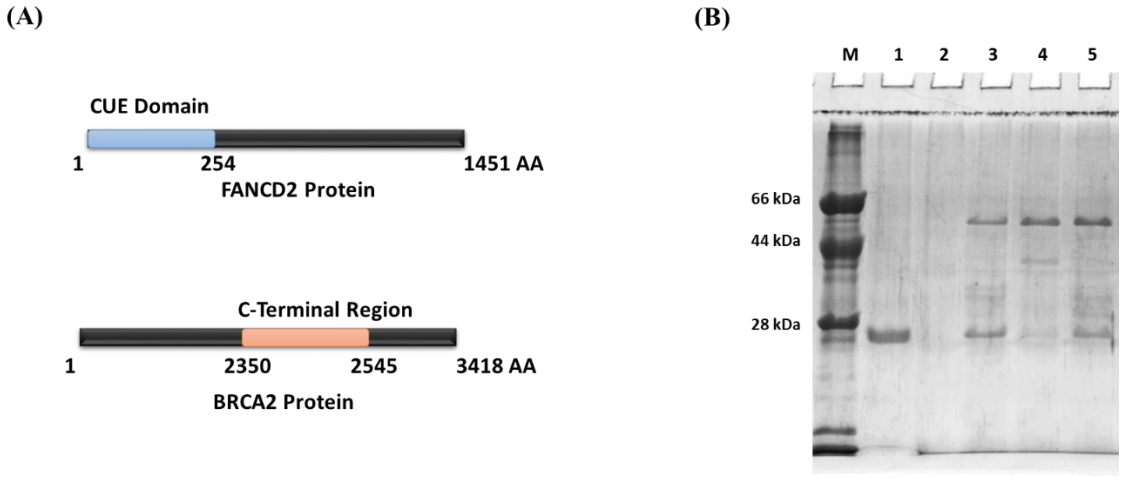


Figure 6.3; (A) Schematic presentation of FANCD2 CUE domain and BRCA2 CTR **(B)** *In-vitro* protein-protein interaction studies of FANCD2 CUE Domain, the supernatant of co-expressed pellet of FANCD2 CUE and BRCA2 CTR were incubated with GST beads and analyzed on SDS-PAGE by using Snapcap column. M= Protein Marker, 1= GST-resin+ GST protein+ BRCA2 CTR, 2= GST-resin+ BRCA2 CTR, 3= GST-resin+ GST-Fusion FD2 CUE *wild-type*+BRCA2 CTR, 4= GST-resin+ GST-Fusion FD2 CUE Ser222Ala+BRCA2 CTR, 5= GST-resin+ GST-Fusion FD2 CUE Leu231Arg+BRCA2 CTR

It has been observed that FANCD2 CUE domain *wild-type* interacts with BRCA2 C-terminal region with high affinity **(Figure 6.3)**. However, FANCD2 Ser222Ala and Leu231Arg mutants were not able to form strong interactions with BRCA2 CTR.

6.3.3 *In-silico* protein-protein interaction between BRCA2 C-terminal region and FANCD2 CUE domain

To further support the *in-vivo* and *in-vitro* results, *in-silico* protein-protein interactions were performed. Binding interfaces were analyzed and reduced interacting surface area were observed in FANCD2 mutant than the FANCD2 CUE domain *wild-type* with mutants (Ser222Ala and Leu231Arg) **(Figure 6.4 and Table 6.2).** Interestingly, from the docked complex of that in the *wild-type,* FANCD2 CUE domain and BRCA2 CTR. Tryptophan at 182 position of the FANCD2 CUE domain is present at the interaction interface, and suggest its important role in PPIs. In general, protein-protein interaction hotspots are known to enrich with tryptophan, raising the possibility that Tryptophan 182 might be playing an important role in PPIs between FANCD2 CUE and BRCA2 CTR.

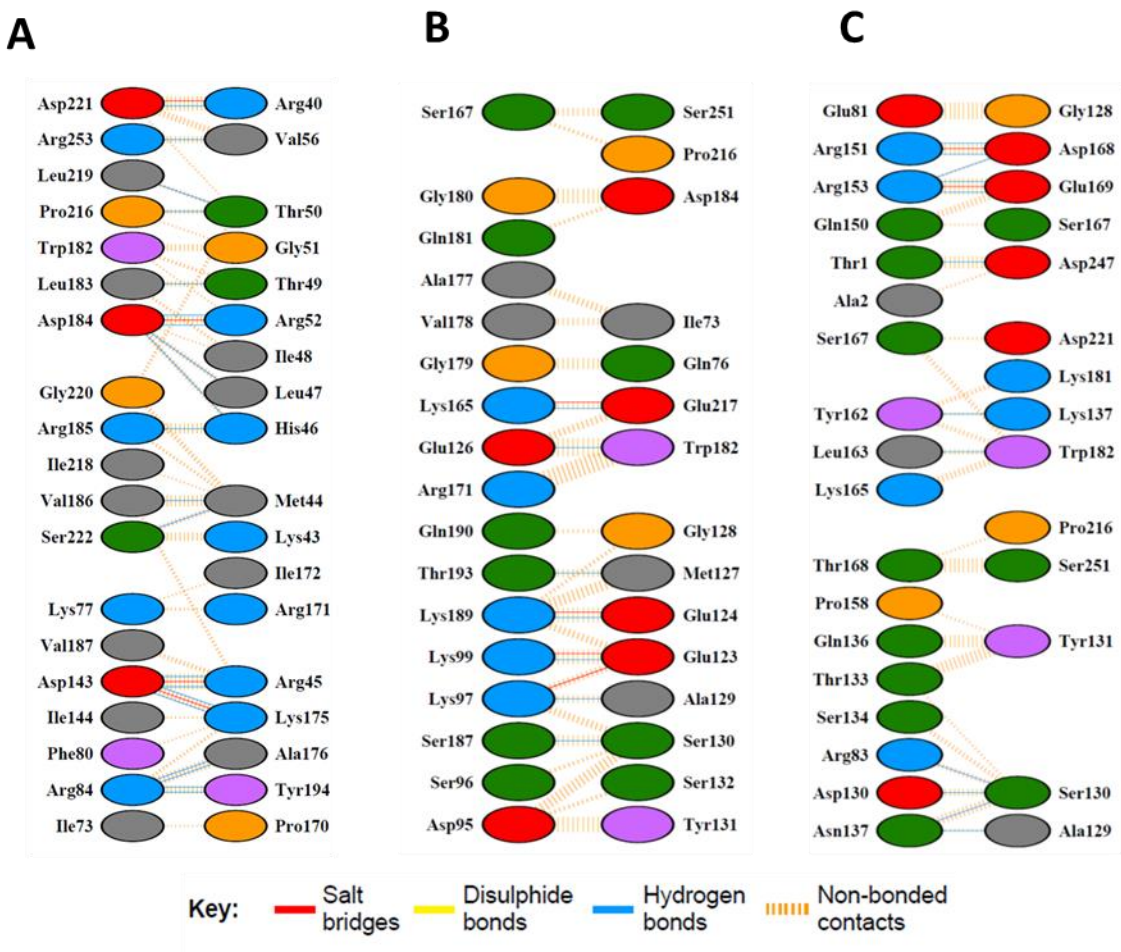


Figure 6.4. *In-silico* protein-protein interactions studies of FANCD2 CUE Domain, FANCD2 CUE Mutants and BRCA2 C- Terminal region, **A**= FANCD2 CUE *wild-type* and BRCA2 CTR **B**= BRCA2 CTR and FAND2 CUE Ser222Ala and **C**= BRCA2 CTR and FANCD2 CUE Leu231Arg

Table 6.2 Interface statistics of BRCA2 CTR with FANCD2 CUE domain and mutants

Interface statistics of BRCA2 (2350-2545) with FD2 Cue domain and mutants			
BRCA2 CTR	**Salt bridges** (In number)	**Hydrogen bond** (In number)	**Interface area** (Å²)
FD2 Cue (*wild type*)	4	20	1041,1046 (Chain A, Chain B)
FD2 Cue (*Ser222Ala*)	4	7	879,909 (Chain A, Chain B)
FD2 Cue (*Leu231Arg*)	2	12	766,795 (Chain A, Chain B)

Furthermore, FANCD2 Ser222Ala mutant showed a significant drop in the number of hydrogen bonds (7 hydrogen bonds) at the protein-protein interface as compared to Leu231Arg (12 hydrogen bonds) mutant. However, both the mutants have less number of hydrogen bonds than the FANCD2 CUE domain *wild-type* **(Table 6.2)**. These data suggest that FANCD2 CUE domain *wild-type* binds effectively to BRCA2 CTR compared to the mutants. Similar findings were also observed from *in-vivo* and *in-vitro* experiments. To our conclusion, FANCD2 CUE domain interacts with BRCA2 C-terminal region whereas the mutants Ser222Ala and Leu231Arg mitigates these interactions.

6.4 Conclusion

FANCI, FANCD2 and BRCA2 play a very important role in DNA ICL repair mechanism [2, 4, 12, 42, 173]. But how the PPIs associated with different functional domains of FA and BRCA proteins have not been explored. We have found that FANCI ARM repeat interacts with FANCD2 CUE domain and BRCA2 C-terminal region. Moreover, FANCD2 CUE domain interacts robustly with BRCA2 C-terminal region. Different mutations present in FANCD2 CUE domain abrogates the binding affinity with BRCA2 CTR. These results suggest that the domains present in FANCD2 and BRCA2 play an important role in PPIs, and mutations cause the impairment. However, it would be worth to perform studies to validate further these findings in a mammalian cell system. Furthermore, studies on specific post-translational modifications such as phosphorylation and monoubiquitylation needed to understand the molecular mechanism of DNA ICL repair. To our conclusion, protein-protein interactions between FANCD2 and BRCA2 might be regulated by these functional domains, which in turn may facilitate DNA ICL damage repair.

Chapter 7

Summary

And

Future

Prospects

The Fanconi anemia pathway comprise the FA/BRCA proteins particularly hereditary breast cancer susceptibility gene products such as BRCA1 and FANCD1/BRCA2 [2, 12, 159]. These proteins function cooperatively in the FA/BRCA pathway to repair ICL damaged DNA and prevent the cellular transformation and tumorigenicity [3]. FA/BRCA pathway activation occurs via site-specific phosphorylation at SQ/TQ motif and mono-ubiquitination on a conserved lysine residue of the FANCI and FANCD2 proteins [3, 41]. These specific modifications target to form nuclear foci where it co-localize with FANCD1/BRCA2, RAD51, PCNA, FAN1 and FAAPs at DNA damage site [2, 3, 5, 343]. The regulation of the mono-ubiquitination of FANCI and FANCD2 as well as its function in DNA ICL repair has not been fully understood. Furthermore, it would be interesting to explore detailed and in-depth knowledge at the structural and functional level for the different conserved domains present in FA/BRCA pathway proteins.

7.1 Fanconi anemia complementation group I (FANCI)

The Fanconi anemia complementation group I (FANCI) protein identified as ATM/ATR substrate facilitates the DNA intercrosslink damage repair [3]. FANCI protein comprises 1328 amino acids [3, 5, 83, 109, 115, 169]. FANCI gene comprises different functional domains including Armadillo repeat (ARM repeat) at C-terminus [3]. The FANCI protein forms complex with FANCD2 also known as the FANCI-FANCD2 (ID) complex. FANCI-FANCD2 (ID) complex are recruited to DNA damage sites to form nuclear foci with other DNA damage repair protein such as RAD51, BRCA2, BRCA1 and FAN1 [3, 59, 85]. FANCI mutant cells are not able to monoubiquitinate FANCD2, and prevent its chromatin localization to repair foci [3]. It was reported that FANCI also has a role in cell -cycle checkpoint. FANCI protein is a leucine-rich protein (LRP) which mediates protein-protein

and protein-DNA interactions [3, 83]. The LRPs are generally composed of ARM, HEAT, leucine-rich repeat and leucine zipper [20, 21, 295, 344].

ARM repeats usually involved in protein-protein and protein-DNA interactions to maintain genomic integrity and regulate different cellular functions [22, 23, 83, 115, 167, 325]. In general, ARM repeats are consist of the tandem copies of degenerate protein sequences [295]. It was first discovered in segment polarity gene of Drosophila and afterward in other proteins like junctional plaque plakoglobin, tumour suppressor adenomatous polyposis coli (APC), nucleocytoplasmic transport factor importin, DNA intercrosslink repair FANCI and FANCD2 proteins [3, 296]. In general, ARM repeat proteins harbours a compact and dynamic region which acts as a molecular recognition component and an interacting module for different binding partners [19, 20]. Considering, the importance of FANCI in DNA ICL repair mechanism, the functional armadillo (ARM) repeat domain of FANCI protein was subjected to interdisciplinary *in-vitro* and *in-silico* approaches to understand the structure and dynamics. FANCI C-terminus harbours ARM repeat, EDGE motif and nuclear localization signal [3]. It was reported that C-terminus of FANCI has a similar DNA-binding affinity as the full length of FANCI [83]. Furthermore, it binds preferentially to double stranded DNA [83]. It has been observed that ARM repeat has structurally dynamic region majorly composed of α-helices. And binding with the major groove of DNA suggest that FANCI ARM repeat harbours HTH structural motif. Molecular modelling, dynamics, and docking studies are in concordance with *in vitro* results derived from limited proteolysis, sequences and structural alignments. Looking at the helix propensity of the amino acids sequences in HTH structural motif and presence of good helix-forming amino acids such as Ala, Glu, Gln and Ser also suggest that the dynamic region of FANCI (100-150) amino acids forms α-helices [318]. It is well known that the HTH structural motif generally binds to the major groove of DNA and conserved

amino acids such as Glu, Gln and Ser are at the binding interface [28, 29, 331, 332]. The *in-silico* docked complex structure of FANCI ARM–DNA interface indicates that the conserved amino acids Gln128, Glu129 and Ser132 are present at the binding region. The molecular dynamics simulation of 80 ns confirms the dynamic behaviour of HTH structural motif. Furthermore, limited proteolysis, mass spectrometry, and principal component analysis results are also supporting that ARM repeat has the compact domain devoid of the HTH motif. In conclusion, FANCI HTH motif might help ARM repeat in DNA binding. However, it would be worth to make different mutants of the conserved amino acids present in HTH region and co-crystallize with the dsDNA. The comparative study of the complexes of *wild-type* and mutants with dsDNA by X -ray crystallography will reveal the atomic level insights of the interactions interface.

7.2 Fanconi anemia complementation group D2 (FANCD2)

Fanconi anemia complementation group D2 (FANCD2) protein is critical for the DNA inter-strand crosslink repair and genome stability [69, 77, 345]. Monoubiquitinated FANCD2 recruited at the DNA damage site and protects the DNA replication forks from nucleolytic degradation [58, 77, 103, 340, 345]. The FANCD2 proteins together with BRCA1 and BRCA2 function cooperatively to repair intercrosslink DNA damage. Thus FA pathway is also known as FA/BRCA pathway [2]. In the FA-BRCA pathway, the activated ATM/ATR kinases phosphorylate the FA core complex that assembles in the nucleus and monoubiquitinates the FANCD2 and FANCI proteins [3]. Monoubiquitinated FANCD2 and FANCI proteins are recruited at DNA damage sites to interact with crucial DNA repair proteins including BRCA1, FANCD1/BRCA2, RAD51, SLX4 and FAN1 endonuclease [3, 173]. FANCD2 mono-ubiquitination at the conserved lysine (Lys 561) is necessary for the recruitment of the FAN1 and SLX4/FANCP endonucleases to the DNA

damage sites. It was found that FANCD2 protein contains an ubiquitin binding domain known as the coupling of ubiquitin conjugation to endoplasmic reticulum (CUE) domain [4]. This ubiquitin binding domain is important for efficient DNA ICL repair and FANCI protein interactions [4]. CUE domain comprises the region from 1-254 amino acids at N-terminus of FANCD2 that regulates the FA pathway by correcting the mitomycin c hypersensitivity and localizing the FANCD2 and FANCI proteins at DNA damage site [4]. This domain is also responsible for the efficient interactions between FANCD2 and FANCI protein which shields the monoubiquitinated FANCI from poly-ubiquitination, as a result regulating protein homeostasis at the time of DNA repair [4].

Ubiquitin-binding domains acts as a regulatory signal in the cell cycle, DNA replication, transcription and DNA repair processes [207, 321]. Due to diverse functions in cellular signaling, different ubiquitin binding regions are characterized in proteins and CUE domains is one of them [322]. Shih *et. al*. reported that CUE domains have a role in ubiquitin binding and intramolecular mono-ubiquitination [336]. Furthermore, two point mutations have been reported in the CUE domain region. First, FANCD2 Ser222Ala mutation responsible for the deregulation of S-phase cell-cycle checkpoint [341, 342] and second is Leu231Arg mutation [90], causing defective DNA ICL repair.

In the present study, we have focused on the important missense mutations and described the structural and functional aspect of FANCD2 CUE domain and mutants using *in- vivo, in-vitro* and *in-silico* biophysical approach. It has been found that CUE domain *wild type* is stable and binds to the mono-ubiquitin efficiently compared to DNA ICL repair mutant Leu231Arg. In addition, the CUE domain has the chaperone-like activity that gets impaired in mutants. Interaction interface of *wild-type* CUE domain and mutants with mono-ubiquitin has been calculated. We have analyzed the amino acids residues relevant to interaction hotspots in CUE domain and mutants with mono-ubiquitin in docked

complexes. To our knowledge, it is the first report to describe the comparison of DNA ICL repair mutant with the *wild-type* CUE domain at the structure, dynamics, and function level. Further comparative interaction studies of the *wild-type*/mutants CUE domain with mono-ubiquitin, di-ubiquitin and tetra-ubiquitin is required to get the molecular insights of ubiquitin signaling.

7.3 Breast cancer susceptibility gene 2 (BRCA2)

BRCA2 has been involved in DNA double-strand break repair through homologous recombination [159, 346]. The DNA repair pathway in BRCA2 deficient cells leads to error prone single-strand annealing which regulates the DNA repair processes such as nucleolytic incision, translesion DNA synthesis, homologous recombination [12, 13, 76, 159, 196, 346]. These processes are coordinated by FA and BRCA proteins. The C-terminus region of BRCA2 harbors BLAT (BRCA2-motif in *Leishmania, Arabidopsis, and Trypanosoma*) domain, which is highly conserved in divergent species [12, 13]. The BLAT domain harbors BRCA2 region encompassing (2350-2545) amino acids that interact with FANCD2 (1-1102) amino acids [347]. However, FANCD2 lacking N-terminus residues (1-325) failed to interact with BRCA2 [12]. Importantly expression of BRCA2 proteins with C-terminal truncations were observed in cell lines derived from Fanconi Anemia patients [2]. This confirms the importance of FANCD2 N-terminus and BRCA2 C-terminus in the protein-protein interactions (PPIs). The FANCI and FANCD2 also known to interact with BRCA2 [2]. Considering the importance of mutations, and functional diversity of different domains present in FANCI, FANCD2, and BRCA2, we systematically evaluated the protein-protein interactions between FANCI ARM repeat, FANCD2 CUE domain (FD2 CUE) and BRCA2 C-terminal region (BRCA2 CTR). We have also analyzed the interactions of mutated FANCD2 CUE domain (Ser222Ala and Leu231Arg) with BRCA2 C-terminal region using *in-vivo*, *in-vitro* and *in-silico* advancement.

The results presented in this thesis suggest that the different regions present in FANCI, FANCD2, and BRCA2 play an important role in PPIs, and mutations cause the impairment. In conclusion, protein-protein interaction network between different domains of FANCI, FANCD2, and BRCA2, play an important role in functions associated to PPIs, which in turn may regulate DNA ICL damage repair mechanism.

7.4 Future Prospects

The structural and functional analysis of FANCI and FANCD2 domains were studied, to get molecular insights of ICL DNA damage repair. Two important missense mutations and their disease correlation are connected to FANCI and FANCD2 protein stability as well as its FANCI-FANCD2 complex integrity. The interaction studies were carried out to explore the biological relevance of the functional domains present in FANCI, FANCD2, and BRCA2. However, the next approach would be the expression of human FANCI and FANCD2 full-length protein in higher expression system such as yeast, mammalian or insect cells to get a well-folded protein with all desired post-translation modifications. The main drawback in the bacterial system is oligomerization which reduces the propensity of FANCI ARM repeat for crystallization. Therefore, a rational approach of cloning into eukaryotic expression system could result in a well folded homogenous protein which most likely can produce the better quality crystal for good diffraction pattern. Further, co-crystallization of FANCI-FANCD2 and or/with other FA/BRCA pathway members such as BRCA1, BRCA2, FAN1, RAD 51 and PALB2 will unravel the atomic association of complexes. The major challenges were the purification of functional and homogenous FANCI, FANCD2 full-length protein at high concentration. The issue can be significantly resolved by its co-expression with other members in insect or mammalian expression system. In addition, the role of FANCD2 needs to be explored more in ubiquitination/proteasome signaling.

References

1. Cohen, M.M., et al., *The identification of fanconi anemia genotypes by clastogenic stress.* Am J Hum Genet, 1982. **34**(5): p. 794-810.
2. D'Andrea, A.D. and M. Grompe, *The Fanconi anaemia/BRCA pathway.* Nat Rev Cancer, 2003. **3**(1): p. 23-34.
3. Smogorzewska, A., et al., *Identification of the FANCI protein, a monoubiquitinated FANCD2 paralog required for DNA repair.* Cell, 2007. **129**(2): p. 289-301.
4. Rego, M.A., et al., *Regulation of the Fanconi anemia pathway by a CUE ubiquitin-binding domain in the FANCD2 protein.* Blood, 2012. **120**(10): p. 2109-17.
5. Castella, M., et al., *FANCI Regulates Recruitment of the FA Core Complex at Sites of DNA Damage Independently of FANCD2.* PLoS Genet, 2015. **11**(10): p. e1005563.
6. Alter, B.P., et al., *Cancer in Fanconi anemia.* Blood, 2003. **101**(5): p. 2072.
7. Rosenberg, P.S., B.P. Alter, and W. Ebell, *Cancer risks in Fanconi anemia: findings from the German Fanconi Anemia Registry.* Haematologica, 2008. **93**(4): p. 511-7.
8. Rosenberg, P.S., M.H. Greene, and B.P. Alter, *Cancer incidence in persons with Fanconi anemia.* Blood, 2003. **101**(3): p. 822-6.
9. Mitra, P., et al., *Evidence for the Nucleo-Apical Shuttling of a Beta-Catenin Like Plasmodium falciparum Armadillo Repeat Containing Protein.* PloS one, 2016. **11**(2): p. e0148446.
10. Ameziane, N., et al., *Lack of large genomic deletions in BRIP1, PALB2, and FANCD2 genes in BRCA1/2 negative familial breast cancer.* Breast Cancer Res Treat, 2009. **118**(3): p. 651-3.
11. Airik, R., et al., *A FANCD2/FANCI-Associated Nuclease 1-Knockout Model Develops Karyomegalic Interstitial Nephritis.* J Am Soc Nephrol, 2016. **27**(12): p. 3552-3559.
12. Hussain, S., et al., *Direct interaction of FANCD2 with BRCA2 in DNA damage response pathways.* Hum Mol Genet, 2004. **13**(12): p. 1241-8.
13. Hussain, S., et al., *Direct interaction of the Fanconi anaemia protein FANCG with BRCA2/FANCD1.* Hum Mol Genet, 2003. **12**(19): p. 2503-10.
14. Hess, B., et al., *GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation.* J Chem Theory Comput, 2008. **4**(3): p. 435-47.
15. Bahar, I., et al., *Global dynamics of proteins: bridging between structure and function.* Annu Rev Biophys, 2010. **39**: p. 23-42.
16. Bakan, A., L.M. Meireles, and I. Bahar, *ProDy: protein dynamics inferred from theory and experiments.* Bioinformatics, 2011. **27**(11): p. 1575-7.
17. van Dijk, M., et al., *Information-driven protein-DNA docking using HADDOCK: it is a matter of flexibility.* Nucleic Acids Res, 2006. **34**(11): p. 3317-25.
18. Macindoe, G., et al., *HexServer: an FFT-based protein docking server powered by graphics processors.* Nucleic Acids Res, 2010. **38**(Web Server issue): p. W445-9.
19. Tewari, R., et al., *Armadillo-repeat protein functions: questions for little creatures.* Trends Cell Biol, 2010. **20**(8): p. 470-81.
20. Tsytlonok, M., et al., *Complex energy landscape of a giant repeat protein.* Structure, 2013. **21**(11): p. 1954-65.
21. Coates, J., *Armadillo repeat proteins: beyond the animal kingdom.* Trends Cell Biol, 2003. **13**(9): p. 463-471.
22. Crist, R.C., et al., *The armadillo repeat domain of Apc suppresses intestinal tumorigenesis.* Mamm Genome, 2010. **21**(9-10): p. 450-7.
23. Drechsel, G., et al., *C-terminal armadillo repeats are essential and sufficient for association of the plant U-box armadillo E3 ubiquitin ligase SAUL1 with the plasma membrane.* J Exp Bot, 2011. **62**(2): p. 775-85.
24. Andreassen, P.R., A.D. D'Andrea, and T. Taniguchi, *ATR couples FANCD2 monoubiquitination to the DNA-damage response.* Genes Dev, 2004. **18**(16): p. 1958-63.
25. Longerich, S., et al., *Regulation of FANCD2 and FANCI monoubiquitination by their interaction and by DNA.* Nucleic Acids Res, 2014. **42**(9): p. 5657-70.

26. Joo, W., et al., *Structure of the FANCI-FANCD2 complex: insights into the Fanconi anemia DNA repair pathway.* Science, 2011. **333**(6040): p. 312-6.
27. Aravind, L., et al., *The many faces of the helix-turn-helix domain: transcription regulation and beyond.* FEMS Microbiol Rev, 2005. **29**(2): p. 231-62.
28. Brennan, R.G., *The winged-helix DNA-binding motif: another helix-turn-helix takeoff.* Cell, 1993. **74**(5): p. 773-6.
29. Brennan, R.G. and B.W. Matthews, *The helix-turn-helix DNA binding motif.* J Biol Chem, 1989. **264**(4): p. 1903-6.
30. Andreassen, P.R. and K. Ren, *Fanconi anemia proteins, DNA interstrand crosslink repair pathways, and cancer therapy.* Curr Cancer Drug Targets, 2009. **9**(1): p. 101-17.
31. Alter, B.P., *Cancer in Fanconi anemia, 1927-2001.* Cancer, 2003. **97**(2): p. 425-40.
32. Akbari, M.R., et al., *Mutations in Fanconi anemia genes and the risk of esophageal cancer.* Hum Genet, 2011. **129**(5): p. 573-82.
33. Esteban-Jurado, C., et al., *The Fanconi anemia DNA damage repair pathway in the spotlight for germline predisposition to colorectal cancer.* Eur J Hum Genet, 2016. **24**(10): p. 1501-5.
34. Gallmeier, E. and S.E. Kern, *Targeting Fanconi anemia/BRCA2 pathway defects in cancer: the significance of preclinical pharmacogenomic models.* Clin Cancer Res, 2007. **13**(1): p. 4-10.
35. Auerbach, A.D. and R.G. Allen, *Leukemia and preleukemia in Fanconi anemia patients. A review of the literature and report of the International Fanconi Anemia Registry.* Cancer Genet Cytogenet, 1991. **51**(1): p. 1-12.
36. Barroso, E., et al., *The Fanconi anemia family of genes and its correlation with breast cancer susceptibility and breast cancer features.* Breast Cancer Res Treat, 2009. **118**(3): p. 655-60.
37. Dokal, I., *Fanconi anemia is a highly penetrant cancer susceptibility syndrome.* Haematologica, 2008. **93**(4): p. 486-8.
38. Berwick, M., et al., *Genetic heterogeneity among Fanconi anemia heterozygotes and risk of cancer.* Cancer Res, 2007. **67**(19): p. 9591-6.
39. Garcia, M.J. and J. Benitez, *The Fanconi anaemia/BRCA pathway and cancer susceptibility. Searching for new therapeutic targets.* Clin Transl Oncol, 2008. **10**(2): p. 78-84.
40. Duan, W., et al., *Assessment of FANCD2 nuclear foci formation in paraffin-embedded tumors: a potential patient-enrichment strategy for treatment with DNA interstrand crosslinking agents.* Transl Res, 2013. **161**(3): p. 156-64.
41. Garcia-Higuera, I., et al., *Interaction of the Fanconi anemia proteins and BRCA1 in a common pathway.* Mol Cell, 2001. **7**(2): p. 249-62.
42. Folias, A., et al., *BRCA1 interacts directly with the Fanconi anemia protein FANCA.* Hum Mol Genet, 2002. **11**(21): p. 2591-7.
43. Tutt, A., et al., *Mutation in Brca2 stimulates error-prone homology-directed repair of DNA double-strand breaks occurring between repeated sequences.* EMBO J, 2001. **20**(17): p. 4704-16.
44. Ceccaldi, R., P. Sarangi, and A.D. D'Andrea, *The Fanconi anaemia pathway: new players and new functions.* Nat Rev Mol Cell Biol, 2016. **17**(6): p. 337-49.
45. King, M.C., et al., *Breast and ovarian cancer risks due to inherited mutations in BRCA1 and BRCA2.* Science, 2003. **302**(5645): p. 643-6.
46. Manguoglu, A.E., et al., *Germline mutations in the BRCA1 and BRCA2 genes in Turkish breast/ovarian cancer patients.* Hum Mutat, 2003. **21**(4): p. 444-5.
47. Ottini, L., et al., *BRCA1 and BRCA2 mutation status and tumor characteristics in male breast cancer: a population-based study in Italy.* Cancer Res, 2003. **63**(2): p. 342-7.
48. Tischkowitz, M., et al., *Mutations in BRCA1 and BRCA2 and predisposition to prostate cancer.* Lancet, 2003. **362**(9377): p. 80; author reply 80.
49. Vikrant, et al., *Mislocalization of BRCA1-complex due to ABRAXAS Arg361Gln mutation.* J Biomol Struct Dyn, 2015. **33**(6): p. 1291-301.

50. Vandenberg, C.J., et al., *BRCA1-independent ubiquitination of FANCD2.* Molecular cell, 2003. **12**(1): p. 247-254.
51. Choudhary, R.K., et al., *Structural basis to stabilize the domain motion of BARD1-ARD BRCT by CstF50.* Sci Rep, 2017. **7**(1): p. 3849.
52. Castillo, A., et al., *The BRCA1-interacting protein Abraxas is required for genomic stability and tumor suppression.* Cell reports, 2014. **8**(3): p. 807-817.
53. Huberman, J.A., *Eukaryotic DNA replication: a complex picture partially clarified.* Cell, 1987. **48**(1): p. 7-8.
54. Knipscheer, P., et al., *The Fanconi anemia pathway promotes replication-dependent DNA interstrand cross-link repair.* Science, 2009. **326**(5960): p. 1698-701.
55. Kais, Z., et al., *FANCD2 Maintains Fork Stability in BRCA1/2-Deficient Tumors and Promotes Alternative End-Joining DNA Repair.* Cell Rep, 2016. **15**(11): p. 2488-99.
56. Langevin, F., et al., *Fancd2 counteracts the toxic effects of naturally produced aldehydes in mice.* Nature, 2011. **475**(7354): p. 53-8.
57. Lewis, A.G., et al., *Mutation analysis of FANCD2, BRIP1/BACH1, LMO4 and SFN in familial breast cancer.* Breast Cancer Res, 2005. **7**(6): p. R1005-16.
58. Liang, C.C., et al., *The FANCD2-FANCI complex is recruited to DNA interstrand crosslinks before monoubiquitination of FANCD2.* Nat Commun, 2016. **7**: p. 12124.
59. Lossaint, G., et al., *FANCD2 binds MCM proteins and controls replisome function upon activation of s phase checkpoint signaling.* Mol Cell, 2013. **51**(5): p. 678-90.
60. Michl, J., et al., *FANCD2 limits replication stress and genome instability in cells lacking BRCA2.* Nat Struct Mol Biol, 2016. **23**(8): p. 755-7.
61. Moldovan, G.L. and A.D. D'Andrea, *FANCD2 hurdles the DNA interstrand crosslink.* Cell, 2009. **139**(7): p. 1222-4.
62. Nakanishi, K., et al., *Interaction of FANCD2 and NBS1 in the DNA damage response.* Nat Cell Biol, 2002. **4**(12): p. 913-20.
63. Couch, F.J., et al., *Germ line Fanconi anemia complementation group C mutations and pancreatic cancer.* Cancer Res, 2005. **65**(2): p. 383-6.
64. MacKay, C., et al., *Identification of KIAA1018/FAN1, a DNA repair nuclease recruited to DNA damage by monoubiquitinated FANCD2.* Cell, 2010. **142**(1): p. 65-76.
65. Mann, W.R., et al., *Fanconi anemia: evidence for linkage heterogeneity on chromosome 20q.* Genomics, 1991. **9**(2): p. 329-37.
66. Thompson, P.W., G.R. Standen, and A.D. Geddes, *Transient t(Y;1)(q12;q21) in a patient with Fanconi anemia and myelodysplastic syndrome.* Cancer Genet Cytogenet, 1991. **52**(2): p. 201-2.
67. Timmers, C., et al., *Positional cloning of a novel Fanconi anemia gene, FANCD2.* Mol Cell, 2001. **7**(2): p. 241-8.
68. Chaudhury, I., et al., *FANCD2 regulates BLM complex functions independently of FANCI to promote replication fork recovery.* Nucleic Acids Res, 2013. **41**(13): p. 6444-59.
69. Chaudhury, I., D.R. Stroik, and A. Sobeck, *FANCD2-controlled chromatin access of the Fanconi-associated nuclease FAN1 is crucial for the recovery of stalled replication forks.* Mol Cell Biol, 2014. **34**(21): p. 3939-54.
70. Sobeck, A., S. Stone, and M.E. Hoatlin, *DNA structure-induced recruitment and activation of the Fanconi anemia pathway protein FANCD2.* Mol Cell Biol, 2007. **27**(12): p. 4283-92.
71. Singh, T.R., et al., *Impaired FANCD2 monoubiquitination and hypersensitivity to camptothecin uniquely characterize Fanconi anemia complementation group M.* Blood, 2009. **114**(1): p. 174-80.
72. Sobeck, A., et al., *The Fanconi anemia protein FANCM is controlled by FANCD2 and the ATR/ATM pathways.* J Biol Chem, 2009. **284**(38): p. 25560-8.

73. Buisson, R., et al., *Breast cancer proteins PALB2 and BRCA2 stimulate polymerase eta in recombination-associated DNA synthesis at blocked replication forks.* Cell Rep, 2014. **6**(3): p. 553-64.
74. Burdak-Rothkamm, S., et al., *BRCA1, FANCD2 and Chk1 are potential molecular targets for the modulation of a radiation-induced DNA damage response in bystander cells.* Cancer Lett, 2015. **356**(2 Pt B): p. 454-61.
75. Chen, X., et al., *A novel role for non-ubiquitinated FANCD2 in response to hydroxyurea-induced DNA damage.* Oncogene, 2016. **35**(1): p. 22-34.
76. Costanzo, V., *Brca2, Rad51 and Mre11: performing balancing acts on replication forks.* DNA Repair (Amst), 2011. **10**(10): p. 1060-5.
77. Chen, X., et al., *The Fanconi anemia proteins FANCD2 and FANCJ interact and regulate each other's chromatin localization.* J Biol Chem, 2014. **289**(37): p. 25774-82.
78. Cox, M.M., et al., *The importance of repairing stalled replication forks.* Nature, 2000. **404**(6773): p. 37-41.
79. Michel, B., et al., *Rescue of arrested replication forks by homologous recombination.* Proc Natl Acad Sci U S A, 2001. **98**(15): p. 8181-8.
80. Sengupta, S., et al., *BLM helicase-dependent transport of p53 to sites of stalled DNA replication forks modulates homologous recombination.* EMBO J, 2003. **22**(5): p. 1210-22.
81. Grompe, M., *FANCD2: a branch-point in DNA damage response?* Nat Med, 2002. **8**(6): p. 555-6.
82. Houghtaling, S., et al., *Fancd2 functions in a double strand break repair pathway that is distinct from non-homologous end joining.* Hum Mol Genet, 2005. **14**(20): p. 3027-33.
83. Longerich, S., et al., *FANCI binds branched DNA and is monoubiquitinated by UBE2T-FANCL.* J Biol Chem, 2009. **284**(35): p. 23182-6.
84. Siddiqui, M.Q., et al., *Structural and biophysical properties of h-FANCI ARM repeat protein.* J Biomol Struct Dyn, 2016: p. 1-11.
85. Sareen, A., et al., *Fanconi anemia proteins FANCD2 and FANCI exhibit different DNA damage responses during S-phase.* Nucleic Acids Res, 2012. **40**(17): p. 8425-39.
86. Sato, K., et al., *Defective FANCI binding by a fanconi anemia-related FANCD2 mutant.* PLoS One, 2014. **9**(12): p. e114752.
87. Kim, H. and A.D. D'Andrea, *Regulation of DNA cross-link repair by the Fanconi anemia/BRCA pathway.* Genes Dev, 2012. **26**(13): p. 1393-408.
88. Lachaud, C., et al., *Ubiquitinated Fancd2 recruits Fan1 to stalled replication forks to prevent genome instability.* Science, 2016. **351**(6275): p. 846-9.
89. Colnaghi, L., et al., *Patient-derived C-terminal mutation of FANCI causes protein mislocalization and reveals putative EDGE motif function in DNA repair.* Blood, 2011. **117**(7): p. 2247-56.
90. Kalb, R., et al., *Hypomorphic mutations in the gene encoding a key Fanconi anemia protein, FANCD2, sustain a significant group of FA-D2 patients with severe phenotype.* Am J Hum Genet, 2007. **80**(5): p. 895-910.
91. Vazharova, R., et al., *Novel Hypomorphic Mutation in FANCD2 Gene Observed in a Fetus with Multiple Congenital Anomalies.* Case Rep Obstet Gynecol, 2016. **2016**: p. 1462818.
92. Houghtaling, S., et al., *Epithelial cancer in Fanconi anemia complementation group D2 (Fancd2) knockout mice.* Genes Dev, 2003. **17**(16): p. 2021-35.
93. Jin, S., et al., *Menin associates with FANCD2, a protein involved in repair of DNA damage.* Cancer Res, 2003. **63**(14): p. 4204-10.
94. Parmar, K., A. D'Andrea, and L.J. Niedernhofer, *Mouse models of Fanconi anemia.* Mutat Res, 2009. **668**(1-2): p. 133-40.
95. Dong, H., et al., *Update of the human and mouse Fanconi anemia genes.* Hum Genomics, 2015. **9**: p. 32.

96. Cappelli, E., D. Vecchio, and G. Frosina, *Delayed formation of FancD2 foci in glioma stem cells treated with ionizing radiation.* J Cancer Res Clin Oncol, 2012. **138**(5): p. 897-9.
97. Collis, S.J., et al., *C. elegans FANCD2 responds to replication stress and functions in interstrand cross-link repair.* DNA Repair (Amst), 2006. **5**(11): p. 1398-406.
98. Kachnic, L.A., et al., *FANCD2 but not FANCA promotes cellular resistance to type II topoisomerase poisons.* Cancer Lett, 2011. **305**(1): p. 86-93.
99. Kuhnert, V.M., et al., *FANCD2-deficient human fibroblasts are hypersensitive to ionising radiation at oxygen concentrations of 0% and 3% but not under normoxic conditions.* Int J Radiat Biol, 2009. **85**(6): p. 523-31.
100. Li, X., et al., *Fancd2 is required for nuclear retention of Foxo3a in hematopoietic stem cell maintenance.* J Biol Chem, 2015. **290**(5): p. 2715-27.
101. Liu, T., et al., *FAN1 acts with FANCI-FANCD2 to promote DNA interstrand cross-link repair.* Science, 2010. **329**(5992): p. 693-6.
102. Lyakhovich, A., et al., *Fanconi anemia protein FANCD2 inhibits TRF1 polyADP-ribosylation through tankyrase1-dependent manner.* Genome Integr, 2011. **2**(1): p. 4.
103. Madireddy, A., et al., *FANCD2 Facilitates Replication through Common Fragile Sites.* Mol Cell, 2016. **64**(2): p. 388-404.
104. Matsushita, N., et al., *A FancD2-monoubiquitin fusion reveals hidden functions of Fanconi anemia core complex in DNA repair.* Mol Cell, 2005. **19**(6): p. 841-7.
105. Murina, O., et al., *FANCD2 and CtIP cooperate to repair DNA interstrand crosslinks.* Cell Rep, 2014. **7**(4): p. 1030-8.
106. Park, E., et al., *FANCD2 activates transcription of TAp63 and suppresses tumorigenesis.* Mol Cell, 2013. **50**(6): p. 908-18.
107. Patil, A.A., et al., *FANCD2 re-expression is associated with glioma grade and chemical inhibition of the Fanconi Anaemia pathway sensitises gliomas to chemotherapeutic agents.* Oncotarget, 2014. **5**(15): p. 6414-24.
108. Pilonetto, D.V., et al., *FANCD2 Western blot as a diagnostic tool for Brazilian patients with Fanconi anemia.* Braz J Med Biol Res, 2009. **42**(3): p. 237-43.
109. Sato, K., et al., *FANCI-FANCD2 stabilizes the RAD51-DNA complex by binding RAD51 and protects the 5'-DNA end.* Nucleic Acids Res, 2016.
110. Shen, C., et al., *Regulation of FANCD2 by the mTOR pathway contributes to the resistance of cancer cells to DNA double-strand breaks.* Cancer Res, 2013. **73**(11): p. 3393-401.
111. Song, X., et al., *FANCD2 protects against bone marrow injury from ferroptosis.* Biochem Biophys Res Commun, 2016. **480**(3): p. 443-449.
112. van der Groep, P., et al., *Loss of expression of FANCD2 protein in sporadic and hereditary breast cancer.* Breast Cancer Res Treat, 2008. **107**(1): p. 41-7.
113. Vilcheck, S.K., et al., *FANCD2 monoubiquitination and activation by hexavalent chromium [Cr(VI)] exposure: activation is not required for repair of Cr(VI)-induced DSBs.* Mutat Res, 2006. **610**(1-2): p. 21-30.
114. Yang, Y., et al., *FANCD2 and REV1 cooperate in the protection of nascent DNA strands in response to replication stress.* Nucleic Acids Res, 2015. **43**(17): p. 8325-39.
115. Yuan, F., et al., *FANCI protein binds to DNA and interacts with FANCD2 to recognize branched structures.* J Biol Chem, 2009. **284**(36): p. 24443-52.
116. Zhang, Q.S., et al., *Fancd2-/- mice have hematopoietic defects that can be partially corrected by resveratrol.* Blood, 2010. **116**(24): p. 5140-8.
117. Zhang, Q.S., et al., *Fancd2 and p21 function independently in maintaining the size of hematopoietic stem and progenitor cell pool in mice.* Stem Cell Res, 2013. **11**(2): p. 687-92.
118. Alter, B.P. and S.B. Olson, *Wilms tumor, AML, and medulloblastoma in a child with cancer prone syndrome of total premature chromatid separation and Fanconi anemia.* Pediatr Blood Cancer, 2010. **54**(3): p. 488; author reply 489.

119. Baris, H.N., et al., *Prevalence of breast and colorectal cancer in Ashkenazi Jewish carriers of Fanconi anemia and Bloom syndrome.* Isr Med Assoc J, 2007. **9**(12): p. 847-50.
120. Barroso, E., et al., *FANCD2 associated with sporadic breast cancer risk.* Carcinogenesis, 2006. **27**(9): p. 1930-7.
121. Borriello, A., et al., *A novel Leu153Ser mutation of the Fanconi anemia FANCD2 gene is associated with severe chemotherapy toxicity in a pediatric T-cell acute lymphoblastic leukemia.* Leukemia, 2007. **21**(1): p. 72-8.
122. Butturini, A., et al., *Hematologic abnormalities in Fanconi anemia: an International Fanconi Anemia Registry study.* Blood, 1994. **84**(5): p. 1650-5.
123. Du, W., et al., *Deletion of Fanca or Fancd2 dysregulates Treg in mice.* Blood, 2014. **123**(12): p. 1938-47.
124. Ferrer, M., et al., *FANCD2 expression in advanced non-small-cell lung cancer and response to platinum-based chemotherapy.* Clin Lung Cancer, 2005. **6**(4): p. 250-4.
125. Patel, K.J., *Fanconi anemia and breast cancer susceptibility.* Nat Genet, 2007. **39**(2): p. 142-3.
126. Romick-Rosendale, L.E., et al., *Defects in the Fanconi Anemia Pathway in Head and Neck Cancer Cells Stimulate Tumor Cell Invasion through DNA-PK and Rac1 Signaling.* Clin Cancer Res, 2016. **22**(8): p. 2062-73.
127. Rosenberg, P.S., et al., *Risk of head and neck squamous cell cancer and death in patients with Fanconi anemia who did and did not receive transplants.* Blood, 2005. **105**(1): p. 67-73.
128. Rudland, P.S., et al., *Significance of the Fanconi anemia FANCD2 protein in sporadic and metastatic human breast cancer.* Am J Pathol, 2010. **176**(6): p. 2935-47.
129. Seal, S., et al., *Evaluation of Fanconi Anemia genes in familial breast cancer predisposition.* Cancer Res, 2003. **63**(24): p. 8596-9.
130. Rochowski, A., et al., *Patients with Fanconi anemia and AML have different cytogenetic clones than de novo cases of AML.* Pediatr Blood Cancer, 2012. **59**(5): p. 922-4.
131. Nalepa, G., et al., *Fanconi anemia signaling network regulates the spindle assembly checkpoint.* J Clin Invest, 2013. **123**(9): p. 3839-47.
132. Ho, G.P., et al., *Phosphorylation of FANCD2 on two novel sites is required for mitomycin C resistance.* Mol Cell Biol, 2006. **26**(18): p. 7005-15.
133. McDonough, E.R., *Fanconi anemia syndrome.* Arch Otolaryngol, 1970. **92**(3): p. 284-5.
134. Michl, J., J. Zimmer, and M. Tarsounas, *Interplay between Fanconi anemia and homologous recombination pathways in genome integrity.* EMBO J, 2016. **35**(9): p. 909-23.
135. Abraham, J., et al., *Alcohol metabolism in human cells causes DNA damage and activates the Fanconi anemia-breast cancer susceptibility (FA-BRCA) DNA damage response network.* Alcohol Clin Exp Res, 2011. **35**(12): p. 2113-20.
136. Ghosh, S., et al., *Hypersensitivities for acetaldehyde and other agents among cancer cells null for clinically relevant Fanconi anemia genes.* Am J Pathol, 2014. **184**(1): p. 260-70.
137. Marietta, C., et al., *Acetaldehyde stimulates FANCD2 monoubiquitination, H2AX phosphorylation, and BRCA1 phosphorylation in human cells in vitro: implications for alcohol-related carcinogenesis.* Mutat Res, 2009. **664**(1-2): p. 77-83.
138. Rulten, S.L., et al., *Alcohol induces DNA damage and the Fanconi anemia D2 protein implicating FANCD2 in the DNA damage response pathways in brain.* Alcohol Clin Exp Res, 2008. **32**(7): p. 1186-96.
139. Ren, X., et al., *The impact of FANCD2 deficiency on formaldehyde-induced toxicity in human lymphoblastoid cell lines.* Arch Toxicol, 2013. **87**(1): p. 189-96.
140. Burkitt, K. and M. Ljungman, *Phenylbutyrate interferes with the Fanconi anemia and BRCA pathway and sensitizes head and neck cancer cells to cisplatin.* Mol Cancer, 2008. **7**: p. 24.
141. Dudek, A.Z., et al., *Neoadjuvant chemotherapy with reduced-dose carboplatin and gemcitabine for non-small cell lung cancer in a patient with Fanconi anemia.* J Thorac Oncol, 2008. **3**(4): p. 447-50.

142. Jacquemont, C., et al., *Non-specific chemical inhibition of the Fanconi anemia pathway sensitizes cancer cells to cisplatin.* Mol Cancer, 2012. **11**: p. 26.
143. Willers, H., et al., *Biomarkers and mechanisms of FANCD2 function.* J Biomed Biotechnol, 2008. **2008**: p. 821529.
144. Alpi, A.F. and K.J. Patel, *Monoubiquitylation in the Fanconi anemia DNA damage response pathway.* DNA Repair (Amst), 2009. **8**(4): p. 430-5.
145. Wang, W., *Emergence of a DNA-damage response network consisting of Fanconi anaemia and BRCA proteins.* Nat Rev Genet, 2007. **8**(10): p. 735-48.
146. Youds, J.L., L.J. Barber, and S.J. Boulton, *C. elegans: a model of Fanconi anemia and ICL repair.* Mutat Res, 2009. **668**(1-2): p. 103-16.
147. Hira, A., et al., *Mutations in the gene encoding the E2 conjugating enzyme UBE2T cause Fanconi anemia.* Am J Hum Genet, 2015. **96**(6): p. 1001-7.
148. Miles, J.A., et al., *The Fanconi Anemia DNA Repair Pathway Is Regulated by an Interaction between Ubiquitin and the E2-like Fold Domain of FANCL.* J Biol Chem, 2015. **290**(34): p. 20995-1006.
149. Takahashi, D., et al., *Human FAN1 promotes strand incision in 5'-flapped DNA complexed with RPA.* J Biochem, 2015. **158**(3): p. 263-70.
150. Zhao, Q., et al., *Structural insights into 5' flap DNA unwinding and incision by the human FAN1 dimer.* Nat Commun, 2014. **5**: p. 5726.
151. Guervilly, J.H., et al., *The SLX4 complex is a SUMO E3 ligase that impacts on replication stress outcome and genome stability.* Mol Cell, 2015. **57**(1): p. 123-37.
152. McHugh, P.J., W.R. Sones, and J.A. Hartley, *Repair of intermediate structures produced at DNA interstrand cross-links in Saccharomyces cerevisiae.* Mol Cell Biol, 2000. **20**(10): p. 3425-33.
153. Sarkar, S., et al., *DNA interstrand crosslink repair during G1 involves nucleotide excision repair and DNA polymerase zeta.* EMBO J, 2006. **25**(6): p. 1285-94.
154. Bomar, M.G., et al., *Structure of the ubiquitin-binding zinc finger domain of human DNA Y-polymerase eta.* EMBO Rep, 2007. **8**(3): p. 247-51.
155. Kannouche, P.L., J. Wing, and A.R. Lehmann, *Interaction of human DNA polymerase eta with monoubiquitinated PCNA: a possible mechanism for the polymerase switch in response to DNA damage.* Mol Cell, 2004. **14**(4): p. 491-500.
156. Budzowska, M., et al., *Regulation of the Rev1-pol zeta complex during bypass of a DNA interstrand cross-link.* EMBO J, 2015. **34**(14): p. 1971-85.
157. Kim, H., et al., *Regulation of Rev1 by the Fanconi anemia core complex.* Nat Struct Mol Biol, 2012. **19**(2): p. 164-70.
158. Kim, Y., et al., *Regulation of multiple DNA repair pathways by the Fanconi anemia protein SLX4.* Blood, 2013. **121**(1): p. 54-63.
159. Howlett, N.G., et al., *Biallelic inactivation of BRCA2 in Fanconi anemia.* Science, 2002. **297**(5581): p. 606-9.
160. Prakash, R., et al., *Homologous recombination and human health: the roles of BRCA1, BRCA2, and associated proteins.* Cold Spring Harb Perspect Biol, 2015. **7**(4): p. a016600.
161. Peng, M., et al., *The FANCJ/MutLalpha interaction is required for correction of the cross-link response in FA-J cells.* EMBO J, 2007. **26**(13): p. 3238-49.
162. Yu, X., et al., *The BRCT domain is a phospho-protein binding domain.* Science, 2003. **302**(5645): p. 639-42.
163. Guillemette, S., et al., *FANCJ localization by mismatch repair is vital to maintain genomic integrity after UV irradiation.* Cancer Res, 2014. **74**(3): p. 932-44.
164. Meetei, A.R., Z. Yan, and W. Wang, *FANCL replaces BRCA1 as the likely ubiquitin ligase responsible for FANCD2 monoubiquitination.* Cell Cycle, 2004. **3**(2): p. 179-81.
165. Striegl, H., M.A. Andrade-Navarro, and U. Heinemann, *Armadillo motifs involved in vesicular transport.* PLoS One, 2010. **5**(2): p. e8991.

166. van de Wetering, M., et al., *Armadillo coactivates transcription driven by the product of the Drosophila segment polarity gene dTCF.* Cell, 1997. **88**(6): p. 789-99.
167. Webb, S.J., et al., *Caspase-mediated cleavage of APC results in an amino-terminal fragment with an intact armadillo repeat domain.* FASEB J, 1999. **13**(2): p. 339-46.
168. Zeng, Y.A., et al., *Inhibition of Drosophila Wg signaling involves competition between Mad and Armadillo/beta-catenin for dTcf binding.* PLoS One, 2008. **3**(12): p. e3893.
169. Sims, A.E., et al., *FANCI is a second monoubiquitinated member of the Fanconi anemia pathway.* Nat Struct Mol Biol, 2007. **14**(6): p. 564-7.
170. Kratz, K., et al., *Deficiency of FANCD2-associated nuclease KIAA1018/FAN1 sensitizes cells to interstrand crosslinking agents.* Cell, 2010. **142**(1): p. 77-88.
171. Lyakhovich, A. and J. Surralles, *FANCD2 depletion sensitizes cancer cells repopulation ability in vitro.* Cancer Lett, 2007. **256**(2): p. 186-95.
172. Sakai, W. and K. Sugasawa, *FANCD2 is a target for caspase 3 during DNA damage-induced apoptosis.* FEBS Lett, 2014. **588**(20): p. 3778-85.
173. Taniguchi, T., et al., *S-phase-specific interaction of the Fanconi anemia protein, FANCD2, with BRCA1 and RAD51.* Blood, 2002. **100**(7): p. 2414-20.
174. Yao, C., et al., *Involvement of Fanconi anemia genes FANCD2 and FANCF in the molecular basis of drug resistance in leukemia.* Mol Med Rep, 2015. **11**(6): p. 4605-10.
175. Zhang, J., et al., *FANCD2 monoubiquitination provides a link between the HHR6 and FA-BRCA pathways.* Cell Cycle, 2008. **7**(3): p. 407-13.
176. Zhu, J., et al., *FANCD2 influences replication fork processes and genome stability in response to clustered DSBs.* Cell Cycle, 2015. **14**(12): p. 1809-22.
177. Pizzolato, J., et al., *FANCD2-associated nuclease 1, but not exonuclease 1 or flap endonuclease 1, is able to unhook DNA interstrand cross-links in vitro.* J Biol Chem, 2015. **290**(37): p. 22602-11.
178. Alavattam, K.G., et al., *Elucidation of the Fanconi Anemia Protein Network in Meiosis and Its Function in the Regulation of Histone Modifications.* Cell Rep, 2016. **17**(4): p. 1141-1157.
179. Rego, M.A., et al., *Regulation of the activation of the Fanconi anemia pathway by the p21 cyclin-dependent kinase inhibitor.* Oncogene, 2012. **31**(3): p. 366-75.
180. Smetsers, S., et al., *Heterozygote FANCD2 mutations associated with childhood T Cell ALL and testicular seminoma.* Fam Cancer, 2012. **11**(4): p. 661-5.
181. Raghunandan, M., et al., *FANCD2, FANCJ and BRCA2 cooperate to promote replication fork recovery independently of the Fanconi Anemia core complex.* Cell Cycle, 2015. **14**(3): p. 342-53.
182. Root, H., et al., *FANCD2 limits BLM-dependent telomere instability in the alternative lengthening of telomeres pathway.* Hum Mol Genet, 2016.
183. Shen, C. and P.J. Houghton, *Targeting FANCD2 for therapy sensitization.* Oncotarget, 2014. **5**(11): p. 3426-7.
184. Singh, M., et al., *FANCD2 is a potential therapeutic target and biomarker in alveolar rhabdomyosarcoma harboring the PAX3-FOXO1 fusion gene.* Clin Cancer Res, 2014. **20**(14): p. 3884-95.
185. Stoepker, C., et al., *Defects in the Fanconi Anemia Pathway and Chromatid Cohesion in Head and Neck Cancer.* Cancer Res, 2015. **75**(17): p. 3543-53.
186. Suresh, B., et al., *Regulation of Fanconi anemia protein FANCD2 monoubiquitination by miR-302.* Biochem Biophys Res Commun, 2015. **466**(2): p. 180-5.
187. Unno, J., et al., *FANCD2 binds CtIP and regulates DNA-end resection during DNA interstrand crosslink repair.* Cell Rep, 2014. **7**(4): p. 1039-47.
188. Parmar, K., et al., *Hematopoietic stem cell defects in mice with deficiency of Fancd2 or Usp1.* Stem Cells, 2010. **28**(7): p. 1186-95.
189. Lee, K.Y., et al., *Developmental stage- and DNA damage-specific functions of C. elegans FANCD2.* Biochem Biophys Res Commun, 2007. **352**(2): p. 479-85.

190. Liu, T.X., et al., *Knockdown of zebrafish Fancd2 causes developmental abnormalities via p53-dependent apoptosis.* Dev Cell, 2003. **5**(6): p. 903-14.
191. Marek, L.R. and A.E. Bale, *Drosophila homologs of FANCD2 and FANCL function in DNA repair.* DNA Repair (Amst), 2006. **5**(11): p. 1317-26.
192. Marek, L.R., et al., *MEN1 and FANCD2 mediate distinct mechanisms of DNA crosslink repair.* DNA Repair (Amst), 2008. **7**(3): p. 476-86.
193. Wang, H.F., et al., *BRCA2 and nucleophosmin coregulate centrosome amplification and form a complex with the Rho effector kinase ROCK2.* Cancer Res, 2011. **71**(1): p. 68-77.
194. Zdzienicka, M.Z. and F. Arwert, *Breast cancer and Fanconi anemia: what are the connections?* Trends Mol Med, 2002. **8**(10): p. 458-60.
195. Zhao, W., et al., *Promotion of BRCA2-Dependent Homologous Recombination by DSS1 via RPA Targeting and DNA Mimicry.* Mol Cell, 2015. **59**(2): p. 176-87.
196. Liu, J., et al., *Human BRCA2 protein promotes RAD51 filament formation on RPA-covered single-stranded DNA.* Nat Struct Mol Biol, 2010. **17**(10): p. 1260-2.
197. Nakashima, S., et al., *BRCA/Fanconi anemia pathway implicates chemoresistance to gemcitabine in biliary tract cancer.* Cancer Sci, 2015. **106**(5): p. 584-91.
198. Wooster, R., et al., *Identification of the breast cancer susceptibility gene BRCA2.* Nature, 1995. **378**(6559): p. 789-92.
199. Reid, S., et al., *Biallelic mutations in PALB2 cause Fanconi anemia subtype FA-N and predispose to childhood cancer.* Nat Genet, 2007. **39**(2): p. 162-4.
200. Holloman, W.K., *Unraveling the mechanism of BRCA2 in homologous recombination.* Nat Struct Mol Biol, 2011. **18**(7): p. 748-54.
201. Oliver, A.W., et al., *Structural basis for recruitment of BRCA2 by PALB2.* EMBO Rep, 2009. **10**(9): p. 990-6.
202. Shahid, T., et al., *Structure and mechanism of action of the BRCA2 breast cancer tumor suppressor.* Nat Struct Mol Biol, 2014. **21**(11): p. 962-8.
203. Li, J., et al., *DSS1 is required for the stability of BRCA2.* Oncogene, 2006. **25**(8): p. 1186-94.
204. Yang, H., et al., *BRCA2 function in DNA binding and recombination from a BRCA2-DSS1-ssDNA structure.* Science, 2002. **297**(5588): p. 1837-48.
205. Pickart, C.M., *Mechanisms underlying ubiquitination.* Annu Rev Biochem, 2001. **70**: p. 503-33.
206. Pickart, C.M., *Back to the future with ubiquitin.* Cell, 2004. **116**(2): p. 181-90.
207. Pickart, C.M. and M.J. Eddins, *Ubiquitin: structures, functions, mechanisms.* Biochim Biophys Acta, 2004. **1695**(1-3): p. 55-72.
208. Chau, V., et al., *A multiubiquitin chain is confined to specific lysine in a targeted short-lived protein.* Science, 1989. **243**(4898): p. 1576-83.
209. Pickart, C.M., *Ubiquitin in chains.* Trends Biochem Sci, 2000. **25**(11): p. 544-8.
210. Ulrich, H.D. and H. Walden, *Ubiquitin signalling in DNA replication and repair.* Nat Rev Mol Cell Biol, 2010. **11**(7): p. 479-89.
211. Wang, X., P.R. Andreassen, and A.D. D'Andrea, *Functional interaction of monoubiquitinated FANCD2 and BRCA2/FANCD1 in chromatin.* Mol Cell Biol, 2004. **24**(13): p. 5850-62.
212. Alter, B.P., et al., *Fanconi anemia: adult head and neck cancer and hematopoietic mosaicism.* Arch Otolaryngol Head Neck Surg, 2005. **131**(7): p. 635-9.
213. Holzel, M., et al., *FANCD2 protein is expressed in proliferating cells of human tissues that are cancer-prone in Fanconi anaemia.* J Pathol, 2003. **201**(2): p. 198-203.
214. Vincze, T., J. Posfai, and R.J. Roberts, *NEBcutter: a program to cleave DNA with restriction enzymes.* Nucleic acids research, 2003. **31**(13): p. 3688-3691.
215. Zhu, H. and R.A. Dean, *A novel method for increasing the transformation efficiency of Escherichia coli-application forbacterial artificial chromosome library construction.* Nucleic Acids Res, 1999. **27**(3): p. 910-1.
216. Hannig, G. and S.C. Makrides, *Strategies for optimizing heterologous protein expression in Escherichia coli.* Trends in biotechnology, 1998. **16**(2): p. 54-60.

217. Peti, W. and R. Page, *Strategies to maximize heterologous protein expression in Escherichia coli with minimal cost.* Protein expression and purification, 2007. **51**(1): p. 1-10.
218. Rosano, G.L. and E.A. Ceccarelli, *Rare codon content affects the solubility of recombinant proteins in a codon bias-adjusted Escherichia coli strain.* Microb Cell Fact, 2009. **8**: p. 41.
219. Caprioli, R.M., T.B. Farmer, and J. Gile, *Molecular imaging of biological samples: localization of peptides and proteins using MALDI-TOF MS.* Anal Chem, 1997. **69**(23): p. 4751-60.
220. Greenfield, N.J., *Using circular dichroism spectra to estimate protein secondary structure.* Nat Protoc, 2006. **1**(6): p. 2876-90.
221. Kelly, S.M., T.J. Jess, and N.C. Price, *How to study proteins by circular dichroism.* Biochim Biophys Acta, 2005. **1751**(2): p. 119-39.
222. Pace, C.N. and K.L. Shaw, *Linear extrapolation method of analyzing solvent denaturation curves.* Proteins, 2000. **Suppl 4**: p. 1-7.
223. Ramos-Fernandez, A., et al., *Generalized method for probability-based peptide and protein identification from tandem mass spectrometry data and sequence database searching.* Mol Cell Proteomics, 2008. **7**(9): p. 1748-54.
224. Henderson, R., et al., *Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy.* J Mol Biol, 1990. **213**(4): p. 899-929.
225. Jaroszewski, L., *Protein structure prediction based on sequence similarity.* Methods Mol Biol, 2009. **569**: p. 129-56.
226. Kopp, J. and T. Schwede, *Automated protein structure homology modeling: a progress report.* Pharmacogenomics, 2004. **5**(4): p. 405-16.
227. Hilbert, M., G. Bohm, and R. Jaenicke, *Structural relationships of homologous proteins as a fundamental principle in homology modeling.* Proteins, 1993. **17**(2): p. 138-51.
228. Rost, B. and S. O'Donoghue, *Sisyphus and prediction of protein structure.* Comput Appl Biosci, 1997. **13**(4): p. 345-56.
229. Gething, M.-J. and J. Sambrook, *Protein folding in the cell.* Nature, 1992. **355**(6355): p. 33.
230. Holm, L. and C. Sander, *Protein structure comparison by alignment of distance matrices.* Journal of molecular biology, 1993. **233**(1): p. 123-138.
231. Sticht, H., et al., *Homology modeling of adenylosuccinate synthetase from Saccharomyces cerevisiae reveals a possible binding region for single-stranded ARS sequences.* J Biomol Struct Dyn, 1997. **14**(6): p. 667-75.
232. Altschul, S.F., et al., *Basic local alignment search tool.* Journal of molecular biology, 1990. **215**(3): p. 403-410.
233. Schwede, T., et al., *SWISS-MODEL: an automated protein homology-modeling server.* Nucleic acids research, 2003. **31**(13): p. 3381-3385.
234. Xiang, Z., *Homology-based modeling of protein structure*, in *Computational Methods for Protein Structure Prediction and Modeling*. 2007, Springer. p. 319-357.
235. Bonneau, R. and D. Baker, *Ab initio protein structure prediction: progress and prospects.* Annu Rev Biophys Biomol Struct, 2001. **30**: p. 173-89.
236. Vogt, G., S. Woell, and P. Argos, *Protein thermal stability, hydrogen bonds, and ion pairs.* Journal of molecular biology, 1997. **269**(4): p. 631-643.
237. Simons, K.T., et al., *Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions.* Journal of molecular biology, 1997. **268**(1): p. 209-225.
238. Bowie, J.U. and D. Eisenberg, *An evolutionary approach to folding small alpha-helical proteins that uses sequence information and an empirical guiding fitness function.* Proceedings of the National Academy of Sciences, 1994. **91**(10): p. 4436-4440.
239. Simons, K.T., et al., *Ab initio protein structure prediction of CASP III targets using ROSETTA.* Proteins: Structure, Function, and Bioinformatics, 1999. **37**(S3): p. 171-176.
240. Das, R., et al., *Structure prediction for CASP7 targets using extensive all-atom refinement with Rosetta@ home.* Proteins: Structure, Function, and Bioinformatics, 2007. **69**(S8): p. 118-128.

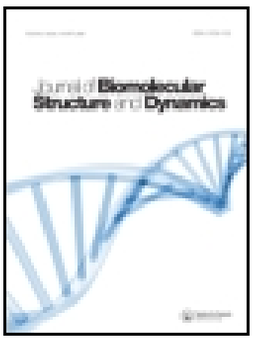
241. Bradley, P., et al., *Free modeling with Rosetta in CASP6.* Proteins: Structure, Function, and Bioinformatics, 2005. **61**(S7): p. 128-134.
242. Li, Z. and H.A. Scheraga, *Monte Carlo-minimization approach to the multiple-minima problem in protein folding.* Proceedings of the National Academy of Sciences, 1987. **84**(19): p. 6611-6615.
243. Fujitsuka, Y., G. Chikenji, and S. Takada, *SimFold energy function for de novo protein structure prediction: consensus with Rosetta.* Proteins: Structure, Function, and Bioinformatics, 2006. **62**(2): p. 381-398.
244. Lee, J., et al., *Prediction of protein tertiary structure using PROFESY, a novel method based on fragment assembly and conformational space annealing.* PROTEINS: Structure, Function, and Bioinformatics, 2004. **56**(4): p. 704-714.
245. van Zundert, G.C., et al., *The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes.* J Mol Biol, 2016. **428**(4): p. 720-5.
246. De Vries, S.J., M. Van Dijk, and A.M. Bonvin, *The HADDOCK web server for data-driven biomolecular docking.* Nature protocols, 2010. **5**(5): p. 883-897.
247. Dominguez, C., R. Boelens, and A.M. Bonvin, *HADDOCK: A protein− protein docking approach based on biochemical or biophysical information.* Journal of the American Chemical Society, 2003. **125**(7): p. 1731-1737.
248. van Dijk, A.D., et al., *Combining NMR relaxation with chemical shift perturbation data to drive protein–protein docking.* Journal of biomolecular NMR, 2006. **34**(4): p. 237-244.
249. Alijabbari, N., et al., *Molecular dynamics modeling of the sub-THz vibrational absorption of thioredoxin from E. coli.* Journal of molecular modeling, 2012. **18**(5): p. 2209-2218.
250. Price, D.J. and C.L. Brooks III, *A modified TIP3P water potential for simulation with Ewald summation.* The Journal of chemical physics, 2004. **121**(20): p. 10096-10103.
251. Verlet, L., *Computer" experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules.* Physical review, 1967. **159**(1): p. 98.
252. Katchalski-Katzir, E., et al., *Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques.* Proceedings of the National Academy of Sciences, 1992. **89**(6): p. 2195-2199.
253. Gabb, H.A., R.M. Jackson, and M.J. Sternberg, *Modelling protein docking using shape complementarity, electrostatics and biochemical information.* Journal of molecular biology, 1997. **272**(1): p. 106-120.
254. Meyer, M., P. Wilson, and D. Schomburg, *Hydrogen bonding and molecular surface shape complementarity as a basis for protein docking.* Journal of molecular biology, 1996. **264**(1): p. 199-210.
255. Moont, G., H.A. Gabb, and M.J. Sternberg, *Use of pair potentials across protein interfaces in screening predicted docked complexes.* Proteins: Structure, Function, and Bioinformatics, 1999. **35**(3): p. 364-373.
256. Fernández-Recio, J., R. Abagyan, and M. Totrov, *Improving CAPRI predictions: Optimized desolvation for rigid-body docking.* PROTEINS: Structure, Function, and Bioinformatics, 2005. **60**(2): p. 308-313.
257. Kozakov, D., et al., *Optimal clustering for detecting near-native conformations in protein docking.* Biophysical journal, 2005. **89**(2): p. 867-875.
258. Murphy, J., et al., *Combination of scoring functions improves discrimination in protein–protein docking.* Proteins: Structure, Function, and Bioinformatics, 2003. **53**(4): p. 840-854.
259. Li, C.H., et al., *Complex-type-dependent scoring functions in protein–protein docking.* Biophysical chemistry, 2007. **129**(1): p. 1-10.
260. Martin, O. and D. Schomburg, *Efficient comprehensive scoring of docked protein complexes using probabilistic support vector machines.* Proteins: Structure, Function, and Bioinformatics, 2008. **70**(4): p. 1367-1378.

261. Grubmüller, H., et al., *Generalized Verlet algorithm for efficient molecular dynamics simulations with long-range interactions.* Molecular Simulation, 1991. **6**(1-3): p. 121-142.
262. Kresse, G. and J. Hafner, *Ab initio molecular dynamics for liquid metals.* Physical Review B, 1993. **47**(1): p. 558.
263. Kim, D.E., D. Chivian, and D. Baker, *Protein structure prediction and analysis using the Robetta server.* Nucleic acids research, 2004. **32**(suppl_2): p. W526-W531.
264. Berendsen, H.J., D. van der Spoel, and R. van Drunen, *GROMACS: a message-passing parallel molecular dynamics implementation.* Computer Physics Communications, 1995. **91**(1-3): p. 43-56.
265. Hess, B., et al., *GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation.* Journal of chemical theory and computation, 2008. **4**(3): p. 435-447.
266. Van Der Spoel, D., et al., *GROMACS: fast, flexible, and free.* Journal of computational chemistry, 2005. **26**(16): p. 1701-1718.
267. Grant, B.J., et al., *Bio3d: an R package for the comparative analysis of protein structures.* Bioinformatics, 2006. **22**(21): p. 2695-2696.
268. Wallace, A.C., R.A. Laskowski, and J.M. Thornton, *LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions.* Protein Engineering, Design and Selection, 1995. **8**(2): p. 127-134.
269. Amadei, A., M.A. Ceruso, and A. Di Nola, *On the convergence of the conformational coordinates basis set obtained by the essential dynamics analysis of proteins' molecular dynamics simulations.* Proteins, 1999. **36**(4): p. 419-24.
270. Amadei, A., A. Linssen, and H.J. Berendsen, *Essential dynamics of proteins.* Proteins: Structure, Function, and Bioinformatics, 1993. **17**(4): p. 412-425.
271. David, C.C. and D.J. Jacobs, *Principal component analysis: a method for determining the essential dynamics of proteins.* Protein Dynamics: Methods and Protocols, 2014: p. 193-226.
272. Alexandrov, V., et al., *Normal modes for predicting protein motions: a comprehensive database assessment and associated Web tool.* Protein science, 2005. **14**(3): p. 633-643.
273. Levitt, M., C. Sander, and P.S. Stern, *Protein normal-mode dynamics: trypsin inhibitor, crambin, ribonuclease and lysozyme.* Journal of molecular biology, 1985. **181**(3): p. 423-447.
274. Streett, W., D. Tildesley, and G. Saville, *Multiple time-step methods in molecular dynamics.* Molecular Physics, 1978. **35**(3): p. 639-648.
275. Hayward, S. and B.L. De Groot, *Normal modes and essential dynamics.* Molecular Modeling of Proteins, 2008: p. 89-106.
276. Hayward, S., A. Kitao, and H.J. Berendsen, *Model-free methods of analyzing domain motions in proteins from simulation: a comparison of normal mode analysis and molecular dynamics simulation of lysozyme.* Proteins: Structure, Function, and Bioinformatics, 1997. **27**(3): p. 425-437.
277. Cizas, P., et al., *Size-dependent neurotoxicity of β-amyloid oligomers.* Archives of biochemistry and biophysics, 2010. **496**(2): p. 84-92.
278. Melcher, K., *New chemical crosslinking methods for the identification of transient protein-protein interactions with multiprotein complexes.* Current Protein and Peptide Science, 2004. **5**(4): p. 287-296.
279. FEENEY, R.E., *Chemical modification of proteins: comments and perspectives.* Chemical Biology & Drug Design, 1987. **29**(2): p. 145-161.
280. George, A. and W.W. Wilson, *Predicting protein crystallization from a dilute solution property.* Acta Crystallographica Section D: Biological Crystallography, 1994. **50**(4): p. 361-365.
281. Mølgaard, A. and S. Larsen, *Crystal packing in two pH-dependent crystal forms of rhamnogalacturonan acetylesterase.* Acta Crystallographica Section D: Biological Crystallography, 2004. **60**(3): p. 472-478.
282. D'arcy, A., et al., *A novel approach to crystallising proteins under oil.* Journal of crystal growth, 1996. **168**(1-4): p. 175-180.

283. Chayen, N.E., et al., *An automated system for micro-batch protein crystallization and screening.* Journal of applied crystallography, 1990. **23**(4): p. 297-302.
284. Zeppezauer, M., H. Eklund, and E.S. Zeppezauer, *Micro diffusion cells for the growth of single protein crystals by means of equilibrium dialysis.* Archives of biochemistry and biophysics, 1968. **126**(2): p. 564-573.
285. Battesti, A. and E. Bouveret, *The bacterial two-hybrid system based on adenylate cyclase reconstitution in Escherichia coli.* Methods, 2012. **58**(4): p. 325-34.
286. Karimova, G., et al., *A bacterial two-hybrid system based on a reconstituted signal transduction pathway.* Proc Natl Acad Sci U S A, 1998. **95**(10): p. 5752-6.
287. Nefedova, V.V., et al., *Physico-chemical properties of R140G and K141Q mutants of human small heat shock protein HspB1 associated with hereditary peripheral neuropathies.* Biochimie, 2013. **95**(8): p. 1582-1592.
288. Pattnaik, P., *Surface plasmon resonance.* Applied biochemistry and biotechnology, 2005. **126**(2): p. 79-92.
289. Homola, J., S.S. Yee, and G. Gauglitz, *Surface plasmon resonance sensors: review.* Sensors and Actuators B: Chemical, 1999. **54**(1): p. 3-15.
290. Magnaghi-Jaulin, L., et al., *Retinoblastoma protein represses transcription by recruiting a histone deacetylase.* Nature, 1998. **391**(6667): p. 601-605.
291. Zhou, B.-B.S. and S.J. Elledge, *The DNA damage response: putting checkpoints in perspective.* Nature, 2000. **408**(6811): p. 433-439.
292. Shiloh, Y., *The ATM-mediated DNA-damage response: taking shape.* Trends in biochemical sciences, 2006. **31**(7): p. 402-410.
293. Chaganti, R.S. and J. Houldsworth, *Fanconi anemia: a pleotropic mutation with multiple cellular and developmental abnormalities.* Ann Genet, 1991. **34**(3-4): p. 206-11.
294. Willert, K. and R. Nusse, *β-catenin: a key mediator of Wnt signaling.* Current opinion in genetics & development, 1998. **8**(1): p. 95-102.
295. Andrade, M.A., et al., *Comparison of ARM and HEAT protein repeats.* J Mol Biol, 2001. **309**(1): p. 1-18.
296. Coates, J.C., *Armadillo repeat proteins: beyond the animal kingdom.* Trends Cell Biol, 2003. **13**(9): p. 463-71.
297. Hatzfeld, M. and C. Nachtsheim, *Cloning and characterization of a new armadillo family member, p0071, associated with the junctional plaque: evidence for a subfamily of closely related proteins.* J Cell Sci, 1996. **109 (Pt 11)**: p. 2767-78.
298. Kruger, N.J., *The Bradford method for protein quantitation.* Basic protein and peptide protocols, 1994: p. 9-15.
299. Louis-Jeune, C., M.A. Andrade-Navarro, and C. Perez-Iratxeta, *Prediction of protein secondary structure from circular dichroism using theoretically derived spectra.* Proteins: Structure, Function, and Bioinformatics, 2012. **80**(2): p. 374-381.
300. Boutet, E., et al., *UniProtKB/Swiss-Prot: the manually annotated section of the UniProt KnowledgeBase.* Plant bioinformatics: methods and protocols, 2007: p. 89-112.
301. Kim, D.E., D. Chivian, and D. Baker, *Protein structure prediction and analysis using the Robetta server.* Nucleic Acids Res, 2004. **32**(Web Server issue): p. W526-31.
302. Fiser, A. and A. Sali, *ModLoop: automated modeling of loops in protein structures.* Bioinformatics, 2003. **19**(18): p. 2500-1.
303. Wiederstein, M. and M.J. Sippl, *ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins.* Nucleic Acids Res, 2007. **35**(Web Server issue): p. W407-10.
304. Kaminski, G.A., et al., *Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides.* The Journal of Physical Chemistry B, 2001. **105**(28): p. 6474-6487.

305. Abraham, M.J. and J.E. Gready, *Optimization of parameters for molecular dynamics simulation using smooth particle-mesh Ewald in GROMACS 4.5.* Journal of computational chemistry, 2011. **32**(9): p. 2031-2040.
306. Ryckaert, J.-P., G. Ciccotti, and H.J. Berendsen, *Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes.* Journal of Computational Physics, 1977. **23**(3): p. 327-341.
307. Kabsch, W. and C. Sander, *Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features.* Biopolymers, 1983. **22**(12): p. 2577-2637.
308. Amadei, A., A.B. Linssen, and H.J. Berendsen, *Essential dynamics of proteins.* Proteins, 1993. **17**(4): p. 412-25.
309. Choudhary, R.K., et al., *Multimodal approach to explore the pathogenicity of BARD1, ARG 658 CYS, and ILE 738 VAL mutants.* J Biomol Struct Dyn, 2016. **34**(7): p. 1533-44.
310. Bahar, I., et al., *Collective motions in HIV-1 reverse transcriptase: examination of flexibility and enzyme function.* J Mol Biol, 1999. **285**(3): p. 1023-37.
311. Grant, B.J., et al., *Bio3d: an R package for the comparative analysis of protein structures.* Bioinformatics, 2006. **22**(21): p. 2695-6.
312. Choudhary, R.K., et al., *Multimodal approach to explore the pathogenicity of BARD1, ARG 658 CYS, and ILE 738 VAL mutants.* J Biomol Struct Dyn, 2015: p. 1-12.
313. Vikrant, et al., *Structural and functional characterization of the MERIT40 to understand its role in DNA repair.* J Biomol Struct Dyn, 2014. **32**(12): p. 2017-32.
314. Fontana, A., et al., *Probing the conformational state of apomyoglobin by limited proteolysis.* J Mol Biol, 1997. **266**(2): p. 223-30.
315. Ishida, T. and K. Kinoshita, *PrDOS: prediction of disordered protein regions from amino acid sequence.* Nucleic acids research, 2007. **35**(suppl 2): p. W460-W464.
316. Gasteiger, E., et al., *Protein identification and analysis tools on the ExPASy server*. 2005: Springer.
317. Holm, L. and P. Rosenström, *Dali server: conservation mapping in 3D.* Nucleic acids research, 2010: p. gkq366.
318. Pace, C.N. and J.M. Scholtz, *A helix propensity scale based on experimental studies of peptides and proteins.* Biophys J, 1998. **75**(1): p. 422-7.
319. Ohlendorf, D.H., et al., *The molecular basis of DNA-protein recognition inferred from the structure of cro repressor.* Nature, 1982. **298**(5876): p. 718-23.
320. Luscombe, N.M. and J.M. Thornton, *Protein-DNA interactions: amino acid conservation and the effects of mutations on binding specificity.* J Mol Biol, 2002. **320**(5): p. 991-1009.
321. Hicke, L., H.L. Schubert, and C.P. Hill, *Ubiquitin-binding domains.* Nat Rev Mol Cell Biol, 2005. **6**(8): p. 610-21.
322. Shih, S.C., et al., *A ubiquitin-binding motif required for intramolecular monoubiquitylation, the CUE domain.* EMBO J, 2003. **22**(6): p. 1273-81.
323. David, C.C. and D.J. Jacobs, *Principal component analysis: a method for determining the essential dynamics of proteins.* Methods Mol Biol, 2014. **1084**: p. 193-226.
324. Funayama, N., et al., *Embryonic axis induction by the armadillo repeat domain of beta-catenin: evidence for intracellular signaling.* J Cell Biol, 1995. **128**(5): p. 959-68.
325. Sharma, M., et al., *Specific armadillo repeat sequences facilitate beta-catenin nuclear transport in live cells via direct binding to nucleoporins Nup62, Nup153, and RanBP2/Nup358.* J Biol Chem, 2012. **287**(2): p. 819-31.
326. Rajpurohit, Y.S. and H.S. Misra, *Structure-function study of deinococcal serine/threonine protein kinase implicates its kinase activity and DNA repair protein phosphorylation roles in radioresistance of Deinococcus radiodurans.* Int J Biochem Cell Biol, 2013. **45**(11): p. 2541-52.
327. Kadonaga, J.T., *Regulation of RNA polymerase II transcription by sequence-specific DNA binding factors.* Cell, 2004. **116**(2): p. 247-257.

328. Wickner, S., *Mechanism of DNA elongation catalyzed by Escherichia coli DNA polymerase III, dnaZ protein, and DNA elongation factors I and III.* Proceedings of the National Academy of Sciences, 1976. **73**(10): p. 3511-3515.
329. Baeuerle, P.A. and D. Baltimore, *Activation of DNA-binding activity in an apparently cytoplasmic precursor of the NF-κB transcription factor.* Cell, 1988. **53**(2): p. 211-217.
330. Brennan, R. and B.W. Matthews, *The helix-turn-helix DNA binding motif.* J. biol. Chem, 1989. **264**(4): p. 1903-1906.
331. Harrison, S.C. and A.K. Aggarwal, *DNA recognition by proteins with the helix-turn-helix motif.* Annual review of biochemistry, 1990. **59**(1): p. 933-969.
332. Gabrielsen, O.S., A. Sentenac, and P. Fromageot, *Specific DNA binding by c-Myb: evidence for a double helix-turn-helix-related motif.* Science, 1991. **253**(5024): p. 1140-3.
333. Sato, K., et al., *Histone chaperone activity of Fanconi anemia proteins, FANCD2 and FANCI, is required for DNA crosslink repair.* EMBO J, 2012. **31**(17): p. 3524-36.
334. Penengo, L., et al., *Crystal structure of the ubiquitin binding domains of rabex-5 reveals two modes of interaction with ubiquitin.* Cell, 2006. **124**(6): p. 1183-95.
335. Hicke, L., H.L. Schubert, and C.P. Hill, *Ubiquitin-binding domains.* Nature Reviews Molecular Cell Biology, 2005. **6**(8): p. 610-621.
336. Shih, S.C., et al., *A ubiquitin-binding motif required for intramolecular monoubiquitylation, the CUE domain.* The EMBO journal, 2003. **22**(6): p. 1273-1281.
337. Nefedova, V.V., et al., *Physico-chemical properties of R140G and K141Q mutants of human small heat shock protein HspB1 associated with hereditary peripheral neuropathies.* Biochimie, 2013. **95**(8): p. 1582-92.
338. Modi, K. and H.S. Misra, *Dr-FtsA, an actin homologue in Deinococcus radiodurans differentially affects Dr-FtsZ and Ec-FtsZ functions in vitro.* PLoS One, 2014. **9**(12): p. e115918.
339. Bogan, A.A. and K.S. Thorn, *Anatomy of hot spots in protein interfaces.* J Mol Biol, 1998. **280**(1): p. 1-9.
340. Klein Douwel, D., et al., *XPF-ERCC1 acts in Unhooking DNA interstrand crosslinks in cooperation with FANCD2 and FANCP/SLX4.* Mol Cell, 2014. **54**(3): p. 460-71.
341. Nakanishi, K., et al., *Human Fanconi anemia monoubiquitination pathway promotes homologous DNA repair.* Proc Natl Acad Sci U S A, 2005. **102**(4): p. 1110-5.
342. Taniguchi, T., et al., *Convergence of the fanconi anemia and ataxia telangiectasia signaling pathways.* Cell, 2002. **109**(4): p. 459-72.
343. Rubinson, E.H. and B.F. Eichman, *Nucleic acid recognition by tandem helical repeats.* Curr Opin Struct Biol, 2012. **22**(1): p. 101-9.
344. Marcotte, E.M., et al., *A census of protein repeats.* J Mol Biol, 1999. **293**(1): p. 151-60.
345. Ohashi, A., et al., *Fanconi anemia complementation group D2 (FANCD2) functions independently of BRCA2- and RAD51-associated homologous recombination in response to DNA damage.* J Biol Chem, 2005. **280**(15): p. 14877-83.
346. Bhatia, V., et al., *BRCA2 prevents R-loop accumulation and associates with TREX-2 mRNA export factor PCID2.* Nature, 2014. **511**(7509): p. 362-5.
347. Warren, M., et al., *Structural analysis of the chicken BRCA2 gene facilitates identification of functional domains and disease causing mutations.* Hum Mol Genet, 2002. **11**(7): p. 841-51.



Taylor & Francis
Taylor & Francis Group

Journal of Biomolecular Structure and Dynamics

ISSN: 0739-1102 (Print) 1538-0254 (Online) Journal homepage: http://www.tandfonline.com/loi/tbsd20

Structural and biophysical properties of h-FANCI ARM repeat protein

Mohd. Quadir Siddiqui, Rajan Kumar Choudhary, Pankaj Thapa, Neha Kulkarni, Yogendra S. Rajpurohit, Hari S. Misra, Nikhil Gadewal, Satish Kumar, Syed K. Hasan & Ashok K. Varma

To cite this article: Mohd. Quadir Siddiqui, Rajan Kumar Choudhary, Pankaj Thapa, Neha Kulkarni, Yogendra S. Rajpurohit, Hari S. Misra, Nikhil Gadewal, Satish Kumar, Syed K. Hasan & Ashok K. Varma (2016): Structural and biophysical properties of h-FANCI ARM repeat protein, Journal of Biomolecular Structure and Dynamics, DOI: 10.1080/07391102.2016.1235514

To link to this article: http://dx.doi.org/10.1080/07391102.2016.1235514

View supplementary material

Accepted author version posted online: 30 Sep 2016.
Published online: 10 Nov 2016.

Submit your article to this journal

Article views: 14

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
http://www.tandfonline.com/action/journalInformation?journalCode=tbsd20

Download by: [14.139.127.177] **Date:** 17 November 2016, At: 06:16

Journal of Biomolecular Structure and Dynamics, 2016
http://dx.doi.org/10.1080/07391102.2016.1235514

Taylor & Francis
Taylor & Francis Group

Structural and biophysical properties of h-FANCI ARM repeat protein

Mohd. Quadir Siddiqui[a], Rajan Kumar Choudhary[a], Pankaj Thapa[a], Neha Kulkarni[a], Yogendra S. Rajpurohit[b], Hari S. Misra[b], Nikhil Gadewal[a], Satish Kumar[c], Syed K. Hasan[a*] and Ashok K. Varma[a*]

[a]*Tata Memorial Centre, Advanced Centre for Treatment, Research and Education in Cancer, Kharghar, Navi Mumbai, Maharashtra 410 210, India;* [b]*Molecular Biology Division, Bhabha Atomic Research Centre, Mumbai 400 085, India;* [c]*Department of Biochemistry & Bioinformatics Centre, Mahatma Gandhi Institute of Medical Sciences, Sevagram (Wardha) 442102, India*

Communicated by Ramaswamy H. Sarma

(*Received 17 May 2016; accepted 2 September 2016*)

Fanconi anemia complementation groups – I (FANCI) protein facilitates DNA ICL (Inter-Cross-link) repair and plays a crucial role in genomic integrity. FANCI is a 1328 amino acids protein which contains armadillo (ARM) repeats and EDGE motif at the C-terminus. ARM repeats are functionally diverse and evolutionarily conserved domain that plays a pivotal role in protein–protein and protein–DNA interactions. Considering the importance of ARM repeats, we have explored comprehensive *in silico* and *in vitro* approach to examine folding pattern. Size exclusion chromatography, dynamic light scattering (DLS) and glutaraldehyde crosslinking studies suggest that FANCI ARM repeat exist as monomer as well as in oligomeric forms. Circular dichroism (CD) and fluorescence spectroscopy results demonstrate that protein has predominantly α- helices and well-folded tertiary structure. DNA binding was analysed using electrophoretic mobility shift assay by autoradiography. Temperature-dependent CD, Fluorescence spectroscopy and DLS studies concluded that protein unfolds and start forming oligomer from 30°C. The existence of stable portion within FANCI ARM repeat was examined using limited proteolysis and mass spectrometry. The normal mode analysis, molecular dynamics and principal component analysis demonstrated that helix-turn-helix (HTH) motif present in ARM repeat is highly dynamic and has anti-correlated motion. Furthermore, FANCI ARM repeat has HTH structural motif which binds to double-stranded DNA.

Keywords: human-FANCI; ARM repeats; helix-turn-helix structural motif

1. Introduction

Fanconi Anaemia (FA) is one of the rare genetic disorders which provide an extra-ordinary opportunity to investigate the biological processes and molecular mechanisms of DNA inter-crosslink (DNA ICL) repair (Cohen et al., 1982). Recent studies about FA have revealed that DNA ICL pathway comprises 18 complementation groups with discrete genes (A,B,C,D1/BRCA2,D2,E,F,G, I,J/BRIP1,L,M,N/PALB2,O/RAD51C,P/SLX4,Q/XPF,S/BRCA1 and T/UBE2T) (Castella et al., 2015). FA patients exhibited a diverse spectrum of clinical phenotypes and showed high sensitivity to inter-crosslinking agents such as mitomycin C and di-epoxybutane (Alter, Greene, Velazquez, & Rosenberg, 2003). Radial chromosomes are the diagnostic hallmark of FA cells. Basic mechanism underlying the formation of the radial chromosome is the inter-crosslinking of the DNA strands (McCabe, Olson, & Moses, 2009). Fanconi anaemia pathway proteins comprise a set of specific proteins that are expressed and recruited to DNA damage sites to facilitate the inter-crosslink DNA repair (Butturini et al., 1994). In eukaryotes, the repair mechanism of intra-strand and inter-strand crosslinks (ICL's) are not yet fully explored. Fanconi anaemia complementation group I (FANCI) is one of the FA protein known to be recruited at DNA damage sites facilitating the DNA ICL repair (Dorsman et al., 2007). FANCI gene comprised different domains including Armadillo repeat (ARM repeat). ARM repeat of FANCI protein is involved in protein–protein and protein–DNA interactions that are indispensable for maintaining genomic integrity and cellular functions (Smogorzewska et al., 2007).

ARM repeats are composed of tandem copies of degenerate protein sequences that form conserved three-dimensional structures (Andrade, Petosa, O'Donoghue, Muller, & Bork, 2001). ARM repeat was first discovered in segment polarity gene of drosophila and afterwards in other proteins like junctional plaque protein plakoglobin, tumour suppressor adenomatous polyposis coli (APC) protein, nucleocytoplasmic transport factor protein importin, FANCI and FANCD2 (Coates, 2003; Smogorzewska et al., 2007). ARM repeat containing proteins are composed of compact and dynamic region which acts as molecular recognition component and an

*Corresponding authors. Email: shasan@actrec.gov.in (S. K. Hasan); avarma@actrec.gov.in (A. K. Varma)

© 2016 Informa UK Limited, trading as Taylor & Francis Group

interacting module for different binding partners (Tsytlonok et al., 2013).

Protein–DNA interactions are essential to maintaining genomic integrity and cell survival (Wang, 2007). Considering the functional diversity of ARM repeats, we have characterized the FANCI ARM repeat to explore the DNA binding, folding pattern, dynamics as well as presence of helix-turn-helix (HTH) motif using *in vitro* and *in silico* approaches. Thus, the presence of HTH structural motif in FANCI ARM repeat that has inherent structural flexibility helps in establishing the interactions with DNA.

2. Results and discussion

Noting the importance of FANCI in DNA ICL repair mechanism, the functional armadillo (ARM) repeat domain of FANCI protein was purified to understand the folding patterns. FANCI protein is a leucine-rich protein (LRP) which mediates protein–protein and protein–DNA interactions (Yuan, El Hokayem, Zhou, & Zhang, 2009). The LRPs are generally composed of ARM, HEAT, leucine-rich repeat and leucine zipper (Tsytlonok et al., 2013). FANCI C-terminus harbours ARM repeat, EDGE motif and nuclear localization signal (Smogorzewska et al., 2007). However, detailed functional characterization of FANCI C-terminus specifically ARM repeat is still unexplored. Longerich et al. reported that C-terminus of FANCI have similar DNA-binding activity as full length of FANCI and binds preferentially to double-stranded DNA (Longerich, San Filippo, Liu, & Sung, 2009). It has been observed that ARM repeat has structurally dynamic region majorly composed of α-helices, and major groove DNA binding suggest that FANCI ARM repeat harbours HTH structural motif. Molecular modelling, dynamics and docking studies are in concordance with *in vitro* results derived from limited proteolysis, sequence alignment and DALI findings. Looking at the helix propensity (Pace & Scholtz, 1998) of the amino acids sequences in HTH structural motif, presence of good helix former amino acids such as Ala, Glu, Gln and Ser also suggest that the dynamic region forms α-helices. It is well known that the HTH structural motif generally binds to major groove of DNA and conserved amino acids such as Glu, Gln and Ser form the binding interface between HTH motif and DNA (Luscombe & Thornton, 2002; Ohlendorf, Anderson, Fisher, Takeda, & Matthews, 1982). Docking study indicates that the conserved amino acids Gln128, Glu129 and Ser132 present in HTH motif are at the binding interface of DNA, which suggest that HTH motif help ARM repeat in DNA binding. The 80 ns molecular dynamincs simulation also confirms that the HTH structural motif is dynamic. Furthermore, limited proteolysis, mass spectrometry and principal component analysis results are also supporting that ARM repeat has the compact domain devoid of the HTH motif. Overall cumulative results suggest that HTH motif might help ARM repeat in DNA binding.

2.1. Presence of HTH-type structural motif in FANCI ARM repeat

It has been reported that FANCI, C-terminal of ARM repeat has tumour suppressor property and might be having DNA-binding motif (Crist et al., 2010; Longerich et al., 2009). The DALI structural alignment exhibited that ARM repeats present in C-terminal of FANCI is structurally conserved and showing significant similarity with DNA-binding transcription factors (Supplementary Figure 1(A)). Since the primary sequences of amino acids dictate protein folds and functions, we have also performed sequence alignment with known HTH structural motif "master sets" (Brennan & Matthews, 1989) and found significant similarity with dynamic part of 20 amino acids of FANCI ARM repeat. The well conserved "PHS" signature (where P is the charged residue mostly glutamate, H is any hydrophobic and S is small residue) of HTH motif (Aravind, Anantharaman, Balaji, Babu, & Iyer, 2005) was also found in the amino acid sequences (Figure 1(A)–(C)). These results suggest that DNA-binding element present in ARM repeat is HTH-type structural motif.

2.2. Oligomeric behaviour and secondary structural characterization of FANCI ARM repeat

Purified FANCI ARM repeat protein was subjected to size exclusion chromatography (SEC), it has been found that protein elutes at different column volumes of 62.7 ± .25, 56.91 ± .16 and 44.9 ± .56 ml has monomeric, dimeric and oligomeric nature, respectively (Supplementary Figure 2(A)). Further to confirm oligomeric property of ARM repeat, time-dependent glutaraldehyde cross-linking and DLS studies (Supplementary Figure 2(C) and (D)) were performed. The results are suggesting that protein exist in monomeric and multi-meric forms. Moreover, molecular weight estimation of protein was calculated by SEC (Supplementary Figure 2(B)) and mass spectrometry (Supplementary Figure 1(B)) (Table 1). The results concluded that ARM repeat protein is forming oligomers at physiological temperature that might be due to presence of some structurally distorted region.

Secondary structure of FANCI ARM repeat was analysed using far-UV, circular dichroism (CD) spectroscopy from monomeric fraction of protein. CD spectra of ARM repeat show millidegree ellipticity at $\lambda = 218$ and 222 nm indicating predominant α-helical content (Figure 2(A)), which is in full agreement with the crystal structure

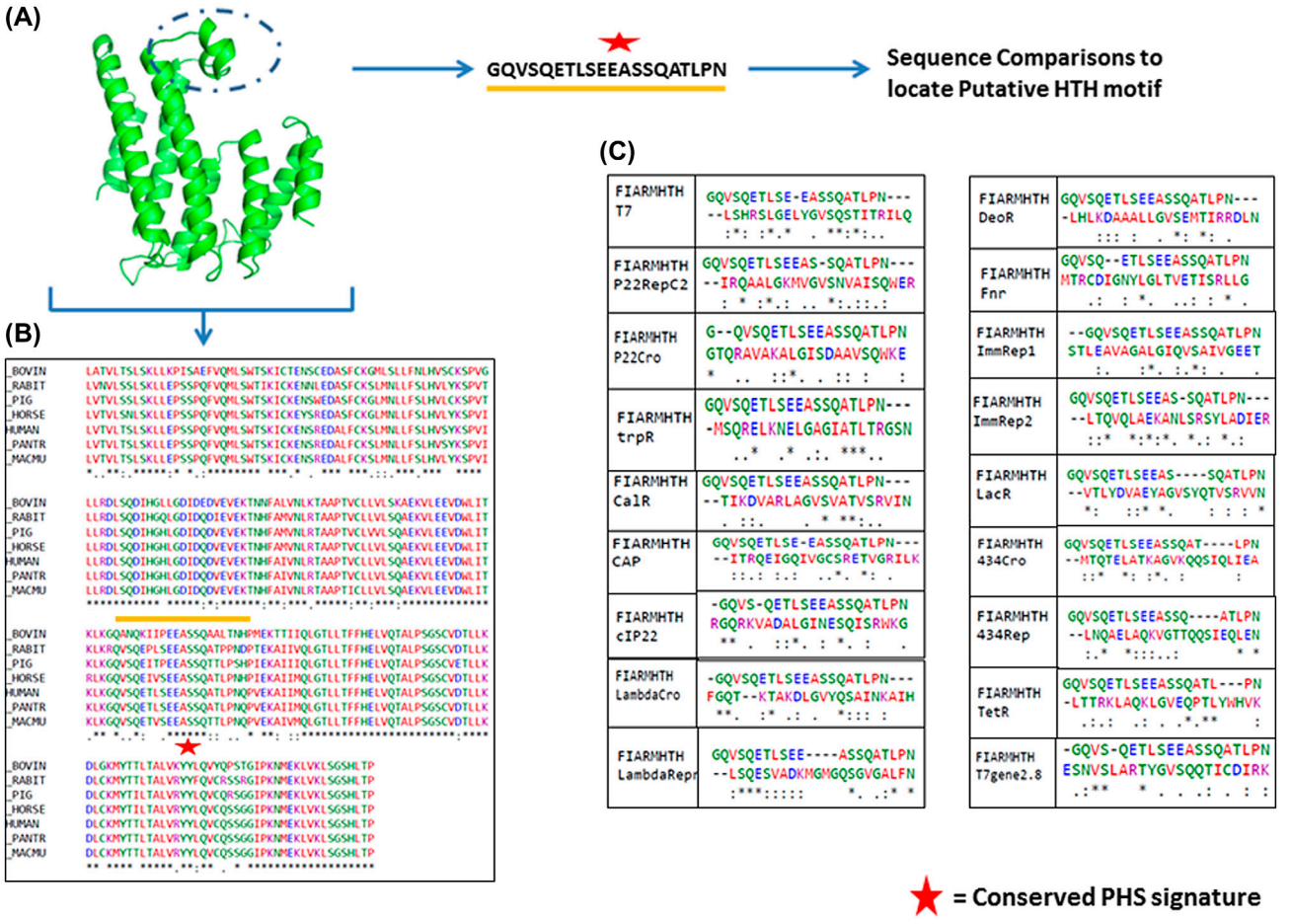


Figure 1. (A) Modelled structure of FANCI ARM repeat protein, (B) Multiple sequence alignment of FANCI ARM repeat, (C) MUSCLE alignment of ARM repeat dynamic region with different known HTH motif present in "Master Sets".

Table 1. Molecular weight estimation of purified protein.

Theoretical Mol. Wt. (kDa)[a]	Experimentally derived Mol. Wt. (kDa)		
	V_e/V_o[b]	Size exclusion chromatography	Mass spectrometry
28.98	1.40 ± .01	26.6 ± 1	29.08 ± .05

Note: V_e/V_o: Elution volume/Void volume ratio in gel filtration chromatography (superdex 75 16/60).
[a]Determined from Protparam, Expasy.
[b]Determined from standards chromate, aprotinin, lysozyme, carbonic anhydrase, ovalbumin, albumin, ferritin, dextran.

(PDB ID; 3S4 W) and *in silico* model (Supplementary Figure 3(B)). To further investigate folding pattern and thermodynamic stability, temperature-dependent unfolding of the protein was performed in the range from 10 to 50°C with 2°C of interval, and fraction unfolded was calculated at $\lambda = 222$ nm. Further melting temperature (T_m) was calculated by fitting into a two-state unfolding pathway (Choudhary et al., 2015). These results suggest that protein is having T_m of about 36°C and follow the unfolding pattern of a two-state transition (Figure 2(B)) (Table 2).

2.3. *Three-dimensional folding of FANCI ARM repeat*

To investigate the overall folding and compactness of ARM repeat, fluorescence spectroscopy was performed with two intrinsic fluorophores 25 and 118 W (tryptophans) residues. We have recorded the scan of both folded as well as unfolded FANCI ARM repeat protein and observed the emission maxima of folded protein at $\lambda = 333$ nm, whereas for unfolded protein which was in 8 M urea, the emission maxima was at $\lambda = 345$ nm. Furthermore, at 4 M urea, FANCI ARM repeat protein shows loss in fluorescence intensity (Supplementary

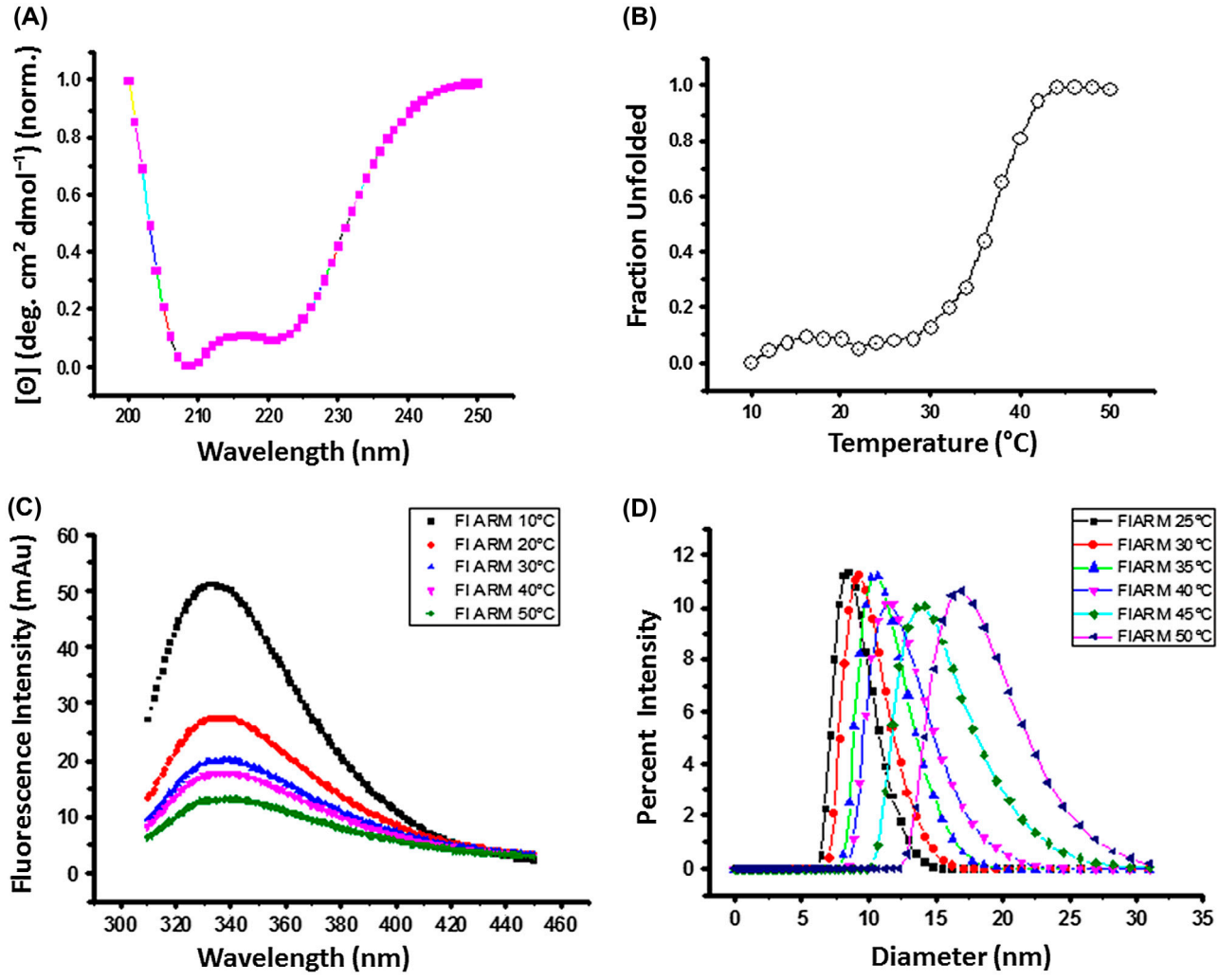


Figure 2. (A) Far-UV, CD spectra of FANCI-ARM repeat, indicating the α-helical nature of protein, (B) Thermal denaturation profile by CD showing protein unfolds at 50°C, (C) Thermal denaturation profile using Fluorescence spectroscopy showing steep decrease in intensity beyond 30°C, (D) Temperature-dependent DLS profile showing oligomer formation beyond 30°C.

Table 2. Thermal stability of protein.

Circular dichroism	Fluorescence spectroscopy	Dynamic light scattering
36.12 ± .5°C (T_m)	>30°C[a]	>30°C[b]

Note: T_m = Melting temperature.
[a]Decrease in intensity and red shift was observed.
[b]Oligomeric species were observed.

Figure 1(D)) confirming the presence of intermediate such as molten globule or oligomeric species which has sustained resemblance with FPLC and DLS data. Interestingly, temperature-dependent unfolding pattern from 10 to 50°C, loss of fluorescence intensity was observed at 30°C. The characteristic red shift of emission maxima suggests that the unfolded molten globule protein fractions are predominant beyond 30°C (Figure 2(C)). Further to validate this observation, temperature-dependent DLS studies (Figure 2(D)) were performed and found that beyond 30°C protein was losing its structural integrity and forming oligomers. Fluorescence spectroscopy, CD and DLS suggest that protein completely unfolds at 50°C (Table 2).

2.4. FANCI ARM repeat has double-stranded DNA-binding properties

DNA-binding activity of FANCI ARM repeat was monitored using radio-labelled double-stranded DNA substrates by electrophoretic mobility shift assay (EMSA). The observed dissociation constant (K_d) values for FANCI ARM repeat were 3.969 ± 1.712 μM (Figure 3(A) and (B)). The K_d values in the range of μM concentration suggest the greater affinity of FANCI ARM repeat to double strand DNA which is important for the FANCI protein to perform the DNA ICL repair.

2.5. Domain stability of FANCI ARM repeat

Compact globular domain of protein resists the protease digestion and it helps to determine the stability and dynamic conformation (Fontana et al., 1997). To confirm the stable region of ARM repeat, peptide mass fingerprinting using MALDI TOF-TOF was performed for ARM repeat as well as the region which withstand with proteolysis against trypsin protease. It has been found that ~95 amino acids at N-terminus of ARM repeat are forming compact region which shows prominent resistivity towards trypsin digestion. The modelled structure is also showing the stable region at N-terminus and dynamic region at C-terminus (Supplementary Figure 3 (A), (B), (D)–(F). Furthermore, *in silico* prediction for ARM repeat disorderness using PrDOS suggests that N-terminus is composed of amino acids forming ordered region (Supplementary Figure 1(E)), which is in agreement with peptide mass fingerprinting data. To rule out the possibility of miscleavage, *in silico* trypsin digestion prediction tool ExPASy peptide cutter was used, and a very good match with peptide mass fingerprinting results was observed. Cumulative results from limited proteolysis, mass fingerprinting concludes that ARM repeat has a stable domain at N-terminus of around 100 amino acids.

2.6. Molecular dynamics simulation and folding pattern of FANCI ARM repeat

To delineate the flexibility and dynamics of the ARM repeat domain, Molecular dynamics simulation (MDS) studies of 80 ns using GROMACS were carried out to understand the dynamic region present in the protein as well as folding pattern of HTH-type motif. MDS data were analysed by plotting Root mean square deviation (RMSD), root mean square fluctuations (RMSF), Rg and solvent accessible surface area (SASA) (Figure 4(A)–(D)). RMSD profile of ARM repeat shows dynamic behaviour with different conformations and stabilized after 60 ns. Rg fluctuation is another determinant of structural flexibility which suggests the presence of structurally disordered regions within the ARM repeat domain. RMSF for C-alpha of ARM repeat domain residues indicates amplitude of fluctuation which unravels the dynamic residual regions. RMSF profile revealed that C-terminus region specifically HTH-type region (124–143) amino acids is highly flexible. It was also evident with the projection of eigenvector 1 vs. residual RMSF ARM repeat (Supplementary Figure 1(F)). High values of SASA at C-terminus indicated that the HTH type structural motif showing high accessible surface area might act as interaction motif in ARM repeat (Figure 4(D)).

Cross-correlation analysis and RMSF sausage plot suggest that HTH structural motif has an anti-correlated motion and high fluctuation (Figure 5(A) and (B)).

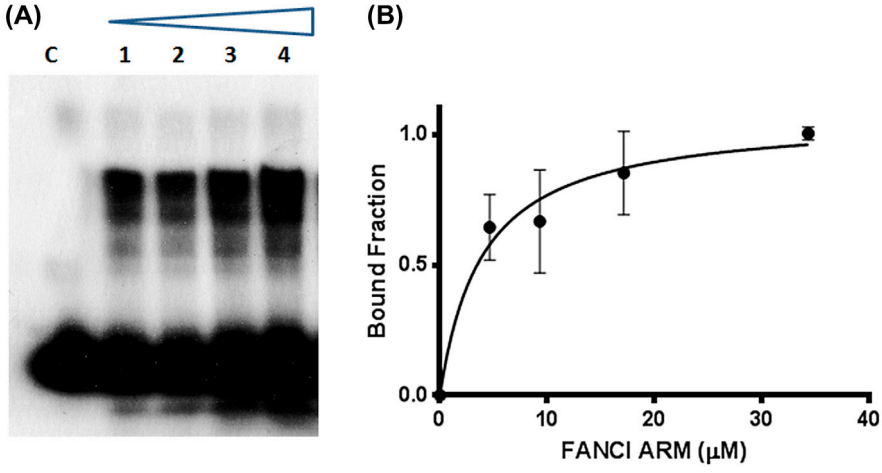


Figure 3. (A) DNA-binding analysis of FANCI ARM repeat with DNA by autoradiography (C = Control probe, 1, 2, 3, and 4 with increasing concentration of protein 4.6, 9.29, 17.14 and 34.29 μM, respectively). (B) Graph plot of bound fractions of protein to DNA.

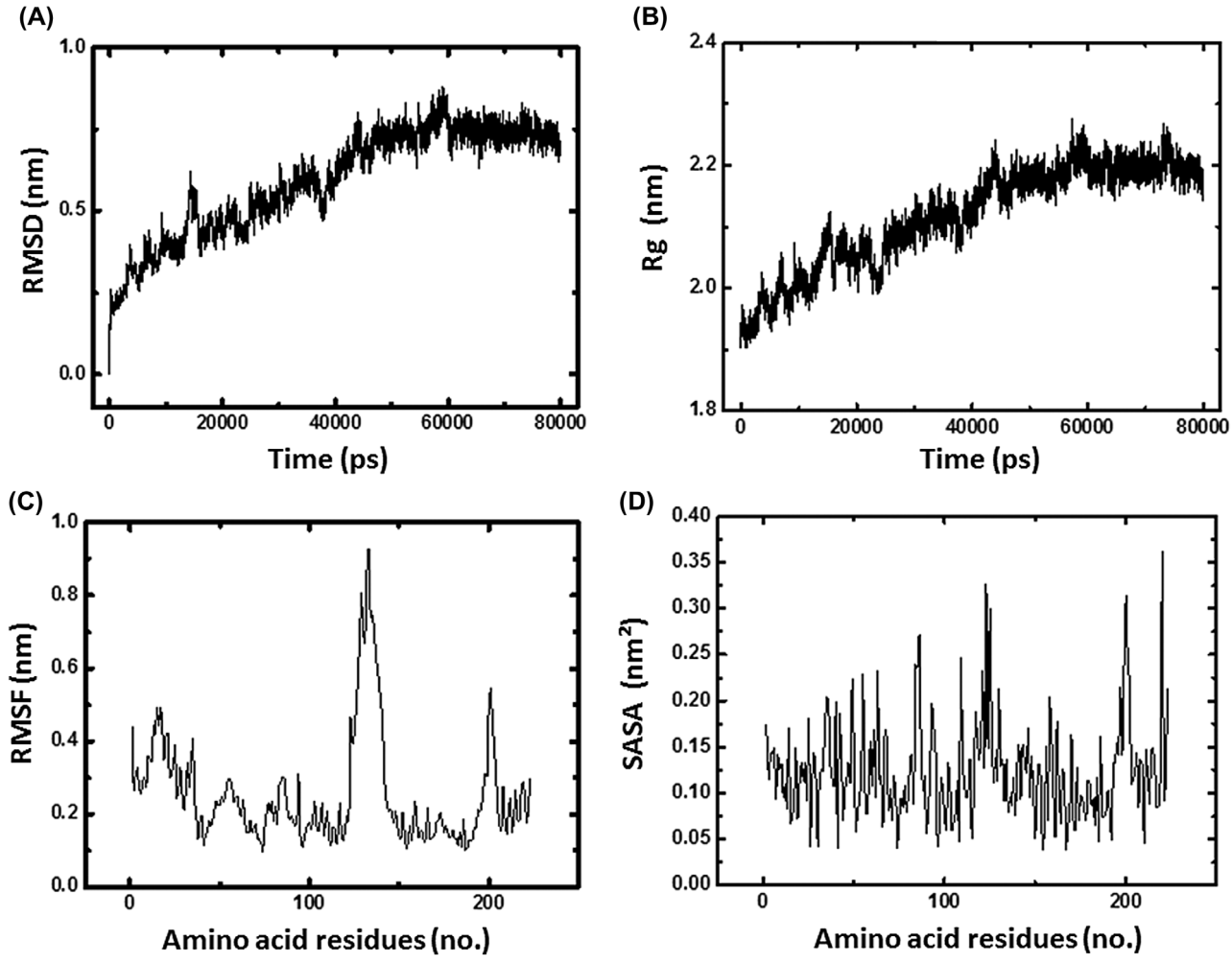


Figure 4. Molecular dynamics simulation profile of FANCI ARM repeat (A) RMSD profile and (B) Radius of gyration (Rg) profile of 80 ns showing structural transitions up to 60 ns, (C) RMSF profile showing large fluctuation of C-terminus, (D) Solvent accessible surface area (SASA) profile of h-FANCI ARM repeat protein.

Principal component analysis (PCA) was further performed for the ARM repeats of N-terminus (1–100) and C-terminus (100–223) amino acids to understand the dynamics in essential subspaces. Scree plot revealed that C-terminus is having the high eigenvalue than N-terminus suggesting large conformational motion of C-terminus domain (Figure 5(C)). The trace of covariance matrix values calculated for N-terminus and C-terminus were to be 64.71 and 131.62 nm², respectively. Hence, C-terminus comprising HTH-type motif is more dynamic than the N-terminus. Projection of eigenvector 2 on 1 indicates a large periodic tertiary structural transition in C-terminus than the N-terminus of ARM repeat protein (Figure 5(D)). PCA results suggest that protein has large concerted motion due to the presence of HTH type structural motif. Furthermore, to look at cross-correlation and domain mobility, Gaussian network modelling (GNM) and normal mode analysis (NMA) were performed and observed that HTH-type motif at C-terminus part has anti-correlated motion, more dynamicity than N-terminus, and both are in opposite direction (Supplementary Figure 3(C), (F)). Essential Dynamics results are in agreement with MDS and suggest that structurally dynamic HTH structural motif might be stabilized by binding to DNA during ICL DNA repair.

2.7. Structural characterization of HTH-motif

Molecular dynamics indicates that HTH structural motif is highly flexible. Secondary structure of ARM repeat was characterized using dictionary of secondary structure of protein (DSSP). It has been observed that large helix turn to loop transition is pre-dominant in the HTH motif. The results obtained from DSSP also corroborates well

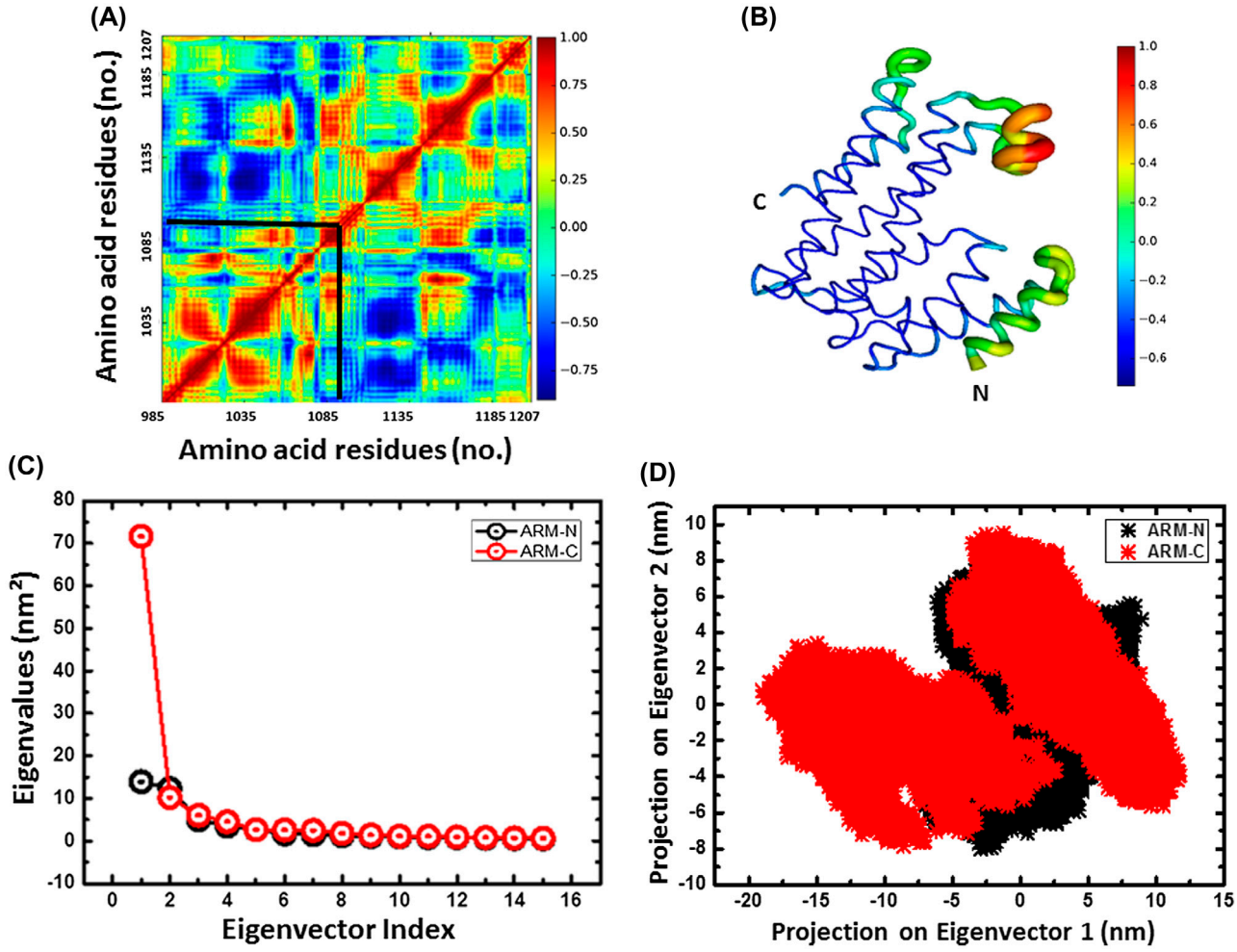


Figure 5. (A) Cross-correlation diagonal matrix of h-FANCI ARM repeat, (B) RMSF sausage profile, (C) Scree plot of N-terminus (1–100) amino acids and C-terminus (100–223) amino acids, (D) PCA profile of Eigenvector 1 and 2 of N-terminus (1–100) amino acids and C-terminus (100–223) amino acids.

with the results obtained from structural alignment over the structures extracted from trajectory at different time points. DSSP analysis of the 80 ns simulation data found that only HTH structural region showing unfolding at 15 ns, and unable to form stable structure in due course of simulation (Figure 6(A) and (B)). Thus, it suggests that HTH-type region has a motif character (Religa et al., 2007). It is well established that HTH motifs bind specifically to major groove of DNA (Ohlendorf et al., 1982). To explore the binding of FANCI ARM repeat with DNA, we have performed docking analysis and found that HTH-type region binds to major groove of DNA (Figure 6(C)). It has also been observed that conserved amino acids such as Gln, Ser, Glu and Thr (Luscombe & Thornton, 2002; Ohlendorf et al., 1982) form interactions interface between DNA and HTH-type motif of ARM repeat (Figure 6(D)).

3. Materials and methods

All the used chemicals were of molecular biology grade and purchased from Sigma–Aldrich, unless otherwise specified. Protein and buffer solutions were filtered well and degassed before use.

3.1. Gene cloning, protein expression and purification

FANCI ARM repeat region (985–1207) amino acids was PCR amplified (Thermocycler, Bio-rad) using cDNA of full length FANCI (kind gift from Dr Stephen J. Elledge, Harvard Medical School, USA) as a template. The forward and reverse primers are 5′-GTCGGATCCGA-GAACCTGTACTTTCAGGGTCTAGTCACGGTTCTT-ACCAG-3′ and 5′-GTCCTCGAGCTATTAGGGGGTCA-GATGAGAACCAG-3’, respectively. The forward primer was designed with the TEV protease site having

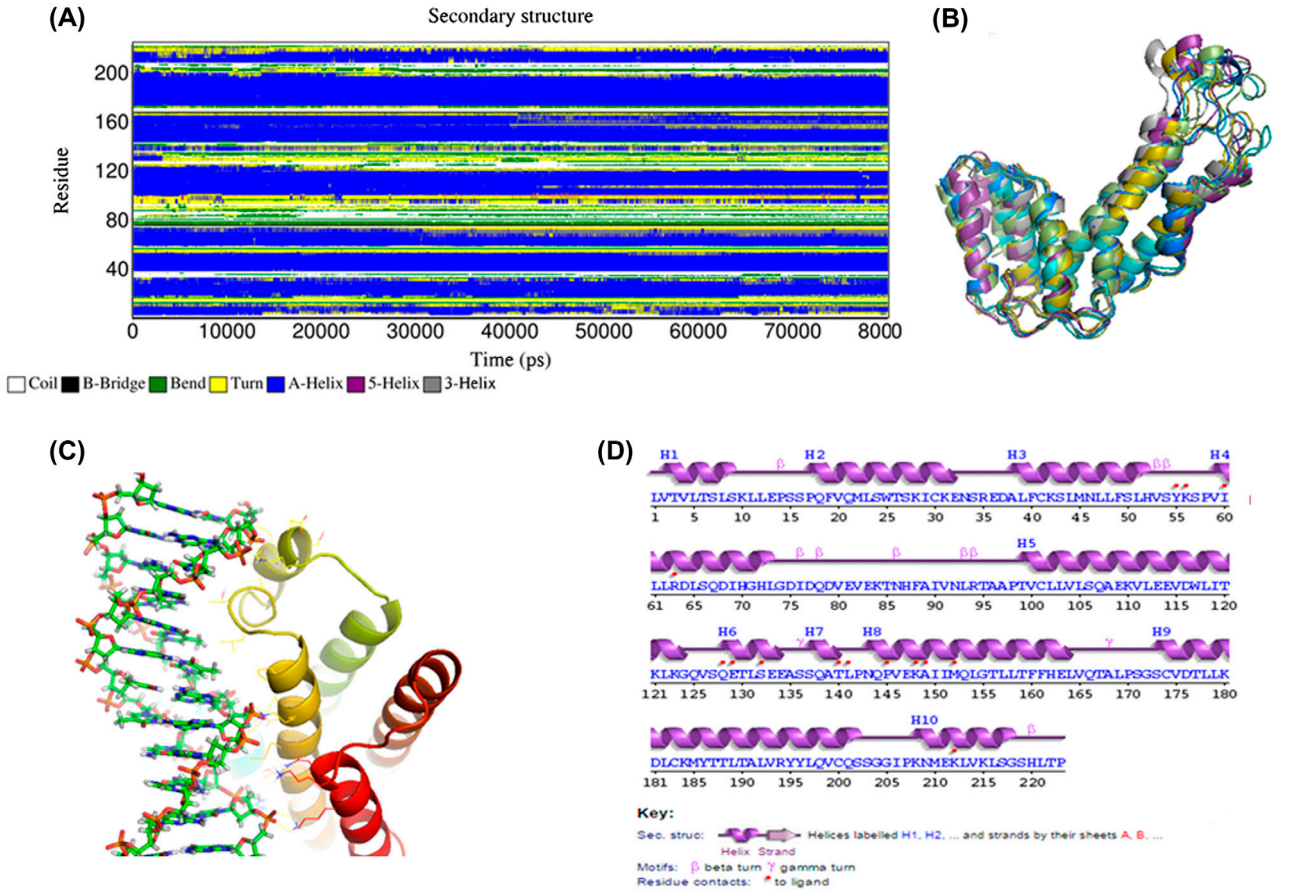


Figure 6. (A) DSSP profile of FANCI ARM repeat protein, (B) Superimposed structures in different time points (10–40 ns) of trajectories showing large structural rearrangement in HTH type motif, (C) Docking profile of FANCI ARM repeat with DNA do-decamer shows major groove binding mode of putative HTH type motif, (D) Intermolecular interactions analysis profile of docked pose indicates the conserve residues present at binding interface marked (in red dots).

ENLYFQG amino acids for native protein purification. PCR product of amplified ARM repeat was sub-cloned into the pET28a vector (Novagen). ARM repeat cloned in pET28a was expressed and purified using the *E. coli* Rosetta (2DE3) cells (Novagen) by inducing at $O.D_{600}$ between .6–.8, with .4 mM IPTG at 22°C overnight. Protein was purified in the buffer containing 300 mM NaCl, 50 mM Tris, .1% Triton-X, 5 mM beta-mercaptoethanol. 6× His-Tag fusion protein was purified by affinity chromatography (Ni-NTA beads, Qiagen) and further passed through AKTA FPLC gel filtration column (Superdex-75) in the 300 mM NaCl, 50 mM Tris, 5 mM β-Mercaptoethanol buffer to get highly purified protein.

Standard protein markers of known molecular weight were used to calculate void volume and the total volume of the AKTA–FPLC column. The experiments were repeated twice and averaged for elution volume calculation.

3.2. *Chemical cross-linking assay*

Purified protein was incubated with .1% glutaraldehyde and the reaction was terminated in a time-dependent manner (0, 2.5, 5, 10, 15, 30, and 60 min, respectively) by adding 5 µl of 1 M Tris pH-8.5. The untreated protein sample was taken as control. Cross-linked product was mixed with equal amount of Laemmli buffer and analysed on 12% SDS-PAGE gel.

3.3. *Limited proteolysis and mass spectrometry*

Purified protein (1 mg/ml) was treated with trypsin in time-dependent manner with their final concentration of 10 pg/µl and untreated protein was taken as control. Reaction mixture was incubated at 37°C (trypsin) for different time period 1, 5, 10, 15, 30, 60 and 180 min. Reactions were terminated by adding 2 µl of 200 mM PMSF (Sigma–Aldrich). Samples were heated with laemmli buffer at 85°C for 5 min and analysed on

SDS-PAGE gel by coomassie staining. Band corresponding to 14 kDa was considered as stable fragment and was subjected to trypsin digestion followed by mass spectrometry (MALDI TOF–TOF Ultraflex-II from Bruker Daltonics, Germany) in which peptides were captured with high sensitivity at attomole range for peptide mass fingerprinting. Domain of interest was identified by Mascot analysis with Bio Tool software (Bruker Daltonics).

3.4. Dynamic light scattering

FANCI ARM repeat at a concentration of (1 mg/ml) was filtered (.22 µm) and degassed at 4°C prior to all DLS measurements. Malvern zeta-sizer was used to study the oligomeric characteristics of protein at different temperatures. Wyatt DynaPro NanoStar was also used for DLS experiment.

3.5. Circular dichroism and fluorescence spectroscopy

CD polarimeter (Jasco J-815, Japan) in the far-UV range (λ = 190–260 nm) was used to characterize the secondary structure pattern of protein. Averages of seven spectra were taken for final representation in mean residual ellipticity. Further, averaged spectrum was used for secondary structure quantification using K2D3 server (Louis-Jeune, Andrade-Navarro, & Perez-Iratxeta, 2012). For thermal denaturation experiment spectra were recorded from 10 to 50°C with 2°C temperature interval.

Micro-environment of tryptophans (intrinsic fluorophore) at the hydrophobic core of the protein was monitored by fluorescence spectrophotometer (Horiba, Japan) at the excitation wavelength of λ = 295 nm. The emission spectra were recorded from λ = 310 to 450 nm. For thermal and chemical unfolding experiments, 10 µM protein was used to record the spectra.

3.6. DNA binding by EMSA

DNA-binding activity of FANCI ARM repeat protein was determined using EMSA as described earlier (Rajpurohit & Misra, 2013). Eighty-two nucleotide long random sequence oligonucleotide was used as dsDNA substrate, and was made by annealing with its complementary strand (Table 3). The dsDNA were labelled with [^{32}P] γATP using polynucleotide kinase and purified by G-25 column. The .2 pmol of labelled probe (dsDNA) was incubated with increasing concentrations of FANCI ARM repeat protein in 10 µl of the reaction containing buffer 35 mM Tris–HCl, pH 7.5, 2.5 mM $MgCl_2$, 25 mM KCl, 1 mM ATP and 1 mM DTT for 20 min at 37°C. Products were analysed on a 6% native polyacrylamide gel, and signals were recorded by autoradiography. DNA band intensity either in free form or bound to protein was quantified using GelQuant software (http://biochemlabsolutions.com/GelQuantNET.html). Per cent bound fraction of DNA was plotted against protein concentration using GraphPad Prism 5 (http://www.graphpad.com/scientific-software/prism/) and Kd for curve fitting of individual plot was determined.

3.7. Molecular modelling, principal component analysis and NMA

FANCI sequences from 985 to 1207 amino acids were retrieved from UniProtKB (Uniprot ID: Q9NVI1) and submitted to Robetta server for 3D model building (Kim, Chivian, & Baker, 2004). Steriochemical refinement of Ramachandran outliers present in the model was performed by ModLoop server (Fiser & Sali, 2003). Further, refined model was validated by SAVES server (http://services.mbi.ucla.edu/SAVES/) and Protein Structure analysis (ProSA) (Wiederstein & Sippl, 2007) (Supplementary Figure 2(E)). Validated model was used for MDS studies by GROMACS 4.5.5 with implementation of OPLS-AA/L force field (Hess, Kutzner, van der Spoel, & Lindahl, 2008; Kaminski, Friesner, Tirado-Rives, & Jorgensen, 2001). The systems were solvated using TIP3P water model in a cubic box with periodic boundary conditions. Furthermore, counter-ions were added to neutralize the system. The systems were first energy minimized using steepest descent algorithm with a tolerance of 1000 kJ/mol/nm. Electrostatic interactions were calculated using particle-mesh Ewald summation (Abraham & Gready, 2011) with 1 nm cut-offs. Columbic interactions and van der Waal's interactions were calculated with a distance cut-off of 1.4 nm. System was equilibrated by applying positional restraints on the structure using NVT followed by NPT ensemble for 100 ps each. Temperature of 300 K was coupled by Berendsen

Table 3. Radioactive labelled probe DNA sequences.

1.	82F	5′GAATTCGGTGCGCATAATGTATATTATGTTAAATCATGTCC-CTGCCCCAATATAAACCAAGCGTATGCAGTAAGCTTCGATC3′	EMSA
2.	82R	5′ GATCGAAGCTTACTGCATACGCTTGGTTTATATTGGGGCAGG-GACATGATTTAACATAATATACATTATGCGCACCGAATTC3′	EMSA

thermostat with pressure of one bar using SHAKE algorithm (Ryckaert, Ciccotti, & Berendsen, 1977). The equilibrated systems were subjected to 80 ns of production run with time-step integration of 2fs. The trajectories were saved at every 2 ps and analysed using Gromacs 4.5.5. RMSD, RMSF, radius of gyration (Rg), hydrogen bonds, SASA and DSSP (Kabsch & Sander, 1983) were analysed. Cross-correlation for PCA (Amadei, Linssen, & Berendsen, 1993) was performed. Eigenvector and Eigenvalues were calculated after diagonalizing the covariance matrix. Trace of co-variance matrix was calculated by adding up all the eigenvalues. The eigenvalue calculated was plotted for each eigenvector to understand the dynamics. NMA, anisotropic network modelling and Gaussian network modelling were performed using R 3.2 package and ProDy (Protein Dynamics 1.7), respectively (Bahar, Erman, Jernigan, Atilgan, & Covell, 1999; Grant, Rodrigues, ElSawy, McCammon, & Caves, 2006)

3.8. Docking studies

Molecular docking studies were performed using HADDOCK server (de Vries, van Dijk, & Bonvin, 2010). The structure of the DNA do-decamer (PDB ID: 1BNA) was downloaded from the protein data bank (http://www.rcsb.org./pdb). Both the ligand and the receptor were made in the PDB format. Prior to docking, all hetero atoms were removed from ligand and receptor. Docked pose of best HADDOCK score was selected. Intermolecular interactions were analysed by LigPlot (Laskowski & Swindells, 2011), PDB SUM generation (http://www.ebi.ac.uk/thornton-srv/databases/pdbsum/Generate.html) and visualized using the PyMOL molecular graphics software. (http://pymol.sourceforge.net/).

Supplementary material

The supplementary material for this paper is available online at http://dx.doi.10.1080/07391102.2016.1235514.

Acknowledgements

We thank DBT-BTIS facility, BARC supercomputing facility UTKARSH and Proteomics facility at ACTREC for providing necessary software to this study.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This study was supported by Seed in Air grant from TMC.

References

Abraham, M. J., & Gready, J. E. (2011). Optimization of parameters for molecular dynamics simulation using smooth particle-mesh Ewald in GROMACS 4.5. *Journal of Computational Chemistry, 32*, 2031–2040.

Alter, B. P., Greene, M. H., Velazquez, I., & Rosenberg, P. S. (2003). Cancer in Fanconi anemia. *Blood, 101*, 2072.

Amadei, A., Linssen, A. B., & Berendsen, H. J. (1993). Essential dynamics of proteins. *Proteins: Structure, Function, and Genetics, 17*, 412–425.

Andrade, M. A., Petosa, C., O'Donoghue, S. I., Muller, C. W., & Bork, P. (2001). Comparison of ARM and HEAT protein repeats. *Journal of Molecular Biology, 309*, 1–18.

Aravind, L., Anantharaman, V., Balaji, S., Babu, M. M., & Iyer, L. M. (2005). The many faces of the helix-turn-helix domain: Transcription regulation and beyond. *FEMS Microbiology Reviews, 29*, 231–262.

Bahar, I., Erman, B., Jernigan, R. L., Atilgan, A. R., & Covell, D. G. (1999). Collective motions in HIV-1 reverse transcriptase: Examination of flexibility and enzyme function. *Journal of Molecular Biology, 285*, 1023–1037.

Brennan, R. G., & Matthews, B. W. (1989). The helix-turn-helix DNA binding motif. *The Journal of Biological Chemistry, 264*, 1903–1906.

Butturini, A., Gale, R. P., Verlander, P. C., Adler-Brecher, B., Gillio, A. P., & Auerbach, A. D. (1994). Hematologic abnormalities in Fanconi anemia: An international Fanconi anemia registry study. *Blood, 84*, 1650–1655.

Castella, M., Jacquemont, C., Thompson, E. L., Yeo, J. E., Cheung, R. S., Huang, J. W., … Taniguchi, T. (2015). FANCI regulates recruitment of the FA core complex at sites of DNA damage independently of FANCD2. *PLoS Genetics, 11*, e1005563.

Choudhary, R. K., Vikrant, Siddiqui, Q. M., Thapa, P. S., Raikundalia, S., Gadewal, N., … Varma, A. K. (2015). Multimodal approach to explore the pathogenicity of BARD1, ARG 658 CYS, and ILE 738 VAL mutants. *Journal of Biomolecular Structure and Dynamics*, *34,* 1–12.

Coates, J. C. (2003). Armadillo repeat proteins: Beyond the animal kingdom. *Trends in Cell Biology, 13*, 463–471.

Cohen, M. M., Simpson, S. J., Honig, G. R., Maurer, H. S., Nicklas, J. W., & Martin, A. O. (1982). The identification of fanconi anemia genotypes by clastogenic stress. *The American Journal of Human Genetics, 34*, 794–810.

Crist, R. C., Roth, J. J., Baran, A. A., McEntee, B. J., Siracusa, L. D., & Buchberg, A. M. (2010). The armadillo repeat domain of Apc suppresses intestinal tumorigenesis. *Mammalian Genome, 21*, 450–457.

de Vries, S. J., van Dijk, M., & Bonvin, A. M. (2010). The HADDOCK web server for data-driven biomolecular docking. *Nature Protocols, 5*, 883–897.

Dorsman, J. C., Levitus, M., Rockx, D., Rooimans, M. A., Oostra, A. B., Haitjema, A., … Joenje, H. (2007). Identification of the Fanconi anemia complementation group I gene, FANCI. *Cellular Oncology, 29*, 211–218.

Fiser, A., & Sali, A. (2003). ModLoop: Automated modeling of loops in protein structures. *Bioinformatics, 19*, 2500–2501.

Fontana, A., Zambonin, M., Polverino de Laureto, P., De Filippis, V., Clementi, A., & Scaramella, E. (1997). Probing the conformational state of apomyoglobin by limited proteolysis. *Journal of Molecular Biology, 266*, 223–230.

Grant, B. J., Rodrigues, A. P., ElSawy, K. M., McCammon, J. A., & Caves, L. S. (2006). Bio3d: An R package for the comparative analysis of protein structures. *Bioinformatics, 22*, 2695–2696.

Hess, B., Kutzner, C., van der Spoel, D., & Lindahl, E. (2008). GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of Chemical Theory and Computation, 4*, 435–447.

Kabsch, W., & Sander, C. (1983). Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers, 22*, 2577–2637.

Kaminski, G. A., Friesner, R. A., Tirado-Rives, J., & Jorgensen, W. L. (2001). Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *The Journal of Physical Chemistry B, 105*, 6474–6487.

Kim, D. E., Chivian, D., & Baker, D. (2004). Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Research, 32*, W526–W531.

Laskowski, R. A., & Swindells, M. B. (2011). LigPlot+: Multiple ligand–protein interaction diagrams for drug discovery. *Journal of Chemical Information and Modeling, 51*, 2778–2786.

Longerich, S., San Filippo, J., Liu, D., & Sung, P. (2009). FANCI binds branched DNA and is monoubiquitinated by UBE2T-FANCL. *Journal of Biological Chemistry, 284*, 23182–23186.

Louis-Jeune, C., Andrade-Navarro, M. A., & Perez-Iratxeta, C. (2012). Prediction of protein secondary structure from circular dichroism using theoretically derived spectra. *Proteins: Structure, Function, and Bioinformatics, 80*, 374–381.

Luscombe, N. M., & Thornton, J. M. (2002). Protein–DNA interactions: Amino acid conservation and the effects of mutations on binding specificity. *Journal of Molecular Biology, 320*, 991–1009.

McCabe, K. M., Olson, S. B., & Moses, R. E. (2009). DNA interstrand crosslink repair in mammalian cells. *Journal of Cellular Physiology, 220*, 569–573.

Ohlendorf, D. H., Anderson, W. F., Fisher, R. G., Takeda, Y., & Matthews, B. W. (1982). The molecular basis of DNA-protein recognition inferred from the structure of cro repressor. *Nature, 298*, 718–723.

Pace, C. N., & Scholtz, J. M. (1998). A helix propensity scale based on experimental studies of peptides and proteins. *Biophysical Journal, 75*, 422–427.

Rajpurohit, Y. S., & Misra, H. S. (2013). Structure-function study of deinococcal serine/threonine protein kinase implicates its kinase activity and DNA repair protein phosphorylation roles in radioresistance of *Deinococcus radiodurans*. *The International Journal of Biochemistry & Cell Biology, 45*, 2541–2552.

Religa, T. L., Johnson, C. M., Vu, D. M., Brewer, S. H., Dyer, R. B., & Fersht, A. R. (2007). The helix–turn–helix motif as an ultrafast independently folding domain: The pathway of folding of Engrailed homeodomain. *Proceedings of the National Academy of Sciences, 104*, 9272–9277.

Ryckaert, J.-P., Ciccotti, G., & Berendsen, H. J. (1977). Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. *Journal of Computational Physics, 23*, 327–341.

Smogorzewska, A., Matsuoka, S., Vinciguerra, P., McDonald 3rd, E. R., Hurov, K. E., Luo, J., … Elledge, S. J. (2007). Identification of the FANCI Protein, a monoubiquitinated FANCD2 paralog required for DNA repair. *Cell, 129*, 289–301.

Tsytlonok, M., Craig, P. O., Sivertsson, E., Serquera, D., Perrett, S., Best, R. B., … Itzhaki, L. S. (2013). Complex energy landscape of a giant repeat protein. *Structure, 21*, 1954–1965.

Wang, W. (2007). Emergence of a DNA-damage response network consisting of Fanconi anaemia and BRCA proteins. *Nature Reviews Genetics, 8*, 735–748.

Wiederstein, M., & Sippl, M. J. (2007). ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Research, 35*, W407–W410.

Yuan, F., El Hokayem, J., Zhou, W., & Zhang, Y. (2009). FANCI protein binds to DNA and interacts with FANCD2 to recognize branched structures. *Journal of Biological Chemistry, 284*, 24443–24452.

International Journal of Biological Macromolecules xxx (2017) xxx–xxx

Contents lists available at ScienceDirect

International Journal of Biological Macromolecules

journal homepage: www.elsevier.com/locate/ijbiomac

Studies of protein–protein interactions in Fanconi anemia pathway to unravel the DNA interstrand crosslink repair mechanism

Mohd Quadir Siddiqui[a,c], Yogendra S. Rajpurohit[b], Pankaj S. Thapa[a,c], Ganesh Kumar Maurya[c], Kuheli Banerjee[a,c], Mudassar Ali Khan[a,c], Pragnya Panda[a], Syed K. Hasan[a], Nikhil Gadewal[a], Hari S. Misra[b,c], Ashok K. Varma[a,c,*]

[a] Advanced Centre for Treatment, Research and Education in Cancer, Kharghar, Navi Mumbai, Maharashtra 410 210, India
[b] Molecular Biology Division, Bhabha Atomic Research Centre, Mumbai 400085, India
[c] Homi Bhabha National Institute, Training School Complex, Anushaktinagar, Mumbai 400 094, India

ARTICLE INFO

Article history:
Received 25 March 2017
Received in revised form 16 May 2017
Accepted 30 May 2017
Available online xxx

Keywords:
Protein–protein interactions
Cue domain
ARM repeat
FANCI-FANCD2

ABSTRACT

Fanconi anemia (FA), a cancer predisposition syndrome exhibits hallmark feature of radial chromosome formation, and hypersensitivity to DNA crosslinking agents. A set of FA pathway proteins mainly FANCI, FANCD2 and BRCA2 are expressed to repair the covalent crosslink between the dsDNA. However, FA, BRCA pathways play an important role in DNA ICL repair as well as in homologous recombination repair, but the presumptive role of FA-BRCA proteins has not clearly explored particularly in context to function associated protein–protein interactions (PPIs). Here, *in-vivo*, *in-vitro* and *in-silico* studies have been performed for functionally relevant domains of FANCI, FANCD2 and BRCA2. To our conclusion, FANCI ARM repeat interacts with FANCD2 CUE domain and BRCA2 C-terminal region. Interestingly, FANCD2 CUE domain also interacts strongly with BRCA2 C-terminal region. Interactions between BRCA2 CTR and functionally relevant mutations Ser222Ala (cell cycle checkpoint mutant) and Leu231Arg (DNA ICL repair mutant) present in FANCD2 CUE domain have been analysed. To our finding, these mutations abrogate the binding between FANCD2 CUE domain and BRCA2 CTR. Furthermore, (1) different domain of FANCI, FANCD2 and BRCA2 are playing important role in PPIs, (2) mutations cause the impairment in the PPIs which in turn may disrupt the DNA ICL repair mechanism.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

Fanconi anemia (FA) is a cancer predisposition disorder, characterized by developmental defects and bone marrow failure [1]. The radial chromosomes and high sensitivity towards DNA crosslinking agents are the hallmark features of FA cells [2]. The DNA inter-crosslink (DNA ICL) repair mechanism helps to elucidate the molecular pathogenesis of FA. The Fanconi anemia proteins repair the DNA inter-crosslink and interact with different proteins involved in DNA double-strand break repair pathway such as BRCA1, BRCA2 (FANCD1), FANCN (PALB2) and FANCJ (BRIP1) [3–5]. BRCA2 has been involved in DNA double-strand break repair through homologous recombination [3,6]. The DNA repair pathway in BRCA2 deficient cells leads to error prone single-strand annealing [6]. Furthermore, DNA repair processes such as nucleolytic incision, translesion DNA synthesis, and homologous recombination have been co-ordinated by FA/BRCA pathway [3,6]. It was reported that FANCI, FANCD2 and BRCA2 functions downstream of the FA core complex, and also perform independent functions in DNA ICL repair mechanism [7,8].

The C-terminus region of BRCA2 harbors BLAT (BRCA2-motif in *Leishmania, Arabidopsis, and Trypanosoma*) domain, which is highly conserved in divergent species [9]. The BLAT domain includes BRCA2 region encompassing from 2350 to 2545 amino acids interacts with FANCD2 (1-1102 amino acids). However, FANCD2 lacking N-terminus residues (1-325) failed to interact with BRCA2 [10]. In addition, BRCA2 proteins with C-terminal truncations were observed in cell lines derived from FA patients [3]. Thus, it confirm the importance of FANCD2 N-terminus and BRCA2 C-terminus in the protein–protein interactions (PPI). FA proteins particularly FANCI and FANCD2 form interdependent complexes required for chromatin association and DNA ICL repair. FANCI and FANCD2 comprise evolutionarily conserved domains such as armadillo repeat

* Corresponding author at: Principal Investigator & Scientific Officer 'F', Tata Memorial Centre, Advanced Centre for Treatment, Research and Education in Cancer, Kharghar, Navi-Mumbai, Maharastra State 410 210, India.
E-mail address: avarma@actrec.gov.in (A.K. Varma).

http://dx.doi.org/10.1016/j.ijbiomac.2017.05.166
0141-8130/© 2017 Elsevier B.V. All rights reserved.

Table 1
Bacterial strains and plasmids used in this study.

Bacterial strains	Description
E. coli BL21 (DE3)	B strain, *fhu*A2 (*lon*) *ompT* gal (λ*DE3*)(*dcm*) Δ*hsdS i21* Δ*nin5*
E. coli Novablue	*end* A1 *hsd* R17(rK12⁻ mK12+) supE44 *thi*-1 *rec*A1 *gyr*A96 *rel*A1 *lac*F′ [*proA*⁺ B⁺ *lac*I^q Z Δ*M15*::Tn*10*] Tet^R
E. coli DH5α	*fhuA2* Δ(*argF-lacZ*)U169 *pho*A *gln*V44 *Φ80* Δ(*lacZ*)M15 *gyrA96 rec*A1 *rel*A1 *end*A1
E. coli Rosetta 2(DE3)	F⁻ *ompT hsdS*$_B$(r$_B$− m$_B$−) *gal dcm* (DE3) pRARE2 (Cam^R)
E. coli BTH101	F′, cya-99, *ara*D139, *gal*E15, *gal*K16, *rps*L1 (*Str*R), *hsd*R2, *mcr*A1 *mcr*B1, *rel*A1
Plasmids	
pKNT25	Origin of replication from pSU40. N-terminal fusion with T25 adenylate cyclase fragment, 3.4 kb, Kanamycin resistance
pUT18	col*E1* origin of replication from pUC19. N-terminal fusion with T18 adenylate cyclase fragment, 3 kb, Ampicillin resistance
pET28a+	pBR322 origin, f1 origin, lacI coding sequence, T7 promoter, T7 transcription start, His-TAG, T7 terminator
pGEX4T-1	pBR322 origin, GST (N-term), tac promoter, size 4969 bp, Ampicillin resistance

(ARM repeat) in FANCI [11], and coupling of ubiquitin conjugation to endoplasmic reticulum domain (CUE domain) in FANCD2 [12]. The FANCD2 CUE domain plays an important role in interaction with monoubiquitinated FANCI, and thereby maintains the protein stability [12].

The FANCI and FANCD2 also known to interact with BRCA2 [3–5,10,12] (Supplementary Fig. 1). Two functionally relevant mutations have been identified in FANCD2 CUE domain, first cell cycle checkpoint mutant (Ser222Ala) [13,14] and second DNA ICL repair mutant (Leu231Arg) [15]. Considering the importance of mutations, and functional diversity of different domains present in FANCI, FANCD2 and BRCA2, we decided to systematically evaluate protein–protein interactions between FANCI ARM repeat, FANCD2 CUE domian (FD2 CUE) and BRCA2C terminal region (BRCA2 CTR). We also analysed the interactions of mutated FANCD2 CUE domain (Ser222Ala and Leu231Arg) with BRCA2 C-terminal region, using multi model *in-vivo*, *in-vitro* and *in-silico* approach.

2. Materials and methods

All molecular biology grade chemicals and enzymes were procured from Sigma Chemicals Company, USA, Roche Biochemicals, Mannheim, Germany, New England Biolabs, USA and Fermentas (Thermo Scientific), or unless otherwise stated. Analytical grade bacterial culture and chemicals: LB medium (Himedia, India, Cat. No. M575), 90 mm (diameter) × 15 mm petri plates (Himedia, India, Cat. No. PW008) Molecular biology grade antibiotics: Ampicillin (Himedia, India, Cat. No. CMS645), Kanamycin (Himedia, India, Cat. No. TC136), Chloramphenicol (Himedia, India, Cat. No. TC204) and β-D-Isopropyl thiogalactopyranoside (IPTG) (MP biomedicals, USA, Cat. No. 102101) and Bacterial plasmid DNA isolation kits: Miniprep kit (Qiagen, Germany, Cat. No. 27106), Plasmid DNA Maxi prep kit (Qiagen, Germany, Cat. No. 12163), Polymerase Chain Reaction extraction kit (Qiagen, Germany, Cat. No. 28106) and Agarose Gel DNA extraction kit (Qiagen, Germany, Cat. No. 28706).

2.1. Bacterial strains, plasmids and materials

cDNA of FANCI, FANCD2 and BRCA2 were generous gift from Prof. Stephen Elledge and Prof. Markus Grompe, USA respectively. The *E. coli* strains DH5α and NOVABLUE were used for molecular cloning. Bacterial strain *E. coli* BL21 (DE3) pLysS and Rosetta 2(DE3) were used for expression the proteins. *E. coli* BTH101 was used for co-expression of the respective interacting partners for *in vivo* protein–protein interaction studies. The pET28a(+), pGEX4T-1, pUT18 and pKNT25 expression vector were preserved in *E. coli* DH5α/NOVABLUE strains (Table 1**).** Selectable antibiotic markers Ampicillin (100 mg/ml), kanamycin (50 mg/ml) and chloramphenicol (34 mg/ml) were used as required. Recombinant *E. coli* harbouring expression vectors and their derivatives were grown in the presence of respective antibiotics. All recombinant techniques, expression and purification of proteins were performed as described earlier [16–18].

2.2. Construction of expression plasmids, cloning and site –directed –mutagenesis

Construction of recombinant plasmids for the expression and purification of proteins were made as described earlier [16–18]. FANCI (985-1207), FANCD2 (1-254) and BRCA2 (2350-2545) amino acids region were PCR amplified using respective human cDNA template and specific primers (Table 2**)**. PCR amplicons and vectors were digested by restriction enzyme (RE) such as *Bam*HI, *Eco*RI, Kpn1 & *Xho*I, as required and sticky ends were ligated, using ligation reaction. Further, recombinant clones were confirmed by double RE digestion and DNA sequencing. Details of bacterial strains, plasmids and primers used in this study are given in Tables 1 and 2. To study the protein–protein interactions, *in vivo* and *in vitro*, FANCI (985-1207), FANCD2 (1-254) and BRCA2 (2350-2545) amino acids region were PCR amplified using gene specific primers with required restriction sites and incorporated in

Table 2
List of the primers used in this study.

S.No.	Name of Primer	Primer Sequence (Forward & Reverse)	Vector
1	FI_ARM_BTH_F	5′GCGGATCCGGCAGCTCTAGTCACGGTTCTTACCAG 3′	pUT18
2	FI_ARM_BTH_R	5′CGGGGTACCGCGGCCGCGGGGGTCAGATGAGAACCAG 3′	pUT18
3	FD2_CUE_BTH_F	5′GCGGATCCGGCAGCTATGGTTTCCAAAAGAAGACTG 3′	pUT18
4	FD2_CUE_BTH_R	5′CCGGAATTCGCGGCCGCAAGTCGGAGGCTTGAAAG 3′	pUT18
5	BRCA2_BTH_F	5′GCGGATCCGGCAGCTACCGCACCTGGTCAAGAATTTCTG 3′	pKNT25
6	BRCA2_BTH_R	5′CCGGAATTCGCGGCCGCAACGCCATACGTATA 3′	pKNT25
7	FD2_CUE_F	5′GTCGGATCCGAGAACCTGTACTTTCAGGGTATGGTTTCCAAAAGAAGACTG 3′	pGEX4T-1
8	FD2_CUE_R	5′GTCCTCGAGCTATTAAAGTCGGAGGCTTGAAAGGAC 3′	pGEX4T-1
9	FD2_CUE_SA_F	5′ ATCCTAGGGGATGCCCAGCACGCTGAT 3′	pGEX4T-1, pUT18
10	FD2_CUE_SA_R	5′ ATCAGCGTGCTGGGCATCCCCTAGGAT 3′	pGEX4T-1 pUT18
11	FD2_CUE_LR_F	5′ GTGGGGAAAGAACGCAGTGACCTACTG 3′	pGEX4T-1, pUT18
12	FD2_CUE_LR_R	5′ CAGTAGGTCACTGCGTTCTTTCCCCAC 3′	pGEX4T-1, pUT18
13	BRCA2_F	5′GTCGGATCCGAGAACCTGTACTTTCAGGGTACCGCACCTGGTCAAGAATTTCTG 3′	pET28a+
14	BRCA2_R	5′GTCGAATTCCTATTAAACGCCATACGTATACAGCTG 3′	pET28a+

Please cite this article in press as: M.Q. Siddiqui, et al., Int. J. Biol. Macromol. (2017), http://dx.doi.org/10.1016/j.ijbiomac.2017.05.166

BTH pUT18 and pKNT25 vectors and pET28a+, pGEX4T-1 vectors (Tables 1 and 2).

2.3. in vivo *protein–protein interaction studies*

In vivo interactions of different proteins were monitored using bacterial two-hybrid system (BACTH) as described earlier [19–21]. In brief, cloned recombinant plasmids FANCI (985-1207), FANCD2 (1-254) and BRCA2 (2350-2545) amino acids from human cDNA template were cloned in bacterial two-hybrid plasmids pUT18 and pKNT25 and a series of recombinant plasmids were generated as given in Table 2. These constructs were co-transformed in *E. coli* BTH101 host in different combinations and interactions between co-expressing recombinant proteins were detected by spotting recombinant cells on LB agar plates supplemented with appropriate antibiotics. These cells were grown overnight in triplicates, and 5 μL of it was spotted on LB agar plate containing 5′ bromo 4 chloro-3-indolyl-β-D-galactopyranoside (X-Gal) (40 μg/ml), IPTG (0.5 mM) in the presence of antibiotics as required. Spotted LB-X-Gal-IPTG plates were further incubated at 30 °C overnight and the appearance of blue-white colour colonies were analysed. In parallel, an aliquot of the same culture was grown in LB broth overnight in 0.5 mM IPTG with appropriate antibiotics, and β-galactosidase activity was measured from liquid cultures. In brief, the cultures were further diluted 1:4 into LB broth medium and optical density at λ = 600 nm was normalized. 100 μL of cultures were mixed with 400 μL Z- buffer (60 mM Na_2HPO_4, 40 mM NaH_2PO_4, 10 mM KCl, 10 mM $MgSO_4$, 50 mM β-mercaptoethanol pH 7.0). Further, 0.01% SDS and 20 μL chloroform were added to permeabilize the cells. Cell debris were removed and enzymatic activities were measured in triplicates with 50 μL of clear supernatant using 0.4% O-nitrophenyl-*β*-D-galactoside (ONPG) as a substrate. For the first set of bacterial two-hybrid experiment Synergy H1 Hybrid Multimode Microplate Reader (Biotek, USA) was used, and for the second set of experiment BMG-Labtech Microplate reader (Germany) was used.

2.4. In vivo *protein–protein interaction statistical analysis*

The β-galactosidase activity was calculated and plotted with P value in Miller units using Graphpad Prism (www.graphpad.com/scientific-software/prism/). β-Galactosidase activity (units/mg protein) was calculated as mean ± SD ($n \geq 2$). The significance differences were analysed using Student's *t*-test, and P values, obtained at 95% confidence intervals, and are shown as (*) for <0.05, (**) for <0.01 and (***) for <0.001.

2.5. in vitro *protein–protein interactions studies*

FANCD2 (1-254), FANCD2 Ser222Ala, FANCD2 Leu231Arg and BRCA2 (2350-2545) cloned in pGEX4T-1 and pET28a+, vectors, as previously described [18,22,23]. Desired constructs were co-transformed for recombinant proteins expression with BRCA2 (2350-2545) in chemically competent *E. coli* Rosetta 2(DE3) cells. The overnight grown culture of *E. coli* Rosetta 2(DE3) harbouring different constructs were diluted 1:100 in fresh LB broth and incubated at 37 °C, using shaking incubator at 200 rpm till OD (λ = 600 nm) reaches 0.6–0.8. The culture was induced by the addition of 0.4 mM IPTG (Isopropyl β-D thio-galactoside) for 16 h at 21 °C. Cells were lysed in buffer containing 300 mM NaCl, 50 mM Tris, 0.1% triton-X, 5 mM beta-mercaptoethanol, pH 8.0, protease inhibitor cocktail (Sigma) and sonicated. Sonicated cell lysate was centrifuged at 15,000 rpm for 40 min at 4 °C. Clear cell supernatant was allowed to pass from GST resin, on Snapcap column (Thermo-Fisher). Bound protein was eluted in the buffer containing 20 mM glutathione, 100 mM NaCl, 25 nM Tris, pH 8.5. Eluted samples from Snapcap column (Thermo-Fisher) were mixed with laemelli loading buffer, and heated at 95 °C for 5 min before loading onto 12% SDS-PAGE gel, the gel was stained using coomassie brilliant blue. Protein identification using mass spectrometry was carried out as described earlier [18,22,23].

2.6. In silico *protein–protein interactions studies*

Amino acids sequences of FANCI (985-1207), BRCA2 (2350-2545) and FANCD2 (1-254) were retrieved from UniprotKB (http://www.uniprot.org/). Molecular modelling of BRCA2 (2350-2545) was performed using FALCON (http://protein.ict.ac.cn/FALCON/), however, FANCI (985-1207) & FANCD2 (1-254) were modelled as described earlier [18,22,23]. The molecular models were validated by SWISS model workspace encompassing the package of Anolea, DFire, QMEAN, Gromos, DSSP, Promotif and ProCheck (http://swissmodel.expasy.org/workspace/). Ramachandran plot ensuring that all the amino acids were present in allowed and generously allowed regions. Stereo-chemical details of the FANCI ARM repeat (Ramachandran plot: 95.1% core region, 3.4% allowed region, 1.5% generously allowed, 0.0% disallowed region), FANCD2 CUE domain (Ramachandran plot: 91.1% core region, 8.9% allowed region, 0.0% generously allowed region, 0.0% disallowed region) and BRCA2 CTR (Ramachandran plot: 81.7% core region, 16.0% allowed region, 2.3% generously allowed, 0.0% disallowed region). HADDOCK were used for molecular docking [24,25] that is based on experimental data driven Monte Carlo approach [24,25]. The best-scored dock complex from HADDOCK was selected and visualizes by PYMOL or UCSF chimera. LigPlot analysis and PDB*Sum* (http://www.ebi.ac.uk/thornton-srv/databases/pdbsum/Generate.html) tools were used for calculating the interface areas in the docked clusters.

3. Results and discussion

Studies of protein–protein interactions are the underlying mechanism through which cells maintain its structural integrity and functions. And when a cell encounters any stress, many proteins which are differentially expressed are recognized as sensors, mediators, and effectors [3,26,27]. In the context of DNA ICL repair, it is known that FANCI, FANCD2, and BRCA2 known to interact as observed in STRING analysis (Supplementary Fig. 1).

3.1. In-vivo *protein–protein interaction studies of BRCA2 C-terminal region with FANCI ARM repeat, FANCD2 CUE domain and mutants*

Bacterial two-hybrid assays were performed to evaluate the interactions between FANCI ARM Repeat (985-1207 aa), FANCD2 Cue domain (1-254 aa) and BRCA2C terminal region (2350-2545 aa). In the first set of experiment, It has been observed that FANCI ARM repeat (FI ARM) shows interactions of 100.10 ± 24.90 Miller Units with FANCD2 CUE domain (FD2 CUE) in β-Galactosidase assay ($p = 0.005$). (Fig. 1A and B). However, in the case of FI ARM and BRCA2 C-terminal region (BRCA2 CTR), the Miller Units was significantly higher (206.09 ± 9.11, $p < 0.0001$). We also observed an increment in Miller Units 1349.27 ± 289.65 between FD2 CUE and BRCA2 CTR ($p < 0.0001$) which suggests that FD2 CUE has a strong interaction with BRCA2 CTR (Fig. 1A and B). Furthermore, we have evaluated the PPIs of functionally relevant mutations Ser222Ala and Leu231Arg present in FD2 CUE domain with BRCA2 CTR. In the second set of experiment, a decrease in the interactions of mutants Ser222Ala, Leu231Arg with BRCA2 CTR compare to *wild type* (21788.06 ± 5473.40 Miller Units, $p = 0.0460$) were observed (Fig. 2A and B). However, binding interaction has been impaired in Ser222Ala mutant (10609.54 ± 1156.80 Miller Units, $p = 0.0169$) compared to Leu231Arg mutant (15785.2 ± 3028.03 Miller Units,

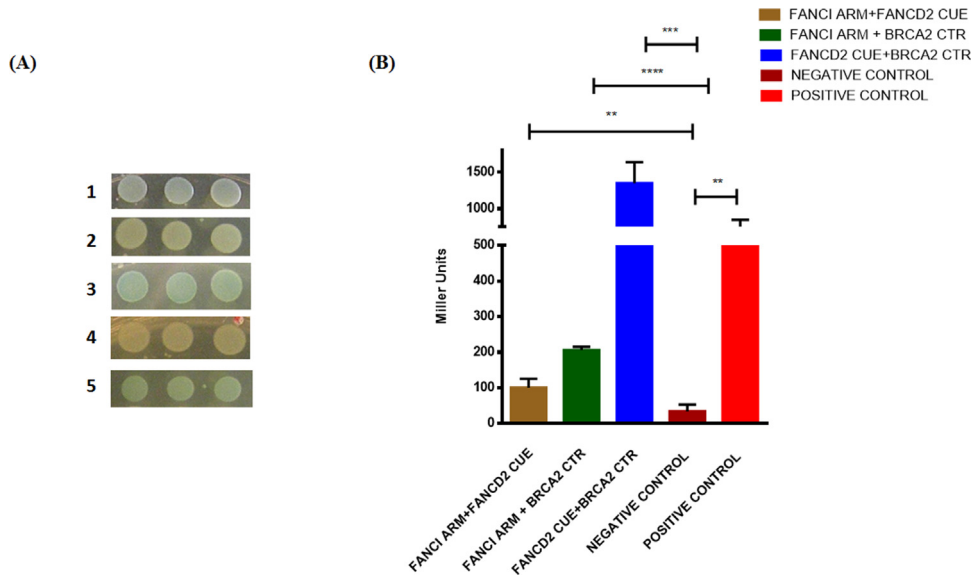


Fig. 1. *In-vivo* protein–protein interaction studies of (A and B) FANCI ARM Repeat, FANCD2 CUE Domain and BRCA2 C- Terminal region (CTR), 1 = FANCI ARM Repeat+ FANCD2 CUE Domain, 2 = FANCI ARM Repeat+ BRCA2 CTR, 3 = FANCD2 CUE Domain+ BRCA2 CTR. 4 = Cells co-expressing pUT18 and pKNT25 tag empty vectors were used as negative control, and 5 = FtsA-T18 and FtsZ-T25 protein was used as positive control, β-Galactosidase activity (units/mg protein) is shown here as mean ± SD ($n \geq 2$) and the significance differences was analysed using Student's *t*-test, and P values obtained at 95% confidence intervals are shown as (*) for <0.05, (**) for <0.01 and (***) for <0.001.

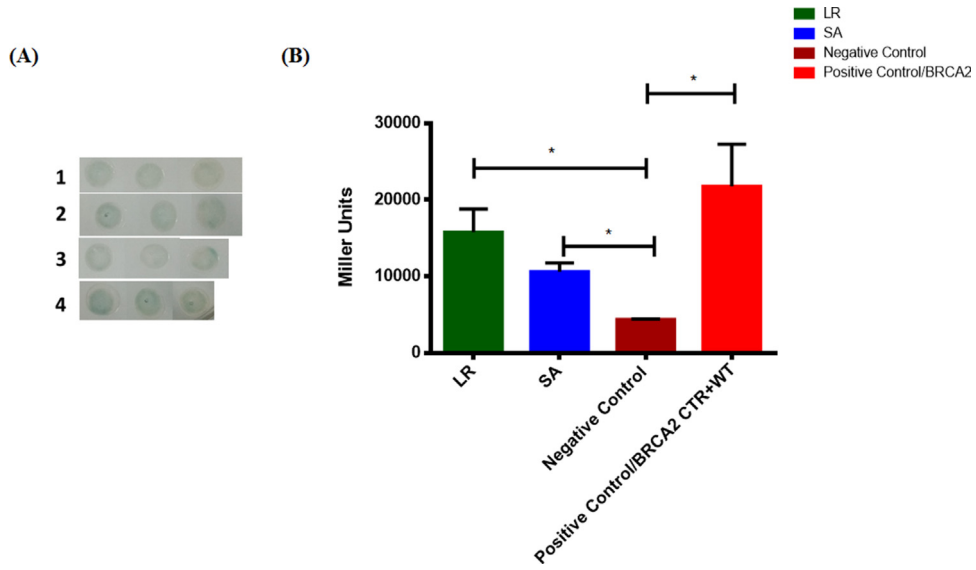


Fig. 2. *In-vivo* protein–protein interaction (A and B)studies of FANCD2 CUE Domain wild type, FANCD2 CUE Ser222Ala, and FANCD2 CUE Leu231Arg with BRCA2 C- Terminal region, 1 = FANCD2 CUE Leu231Arg + BRCA2 CTR, 2 = FANCD2 CUE Ser222Ala+ BRCA2 CTR, 3 = Negative control (pUT18 and pKNT25 tag empty vectors), and 4 = FANCD2 CUE wild type+ BRCA2 CTR (in and as positive control). Cells co-expressing pUT18 and pKNT25 tag empty vectors were used as negative control, and FANCD2 CUE Domain wild type pUT18 and BRCA2-CTR-pKNT25 protein was used as positive control, b-Galactosidase activity (units/mg protein) is shown here as mean ± SD ($n \geq 2$) and the significance differences was analysed using Student's *t*-test, and P values, obtained at 95% confidence intervals are shown as (*) for <0.05, (**) for <0.01 and (***) for <0.001.

p = 0.0335) (Fig. 2B). The data revealed that point mutations abrogate the protein–protein interaction between FD2 CUE domain and BRCA2 CTR. The first and second sets of experiments were performed independently twice for biological and technical duplicates.

3.2. In-vitro *protein–protein interaction studies of BRCA2 C-terminal region with FANCD2 CUE domain and mutants*

In DNA ICL repair, different proteins of FA-BRCA pathway are involved, however, FANCD2 and BRCA2 proteins have important and multiple functions [3]. BRCA2C terminal site also binds with FANCG/XRCC. Furthermore, BRCA2 and FANCD2, was known to co-immunoprecipitated in the cellular extracts of Chinese hamster wild type as well as in human cells [10]. Considering the importance of BRCA2 CTR and FANCD2 CUE, GST-pull down assays were performed to support the *in-vivo* protein–protein interactions between FANCD2 CUE domain and mutants with BRCA2C-terminal region. It has been observed that FANCD2 CUE domain *wild type* interacts with BRCA2C-terminal region with high affinity as prominent bands were observed (Fig. 3B). However, mutants Ser222Ala and Leu231Arg were not able to form strong interactions compared to *wild type* (Fig. 3B). Noting the similar molecular weight of FANCD2 CUE (fusion protein ~54 kDa), native protein (~28 kDa), Glutathione-S- Transferase protein (GST) was ~26 kDa and BRCA2 CTR (~25 kDa), peptide mass fingerprinting using mass spectrome-

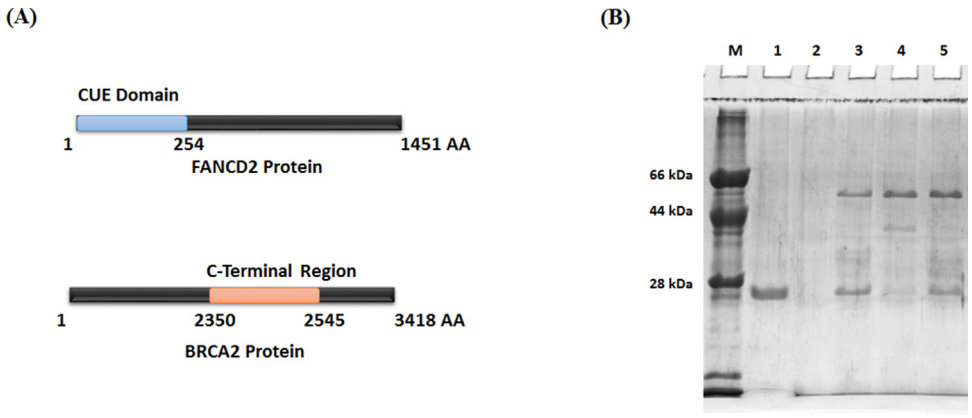


Fig. 3. (A) Schematic presentation of FANCD2 CUE domain and BRCA2 CTR (B) *In-vitro* protein–protein interaction studies of FANCD2 CUE Domain, FANCD2 CUE Mutants and BRCA2 C- Terminal region, M = Protein Marker, 1 = GST-resin+ GST protein+ BRCA2 CTR, 2 = GST-resin+ BRCA2 CTR, 3 = GST-resin+ GST-Fusion FD2 CUE wild type + BRCA2 CTR, 4 = GST-resin+ GST-Fusion FD2 CUE Ser222Ala + BRCA2 CTR, 5 = GST-resin+ GST-Fusion FD2 CUE Leu231Arg + BRCA2 CTR.

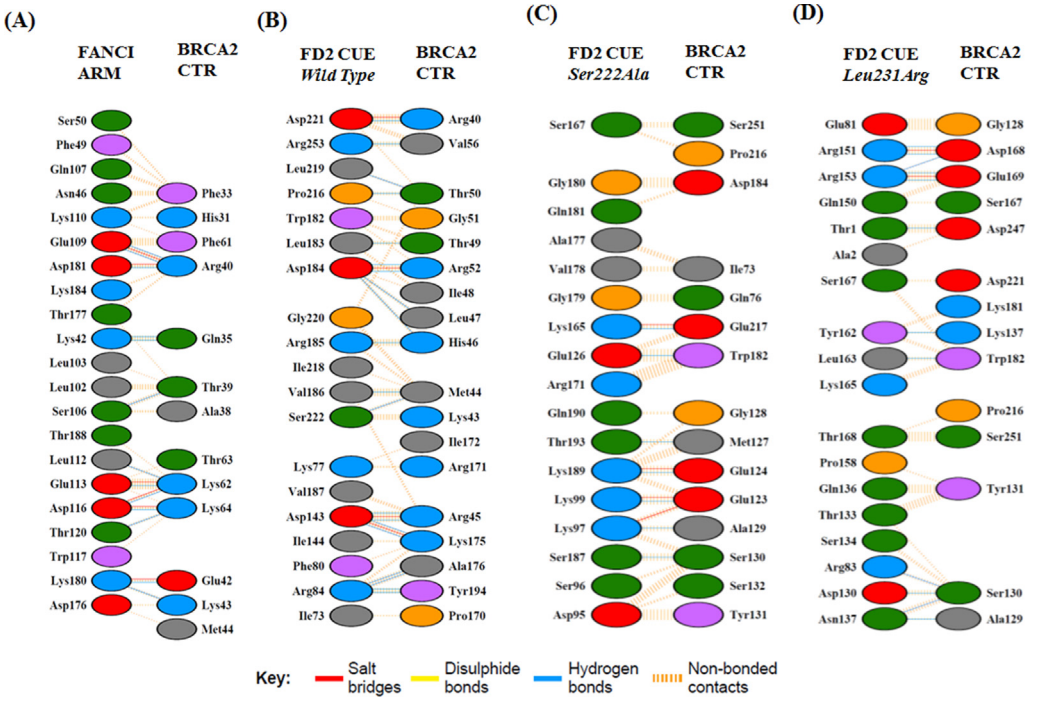


Fig. 4. *In-silico* protein–protein interaction studies of FANCI ARM Repeat, FANCD2 CUE Domain, FANCD2 CUE Mutants and BRCA2 C- Terminal region, (A) = FANCI ARM+ BRCA2 CTR, (B) = FANCD2 CUE wild type+ BRCA2 CTR, (C) = FAND2 CUE Ser222Ala + BRCA2 CTR and (D) = FANCD2 CUE Leu231Arg + BRCA2 CTR.

try was performed to confirm the protein idenitity (Supplementary Fig. 2).

3.3. In-silico *protein–protein interaction between BRCA2 C-terminal region with FANCI ARM repeat, FANCD2 CUE domain and mutants*

To further support the *in-vivo* and *in-vitro* results, *in-silico* protein–protein interactions were performed. Binding interfaces were analysed and reduction in the interacting surface area of BRCA2 with mutants (Ser222Ala and Leu231Arg) than the *wild-type* FANCD2 CUE domain were observed (Fig. 4B–D). Interestingly, from the docked complex between *wild type* FANCD2 CUE domain and BRCA2 CTR, it has been observed that Trp182 of FANCD2 CUE domain is present specifically at the interaction interface, however it is absent in both Ser222Ala and Leu231Arg mutants. In general, protein–protein interaction hotspots were known to enrich with tryptophan [28], raising the possibility that Trp182 might be playing an important role in PPIs. Furthermore, Ser222Ala mutant showed a significant drop in the number of hydrogen bonds (7 hydrogen bonds) at the protein–protein interface as compared to Leu231Arg (12 hydrogen bonds) mutant. However, comparatively, both the mutants showed significantly lower number of hydrogen bonds as compared to wild-type FANCD2 CUE domain (Table 3A). These data suggest that wild-type FANCD2 CUE domain binds more effectively to BRCA2 CTR compared to Ser222Ala and Leu231Arg the mutants. Similar findings were also observed from *in-vivo* and *in-vitro* experiments. Since FANCI protein, particularly through ARM repeat is also known to play an important role in protein–protein interactions in FA-BRCA pathway. Therefore, FANCI ARM repeat interaction was also analysed with BRCA2C-terminal region. It has been observed that FANCI ARM repeat interacts with BRCA2C-terminal region (Fig. 4A). These cumulative results suggest that FANCI ARM repeat and FANCD2 CUE domain interact with BRCA2 C-terminal region while mutants (Ser222Ala and Leu231Arg) abrogates these interactions significantly which

Table 3
Calculations of protein interfaces.

(A) Interface statistics of BRCA2 (2350–2545) with FD2 Cue domain and mutants

BRCA2 (2350–2545)	Salt bridges (In number)	Hydrogen bond (In number)	Interface area (Å^2)
FD2 Cue (wild type)	4	20	1041x1046
FD2 Cue (Ser222Ala)	4	7	879x909
FD2 Cue (Leu231Arg)	2	12	766x795

(B) Interface statistics of BRCA2 (2350–2545) with FANCI ARM Repeat

BRCA2 (2350–2545)	Salt bridges (In number)	Hydrogen bond (In number)	Interface area (Å^2)
FANCI ARM	6	13	778x891

in-turn may result in cell cycle checkpoint and DNA ICL repair failure.

4. Conclusion

The interactions between full-length FANCI and FANCD2, and FANCD2 with BRCA2 protein have been reported observed earlier by different groups [3–5,10,12], but how the protein–protein interactions associated to different domains of FA-BRCA pathway proteins have not been characterized. We have found that FANCI ARM repeat interacts with FANCD2 CUE domain and BRCA2 C-terminal region. Moreover, FANCD2 CUE domain interacts potently with BRCA2 C-terminal region. The interactions between BRCA2 and functionally relevant mutation present in FANCD2 CUE domain, Ser222Ala (cell cycle checkpoint mutant) and Leu231Arg (DNA ICL repair mutant), abrogates the binding ability. However, Ser222Ala mutant showed a significant drop in interactions compared to Leu231Arg mutant. These results suggest that, the domain and regions present in FANCI, FANCD2 and BRCA2 play an important role in PPIs, and mutations cause the impairment in the PPIs. However, further study needed to validate it in mammalian cell system, with specific post translational modifications such as phosphorylation and monoubiquitylation. To our conclusion, protein–protein interactions network between different domains of FANCI, FANCD2 and BRCA2, play an important role in functions associated to PPIs, which in turn may regulate DNA ICL damage repair mechanism.

Acknowledgements

Funding for this study was supported by annual scientific fund from TMC, ACTREC. We thank DBT-BTIS facility at ACTREC for providing necessary software to this study. We thank Dr. A Das –BARC for help in BTH experiment.

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.ijbiomac.2017.05.166.

References

[1] B.P. Alter, M.H. Greene, I. Velazquez, P.S. Rosenberg, Cancer in Fanconi anemia, Blood 101 (5) (2003) 2072.
[2] M.M. Cohen, S.J. Simpson, G.R. Honig, H.S. Maurer, J.W. Nicklas, A.O. Martin, The identification of Fanconi anemia genotypes by clastogenic stress, Am. J. Hum. Genet. 34 (5) (1982) 794–810.
[3] A.D. D'Andrea, M. Grompe, The Fanconi anaemia/BRCA pathway, Nat. Rev. Cancer 3 (1) (2003) 23–34.
[4] 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 (21) (2002) 2591–2597.
[5] T. Taniguchi, I. Garcia-Higuera, P.R. Andreassen, R.C. Gregory, M. Grompe, A.D. D'Andrea, S-phase-specific interaction of the Fanconi anemia protein, FANCD2, with BRCA1 and RAD51, Blood 100 (7) (2002) 2414–2420.
[6] A. Tutt, D. Bertwistle, J. Valentine, A. Gabriel, S. Swift, G. Ross, C. Griffin, J. Thacker, A. Ashworth, Mutation in Brca2 stimulates error-prone homology-directed repair of DNA double-strand breaks occurring between repeated sequences, EMBO J. 20 (17) (2001) 4704–4716.
[7] A. Ohashi, M.Z. Zdzienicka, J. Chen, F.J. Couch, Fanconi anemia complementation group D2 (FANCD2) functions independently of BRCA2- and RAD51-associated homologous recombination in response to DNA damage, J. Biol. Chem. 280 (15) (2005) 14877–14883.
[8] M. Castella, C. Jacquemont, E.L. Thompson, J.E. Yeo, R.S. Cheung, J.W. Huang, A. Sobeck, E.A. Hendrickson, T. Taniguchi, FANCI regulates recruitment of the FA core complex at sites of DNA damage independently of FANCD2, PLoS Genet. 11 (10) (2015) e1005563.
[9] M. Warren, A. Smith, N. Partridge, J. Masabanda, D. Griffin, A. Ashworth, Structural analysis of the chicken BRCA2 gene facilitates identification of functional domains and disease causing mutations, Hum. Mol. Genet. 11 (7) (2002) 841–851.
[10] S. Hussain, J.B. Wilson, A.L. Medhurst, J. Hejna, E. Witt, S. Ananth, A. Davies, J.Y. Masson, R. Moses, S.C. West, J.P. de Winter, A. Ashworth, N.J. Jones, C.G. Mathew, Direct interaction of FANCD2 with BRCA2 in DNA damage response pathways, Hum. Mol. Genet. 13 (12) (2004) 1241–1248.
[11] A. Smogorzewska, S. Matsuoka, P. Vinciguerra, E.R. McDonald 3rd, K.E. Hurov, J. Luo, B.A. Ballif, S.P. Gygi, K. Hofmann, A.D. D'Andrea, S.J. Elledge, Identification of the FANCI protein, a monoubiquitinated FANCD2 paralog required for DNA repair, Cell 129 (2) (2007) 289–301.
[12] M.A. Rego, F.W.T. Kolling, E.A. Vuono, M. Mauro, N.G. Howlett, Regulation of the Fanconi anemia pathway by a CUE ubiquitin-binding domain in the FANCD2 protein, Blood 120 (10) (2012) 2109–2117.
[13] K. Nakanishi, Y.G. Yang, A.J. Pierce, T. Taniguchi, M. Digweed, A.D. D'Andrea, Z.Q. Wang, M. Jasin, Human Fanconi anemia monoubiquitination pathway promotes homologous DNA repair, Proc. Natl. Acad. Sci. U. S. A. 102 (4) (2005) 1110–1115.
[14] T. Taniguchi, I. Garcia-Higuera, B. Xu, P.R. Andreassen, R.C. Gregory, S.T. Kim, W.S. Lane, M.B. Kastan, A.D. D'Andrea, Convergence of the Fanconi anemia and ataxia telangiectasia signaling pathways, Cell 109 (4) (2002) 459–472.
[15] R. Kalb, K. Neveling, H. Hoehn, H. Schneider, Y. Linka, S.D. Batish, C. Hunt, M. Berwick, E. Callen, J. Surralles, J.A. Casado, J. Bueren, A. Dasi, J. Soulier, E. Gluckman, C.M. Zwaan, R. van Spaendonk, G. Pals, J.P. de Winter, H. Joenje, M. Grompe, A.D. Auerbach, H. Hanenberg, D. Schindler, Hypomorphic mutations in the gene encoding a key Fanconi anemia protein, FANCD2, sustain a significant group of FA-D2 patients with severe phenotype, Am. J. Hum. Genet. 80 (5) (2007) 895–910.
[16] K. Modi, H.S. Misra, Dr-FtsA, an actin homologue in Deinococcus radiodurans differentially affects Dr-FtsZ and Ec-FtsZ functions in vitro, PLoS One 9 (12) (2014) e115918.
[17] Y.S. Rajpurohit, H.S. Misra, Structure-function study of deinococcal serine/threonine protein kinase implicates its kinase activity and DNA repair protein phosphorylation roles in radioresistance of Deinococcus radiodurans, Int. J. Biochem. Cell Biol. 45 (11) (2013) 2541–2552.
[18] M.Q. Siddiqui, R.K. Choudhary, P. Thapa, N. Kulkarni, Y.S. Rajpurohit, H.S. Misra, N. Gadewal, S. Kumar, S.K. Hasan, A.K. Varma, Structural and biophysical properties of h-FANCI ARM repeat protein, J. Biomol. Struct. Dyn. (2016) 1–11.
[19] A. Battesti, E. Bouveret, The bacterial two-hybrid system based on adenylate cyclase reconstitution in Escherichia coli, Methods 58 (4) (2012) 325–334.
[20] G. Karimova, J. Pidoux, A. Ullmann, D. Ladant, A bacterial two-hybrid system based on a reconstituted signal transduction pathway, Proc. Natl. Acad. Sci. U. S. A. 95 (10) (1998) 5752–5756.
[21] S. Kota, V.K. Charaka, S. Ringgaard, M.K. Waldor, H.S. Misra, PprA contributes to Deinococcus radiodurans resistance to nalidixic acid, genome maintenance after DNA damage and interacts with deinococcal topoisomerases, PLoS One 9 (1) (2014) e85288.
[22] R.K. Choudhary Vikrant, Q.M. Siddiqui, P.S. Thapa, S. Raikundalia, N. Gadewal, N.S. Kumar, M.V. Hosur, A.K. Varma, Multimodal approach to explore the pathogenicity of BARD1, ARG 658 CYS, and ILE 738 VAL mutants, J. Biomol. Struct. Dyn. 34 (7) (2016) 1533–1544.
[23] R. Vikrant, Q. Kumar, N. Siddiqui, S.K. Singh, A.K.Varma Waghmare, Mislocalization of BRCA1-complex due to ABRAXAS Arg361Gln mutation, J. Biomol. Struct. Dyn. 33 (6) (2015) 1291–1301.

[24] M. van Dijk, A.D. van Dijk, V. Hsu, R. Boelens, A.M. Bonvin, Information-driven protein-DNA docking using HADDOCK: it is a matter of flexibility, Nucleic Acids Res. 34 (11) (2006) 3317–3325.
[25] G.C. van Zundert, J.P. Rodrigues, M. Trellet, C. Schmitz, P.L. Kastritis, E. Karaca, A.S. Melquiond, M. van Dijk, S.J. de Vries, A.M. Bonvin, The HADDOCK2.2 web server: user-friendly integrative modeling of biomolecular complexes, J. Mol. Biol. 428 (4) (2016) 720–725.
[26] K.W. Caldecott, Protein-protein interactions during mammalian DNA single-strand break repair, Biochem. Soc. Trans. 31 (Pt. 1) (2003) 247–251.
[27] P. Friedhoff, P. Li, J. Gotthardt, Protein-protein interactions in DNA mismatch repair, DNA Repair 38 (2016) 50–57.
[28] A.A. Bogan, K.S. Thorn, Anatomy of hot spots in protein interfaces, J. Mol. Biol. 280 (1) (1998) 1–9.

Please cite this article in press as: M.Q. Siddiqui, et al., Int. J. Biol. Macromol. (2017), http://dx.doi.org/10.1016/j.ijbiomac.2017.05.166

SCIENTIFIC REPORTS

OPEN

Structural basis to stabilize the domain motion of BARD1-ARD BRCT by CstF50

Received: 4 October 2016
Accepted: 5 May 2017
Published online: 20 June 2017

Rajan Kumar Choudhary[1,2], Mohd Quadir Siddiqui[1,2], Pankaj S. Thapa[1,2], Nikhil Gadewal[1,2], Senthil Kumar Nachimuthu[3] & Ashok K. Varma[1,2]

BRCA1 associated ring domain protein 1(BARD1) is a tumor suppressor protein having a wide role in cellular processes like cell-cycle checkpoint, DNA damage repair and maintenance of genomic integrity. Germ-line mutation Gln 564 His discovered in linker region of BARD1 leads to loss of binding to Cleavage stimulating factor (CstF50), which in turn instigates the premature mRNA transcript formation and apoptosis. We have studied the dynamics of ARD domain present in the BARD1 *wild-type* and mutant protein in association with CstF50 using biophysical, biochemical and molecular dynamics simulations. It has been observed that the ARD domain is relatively more flexible than the BRCT domain of BARD1. Further relative orientations of both the ARD and BRCT domains varies due to the highly flexible nature of the connecting linker region present between the domains. It has been observed that mutant ARD domain is more dynamic in nature compared to *wild-type* protein. Molecular docking studies between BARD1 Gln 564 His mutant and CstF50 shows the loss of interactions. Furthermore, domain motion of ARD present in BARD1 was stabilized when complexed with CstF50.

BRCA1 associated ring domain 1 (BARD1) is a binding partner of breast cancer associated protein (BRCA1)[1,2]. BARD1 comprises different domains including N-terminus ring domain, an ankyrin repeat domain (ARD) and two-tandem BRCT domains[3]. BARD1-BRCA1, RING-RING domain complex is an E3 ubiquitin ligase complex which is important for many cellular processes like DNA damage repair, cell cycle checkpoint and genomic integrity[1,4–8]. Mutations associated to breast and ovarian tumors present in the ring domain of BRCA1 impairs the protein-protein interactions and complex formation with BARD1[6,9,10]. ARD and BRCT domains are connected via 14 residue linker region and provide binding site for CstF50[11,12]. The BARD1 ARD-BRCT domain is essentially required for homology-directed DNA (HDR) repair, as cells transfected with truncated BARD1 are unable to carry out HDR[13]. BRCA1–BARD1 complex with Cleavage stimulating factor (CstF50) has pivotal role in transcription-coupled DNA-damage repair (TCR)[11,12]. CstF50 has 7 Tryptophan-Aspartate (WD-40) repeats that interacts to RNA polymerase II (RNAP II) through its N-terminal region[14,15] whereas BARD1 interacts through its 7th WD-40 repeat[16,17]. BARD1/BRCA1-CstF50 complex, after UV exposure leads to RNAP II CTD degradation[17–19] and inhibition of 3'end- mRNA processing[16,17,20]. Missense mutations in the BARD1 have been reported in sporadic cases of ovarian, uterine, and breast carcinoma as well as in patients with familial breast or ovarian cancer[21–25]. Germ line mutation Gln 564 His reported in BARD1 disrupts the interactions with CstF50 and p53. BARD1 Gln 564 His mutation affects BARD1 mediated p53-dependent apoptosis and 3'- mRNA processing[12,21,26].

Multidisciplinary approaches were explored to study the effect of Gln 564 His mutation on the stability of BARD1 ARD-BRCT domain and interactions with CstF50. Furthermore, protein-protein docking and molecular dynamics simulation approach were employed to unravel the effect of mutation on domain dynamics of BARD1 ARD-BRCT, Gln 564 His mutant and BARD1 ARD-BRCT-CstF50 complex.

Results and Discussion

Oligomeric behavior of BARD1 ARD-BRCT. To delineate the effect of mutation on the oligomeric behavior of BARD1 ARD-BRCT, the purified BARD1 ARD-BRCT *wild-type* and mutant proteins (Figure 1A,B

[1]Advanced Centre for Treatment, Research and Education in Cancer, Kharghar, Navi Mumbai, Maharashtra, 410 210, India. [2]Homi Bhabha National Institute, Training School Complex, Anushaktinagar, Mumbai, 400 094, India. [3]Department of Biotechnology, Mizoram University (A Central University) Aizawl, 796 004, Mizoram, India. Correspondence and requests for materials should be addressed to A.K.V. (email: avarma@actrec.gov.in)

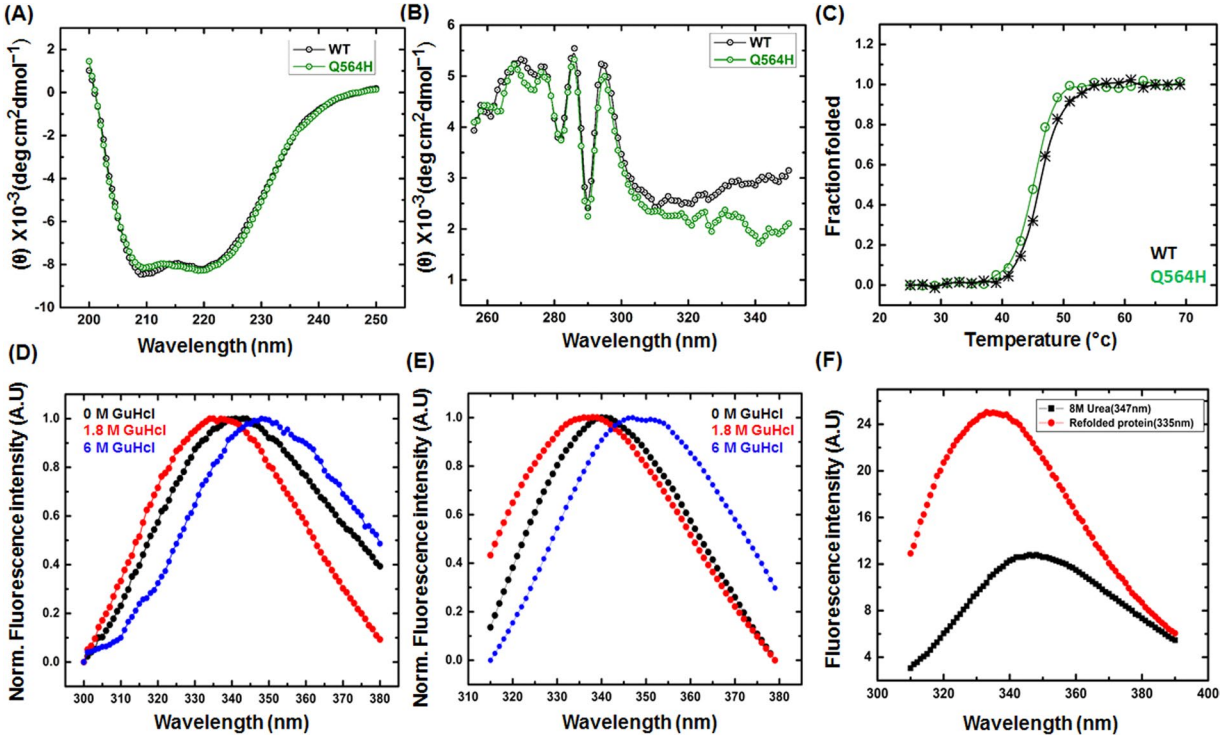


Figure 1. (**A** and **B**) Comparative Far -UV and Near -UV, CD spectroscopy profile of BARD1 ARD-BRCT *wild-type* and Gln 564 His mutant respectively, (**C**) Comparative thermal denaturation profile of *wild-type* and Gln 564 His mutant protein, (**D** and **E**) Comparative chemical denaturation profile of profile of *wild-type* and Gln 564 His mutant protein respectively and (**F**) Refolding profile of CstF50.

Supplementary) were subjected to size exclusion chromatography and dynamic light scattering (DLS). The hydrodynamic radii of BARD1 ARD-BRCT *wild-type* and mutant protein from DLS experiment were 3.3 ± 0.45 nm and 3.38 ± 0.15 nm respectively which suggest that BARD1 Gln 564 His mutation does not affect the overall packing within the hydrophobic core of the protein (Figure 1E Supplementary). Both BARD1 ARD-BRCT *wild-type* and mutant proteins were eluted in the monomeric elution volume, suggesting their monomeric nature (Figure 2 Supplementary). Glutaraldehyde cross-linking assay for *wild-type* and mutant proteins do not show higher molecular weight conjugates on the 12% SDS-PAGE gel which further support the monomeric nature of proteins (Figure 3A and B Supplementary).

ARD and BRCT domains of BARD1 are compact and folds independently. CD Spectroscopy (CD) was performed to explore the effect of mutation on the secondary structure of BARD1 ARD-BRCT. The CD spectra for *wild-type* and mutant protein shows that secondary structure of the protein is majorly stabilized by α-helices because of prominent ellipticity minima at $\lambda = 208$ nm and 222 nm (Fig. 1A). However, the CD spectra do not show change in the secondary structure of mutant protein which indicates that mutation has no effect on the secondary structure of the *wild-type* BARD1 ARD-BRCT protein. However, the absence of characteristic random coils and β-sheets are due to the dominant contribution of CD signal from the α-helix portion of the structure. Additionally, this masking of random coil and β-sheets can also be due to the disulfide chromophores present in ARD-BRCT region which strongly contributes to the CD signal in the range of $\lambda = 215$–235 nm[27,28]. CD spectroscopy was further performed at near-UV range to investigate the effect of mutation on the tertiary structure of the protein. Near-UV spectra of *wild-type* and mutant proteins show peaks at $\lambda_{max} = 295$ nm and 285 nm respectively, which indicates the hydrophobic microenvironments of intrinsic fluorophores substantiating the compact 3D structure of the proteins[29](Fig. 1B). Furthermore, limited proteolysis profile of *wild-type* and mutant protein indicates that ARD and BRCT domain of BARD1 are two compact and independently folding domains (Figure 3C and D Supplementary).

Thermal stability of BARD1 ARD-BRCT. Thermal denaturation was performed to delineate the effect of mutation on the thermal stability of the BARD1 ARD-BRCT *wild-type*. Thermal induced denaturation was monitored from 25 °C–65 °C using CD spectroscopy. The loss in secondary structure was observed at 65 °C for both the *wild-type* and mutant proteins (Figure 4A and B Supplementary). The unfolded fraction was calculated by using change in millidegree ellipticity at $\lambda = 222$ nm, which suggest that *wild-type* and mutant protein unfold via a two state pathway in a co-operative manner. The T_m calculated for *wild-type* and mutant were 45.6 ± 0.60 °C and 45.2 ± 0.42 °C respectively, which further suggest that mutation does not drastically affect the thermal stability of the BARD1 ARD-BRCT *wild-type* (Fig. 1C). Chemical denaturation induced by GuHcl was monitored using fluorescence spectroscopy. Fluorescence emission maximum at $\lambda_{max} = 340$ nm for *wild-type* and at $\lambda_{max} = 341$ nm

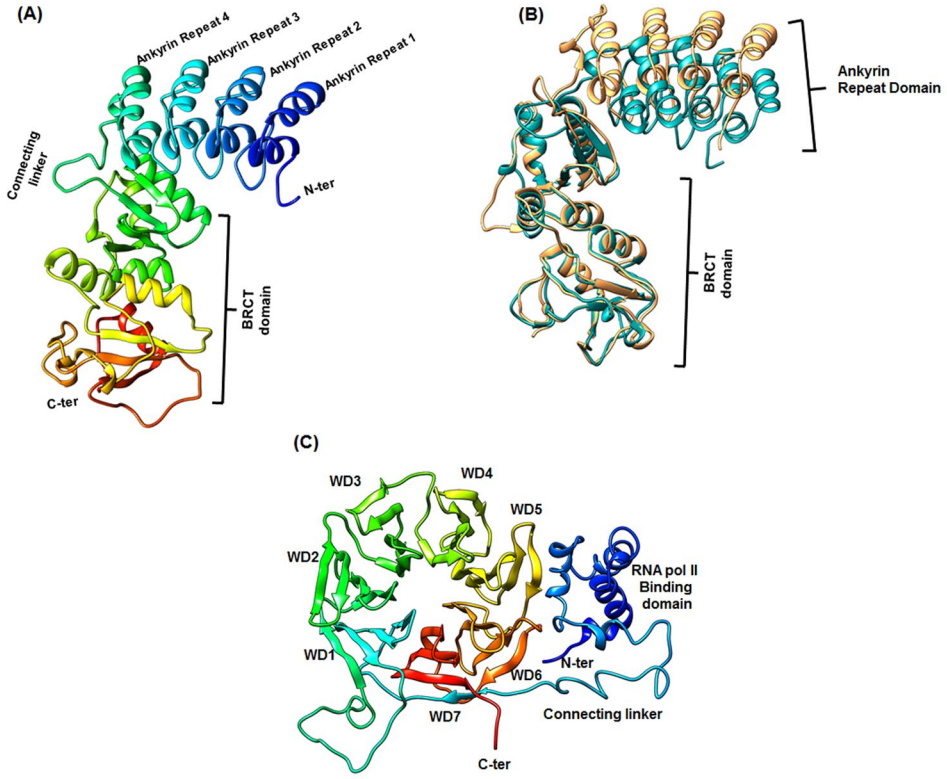


Figure 2. (**A**) Representative molecular model of BARD1 ARD-BRCT *wild-type*, (**B**) Comparative structural alignment profile of BARD1 ARD-BARD1BRCT wild-type (Navy Blue) and Gln 564 His Mutant (Yellow) and (**C**) Molecular Model structure of CstF50.

for mutant indicate that proteins are in folded conformation. However, at 1.8 M GuHcl concentration significant blue-shifts in the emission maxima was observed for *wild-type* at $\lambda_{max} = 335$ nm and mutant at $\lambda = 336$ nm respectively. *Wild-type* and mutant completely unfold at 6 M GuHcl as indicated by the emission maximum at $\lambda_{max} = 348$ nm. These observations suggest that *wild-type* and mutant unfold via an intermediate molten globule formation (Fig. 1D and E), and (Figure 4C and D Supplementary).

Furthermore, to investigate the molecular interactions between CstF50 and BARD1, we have expressed His-tagged CstF50 in bacterial system that expressed mostly in inclusion bodies (Figure 1C and D Supplementary). Hence 6XHis tagged fused CstF50 protein was extracted using 8 M urea, and then further refolded in the buffer D (50 mM Tris, pH 6.2, 500 mM NaCl, 0.1% triton, 10% glycerol pH 7.2, 2 mM EDTA, 700 mM arginine, 100 mM KCl, 50 mM $MgCl_2$, 100 mM $CaCl_2$ and 5 mM DTT) containing Arginine. Refolding pattern of CstF50 protein was further confirmed by fluorescence spectroscopy. The fluorescence spectroscopy profile for denatured CstF50 in 8 M urea shows an emission maximum at $\lambda_{max} = 347$ nm, however, folded protein shows a blue-shifted emission maximum at $\lambda_{max} = 337$ nm, which indicates complete burial of tryptophans in the hydrophobic core of the protein (Fig. 1F). Furthermore, good quality CD spectra could not be achieved because arginine hydrochloride present in buffer D interfered with the measurement[30,31]. An attempt to dialyze out arginine hydrochloride resulted in precipitation of CstF50 protein.

CstF50 stabilizes the ARD domain of BARD1: *In-silico* approach. To understand the function associated to protein-protein interactions for BARD1-CstF50 complex, BARD1 ARD-BRCT region and CstF50 were modeled using Robetta server[32,33]. This Model of BARD1 ARD-BRCT and CstF50 were further validated by MolProbity server[34] (Table 1 and 2 Supplementary). The 3-D structure of BARD1 ARD-BRCT which comprises 4, N-terminal ankyrin repeats containing α-helices, a linker, two BRCT repeats containing 3-α helices and three β- strands (Fig. 2A). BARD1 ARD-BRCT Gln 564 His mutation was generated using SPDB viewer[35]. The BARD1 ARD-BRCT and Gln 564 His mutant structures were superimposed for direct comparison between *wild-type* and Mutant models (Fig. 2B). Modeled structure of CstF50 shows N-terminal helical region and C-terminal 7 WD-40 repeat comprised of beta sheets that fold into a propeller like structure (Fig. 2C).

Normal Mode Analysis (NMA) was performed on the BARD1 ARD-BRCT *wild-type* and Gln 564 His mutant structures to investigate the effect of mutation on protein dynamics. First few modes known as a low frequency modes describe the large domain motion within the protein[36]. Hence, to understand the structural dynamics between ARD and BRCT domain, we have selected the 1st lowest frequency mode for BARD1 ARD-BRCT wild-*type* and Gln 564 His mutant protein structures (Fig. 3A–D). A large inter-domain motion was observed between the ARD and BRCT domain of BARD1. The inter-domain motion is often observed in multidomain proteins having hinge bending[37]. The residues at ARD and BRCT domain show high positive correlated motion at

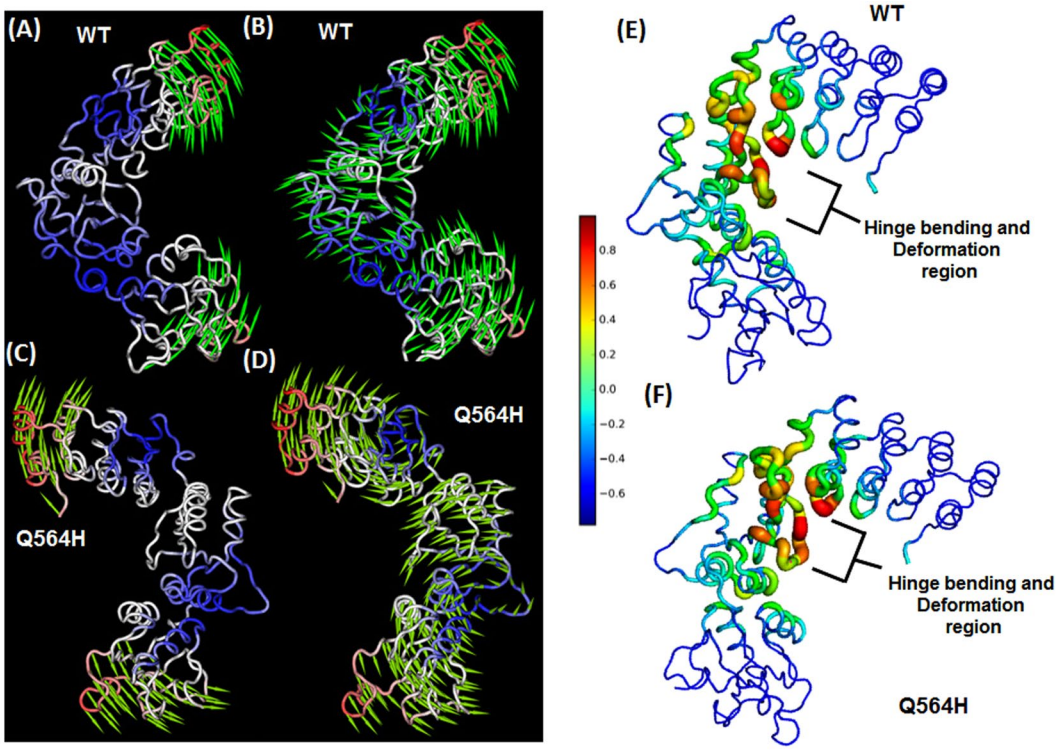


Figure 3. (**A** and **B**) Comparative Normal Mode Analysis and representation of mode 1 profile of BARD1 ARD-BRCT *wild-type*, (**C** and **D**) Gln 564 His mutant protein. (**E** and **F**) Comparative deformation analysis profile of BARD1 ARD-BRCT *wild-type* and Gln 564 His mutant protein.

the terminal regions. Interestingly, the linker region residues show high anti-correlated motion for the ARD and BRCT domain residues in *wild-type* and mutant protein structures. The region of hinge bending and local flexibility were investigated using deformation analysis (Fig. 3E and F). The linker region was found to possess highest flexibility in both *wild-type* and mutant (thickness and color intensity) which shows that the linker between inter-domain region is responsible for large dynamics between ARD domain and BRCT domains.

BARD1- ARD domain motion and CstF50 complex. To understand the function associated to protein-protein interactions of BARD1-CstF50 complex, and inter-domain motion between ARD and BRCT domains, the model structure of BARD1 ARD-BRCT, BARD1 Gln 564 His mutant, CstF50 and BARD1 ARD-BRCT-CstF50 complex were simulated for 100 ns under periodic boundary conditions. The RMSD (Root Mean Square Deviation) profile for BARD1 shows that mutant protein structure exhibits higher RMSD than the *wild-type* (Fig. 4A). During the first 5 ns of simulation, RMSD values of *wild-type* and mutant have increased up to a value of 5 Å. However, after 10 ns, RMSD of *wild-type* decreases to 3.5 Å and mutant increases up to 6.5 Å. At 16 ns, RMSD value for *wild-type* show sharp increase and reaches up to 8 Å. Hence, the mutant structure RMSD shows only a gradual rise to reach the level of *wild-type* RMSD (Fig. 4A). Beyond 16 ns, the decrease in the RMSD up to 4 Å for *wild-type* and mutant structure is indicative of structural stabilization. However, the RMSD of mutant protein structure remained high as compared to *wild-type*. After 35 ns of simulation, *wild-type* and mutant structures are stabilized for next 50 ns. Differences in RMSD were again observed at 98 ns and 96 ns of simulation for mutant and *wild-type* structures respectively. To investigate change in the RMSD, 0 ns input structures were superimposed onto the structure of 16 ns and 98 ns time frame for *wild-type* and mutant respectively. The two superimposed structures (Fig. 5A and B) show large ARD domain movement with respect to BRCT domain for *wild-type* and mutant protein structures. Superimposed structure of *wild-type* with input structure shows that ARD domain rotates about 70° relative to BRCT domain contributing to the RMSD value of 8.55 Å. However, the mutant shows that rotation of ARD domain by 84.4° leading to a RMSD of 9.27 Å with respect to the structure of 0 ns (Fig. 5A and B). The RMSD attained during the entire simulation by ARD domain of *wild-type* mutant shows that ARD domain is dynamic in nature comparative to the BRCT domain, which is stable during entire simulation (Fig. 4C). The RMSD change for ARD domain in mutant protein shows gradual increase but *wild-type* ARD domain is stable. Furthermore, 10 superimposed structures extracted for *wild-type* and mutant protein at different time points of trajectory show highly dynamic behavior of ARD domain, whereas the BRCT domain is stable in nature. The ARD domain in mutant shows conformational changes compared to *wild-type*, signifying that ARD domain in mutant protein is more dynamic than in *wild-type* (Fig. 4F and G). However, RMSD profile for CstF50 structure indicates that CstF50 is compact protein and does not show any changes in the model structure (Fig. 4B).

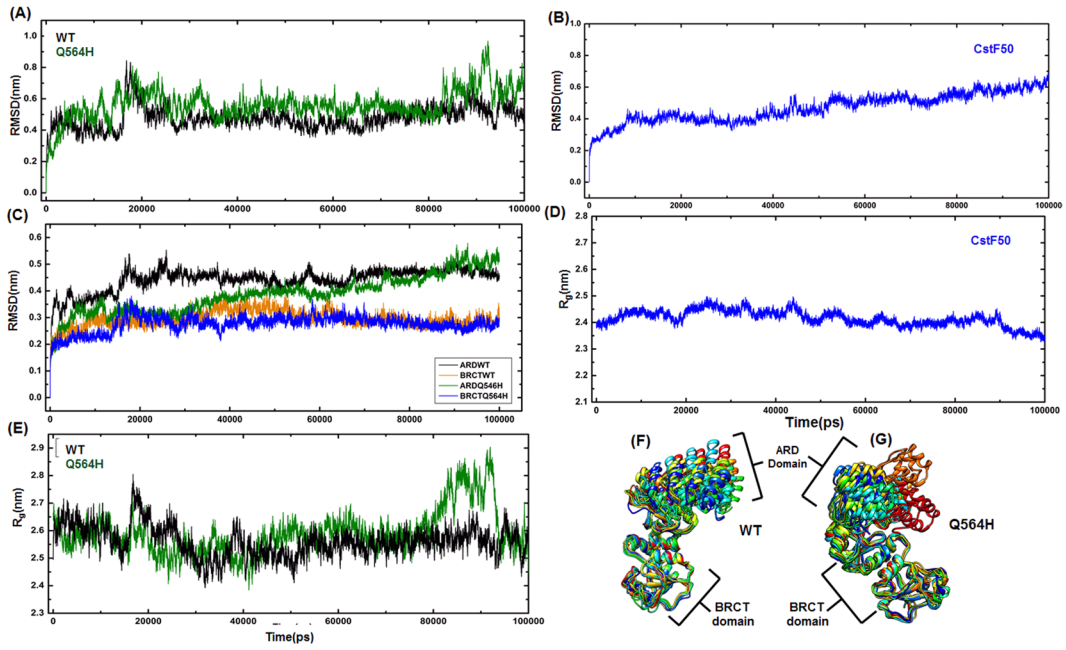


Figure 4. (**A** and **C**) Comparative RMSD and R_g profile of *wild-type* and Gln 564 His mutant protein respectively (**B** and **D**) RMSD and R_g profile of CstF50 respectively. (**E**) Individual RMSD profile of *wild-type* and Gln 564 His mutant protein ARD and BRCT domain respectively. (**F** and **G**) Comparative superimposition of 10 frames extracted from trajectory of *wild-type* and Gln 564 His mutant protein respectively.

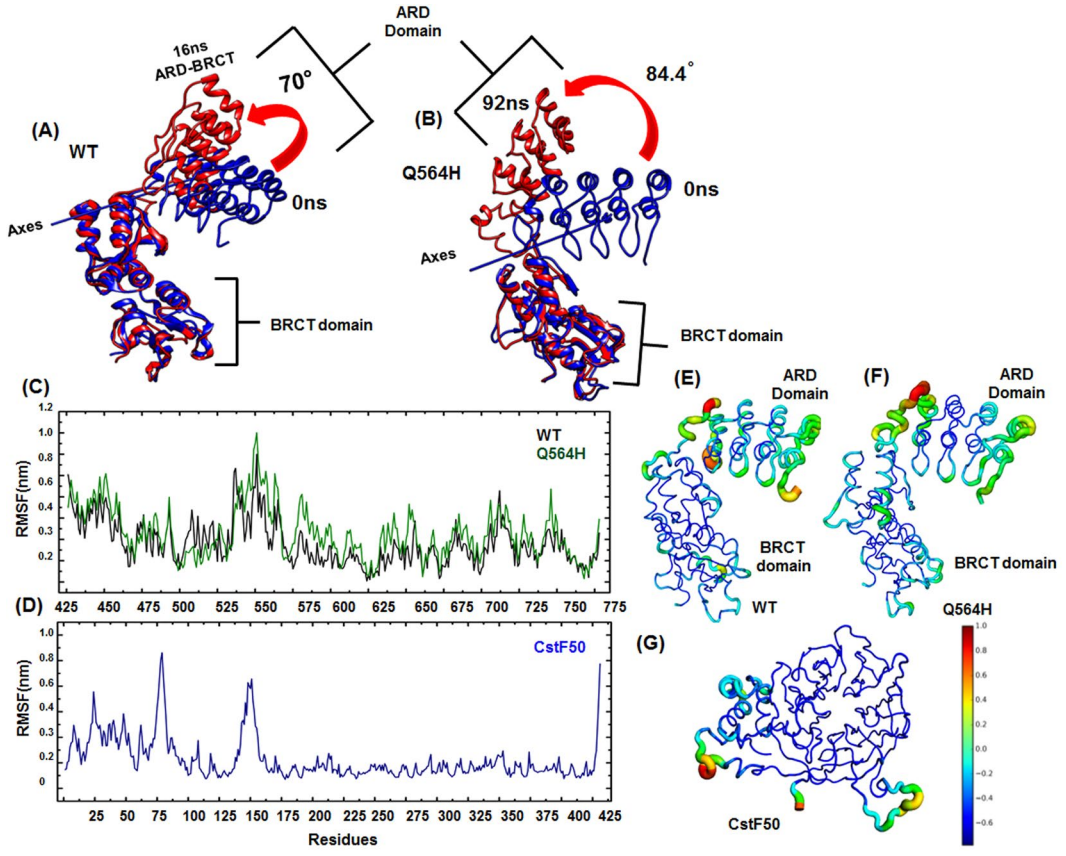


Figure 5. (**A** and **B**) Comparative superimposition of lowest RMSD and highest RMSD attained structure for *wild-type* and Gln 564 His mutant protein respectively (**C**) RMSF of *wild-type* and Gln 564 His mutant protein respectively. (**D**) RMSF profile of CstF50. (**E**,**F** and **G**) RMSF structure extracted from trajectory of *wild-type*, Gln 564 His mutant and CstF50 respectively.

Furthermore, changes in the radius of gyration (R_{gyr}) in a time dependent manner during the course of simulation determines the protein structure compactness[38]. During the early phase of simulation *wild-type* and mutant did not show any sharp changes in the R_{gyr}, but steep rise in the R_{gyr} was observed at 16 ns in *wild-type* and at 98 ns in mutant structure of BARD1. In mutant structure, R_{gyr} change is significantly higher than that of *wild-type*, substantiating the inference that mutation has made the mutant protein dynamic than the *wild-type* structure (Fig. 4E).Change in R_{gyr} is largely due to change in the position of ARD domain relative to the BRCT domain of BARD1. However, R_{gyr} profile for CstF50 structure indicates that CstF50 has overall compact structure and does not show any changes in the conformation (Fig. 4D).

Comparative residual RMSF (Root Mean Square Fluctuation) values attained by *wild-type* and mutant suggest that the N-terminal ARD domain of BARD1 is highly flexible. The linker region (550–568)amino acids in BARD1 ARD-BRCT shows highest RMSF values in *wild-type* and Gln 564 His mutant structures (Fig. 5C). Hence, to explore the residual flexibility of BARD1 ARD-BRCT, RMSF structures showing maximum fluctuation from the trajectory were extracted (Fig. 5D and E). The RMSF structures for *wild-type* and mutant indicate that ARD domain is highly flexible, and high value of RMSF is observed for the residues in N-terminal, linker and loop regions of second BRCT repeat (Fig. 5C,E and F). However, in the CstF50 structure, RMSF indicated two flexible regions, first at the connecting loops between N-terminal and WD-40 domain, and second at the loop region between WD1 and WD2 repeat which also corroborates with the residual RMSF plotted for the CstF50 (Fig. 5D and G).

Furthermore, domain movement of the protein structures have been studied in relation to possible involvement with protein function[39]. To probe the collective motion during the simulation, the principal component analysis (PCA) was performed over the BARD1-ARD-BRCT, Gln 564 His mutant, CstF50 and BARD1-ARD-BRCT-CstF50 complex trajectories. The Eigenvalues and eigenvectors were calculated after the diagonalization of the covariance matrix. The area of trace of covariance matrix calculated for BARD1-ARD-BRCT, Gln 564 His mutant and CstF50 were 487.990 nm^2, 714.963 nm^2 and 540.810 nm^2 respectively, which indicates that mutant shows more dynamic tertiary structural conformations than the *wild-type* protein. Hence, the trajectories for domain motion of BARD1 ARD-BRCT *wild-type*, Gln 564 His mutant and CstF50 were projected on the first three eigenvectors. Projection on eigenvector 2 and 1 for *wild-type* and mutant protein indicates that mutant protein is more dynamic in nature than the *wild-type*. Trajectory Projection on eigenvectors combination 3 and 1, 2 and 1 and 3 and 2 for *wild-type* (Fig. 6A,B and C) and mutant (Figure 6D,E and F) indicates that both the structures can attain three major tertiary structure conformations whereas mutant structure shows higher periodic jump. Further, projection of eigenvectors 3 and 1, 2 and 1 for CstF50 shows that two major groups of tertiary structural conformation can be attained by the CstF50. The large transitions and periodic fluctuation between these conformers throughout the trajectory have been observed (Fig. 6G,H and I).

Comparative residual displacement along the Eigenvectors. To investigate the residual fluctuation, first two eigenvector extracted from trajectories of BARD1 ARD-BRCT *wild-type*, Gln 564 His mutant and CstF50 were projected on the residues. The plot of projection on eigenvector 1 and 2 of residues for *wild-type* (Figure 5A,B Supplementary) and mutant (Figure 5C,D Supplementary) demonstrates concerted residual fluctuation, which shows high flexibility in ARD domain, linker region and loop regions of BARD1 BRCT domain. Furthermore, eigenvector 1 and 2 were also projected on residue for CstF50 (Figure 5E,F Supplementary). The projection of eigenvector 1 and 2 show different profile of residual fluctuation and do not shows any similarity. The projection of eigenvector 1 on residue shows flexible loop regions connecting N-ter, WD-40 domain. On the other hand projection of eigenvector 2 on residues shows stable WD-40 domain but fluctuations in the N-terminal domain that may be due to the connecting loop regions and untethered N-terminal (Figure 5E,F Supplementary).

Positive correlation motion of BARD1 ARD domain. Cross correlation for PCA of BARD1 *wild-type*, Gln 564 His mutant protein, CstF50 and complex was plotted to investigate the correlated motion within the proteins. Cross-correlation of PCA indicates high positive correlation within the ARD domain. The residues within the ARD domain show strong negative correlation with the linker region which is in consistent with the NMA results (Fig. 7A,B). Comparatively, BARD1 *wild-type* shows higher positive correlation motion within the ARD domain and linker region residues as compared to the mutant protein. Furthermore, mutation has also affected the positive correlation within the BRCT repeat as *wild-type* shows higher positive correlation within the N-terminal BRCT and C-terminal repeat residues. CstF50 cross-correlation for PCA indicates that the N-terminal residues show little positive correlation within themselves as compared to the WD-40 repeat domain, which shows very high positive correlated motion within their domain (Fig. 7C). However, BARD1 ARD-BRCT-CstF50 complex cross-correlation for PCA indicates increased positive correlated motion within the N-terminal residues and WD-40 repeat for CstF50 in complex as compared to native condition. Furthermore, higher positive correlation is also observed in the ARD domain as well as BRCT domain in complex compared to the BARD1 ARD-BRCT alone (Fig. 7D).

Protein-Protein Interactions between BARD1 ARD-BRCT and CstF50. The minimum free energy model structures of BARD1 ARD-BRCT, Gln 564 His mutant and CstF50 were selected for protein-protein docking studies (Figure 6A,B and C Supplementary). It is well established that 7th WD 40 (395–431) repeat domain of CstF50 is required to establish BARD1 complex[26]. The Gln at 564 to His mutation reduces the binding affinity of CstF50 to BARD1[17]. Therefore, we considered BARD1 Gln 564 as an important residue for interaction study with CstF50. In-silico docking was performed to understand the residual interactions between BARD1 ARD-BRCT, Gln 564 His mutant with CstF50. In the intermolecular interface of BARD1-CstF50 complex, glutamine 564 of BARD1 ARD-BRCT forms hydrogen bond with Asn 409 and Tyr 425 of CstF50, and also participates in

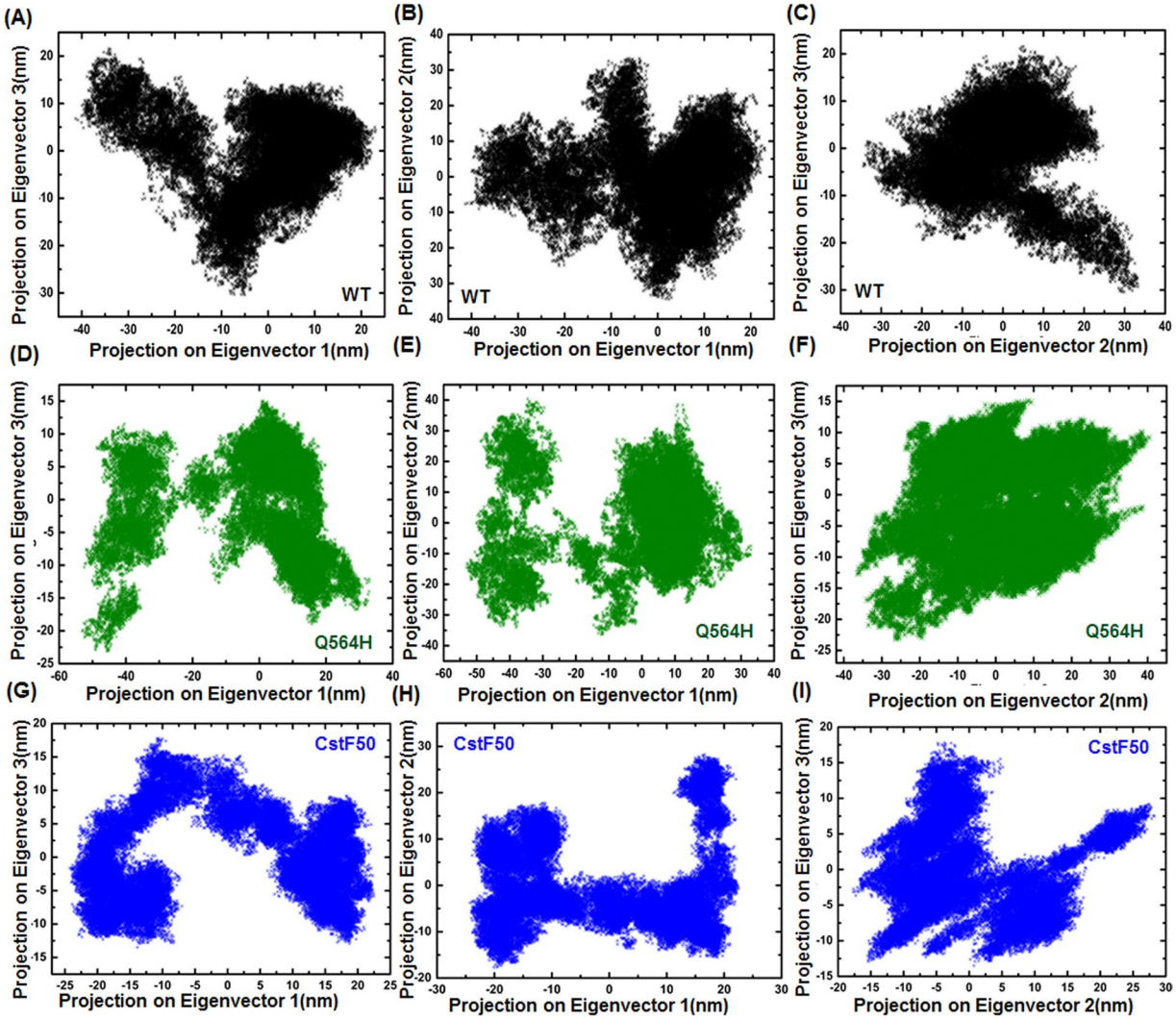


Figure 6. (**A**,**B** and **C**) Comparative Eigenvector projection profile of BARD1 ARD-BRCT *wild-type* (**D**,**E** and **F**) Gln 564 His mutant protein respectively (**G**,**H** and **I**) CstF50 respectively.

non-hydrogen bonding interactions with Tyr 425, Asn 409, Pro 410, and Thr 408 of CstF50. The other BARD1 residues Thr 562 and Gly 563 also form hydrogen bonds with Thr 430, Tyr 425 and Ser 428 of CstF50 respectively (Figure 7A and B Supplementary). However, in the BARD1 Gln 564 His mutant-CstF50 complex, BARD1 His 564 lost hydrogen bonding interactions but forms non-covalent interactions with Gly 394 and Leu 395 of the CstF50 (Figure 7C and D Supplementary). The BARD1 Thr 562 was found to be involved in the non-covalent interactions with CstF50 Glu 100. Loss of weak intermolecular interactions in case of mutant rationalizes the loss of binding affinity between mutant BARD1 and CstF50 (Figure 7D Supplementary).

Binding Interface of BARD1 ARD-BRCT and CstF50 complex. To determine the stability of BARD1 ARD-BRCT-CstF50 complex molecule we have simulated the structure for 100 ns to analyze the comparative RMSD, RMSF, R_{gyr} and hydrogen bonding (Fig. 8). Comparative RMSD and R_{gyr} indicate that binding of CstF50 to the BARD1 ARD-BRCT restricts the ARD domain fluctuation relative to the BRCT domain (Fig. 8A and C). Hydrogen bonding analysis between the linker region and 7th WD-40 domain was performed to analyze the domain stability of complex (Fig. 8B). It was found that BARD1 ARD-BRCT *wild-type* linker and 7th WD-40 repeat form stable hydrogen bonds throughout the simulation, thus enabling formation of a stable complex (Fig. 8D). Further, binding energy and other energy component calculated for BARD1-CstF50 complex for entire simulation indicated that interaction between BARD1 and CstF50 is stable (Table 4 Supplementary). Interestingly, interactions between CstF50 and the BARD1 linker rigidifies the linker flexibility and thereby controls the architectural placement of the BARD1 C-terminal domain relative to the N-terminal ARD domain (Fig. 8E). This phenomenon thereby reduce the relative inter domain flexibility of the ankyrin repeats and BRCT domains. The short length of the linker and germ-line cancer-predisposing mutations suggest that the relative positioning of ARD and BRCT domains within BARD1 is critical in the binding to proteins and functionality of BARD1 (Fig. 8E and F).

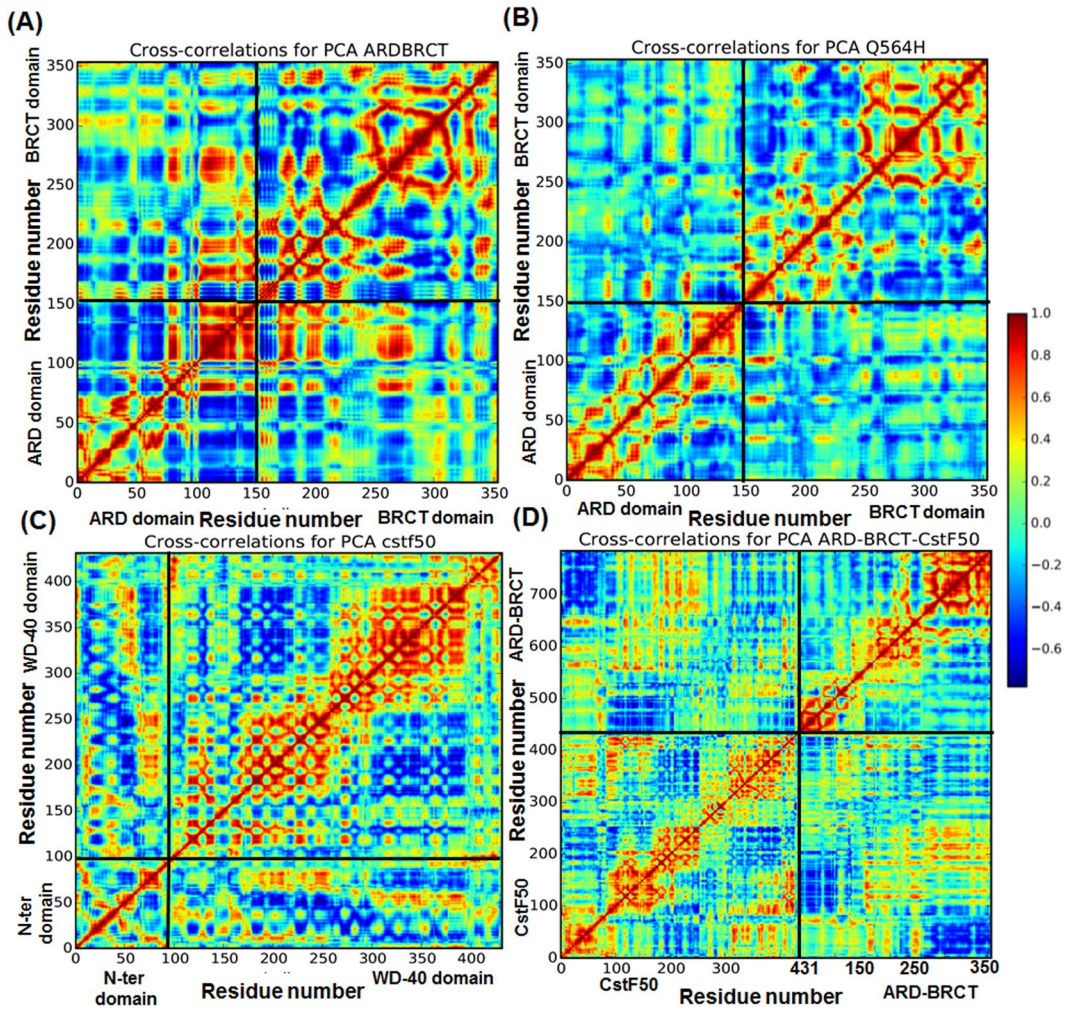


Figure 7. Comparative residual cross-correlation for PCA of (**A**) BARD1 ARD-BRCT *wild-type*, (**B**) Gln 564 His mutant, **(C)** CstF50 respectively and (**D**) BARD1 ARD-BRCT *wild-type* -CstF50 complex respectively. Red and yellow regions show highly positively correlated regions while dark blue represent anti-correlated movement of residues.

Conclusion

Multidisciplinary approaches were applied to characterize the domain motion and binding affinity between BARD1 ARD-BRCT, Gln 564 His mutant and CstF50 proteins. It has been observed that the BARD1 ARD-BRCT Gln 564 His mutant is not affecting the monomeric property, secondary structure, hydrodynamic radii and thermal stability of the *wild-type* protein. The limited proteolysis of *wild-type* and mutant confirms that ARD and BRCT domain are stable and independent folded domains of BARD1 connected by a short stretch of a linker region. These results from limited proteolysis are consistent with earlier NMR studies which conclude that in the absence of any interacting protein, the ARD and tandem BRCT region behave as an independent domain in solution[40,41]. The thermal unfolding study indicates that BARD1 ARD-BRCT *wild-type* and mutant unfold via a two state pathway. However, chemical unfolding of *wild-type* and mutant suggest that both proteins unfold via a molten globule intermediate.

The results from NMA and MD simulation substantiates well with the earlier reported short angle X-ray scattering (SAXS) findings. In solution ARD and BRCT domains sample a wide range of relative orientations with respect to one another[42]. The Small Angle X-ray Scattering (SAXS) for BARD1 ARD-BRCT domain shows two additional features in the p(r) function: a shoulder at r ~40 Å and tailing-off of the at r > 90 Å[42]. SAXS for ARD-BARD1BRCT suggests that Dmax of the domain is around 90 Å[42]. The RMSD values calculated for *wild-type* and mutant are very close (87 Å) to the experimental values (r > 90 Å)[42]. The distance calculated for *wild-type* N-terminal ARD domain and C-terminal BRCT domain shows a maximum change of 40 Å, which shows dynamic behavior of the protein due to the flexible linker region. The NMA and deformation analysis indicate that the linker connecting ARD and BRCT domains makes the *wild-type* and mutant protein flexible. MD simulation studies show that ARD domain is highly dynamic than the BRCT domain due to flexible linker which acts as the connecting bridge between the ARD and BRCT domain of BARD1. Flexibility in linker segments of the BARD1 ARD-BRCT backbone allow corresponding domain motions to occur with minor structural

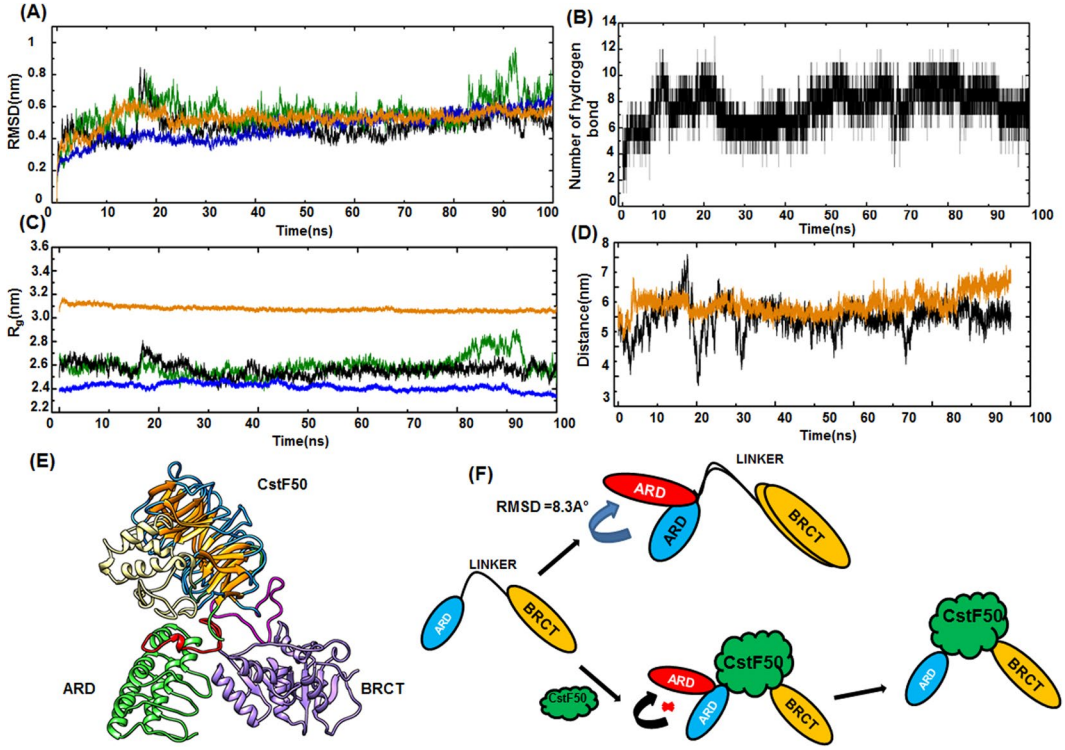


Figure 8. (**A** and **C**) RMSD and R_g profile of BARD1 ARD-BRCT *wild-type*, Gln 564 His mutant, CstF50 and *wild-type*-CstF50 complex respectively (**B**) H bonding profile of BARD1 ARD-BRCT *wild-type*-CstF50 complex (**D**) distance fluctuation profile of ARD domain in BARD1 ARD-BRCT *wild-type* and complex (**E**) Structure of *wild-type*-CstF50 complex and (**F**) Model for mechanism of *wild-type*-CstF50 complex stabilization.

perturbations. The RMSD, R_{gyr} and RMSF profiles show that BARD1 Gln 564 His mutant protein structure is more flexible than the *wild-type*. The relative orientation of ARD domain is significantly affected due to the Gln 564 His mutation. In BARD1 ARD-BRCT domain, flexible linker has different orientation that allows free twist and rotation of the ARD domain through space which can facilitates the recruitment of CstF50. The mutation Gln 564 His is present in the disordered flexible linker region of BARD1 ARD-BARD1BRCT domain. Lack of specific secondary structure provides the accessibility of different orientations to the linker region. The increase in the relative flexibility may be due to twist in the linker region by histidine. The linker region acts as a binding region for two large proteins CstF50 and p53. The difference in hydrogen bond donor group in glutamine and histidine may be the other reason in the loss of BARD1-CstF50 interactions. The results from *wild-type*, mutant and CstF50 docking study indicates that glutamine 564 to histidine mutation shows loss in hydrogen bonding interactions between *wild-type* and CstF50. To our conclusion, BARD1 Gln at 564 position plays a pivotal role in the BARD1 ARD-BRCT-CstF50 complex formation. The study further shows CstF50 binding provides rigidity to the ARD domain of BARD1. Therefore, flexibility in the linker region is immensely required to adopt such unique orientation to accommodate CstF50 protein.

Material and Methods

All the chemicals used in this study were purchased form Sigma-Aldrich, unless otherwise specified. The buffers were prepared in double distilled, 0.44 µM filtered Milli-Q water (Millipore, USA). Filtered 0.44 µM (Millipore, USA) protein and buffer solutions were degassed prior to use.

Protein expression and purification. BARD1 ARD-BRCT (425–777) cloned into a modified pET-6H vector (generous gift by Dr. Richard Baer, Institute for Cancer Genetics, Columbia University, USA). BARD1 ARD-BRCT *wild-type* domain was further PCR amplified and cloned into pGEX-KT vector (Amersham) in such a way that TEV protease site was incorporated between the GST and BARD1 ARD-BRCT. Site directed mutagenesis was performed to engineer the Gln 564 His mutation in BARD1 ARD-BRCT, and further sequenced for confirmation. BARD1 ARD-BRCT *wild-type* and Gln 564 His mutant were expressed in *Escherichia coli* BL21 (DE3). The GST-BARD1 ARD-BRCT *wild-type* and Gln 564 His mutant fusion protein were purified using affinity chromatography (GST-sepharose 4B column) in 50 mM Borate buffer, pH 9.0, 300 mM NaCl(Buffer A). The affinity purified proteins were treated with TEV to remove the fusion tag. Molecular exclusion chromatography was further performed to achieve high purity and homogenous monomeric population of BARD1 ARD-BRCT *wild-type* and Gln 564 His mutant proteins. Superdex-75 column was used for the purification of the BARD1 ARD-BRCT *wild-type* protein. The Mutant protein was purified using the superdex-200 analytical column.

Human-CstF50 cDNA was bought from DNASU cDNA repository (Arizona state university) was PCR amplified and cloned in to pET-28a vector (Invitrogen). CstF50 cloned in pET28a vector was expressed in *Escherichia coli BL21* (*DE3*) cells grown at 37 °C until absorbance at $\lambda = 600$ nm reached a value in between 0.6–0.8, followed by induction with 0.1 mM IPTG at 18 °C for 18 hrs. Most of the CstF50 was found as insoluble aggregates at 18 °C when induced with 0.1 mM (IPTG). The insoluble aggregate of CstF50 was washed with buffer composition 50 mM, 500 mM NaCl, 5% glycerol, 1% sarkosyl, 1% SDS, pH 8 (Buffer B) by resuspending the pellet and again centrifuging at 18000 r.p.m. After washing, pellet was resuspended in 50 mM, 500 mM NaCl, 8 M urea, 5% glycerol (Buffer C), pH 8 for 5 hours to solubilize completely. Then 8M-urea-solubilized-CstF50 was refolded using buffer 50 mM Tris, 500 mM NaCl, 0.1% triton, 10% glycerol pH 7.2, 2 mM EDTA, 700 mM arginine, 100 mM KCl, 50 mM $MgCl_2$, 100 mM $CaCl_2$, and 5 mM DTT, pH 6.2 (Buffer D). In the refolding process we have added 10% glycerol and 0.5 M arginine hydrochloride to improve the refolding process[43,44]. To overcome the aggregate formation, high concentration of arginine was used to suppress aggregation of folding intermediates.

Circular - Dichroism spectroscopy. Circular dichroism spectra for BARD1 ARD-BRCT *wild-type* and Gln 564 His mutant were collected using 0.1 cm path length sealed quartz cuvette on a JASCO J-715 spectropolarimeter (Jasco, Easton, MD), equipped with a JASCO PTC 348 WI temperature controller. BARD1 ARD-BRCT and Gln 564 His mutant at a concentration of 15 μM were scanned in the far- UV ($\lambda = 180$–260 nm) and near-UV range ($\lambda = 350$–260 nm). Seven spectra were collected with 20 nm scan speed at a resolution of 1 nm and, response time of 1 s was averaged for each experiment. Blank spectra using buffer E (50 mM Borate buffer, pH 9.0, 150 mM NaCl) have been taken under the similar conditions and subtracted from the raw data. The spectra was recorded at the 25 °C. The results have been expressed terms of molar ellipticity [θ] (deg cm^2 $dmol^{-1}$).

Limited proteolysis. BARD1 ARD-BRCT *wild-type* and mutant at a concentration of 2 mg/ml were incubated with the 10 Pg/μl trypsin in different time intervals of 0, 5, 10, 30, 60 and 120 mins at 37 °C. After incubation, reaction was terminated by adding 1 mM PMSF (sigma Aldrich). Samples collected at different time intervals were analyzed over SDS-PAGE gel. BARD1 ARD-BRCT *wild-type* and Gln 564 His mutant protein untreated with trypsin were taken as control.

Glutaraldehyde crosslinking. BARD1 ARD-BRCT *wild-type* and Gln 564 His mutant at a concentration of 0.5 mg/ml in buffer A (50 mM Borate buffer, pH 9.0, 300 mM NaCl) was incubated with freshly prepared solution of glutaraldehyde (final concentration 0.1%) for 0, 1, 2, 3, 5, 10, 15 and 30 minutes at 37 °C. Crosslinking reaction was terminated by adding 5 μl of 1 M Tris-HCl, pH 8.0, and the samples were analyzed on 12% SDS-PAGE gel.

Dynamic light scattering. Molecular size measurement and oligomeric behavior of BARD1 ARD-BRCT *wild-type* was performed using Malvern Zetasizer (Malvern). BARD1 ARD-BRCT *wild-type*, Gln 564 His mutant protein, and buffer were filtered (0.22 μm), degassed thoroughly prior to every measurement. BARD1 ARD-BRCT *wild-type* protein and mutant protein at a concentration of 1 mg/ml were scanned at an interval of 5 minutes for 15 minutes. The scanning were performed at the 25 °C. The DLS experiments were repeated in three independent sets.

Thermal and chemical denaturation. Thermal denaturation was monitored by circular dichroism spectroscopy. BARD1 ARD-BRCT *wild-type* and Gln 564 His mutant were allowed to unfold over the temperature ranges from 20 °C to 70 °C. Further fraction unfolded was calculated at each temperature by using millidegree ellipticity at λ_{222} and data was fitted in to a two state unfolding pathway. BARD1 ARD-BRCT *wild-type* and mutant Gln 564 His were allowed to unfold by incubating 2 μM protein in different concentrations of GuHcl at 10 °C for 18 hrs. Chemical denaturation of BARD1 ARD-BRCT *wild-type* and Gln 564 His mutant were performed at 10 °C and the fluorescence from frequently used fluorophores was monitored using fluorescence spectrophotometer (Horiba, USA) at excitation wavelength of $\lambda = 295$ nm. BARD1 ARD-BRCT *wild-type* and Gln 564 His mutant at a concentration of 2 μM was mixed with the GuHcl. The concentration of GuHcl varied from 0 M to 6 M, while the protein concentration was fixed at 2 μM. Fluorescence emission spectra were recorded over the range of wavelength $\lambda = 310$–400 nm in order of increasing GuHcl concentration, and blank subtraction was done to increase signal to noise ratio.

Molecular dynamics simulation. The validated Model structures of BARD1 ARD-BRCT *wild-type*, mutant Gln 564 His, CstF50 and BARD1-CstF50 complex were subjected to MD simulations using GROMACS 4.5.5[45–47] software. Protein were solvated with explicit solvent Single Point Charge water, in a cubic box which left 2 Å space nearby the solute. The counter ions in the form of Na^+ and Cl^- ions were added to make the system electrically neutral. Further energy minimization by the steepest descents method was used, and system was coupled to an external bath using Berendsen's method. The reference temperature for all simulation was fixed at 300 K. LINC method was used to constrain all bonds[48,49]. OPLS-AA force field in GROMACS was used and grid type neighbor searching was done, and long range electrostatics was controlled using PME[50,51]. During simulations, the respective energy minimized structures were subjected to a position restrained MD Simulation for 100 ns with a time step of 2 fs. The trajectories were saved at every 2 ps to form respective trajectories. GROMACS built in tools were used to calculate RMSD (Root Mean Square Deviation), RMSF (Root Mean Square Fluctuations) and R_{gyr} (radius of gyration).

Normal mode, Deformation and fluctuation analysis. Normal mode, deformation and fluctuation analysis were performed to study the domain motion in BARD1 ARD-BRCT *wild-type* and mutant protein using R 3.2 bio3D package[52]. First five modes were calculated by building the hessian matrix, and further extracted to explain the relative fluctuation in ARD domain of *wild-type* and mutant protein.

Principal component analysis (PCA). PCA reduces the dimensionality of the MD trajectory data and provides a concise way to visualize, analyze and compare large-scale concerted motions observed over the course of the simulation. PCA was performed using GROMACS in-built tool g_covar, g_anaeig and covariance matrix was built after removing net translational or rotational motion of the system by fitting the coordinate data to a reference structure. Further, diagonalization of the symmetric $3N \times 3N$ covariance matrix was performed via eigenvector decomposition method and an orthogonal set of eigenvectors which are also called as functional "modes" were calculated with eigenvalues.

Molecular docking. BARD1 ARD-BRCT *wild-type*, Gln 564 His and CstF50 docking was performed using HADDOCK server 2.0 which uses data-driven method of docking, with support from an extensive variety of experimental data obtained from diverse biophysical and biochemical experimental method[53].

Binding energy calculation. In common term binding free of complex in solvent can be calculated by $\mathbf{\Delta G_{Binding} = G_{complex} - (G_{protein} + G_{Ligand})}$. Free energy and other energy parameters of the complex were calculated by using molecular mechanics Poisson-Boltzmann surface area (MM-PBSA) method in GROMACS 4.5.5 package. *g_mmpbsa* tool was used to integrate high-throughput molecular dynamics simulation with the estimation of free energy of interaction[54].

References

1. Wu, L. C. *et al*. Identification of a RING protein that can interact *in vivo* with the BRCA1 gene product. *Nat Genet* **14**, 430–440, doi:10.1038/ng1296-430 (1996).
2. Yu, X. & Baer, R. Nuclear localization and cell cycle-specific expression of CtIP, a protein that associates with the BRCA1 tumor suppressor. *J Biol Chem* **275**, 18541–18549, doi:10.1074/jbc.M909494199M909494199 (2000).
3. Li, M. & Yu, X. Function of BRCA1 in the DNA damage response is mediated by ADP-ribosylation. *Cancer Cell* **23**, 693–704 (2013).
4. Jin, Y. *et al*. Cell cycle-dependent colocalization of BARD1 and BRCA1 proteins in discrete nuclear domains. *Proc Natl Acad Sci USA* **94**, 12075–12080 (1997).
5. Yarden, R. I., Pardo-Reoyo, S., Sgagias, M., Cowan, K. H. & Brody, L. C. BRCA1 regulates the G2/M checkpoint by activating Chk1 kinase upon DNA damage. *Nat Genet* **30**, 285–289, doi:10.1038/ng837ng837 (2002).
6. Baer, R. & Ludwig, T. The BRCA1/BARD1 heterodimer, a tumor suppressor complex with ubiquitin E3 ligase activity. *Curr Opin Genet Dev* **12**, 86–91, doi:S0959437X01002696 (2002).
7. Khanna, K. K. & Jackson, S. P. DNA double-strand breaks: signaling, repair and the cancer connection. *Nat Genet* **27**, 247–254, doi:10.1038/85798 (2001).
8. Wu-Baer, F., Lagrazon, K., Yuan, W. & Baer, R. The BRCA1/BARD1 heterodimer assembles polyubiquitin chains through an unconventional linkage involving lysine residue K6 of ubiquitin. *J Biol Chem* **278**, 34743–34746, doi:10.1074/jbc.C300249200C300249200 (2003).
9. Ruffner, H., Joazeiro, C. A., Hemmati, D., Hunter, T. & Verma, I. M. Cancer-predisposing mutations within the RING domain of BRCA1: loss of ubiquitin protein ligase activity and protection from radiation hypersensitivity. *Proc Natl Acad Sci USA* **98**, 5134–5139, doi:10.1073/pnas.08106839898/9/5134 (2001).
10. Hashizume, R. *et al*. The RING heterodimer BRCA1-BARD1 is a ubiquitin ligase inactivated by a breast cancer-derived mutation. *J Biol Chem* **276**, 14537–14540, doi:10.1074/jbc.C000881200C000881200 (2001).
11. Nazeer, F. I. *et al*. p53 inhibits mRNA 3' processing through its interaction with the CstF/BARD1 complex. *Oncogene* **30**, 3073–3083, doi:10.1038/onc.2011.29onc201129 (2011).
12. Irminger-Finger, I. *et al*. Identification of BARD1 as mediator between proapoptotic stress and p53-dependent apoptosis. *Mol Cell* **8**, 1255–1266, doi:S1097-2765(01)00406-3 (2001).
13. Laufer, M. *et al*. Structural requirements for the BARD1 tumor suppressor in chromosomal stability and homology-directed DNA repair. *J Biol Chem* **282**, 34325–34333, doi:M70519820010.1074/jbc.M705198200 (2007).
14. Fong, N., Bird, G., Vigneron, M. & Bentley, D. L. A 10 residue motif at the C-terminus of the RNA pol II CTD is required for transcription, splicing and 3′ end processing. *EMBO J* **22**, 4274–4282, doi:10.1093/emboj/cdg396 (2003).
15. Fong, N. & Bentley, D. L. Capping, splicing, and 3′ processing are independently stimulated by RNA polymerase II: different functions for different segments of the CTD. *Genes Dev* **15**, 1783–1795, doi:10.1101/gad.889101 (2001).
16. Cevher, M. A. *et al*. Nuclear deadenylation/polyadenylation factors regulate 3′ processing in response to DNA damage. *EMBO J* **29**, 1674–1687, doi:emboj20105910.1038/emboj.2010.59 (2010).
17. Kleiman, F. E. & Manley, J. L. Functional interaction of BRCA1-associated BARD1 with polyadenylation factor CstF-50. *Science* **285**, 1576–1579, doi:7797 (1999).
18. Kleiman, F. E. *et al*. BRCA1/BARD1 inhibition of mRNA 3′ processing involves targeted degradation of RNA polymerase II. *Genes Dev* **19**, 1227–1237, doi:19/10/122710.1101/gad.1309505 (2005).
19. Kim, H. S. *et al*. DNA damage-induced BARD1 phosphorylation is critical for the inhibition of messenger RNA processing by BRCA1/BARD1 complex. *Cancer Res* **66**, 4561–4565, doi:66/9/456110.1158/0008-5472.CAN-05-3629 (2006).
20. Li, L., Chen, R. & Weng, Z. RDOCK: refinement of rigid-body protein docking predictions. *Proteins* **53**, 693–707, doi:10.1002/prot.10460 (2003).
21. Thai, T. H. *et al*. Mutations in the BRCA1-associated RING domain (BARD1) gene in primary breast, ovarian and uterine cancers. *Hum Mol Genet* **7**, 195–202, doi:ddb030 (1998).
22. Sauer, M. K. & Andrulis, I. L. Identification and characterization of missense alterations in the BRCA1 associated RING domain (BARD1) gene in breast and ovarian cancer. *J Med Genet* **42**, 633–638, doi:42/8/63310.1136/jmg.2004.030049 (2005).
23. Ghimenti, C. *et al*. Germline mutations of the BRCA1-associated ring domain (BARD1) gene in breast and breast/ovarian families negative for BRCA1 and BRCA2 alterations. *Genes Chromosomes Cancer* **33**, 235–242 (2002).
24. Lee, C. *et al*. Functional Analysis of BARD1 Missense Variants in Homology-Directed Repair of DNA Double Strand Breaks. *Human mutation* **36**, 1205–1214 (2015).
25. Klonowska, K. *et al*. Analysis of large mutations in BARD1 in patients with breast and/or ovarian cancer: the Polish population as an example. *Scientific reports* **5** (2015).
26. Kleiman, F. E. & Manley, J. L. The BARD1-CstF-50 interaction links mRNA 3′ end formation to DNA damage and tumor suppression. *Cell* **104**, 743–753, doi:S0092-8674(01)00270-7 (2001).
27. Woody, R. W. Contributions of tryptophan side chains to the far-ultraviolet circular dichroism of proteins. *Eur Biophys J* **23**, 253–262 (1994).
28. Myers, J. K., Morris, D. P., Greenleaf, A. L. & Oas, T. G. Phosphorylation of RNA polymerase II CTD fragments results in tight binding to the WW domain from the yeast prolyl isomerase Ess1. *Biochemistry* **40**, 8479–8486 (2001).
29. Lakowicz, J. R. *Principles of fluorescence spectroscopy*. (Springer Science & Business Media, 2007).
30. Ishibashi, M. *et al*. Is arginine a protein-denaturant? *Protein Expr Purif* **42**, 1–6, doi:10.1016/j.pep.2005.03.028 (2005).

31. Srinivas, V., Raman, B., Rao, K. S., Ramakrishna, T. & Rao Ch, M. Structural perturbation and enhancement of the chaperone-like activity of alpha-crystallin by arginine hydrochloride. *Protein Sci* **12**, 1262–1270, doi:10.1110/ps.0302003 (2003).
32. Fiser, A. & Sali, A. ModLoop: automated modeling of loops in protein structures. *Bioinformatics* **19**, 2500–2501 (2003).
33. Fiser, A., Do, R. K. & Sali, A. Modeling of loops in protein structures. *Protein Sci* **9**, 1753–1773, doi:10.1110/ps.9.9.1753 (2000).
34. Chen, V. B. *et al*. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D: Biological Crystallography* **66**, 12–21 (2009).
35. Kaplan, W. & Littlejohn, T. G. Swiss-PDB Viewer (Deep View). *Brief Bioinform* **2**, 195–197 (2001).
36. Yang, L., Song, G. & Jernigan, R. L. How well can we understand large-scale protein motions using normal modes of elastic network models? *Biophysical Journal* **93**, 920–929 (2007).
37. Brooks, B. & Karplus, M. Normal modes for specific motions of macromolecules: application to the hinge-bending mode of lysozyme. *Proceedings of the National Academy of Sciences* **82**, 4995–4999 (1985).
38. Paladino, A., Costantini, S., Colonna, G. & Facchiano, A. M. Molecular modelling of miraculin: Structural analyses and functional hypotheses. *Biochem Biophys Res Commun* **367**, 26–32, doi:10.1016/j.bbrc.2007.12.102 (2008).
39. Kurplus, M. & McCammon, J. Dynamics of proteins: elements and function. *Annual review of biochemistry* **52**, 263–300 (1983).
40. Irminger-Finger, I. & Jefford, C. E. Is there more to BARD1 than BRCA1? *Nat Rev Cancer* **6**, 382–391, doi:nrc187810.1038/nrc1878 (2006).
41. Fox, D. *et al*. Crystal structure of the BARD1 ankyrin repeat domain and its functional consequences. *Journal of Biological Chemistry* **283**, 21179–21186 (2008).
42. Edwards, R. A. *et al*. The BARD1 C-terminal domain structure and interactions with polyadenylation factor CstF-50. *Biochemistry* **47**, 11446–11456, doi:10.1021/bi801115g (2008).
43. Arora, D. & Khanna, N. Method for increasing the yield of properly folded recombinant human gamma interferon from inclusion bodies. *J Biotechnol* **52**, 127–133 (1996).
44. Buchner, J. & Rudolph, R. Renaturation, purification and characterization of recombinant Fab-fragments produced in Escherichia coli. *Biotechnology (NY)* **9**, 157–162 (1991).
45. Berendsen, H. J., van der Spoel, D. & van Drunen, R. GROMACS: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications* **91**, 43–56 (1995).
46. Hess, B., Kutzner, C., Van Der Spoel, D. & Lindahl, E. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of chemical theory and computation* **4**, 435–447 (2008).
47. Van Der Spoel, D. *et al*. GROMACS: fast, flexible, and free. *J Comput Chem* **26**, 1701–1718, doi:10.1002/jcc.20291 (2005).
48. Hess, B., Bekker, H., Berendsen, H. J. & Fraaije, J. G. LINCS: a linear constraint solver for molecular simulations. *Journal of computational chemistry* **18**, 1463–1472 (1997).
49. Miyamoto, S. & Kollman, P. A. SETTLE: an analytical version of the SHAKE and RATTLE algorithm for rigid water models. *Journal of computational chemistry* **13**, 952–962 (1992).
50. Essmann, U. *et al*. A smooth particle mesh Ewald method. *The Journal of chemical physics* **103**, 8577–8593 (1995).
51. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An N· log (N) method for Ewald sums in large systems. *The Journal of chemical physics* **98**, 10089–10092 (1993).
52. Grant, B. J., Rodrigues, A. P., ElSawy, K. M., McCammon, J. A. & Caves, L. S. Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics* **22**, 2695–2696 (2006).
53. De Vries, S. J. *et al*. HADDOCK versus HADDOCK: new features and performance of HADDOCK2. 0 on the CAPRI targets. *Proteins: Structure, Function, and Bioinformatics* **69**, 726–733 (2007).
54. Kumari, R., Kumar, R. & Lynn, A. g_mmpbsa A GROMACS Tool for High-Throughput MM-PBSA Calculations. *Journal of chemical information and modeling* **54**, 1951–1962 (2014).

Acknowledgements

Funding for this study was supported by Seed in Air grant from TMC. We thank DBT-BTIS facility at ACTREC and Mizoram University for providing necessary software to this study. We thanks M.V. hosur for critically reading the manuscript.

Author Contributions

R.K.C., M.Q.S., P.S.T. and N.G. performed the calculations, R.K.C., S.K.N. and A.K.V. designed the experiments and wrote the paper.

Additional Information

Supplementary information accompanies this paper at doi:10.1038/s41598-017-03816-4

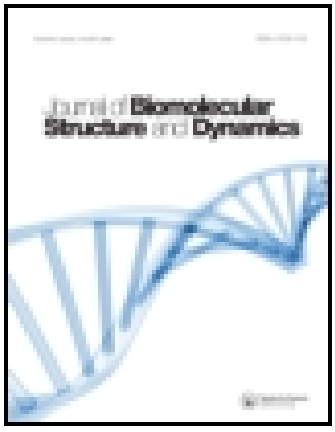
Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2017

This article was downloaded by: [Umeå University Library]
On: 07 October 2014, At: 02:41
Publisher: Taylor & Francis
Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Journal of Biomolecular Structure and Dynamics
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tbsd20

Mislocalization of BRCA1-complex due to ABRAXAS Arg361Gln mutation

Vikrant[a], Rajan Kumar[a], Quadir Siddiqui[a], Nalini Singh[a], Sanjeev K. Waghmare[a] & Ashok K. Varma[a]

[a] Advanced Centre for Treatment, Research and Education in Cancer, Tata Memorial Centre, Kharghar, Navi Mumbai, Maharashtra 410 210, India

Published online: 08 Aug 2014.

To cite this article: Vikrant, Rajan Kumar, Quadir Siddiqui, Nalini Singh, Sanjeev K. Waghmare & Ashok K. Varma (2014): Mislocalization of BRCA1-complex due to ABRAXAS Arg361Gln mutation, Journal of Biomolecular Structure and Dynamics, DOI: 10.1080/07391102.2014.945484

To link to this article: http://dx.doi.org/10.1080/07391102.2014.945484

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at http://www.tandfonline.com/page/terms-and-conditions

Journal of Biomolecular Structure and Dynamics, 2014
http://dx.doi.org/10.1080/07391102.2014.945484

Taylor & Francis
Taylor & Francis Group

Mislocalization of BRCA1-complex due to ABRAXAS Arg361Gln mutation

Vikrant, Rajan Kumar, Quadir Siddiqui, Nalini Singh, Sanjeev K. Waghmare and Ashok K. Varma*

Advanced Centre for Treatment, Research and Education in Cancer, Tata Memorial Centre, Kharghar, Navi Mumbai, Maharashtra 410 210, India

Communicated by Ramaswamy H. Sarma

(*Received 10 March 2014; accepted 14 July 2014*)

ABRAXAS is an integral member of BRCA1-complex, which helps in its recruitment to the DNA damage site. It interacts with BRCA1 via its C-terminal phospho-peptide binding motif while the N-terminal associates with RAP80, and thereby recruits the BRCA1-complex at the site of DNA damage. Nonetheless, how ABRAXAS helps in the structural integrity of BRCA1-complex, and its DNA repair mechanism remains elusive. To elucidate the role of ABRAXAS in the DNA repair process, we characterized the ABRAXAS wild type and Arg361Gln mutant using *in silico* and *in vitro* approach. It has been observed that ABRAXAS Arg361Gln mutant is responsible for defective nuclear localization of BRCA1-complex, and hence important for DNA repair function. We found conformational changes in ABRAXAS mutant, which impaired binding to RAP80 and further disturb BRCA1-complex localization. The results presented in this paper will help to understand the cause of BRCA1 mislocalization, and various DNA repair defects that occur due to substitution. Comparative study of ABRAXAS wild type and mutant will provide helpful perspective for inhibitor designing that can potentially recompense the deleterious effect(s) of Arg361Gln mutation, and have therapeutic application.

Keywords: structural stability; protein–protein interactions; ABRAXAS; BRCA1; DNA repair

Downloaded by [Umeå University Library] at 02:41 07 October 2014

1. Introduction

DNA damage repair is an indispensable phenomenon requisite for maintaining genomic integrity and survival of living organisms in the biological system (Rouse & Jackson, 2002; Zhou & Elledge, 2000). Cells repair its damaged DNA by integrating several processes, including DNA replication, gene regulation, cell cycle, and apoptosis in a complex network of intercommunicating molecules, called DNA damage response (DDR) (Bartek, Bartkova, & Lukas, 2007; Bartek & Lukas, 2007). DDR primarily involves three groups of molecules such as sensor, mediator, and effector (Bartkova et al., 2007; Celeste et al., 2003; Sartori et al., 2007; Shiotani & Zou, 2009a, 2009b). Genotoxic stress causes the double-strand break (DSB) which leads to activation of ATM/ATR kinases. This event executes phosphorylation of histone variant such as H2AX, and MDC1, and promotes their phosphorylation-dependent interaction (Burma, Chen, Murphy, Kurimasa & Chen, 2001; Lou, Minter-Dykhouse, Wu, & Chen 2003; Paull et al., 2000; Rogakou, Pilch, Orr, Ivanova, & Bonner, 1998). γ-H2AX and MDC1 assembly creates a docking site for E3 ubiquitin ligase complex UBC13/RNF8, which further poly-ubiquitinates γ-H2AX (Huen et al., 2007; Kolas et al., 2007; Mailand et al., 2007). The K-63-linked polyubiquitin chain formed on γ-H2AX is recognized by two tandem Ubiquitin Interacting Motif (UIMs) of adaptor molecule, RAP80 (Huen et al., 2007; Kim, Chen, & Yu, 2007; Kolas et al., 2007; Sato et al., 2009; Vikrant, Kumar et al., 2013). RAP80 has a middle ABRAXAS Interacting Region (AIR) domain through which it interacts with ABRAXAS. ABRAXAS acts as a bridging molecule in the BRCA1-complex which comprises RAP80, ABRAXAS, BRCA1, MERIT40, BRCC36 and BRCC45 (Badgujar, Sawant, Yadav, Hosur, & Varma, 2013; Celeste et al., 2003; Mailand et al., 2007; Sobhian et al., 2007; Vikrant, Sawant, & Varma, 2014; Wang et al., 2000, 2007).

BRCA1-complex foci formation is essential for execution of homologous recombination repair process after DNA damage. siRNA-mediated knockdown of ABRAXAS reduces the BRCA1-complex foci formation after IR-induced DNA damage (Wang & Elledge, 2007). RAP80 double mutant lacking UIM and AIR completely abolishes foci formation, inferring a potential role of ABRAXAS in BRCA1-complex formation (Wang & Elledge, 2007). ABRAXAS and RAP80 knockdown cells showed defective homologous recombination repair and become hypersensitive to IR and UV (Kim, Huang, & Chen, 2007; Wang et al., 2007). Role of ABRAXAS is also suspected in G2/M check point activation since

*Corresponding author. Email: avarma@actrec.gov.in

© 2014 Taylor & Francis

Downloaded by [Umeå University Library] at 02:41 07 October 2014

depletion of RAP80 and ABRAXAS displayed defective cell- cycle check points (Kim, Chen et al., 2007; Wang et al., 2007). ABRAXAS harbors consensus sequence of pS-X-X-F (pS- phosphoserine, X-any amino acid) binding motif at its C-terminal, which interact with BRCA1-BRCT and bring about foci formation (Badgujar et al., 2013; Vikrant, Kumar et al., 2013). ABRAXAS acts upstream of BRCA1 and its knockdown significantly reduce accumulation of BRCA1-complex at the DNA damage site. There are reports that loss of ABRAXAS or BRCA1 could lead to similar phenotype, thereby projecting ABRAXAS as an excellent candidate gene for familial breast cancer study (Liu, Wu & Yu, 2007; Osorio et al., 2009; Sobhian et al., 2007; Wang et al., 2007).

Different mutations that disrupt interactions with phospho-binding partners of BRCA1-BRCT have been reported in the BIC (Breast Cancer Information Core) database (Coquelle, Green, & Glover, 2011; Nikkilä et al., 2009; Pylkas, Erkko, Nikkila, Solyom, & Winqvist, 2008; Williams & Glover, 2003). Familial mutation at the C-terminus of ABRAXAS in Finnish population (Arg361Gln) is suspected to disassemble the BRCA1-complex (Solyom et al., 2012). c.1082G > A alteration was observed among 3 of 125 (2.4%) breast cancer families; however, it was absent in 868 healthy female volunteers (Solyom et al., 2012). This variant impairs nuclear localization of BRCA1-complex and its DNA damage repair function, thereby can predispose an individual to repair defects (Solyom et al., 2012). Immunoprecipitation of epitope tagged ABRAXAS and its mutant Arg361Gln with BRCA1 and other core complex in the cytoplasm displays its failure to form foci in the nucleus after DNA damage, while nuclear localization of wild type is retained (Solyom et al., 2012). Failure to achieve nuclear retention adversely affects the G2/M checkpoint and homology-directed DNA repair, which reduces nuclear retention of ABRAXAS-interacting partners to DSB site (Solyom et al., 2012). Moreover, expression of the Arg361Gln variant causes hypersensitivity to ionizing radiation and reduces BRCA1 localization at the sites of DNA damage (Solyom et al., 2012).

ABRAXAS, as a newly emerging susceptibility gene to cancer predisposition, opens the vast perspective of studying the role and impact of genetic alteration in disease progression. In order to understand the deleterious effects of mutation and possible consequences, we have performed bio-physicochemical characterization and interaction analysis of wild type and mutant. To our knowledge, it is the first multi-model approach which elucidates the structural and functional implications of ABRAXAS and its mutant Arg361Gln. ABRAXAS mutant exhibited relatively less structure distortion and stability; however, sufficient to impair its binding with RAP80. The results presented here will help to understand the role of Arg361Gln variant of ABRAXAS in BRCA1-complex recruitment and DNA damage repair function. It might further assist in classifying the severity of ABRAXAS Arg361Gln variant in breast cancer development.

2. Results and discussion

2.1. In silico *analysis*

ABRAXAS is the key member of BRCA1-complex, and share least structural homology with reported structures in the database, which limits full-length *in silico* homology modeling. Consequently, ABRAXAS (351-370) peptide was modeled using the structure of a short region of Amidohydrolase-2 as a template (PDB: 3S4T) which showed highest similarity with ABRAXAS mutation containing region (Kim, Chivian & Baker, 2004; Leaver-Fay et al., 2011). Modeled structure was validated through "SAVES" server, which showed acceptable Ramachandran plot and structural parameters (Ramachandran, Ramakrishnan, & Sasisekharan, 1963; Ramachandran & Sasisekharan, 1968) (Figure 1(A), Figure 1 supplementary). ABRAXAS wild type peptide has shown more of α-helices while the three-dimensional structure (tertiary structure) of mutant peptide appeared different with reduced helicity (Figure 1(A)). Since we have observed localized changes in the structural profile of mutant relative to wild type, it was intriguing to look into such amendments using *in silico* approach.

2.2. *Unaltered oligomeric behavior of ABRAXAS mutant*

Oligomerization is a common phenomenon observed among disordered and exposed hydrophobic patches containing proteins (Zerovnik, 2011). Substitution mutation generally causes distortion in secondary and tertiary structures, which may lead to the disorderness (Kolchanov, Soloviov, & Zharkikh, 1983; Zhukov, Jaroszewski, &

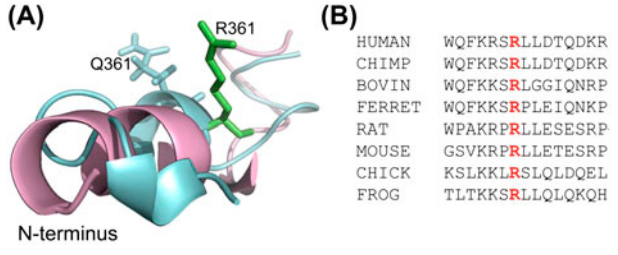


Figure 1. *In silico* structural analysis of ABRAXAS wild type and mutant.
Notes: (A) Superimposed structures of ABRAXAS wild type (pink) and mutant (cyan) peptides. (B) Multiple sequence alignment of ABRAXAS showed highly conserved nature of Arg361 (highlighted in red color) residue among different species.

Downloaded by [Umeå University Library] at 02:41 07 October 2014

Bierzynski, 2000). The exposed hydrophobic patches of protein are highly prone to form nonspecific interaction with each other and may result high-order oligomer. ABRAXAS Arg361Gln mutation results in substitution of a basic amino acid residue with nonpolar charged residue that might affect the monomeric nature of protein. However, results obtained from *in silico* study illustrate localized structural changes in mutant, which may affect the protein folding and oligomerization. To explore this prospect, we purified the ABRAXAS wild type and mutant using two-step purification involving affinity and gel filtration chromatography (Figure 2(A) and (B)). Most often, a mutant protein tends to show different solubility, binding, and elution profiles during purification than the wild type provided that the structural conformation of the protein is drastically altered. Interestingly, we did not observe any significant differences in solubility, binding, and elution profiles between wild type and mutant. In gel filtration chromatography, both the proteins eluted as a monomer at the same elution volume indicating their similar molecular weight (Figure 2(B)) (Table 1). The experimentally determined molecular weight using gel filtration chromatography (wild type: 40KDa, Arg361Gln, 40 KDa) and theoretically derived molecular weight (wild type; 42.9827KDa, Arg361Gln; 42.9547) (Table 1) showed good concurrence, suggesting monomeric nature of both wild type and mutant. Coomassie stained SDS-PAGE of ABRAXAS wild type and mutant showed a single band corresponding to 43 KDa (Figure 2(A)). Therefore, it can be concluded that the Arg361Gln mutation does not affect the monomeric nature of protein.

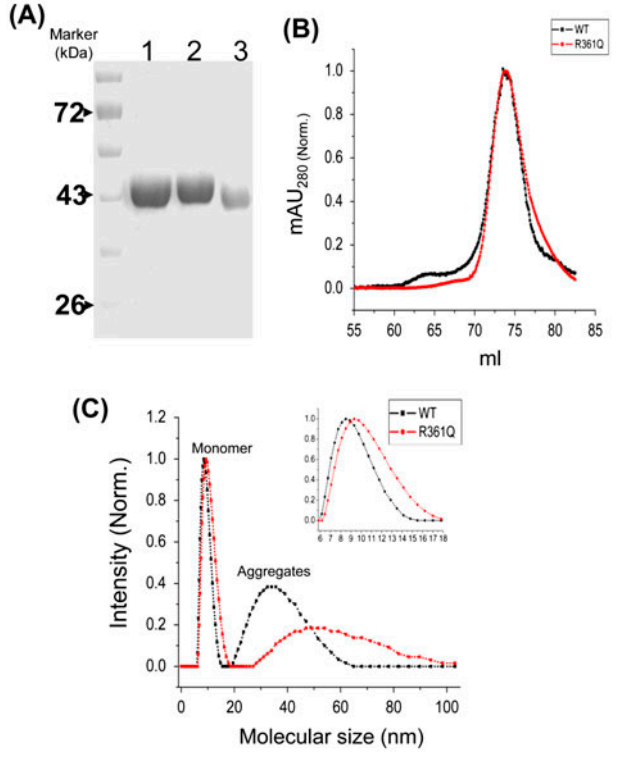


Figure 2. Purification and hydrodynamic diameter measurement of ABRAXAS wild type and mutant.
Notes: (A) SDS-PAGE showing purified protein of ABRAXAS (6-373) wild-type (lane 1), mutant (lane 2), and RAP80 (lane 3). (B) Overlay of gel filtration spectra of wild type and mutant (Superdex 200). Elution profile of both the protein was similar and showed the presence of single peak corresponding to monomer. (C) Overlay of DLS profile of wild type and mutant displayed different hydrodynamic radii, inset is showing the magnified region of the peak corresponding to monomeric population.

To determine the oligomeric heterogeneity between wild type and mutant, and the effect of mutation on molecular dimensions, dynamic light scattering (DLS) was used, which provides precise measurement of the molecular dimension of biomolecules in the solution. We observed that the mutant and wild-type protein samples were composed of two different clusters, the first corresponded to the monomeric populations, whereas second for high-order oligomer or aggregates (Figure 2(C)). The observed hydrodynamic diameter of wild type was found to be $9.0 \pm .3$ nm, while the mutant showed increment up to $10.0 \pm .5$ nm. The relative increase in effective diameter of mutant compared to wild type indicates molecular expansion due to substitution mutation. However, the increment in hydrodynamic diameter did not affect the monomeric nature of mutant, as confirmed through size exclusion chromatography (Figure 2(B)). These findings support that Arg361Gln mutation probably incorporates a short- or long-range structural alteration in protein without changing its oligomeric characteristics. The alterations could be either at the secondary or tertiary structural level, and might have a local or global effect.

2.3. *Structural organization*

Limited-proteolysis based approach was used to determine the structural alterations in ABRAXAS wild type and mutant (Fontana, de Laureto, Spolaore, & Frare, 2012). In general, a compact globular domain of protein considerably resists the protease digestion, while a disordered or unstructured region, which might have arisen due to mutation, undergoes rapid digestion. If a mutation causes severe damage in the α-helices or β-sheets, it might incorporate disorderedness during protein folding. These unstructured regions undergo rapid digestion due to more accessibility of protease sites (Fontana et al., 2012). To understand domain integrity and determine the stability of ABRAXAS wild type and mutant against the protease digestion, limited digestion with trypsin and chymotrypsin proteases was performed. Wild-type and mutant proteins were treated with equal concentration of proteases for limited time, and the results were analyzed

Table 1. Molecular weight estimation of ABRAXAS wild type and mutant.

	Theoretical Mol. Wt. (kDa)[a]	Ve/Vo[b]	Experimental derived Mol. Wt. (kDa) Gel filtration chromatography
Wild type	42.98	1.6344	40 ± 3.2
Mutant	42.95	1.6344	40 ± 3.2

Note: Ve/Vo: Elution volume/Void volume ratio in gel filtration chromatography (superdex 200 16/60).
[a]Determined from Protparam, Expasy.
[b]Compared with standard myoglobin, ovalbumin, albumin, IgG, Ferritin.

on SDS-PAGE (Figure 3(A–D)). ABRAXAS wild-type resistance toward protease digestion was analogous to mutant, which indicates the existence of similar structural domain(s). Furthermore, equivalent susceptibility of mutant protein toward protease digestion suggests that substitution of Arg361Gln does not destabilize the domain integrity of ABRAXAS. This certainly indicates that mutant and wild type are having similar structural pattern with equivalent number of α-helices and β-sheets. It also negates the possibility of major structural changes in the mutant protein, which otherwise could have exposed more protease sites. However, it does not eliminate the possibility of a long- or short-range structural alteration in mutant. Since the observed structural domain (s) organization in limited proteolysis was similar, it appears that wild type and mutant might be having equivalent or modestly changed secondary and tertiary structures. To explore this likelihood, we compared the secondary structure of ABRAXAS wild type and mutant using far-UV Circular Dichroism (CD) (Figure 4(A)). It was observed that ABRAXAS wild type and mutant have well-defined α/β structure, with β-sheets characteristic being more prominent. Data analysis using Dichroweb server showed that wild type and mutant are having α-helices (wild type 15%, mutant 15.5%) and β-sheets (wild type 24%, mutant 23%) (Lobley, Whitmore & Wallace, 2002). Secondary structure comparison of wild type and mutant deciphered similar α-helical and β-sheets characteristics illustrating minor effect of ABRAXAS Arg361Gln mutation on overall protein secondary structure. These observations suggest that Arg361Gln mutation might be inducing short-range secondary structural changes in mutant, which might in turn collectively responsible for tertiary structure modification.

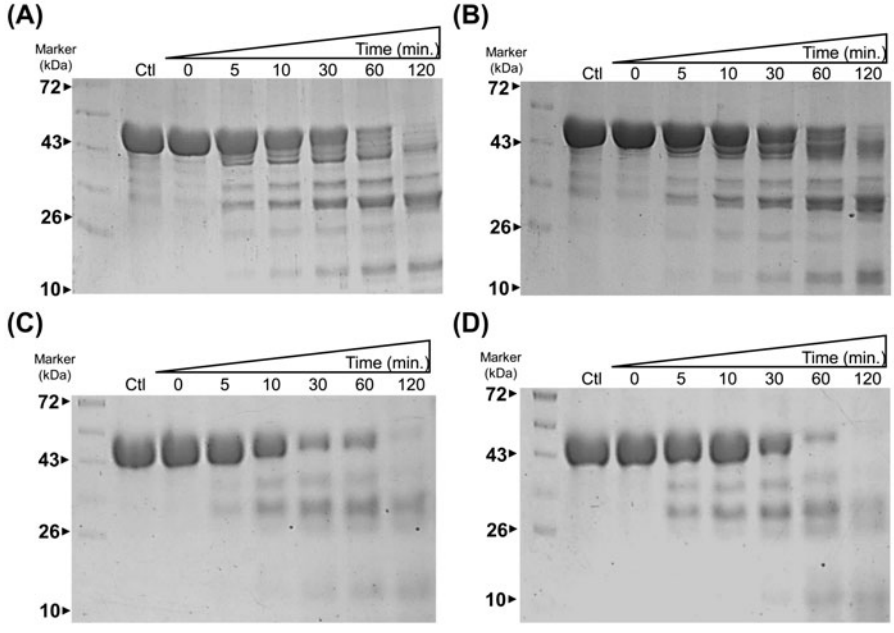


Figure 3. Resistance profile of ABRAXAS wild type and mutant towards protease digestion.
Notes: Limited proteolysis of ABRAXAS (6-373) wild type (A, C) and mutant (B, D) using trypsin (A, B) and Chymotrypsin (C, D) as proteases. Wild type and mutant showed relatively equal resistance toward proteolysis as indicated by similar rate of disappearance in band intensity w.r.t time. Ctl- control as untreated with trypsin and chymotrypsin, respectively.

Downloaded by [Umeå University Library] at 02:41 07 October 2014

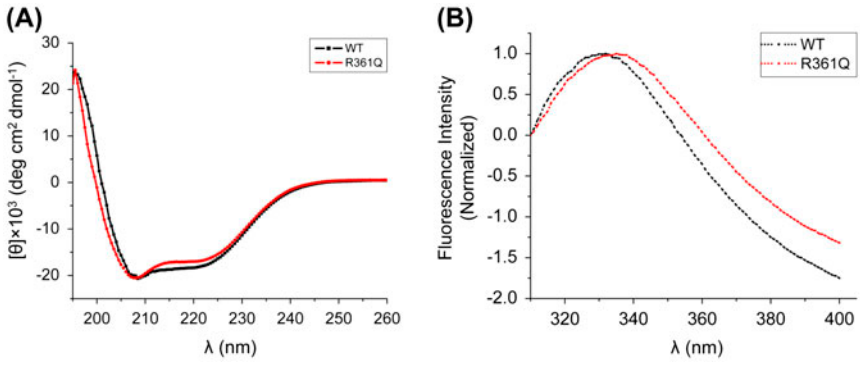


Figure 4. Secondary and tertiary structure evaluation of ABRAXAS wild type and mutant.
Notes: Comparison of secondary structural components of ABRAXAS (6-373) wild type and mutant. (A) Overlay of Far-UV CD spectrum of wild type and mutant. Wild type as well as mutant showed a well defined α/β characteristics and similar structural components. (B) Overlay of fluorescence emission spectrum of wild type and mutant. Wild type showed an emission maxima of 332 nm while mutant undergoes a red shift up to 334 nm.

Downloaded by [Umeå University Library] at 02:41 07 October 2014

Intrinsic fluorophores such as tryptophan and tyrosine are most frequently used to monitor the microenvironmental changes induced in the protein tertiary structure (Muniz et al., 2011). These aromatic residues are usually buried inside the protein hydrophobic core and are extremely sensitive to modification in their microenvironment. Since we observed similar secondary structural components in wild type and mutant, it would be interesting to look into the tertiary structure, which might have altered due to mutation. To study the three-dimensional structure of ABRAXAS wild-type and mutant, Trp and Tyr microenvironments were monitored using fluorescence spectroscopy. Protein's emission spectra were recorded at 280 nm to observe the cumulative effect of Trp and Tyr residues' position on the tertiary structure. An emission maxima at λ 332 nm for wild type and 334 nm for Arg361Gln suggested slight changes in Trp and Tyr microenvironments (Figure 4(B)). A red shift in emission maxima of mutant indicates relatively more exposure of Trp and Tyr toward protein hydrophilic surface (i.e. toward protein exterior). However, the shift was very small and might not be sufficient for large structural amendment. Thus, it can be speculated that tertiary structural components become moderately altered without causing any drastic conformational changes.

2.4. Thermal and chemical stability

ABRAXAS Arg361Gln mutation probably does not amend the secondary structural features of protein drastically, as revealed through CD. A mild alteration in tertiary structure was observed, which may or may not affect the stability profile of protein. However, topology of the native state can be the most imperative factor determining the folding profile, and similar structure proteins may unfold through a different pathway (Zarrine-Afsar, Larson & Davidson, 2005). To evaluate this, thermal stability of ABRAXAS wild type and mutant was compared at secondary (CD) and tertiary structure (fluorescence) level. The spectrum obtained from CD corresponding to λ_{222} nm showed the maximum change in ellipticity and high signal-to-noise ratio. Change in molar ellipticity value was plotted against different temperatures (Figure 5(A)). Melting temperature (T_m), a temperature at which half of the protein assumed to be unfolded, and free energy ($\Delta G°_{H2O}$) of unfolding process were calculated using linear extrapolation method of curve fitting at temperature 273 K (Pace, 1986; Pace & Shaw, 2000). A higher T_m value usually indicates a more stable protein, and the path of unfolding curve describes about the folding pathway. Thermal stability of ABRAXAS wild-type (T_m 35.0 °C, $\Delta G°_{H2O}$ 2.1 ± .50 kcal/mol) was found appreciably equivalent to mutant (T_m 34.8 °C, $\Delta G°_{H2O}$ 1.53 ± .60 kcal/mol), and both endured similar folding pathway (Figure 2A, supplementary). For stability assessment of tertiary structure, emission maximum of wild type and mutant was monitored at different temperatures corresponding to λ_{280} and fraction unfolded was calculated. A plot of fraction unfolded against the temperature or unfolding agent concentration gives the information about folding pathway and stability. Fluorescence spectroscopy revealed comparable unfolding pattern but change in stability. The derived Tm value at 298 K temperature was 23.0 ± 3.1 °C for mutant ($\Delta G°_{H2O}$ 1.25 ± .50 kcal/mol) and 27.0 ± 2.50 °C for wild-type ($\Delta G°_{H2O}$ 1.08 ± .38 kcal/mol) (Figure 5(A); Figure 2A, supplementary). As reported earlier, the presence of a biphasic unfolding curve with two transitions indicates the formation of an intermediate species during unfolding process (Walters, Milam, & Clark, 2009). The resultant unfolding pattern can be considered as three-state transition (Walters et al., 2009).

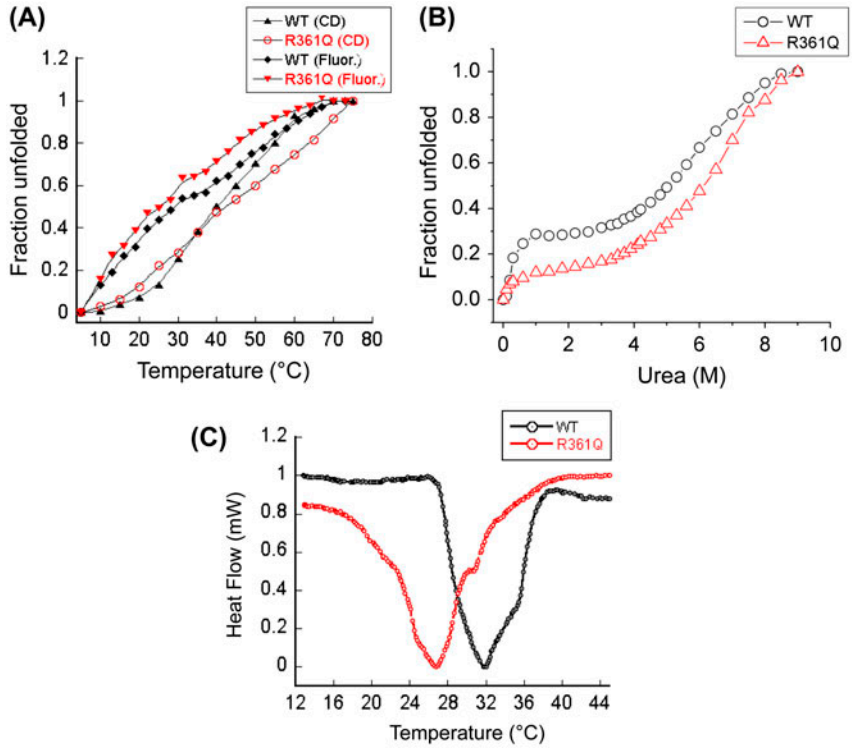


Figure 5. Thermal and chemical stability of ABRAXAS wild type and mutant.
Notes: (A) Thermal stability profile of wild type and mutant showing overlay of protein fraction unfolded at different temperatures. Wild type showed relatively higher thermal stability. (B) Chemical stability profile of wild type and mutant showing overlay of protein fraction unfolded at different urea concentrations. Wild type showed relatively high propensity of intermediate species existence as compared to mutant indicating different unfolding pathways. (C) DSC of wild type and mutant proteins showing a well defined transition around 32 and 27 °C, respectively.

Fluorescence-based thermal denaturation revealed that both wild type and mutant most likely unfold through an intermediate species, and undergo three-state transitions (Figure 5(A)). To determine the chemical stability of wild type and mutant, protein samples (2 μM) were incubated at different concentrations of urea (0–8 M) until equilibrium achieved, and emission maxima were recorded at a excitation wavelength of λ_{280}. Thermodynamic parameters for ABRAXAS wild type $\Delta G°_{H2O}$ 2.54 ± .36 kcal/mol and mutant $\Delta G°_{H2O}$ 2.01 ± .41 kcal/mol were calculated by plotting fraction unfolded against urea concentration at temperature 298 K (Figure 2B, supplementary). Chemical stability of mutant was found different, nevertheless, both unfold through acquiring an intermediate species which was more predominate in case of wild-type (Figure 5(B)). Fluorescence-based unfolding study using thermal and chemical denaturation methods suggests a higher stability of wild type as compared to mutant. To substantiate these findings and evaluate the reversibility of unfolding of wild type and mutant, Differential Scanning Calorimetry (DSC) was performed. A peak maximum was considered as temperature of melting and area under the curve was used to determine enthalpy. Folding pathway was traced according to the pattern of curve progression. DSC data showed overlapping of two transitions, which could be corresponding to an intermediate formed during unfolding process, and displayed the three-state reversible folding pathways of wild type and mutant. The observed Tm value for wild type was significantly higher than the mutant (wild type T_m 32.0 ± 2.5 °C, ΔH 440 ± 12 Kcal/mol; mutant T_m 27.0 ± 3.0 °C, ΔH 660 ± 15 kcal/mol) (Figure 5(C)). Furthermore, the observed intermediate species was reluctant to exist in case of mutant while predominantly shown by wild type. These results suggest that overall stability of ABRAXAS was different due to existence of Arg361Gln mutation, and undergo dissimilar folding pathway. Altogether, these findings suggest the relative existence of wild type in a more stable form. However, the type of intermediate species formed during unfolding transition, and its characterization is beyond the scope of study.

2.5. *Binding interaction*

ABRAXAS is a member of BRCA1-complex and involves in direct interaction with RAP80 and BRCA1 (Kim, Huang et al., 2007). ABRAXAS binds directly to RAP80

Downloaded by [Umeå University Library] at 02:41 07 October 2014

Downloaded by [Umeå University Library] at 02:41 07 October 2014

and BRCA1 BRCT, thus facilitate the recruitment of BRCA1-complex at the DNA damage site. In our study, we found that the secondary structural components of wild type and mutant remained same irrespective of the mutation. However, the three-dimensional folding pattern and unfolding pathway indicated less stability for the mutant protein. Since structural components and folding pathway monitored at the global level showed differences, there is a possibility that the binding interactions between ABRAXAS wild type or mutant with RAP80 would likely be affected. To test this assumption, His pull-down assay of native RAP80 (1-405) with His-tagged ABRAXAS wild type and mutant was performed. Interestingly, the binding of ABRAXAS wild type and RAP80 was found significantly higher compared to the mutant (Figure 6). This difference in binding profile suggested that structural alteration in mutant hinders the association between RAP80 and ABRAXAS.

ABRAXAS is the key member of BRCA1-complex and acts as a bridging molecule among various members. ABRAXAS expression was significantly correlated with lower chance of tumorigenesis in patients with advanced no small-cell lung cancer receiving firstline platinum–gemcitabine chemotherapy (Joerger et al., 2011). Knockout studies of ABRAXAS result in defective recruitment of BRCA1-complex, and hence the DNA repair defect (Kim, Huang et al., 2007; Wang et al., 2007). Thus, it is a multifaceted molecule that plays an important role in BRCA1-mediated homologous recombination repair and cancer progression.

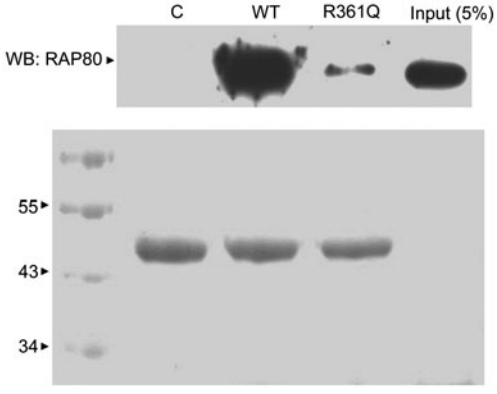


Figure 6. Binding analysis of ABRAXAS wild type and mutant with RAP80.
Notes: Histidine pull down assay followed by western blotting (Aygun, Svejstrup & Liu, 2008; Wang & Kirschner, 2013). Upper panel-ABRAXAS wild type or mutant was used as a bait and RAP80 as prey. RAP80 was probed with anti-RAP80 antibody. Heat denatured RAP80 was taken as control. Lower panel-Ponceau stained PVDF membrane showing the ABRAXAS His-tagged fusion protein as bait(s). Loading sequence is the same as in case of the upper panel. Wild type showed higher binding proportion compared to mutant.

In the present study, we have carried out a comprehensive examination of structural and functional properties of ABRAXAS, and their alterations due to mutation. The modeled structure of ABRAXAS wild type and mutant showed modest structural alteration. However, the observed structural changes in mutant do not alter its monomeric nature. The mutation containing region is a part of coiled-coil domain, which mainly involves in interaction with other BRCA1-complex members. Multiple sequence alignment of ABRAXAS considering various species in phylogenic order showed highly conserved nature of Arg361 residue (Figure 1(B)) which signifies its importance in disease susceptibility. Wild type and mutant showed similar secondary structural composition, whereas the relative orientation of Trp and Tyr was slightly disturbed. This indicates that Arg361Gln mutation brings several localized changes in structural pattern of ABRAXAS, which altogether furnish a different conformational stability in a cumulative manner. Albeit, these conformation changes are very minor and could not be detected at the secondary structure level, but their relative positions were traced during global unfolding. The relative redundancy of intermediate species in case of wild type suggests the existence of different unfolding pathways which are partially followed by mutant as well as unfolding of wild type that was found to be more cooperative. Altogether, the modest localized changes in the mutant structure bring down its thermal and chemical stability which further perturbs the interaction with RAP80. The cumulative global changes in mutant structure were sufficient to disturb critical interaction necessary for BRCA1-complex integrity and localization. Therefore, in the presence of Arg361Gln mutation, ABRAXAS could not extend its bridging interaction(s) through RAP80, which perhaps prevent the recruitment of BRCA1-complex to the DNA damage site (Kim, Chen et al., 2007; Sobhian et al., 2007; Wang et al., 2007). Consequently, the nuclear retention of BRCA1 is adversely affected which further agitates G2/M checkpoint and homology-directed DNA repair (Figure 7(A) and (B)) (Solyom et al., 2012). To best of our knowledge, it is the first comprehensive report elucidating the structural and functional mechanism of mutational consequences and disease susceptibility. These findings would substantially list ABRAXAS as a new susceptibility gene to cancer predisposition. It also opens the vast perspective of considering the Arg361Gln mutation role in disease progression, such as cancer. It will further explore the opportunity of inhibitor design for therapeutic application that can recompense the effect of such deleterious mutation.

3. Materials and methods

High-quality molecular biology or analytical grade chemicals (>99.9%) were purchased from Sigma Aldrich

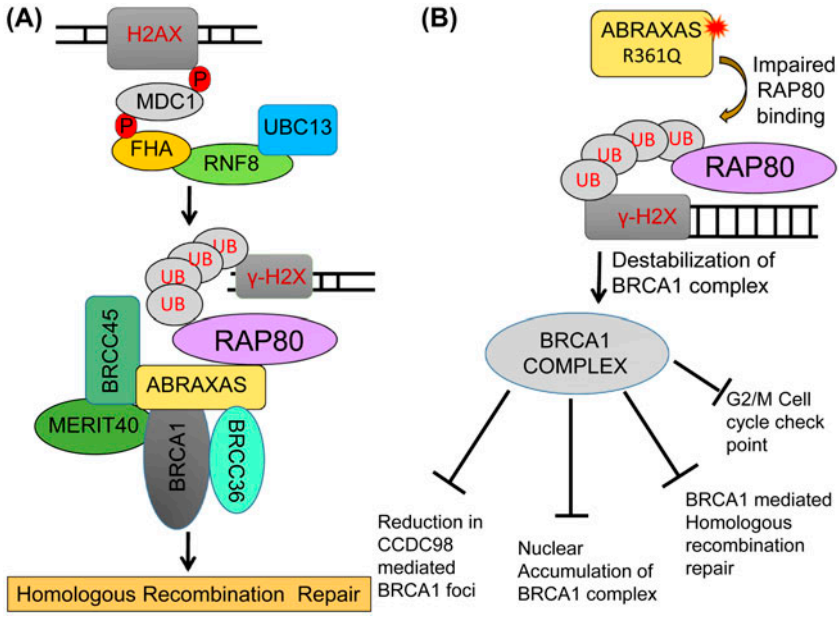


Figure 7. Proposed mechanism of consequence due to ABRAXAS Arg361Gln mutation.
Notes: (A) The figure illustrates the plausible mechanism of defective homologous recombination repair and other consequences due to Arg361Gln mutation. In case of wild-type, after the DNA damage, H2AX becomes phosphorylated by ATM/ATR kinase at nucleosome site which further recruits MDC1, RNF8, and UBC13. This assembly followed by formation of polyubiquitin chain(s) on histone(s) (H2AX) which is further recognized by tandem UIM motifs of RAP80. Since RAP80 is directly associated with ABRAXAS which in turn interacts with other BRCA1-complex members, thereby recruits the entire complex to the DNA damage site. (B) However, in case of Arg361Gln mutation, interaction between ABRAXAS and RAP80 is altered, which leads to defective recruitment of BRCA1-complex. Consequently, nuclear accumulation of BRCA1-complex significantly reduces which further leads to defective G2/M cell-cycle checkpoint, and an error prone homologous recombination repair.

(USA), unless specified. Restriction enzymes were purchased from Fermentas (USA).

3.1. *Gene cloning, protein expression, and purification*

PCR-amplified (Thermo cycler, Bio-Rad) gene product of ABRAXAS (6-373) (kind gift from J. Chen) was purified and digested with Nco1/Xho1. Nco1/Xho1 digested vector pET-28a (Novagen) was ligated with PCR product at 1:3 molecular ratio and transformed into *E. coli* DH5α. Positive clones were screened for insert release followed by confirmation with DNA sequencing. Site-directed mutagenesis was performed on wild-type template using mutagenic primers (Integrated DNA Technology) having c.1082G > A substitution. Amplicons were digested with Dpn1 (Fermentas, USA) and transformed into *E. coli* DH5α bacterial strain. c.1082G > A substitution was confirmed using DNA sequencing. For recombinant protein expression, vector construct was transformed into *E. coli* Rosetta 2 (DE3) bacterial strain (Novagen). Single colony was inoculated and culture was grown till $O.D_{600}$ reached .6–.8, thereafter it was induced with IPTG and grown overnight at 25 °C. Culture was harvested and pellet was re-suspended in 10 mM HEPES buffer containing 300 mM NaCl, 10 mM BME, .1 mM EDTA, pH 7.5 (HNBE buffer). Cells were disrupted by sonication (Branson Sonifier) and soluble fraction was incubated on pre-equilibrated resin (Ni-NTA in case of ABRAXAS, Novagen). Resin was washed thrice with HNBE buffer to remove non-specifically bound impurities. On bead cleavage was performed using TEV protease to remove affinity tag from native protein. Protein of interest was eluted gradientally with 100, 250, and 500 mM of imidazole in HNBE buffer pH 7.5. Protein was further purified on FPLC (AKTA) using size exclusion chromatography (Superdex 200, GE) to remove aggregates, etc. and analyzed on SDS-PAGE for purity.

3.2. *Molecular modeling and docking*

ABRAXAS (351-370) peptide was modeled using homology modeling by considering the template structure from the protein data bank (PDB ID: 3S4T). Authentic models were selected based on overall geometry and stereo-chemistry. The model was validated through protein structure validation server "SAVES" (Metaserver for analyzing and validating protein structures, http://nihserver.mbi.ucla.edu/SAVES/) (Ramachandran et al., 1963;

Downloaded by [Umeå University Library] at 02:41 07 October 2014

Downloaded by [Umeå University Library] at 02:41 07 October 2014

Ramachandran & Sasisekharan, 1968). Modeled structure was simulated using Schrodinger Desmond software. Wild type model was overlaid with the mutant to make a qualitative assumption about the structural changes.

3.3. Protein estimation

Purified protein was quantified with Bradford's protein estimation protocol as per manufacturer instructions (Expedon). Different dilutions of BSA were prepared as a standard reference, and absorbance was recorded using a spectrophotometer (Shimadzu) at $O.D_{595}$. Concentration of protein was determined by interpolation of a standard curve. Experiment was performed in triplicates (Bradford, 1976; Noble & Bailey, 2009).

3.4. Limited proteolysis

Wild type (1 mg/ml) and mutant (1 mg/ml) proteins were digested with trypsin (37 °C) and chymotrypsin (25 °C), such that the final concentration of proteases was 20 and 10 ρg/μl in solution, respectively. Reaction was terminated at the different time point (0, 5, 10, 30, 60, 180, 240 min) by addition of 1 mM PMSF. Reaction product was mixed with equal volume of Laemmli buffer and heated prior to loading on SDS-PAGE. The experiment was repeated thrice by taking untreated wild type and mutant as controls (Havliš, Thomas, Šebela, & Shevchenko, 2003; Jimenez, Huang, Qiu, & Burlingame, 1998, chapter 16).

3.5. Fluorescence spectroscopy

Trp and Tyr residues' position in wild type and mutant was monitored using a Fluorescence spectrophotometer (Horiba, USA). Fluorescence emission scans were recorded from 310 to 400 nm wavelength range at a excitation wavelength of 280 nm, and the temperature 15 °C. For thermal denaturation, 2 μM protein was unfolded in a temperature range of 10–75 °C and emission spectra were acquired. Averaged blank corrected data were considered, and were analyzed by fitting in a three-state transition model to calculate thermal parameters (Pace, 1986).

For chemical denaturation, 10 M urea (Sigma–Aldrich) stock was prepared in 2.5 mM HEPES buffer containing 50 mM NaCl pH 7.5, and protein dilutions were prepared as described previously (Walters et al., 2009). Emission scans were collected and were further blank corrected. Thermodynamic parameters were obtained by curve fitting as per three-state transition model. The experiment was repeated three times, independently.

3.6. CD spectroscopy

CD polarimeter (Jasco J-810, Japan) was used to acquired Far-UV CD spectrum of purified proteins (10 μM) in a wavelength range 200–240 nm at 10 °C (buffer 2.5 mM HEPES pH 7.5, 50 mM NaCl). The average blank corrected data of three independent scans were considered, and Dichroweb server (http://dichroweb.cryst.bbk.ac.uk) (Lobley et al., 2002; Sreerama, Venyaminov, & Woody, 2000; Stephens, McKenna, McKenna, Nguyen & Devlin, 1981; Whitmore & Wallace, 2004, 2008) was used for estimation of secondary structure composition. For thermal denaturation, wild-type and mutant (10 μM) proteins were unfolded in a temperature range from 10 to 75 °C at 222 nm wavelength. Unfolding profile was obtained and curve fitting was performed. The experiment was replicated thrice.

3.7. Thermal and chemical denaturation data analysis and curve fitting

Thermal parameters were calculated by considering native (N) to intermediate (I) and intermediate to unfolded (U) transitions during unfolding process. Data analysis, curve fitting, and thermodynamic parameters determination were performed as described earlier (Pace & Shaw, 2000; Vikrant, Nakhwa et al., 2013).

3.8. Differential scanning calorimetry

Unfolding transition and enthalpy of reaction were determined using DSC (Setaram μDSC3 evo, USA). Samples were filtered and degassed prior to scanning. A total of 2 mg protein, wild type and mutant, was allowed to unfold from 10 to 75 °C temperature range with a temperature increment rate of 1 °C/minute. Data fitting was done using "CALISTO" software and enthalpy value was obtained. To determine the thermodynamic reversibility, samples were heated just above the transition maximum, cooled instantaneously, and reheated. The experiment was repeated thrice, independently.

3.9. Pull down assay

ABRAXAS (6-373) wild type and mutant were re-suspended in HNBEEG buffer (HNBE buffer containing 5% ethylene glycol and 5% glycerol) and lysed by sonication as mentioned above in section 1. Soluble His-tagged fusion protein ABRAXAS (6-373) wild type and mutant were allowed to bind on Ni-NTA resin (Novagen) and thereafter washed to remove impurities. Equal concentration of bound fusion protein (.5 mg/ml) was used as bait and incubated with purified RAP80 (1-405) (prey protein, 50 μM concentration) for 5 h at 4 °C. Resin was washed with same buffer and loaded on SDS-PAGE. Protein complex was transferred over PVDF membrane

(Millipore) and probed with anti-RAP80 antibody (Abcam). Experiment was repeated thrice by taking heat denatured RAP80 (1-405) as control.

3.10. Dynamic light scattering

Molecular size measurement was done using a Malvern particle size analyzer (Zetasizer μV). Samples were filtered (.44 μm) and degassed prior to scanning. Two milligram per milliliter of protein was scanned at a 5-min interval for 15 min, and effective diameter of peak of interest was considered. The experiment was repeated in three independent sets.

Supplementary material

The supplementary material for this paper is available online at http://dx.doi.10.1080/07391102.2014.945484.

Acknowledgements

We thank DBT-BTIS facility at ACTREC for providing necessary software to this study. Mr Vikrant thanks CSIR for fellowship (09/513 (0072)/2008-EMR-I).

Funding

Funding for this study was supported by DBT [grant number BT/PR10765/BRB/664/2008], [grant number BT/PR12565/BID/07/303/2009]; Seed in Air grant from TMC.

References

Aygun, O., Svejstrup, J., & Liu, Y. (2008). A RECQ5-RNA polymerase II association identified by targeted proteomic analysis of human chromatin. *Proceedings of the National Academy of Sciences, 105*, 8580–8584.

Badgujar, D. C., Sawant, U., Yadav, L., Hosur, M. V., & Varma, A. K. (2013). Preliminary crystallographic studies of BRCA1 BRCT-ABRAXAS complex. *Acta Crystallographica Section F Structural Biology and Crystallization Communications, 69*, 1401–1404.

Bartek, J., & Lukas, J. (2007). DNA damage checkpoints: From initiation to recovery or adaptation. *Current Opinion in Cell Biology, 19*, 238–245.

Bartek, J., Bartkova, J., & Lukas, J. (2007). DNA damage signalling guards against activated oncogenes and tumour progression. *Oncogene, 26*, 7773–7779.

Bartkova, J., Horejsi, Z., Sehested, M., Nesland, J. M., Rajpert-De Meyts, E., Skakkebaek, N. E., … Bartek, J. (2007). DNA damage response mediators MDC1 and 53BP1: Constitutive activation and aberrant loss in breast and lung cancer, but not in testicular germ cell tumours. *Oncogene, 26*, 7414–7422.

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry, 72*, 248–254.

Burma, S., Chen, B. P., Murphy, M., Kurimasa, A., & Chen, D. J. (2001). ATM phosphorylates histone H2AX in response to DNA double-strand breaks. *Journal of Biological Chemistry, 276*, 42462–42467.

Celeste, A., Fernandez-Capetillo, O., Kruhlak, M. J., Pilch, D. R., Staudt, D. W., Lee, A., … Nussenzweig, A. (2003). Histone H2AX phosphorylation is dispensable for the initial recognition of DNA breaks. *Nature Cell Biology, 5*, 675–679.

Coquelle, N., Green, R., & Glover, J. N. (2011). Impact of BRCA1 BRCT domain missense substitutions on phosphopeptide recognition. *Biochemistry, 50*, 4579–4589.

Fontana, A., de Laureto, P. P., Spolaore, B., & Frare, E. (2012). Identifying disordered regions in proteins by limited proteolysis. *Methods in Molecular Biology, 896*, 297–318.

Havliš, J., Thomas, H., Šebela, M., & Shevchenko, A. (2003). Fast-response proteomics by accelerated in-gel digestion of proteins. *Analytical Chemistry, 75*, 1300–1306.

Huen, M. S., Grant, R., Manke, I., Minn, K., Yu, X., Yaffe, M. B., & Chen, J. (2007). RNF8 transduces the DNA-damage signal via histone ubiquitylation and checkpoint protein assembly. *Cell, 131*, 901–914.

Jimenez, C. R., Huang, L., Qiu, Y., & Burlingame, A. L. (1998). In-gel digestion of proteins for MALDI-MS fingerprint mapping. *Current Protocols in Protein Science*, Chapter 16, Unit 16.4, 16.4.1–16.4.5.

Joerger, M., deJong, D., Burylo, A., Burgers, J. A., Baas, P., Huitema, A. D., … Schellens, J. H. (2011). Tubulin, BRCA1, ERCC1, Abraxas, RAP80 mRNA expression, p53/p21 immunohistochemistry and clinical outcome in patients with advanced non small-cell lung cancer receiving first-line platinum-gemcitabine chemotherapy. *Lung Cancer, 74*, 310–317.

Kim, D. E., Chivian, D., & Baker, D. (2004). Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Research, 32*, W526–531.

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

Kim, H., Huang, J., & Chen, J. (2007). CCDC98 is a BRCA1-BRCT domain-binding protein involved in the DNA damage response. *Nature Structural and Molecular Biology, 14*, 710–715.

Kolas, N. K., Chapman, J. R., Nakada, S., Ylanko, J., Chahwan, R., Sweeney, F. D., … Durocher, D. (2007). Orchestration of the DNA-damage response by the RNF8 ubiquitin ligase. *Science, 318*, 1637–1640.

Kolchanov, N. A., Soloviov, V. V., & Zharkikh, A. A. (1983). The effects of mutations, deletions and insertions of single amino acids on the three-dimensional structure of globins. *FEBS Letters, 161*, 65–70.

Leaver-Fay, A., Tyka, M., Lewis, S. M., Lange, O. F., Thompson, J., Jacak, R., … Bradley, P. (2011). ROSETTA3: An object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymology, 487*, 545–574.

Liu, Z., Wu, J., & Yu, X. (2007). CCDC98 targets BRCA1 to DNA damage sites. *Nature Structural and Molecular Biology, 14*, 716–720.

Lobley, A., Whitmore, L., & Wallace, B. A. (2002). DICHROWEB: An interactive website for the analysis of protein secondary structure from circular dichroism spectra. *Bioinformatics, 18*, 211–212.

Lou, Z., Minter-Dykhouse, K., Wu, X., & Chen, J. (2003). MDC1 is coupled to activated CHK2 in mammalian DNA damage response pathways. *Nature, 421*, 957–961.

Mailand, N., Bekker-Jensen, S., Faustrup, H., Melander, F., Bartek, J., Lukas, C., & Lukas, J. (2007). RNF8 ubiquitylates histones at DNA double-strand breaks and promotes assembly of repair proteins. *Cell, 131*, 887–900.

Muñiz, V. A., Srinivasan, S., Boswell, S. A., Meinhold, D. W., Childs, T., Osuna, R., & Colón, W. (2011). The role of the

Downloaded by [Umeå University Library] at 02:41 07 October 2014

Downloaded by [Umeå University Library] at 02:41 07 October 2014

local environment of engineered Tyr to Trp substitutions for probing the denaturation mechanism of FIS. *Protein Science, 20*, 302–312.

Nikkilä, J., Coleman, K. A., Morrissey, D., Pylkäs, K., Erkko, H., Messick, T. E., … Greenberg, R. A. (2009). Familial breast cancer screening reveals an alteration in the RAP80 UIM domain that impairs DNA damage response function. *Oncogene, 28*, 1843–1852.

Noble, J. E., & Bailey, M. J. (2009). Quantitation of protein. *Methods Enzymology, 463*, 73–95.

Osorio, A., Barroso, A., García, M. J., Martínez-Delgado, B., Urioste, M., & Benítez, J. (2009). Evaluation of the BRCA1 interacting genes RAP80 and CCDC98 in familial breast cancer susceptibility. *Breast Cancer Research and Treatment, 113*, 371–376.

Pace, C. N. (1986). Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods Enzymology, 131*, 266–280.

Pace, C. N., & Shaw, K. L. (2000). Linear extrapolation method of analyzing solvent denaturation curves. *Proteins: Structure, Function, and Genetics, 41*, 1–7.

Paull, T. T., Rogakou, E. P., Yamazaki, V., Kirchgessner, C. U., Gellert, M., & Bonner, W. M. (2000). A critical role for histone H2AX in recruitment of repair factors to nuclear foci after DNA damage. *Current Biology, 10*, 886–895.

Pylkäs, K., Erkko, H., Nikkilä, J., Sólyom, S., & Winqvist, R. (2008). Analysis of large deletions in BRCA1, BRCA2 and PALB2 genes in Finnish breast and ovarian cancer families. *BMC Cancer, 8*, 146.

Ramachandran, G. N., Ramakrishnan, C., & Sasisekharan, V. (1963). Stereochemistry of polypeptide chain configurations. *Journal of Molecular Biology, 7*, 95–99.

Ramachandran, G. N., & Sasisekharan, V. (1968). Conformation of polypeptides and proteins. *Advances in Protein Chemistry, 23*, 283–437.

Rogakou, E. P., Pilch, D. R., Orr, A. H., Ivanova, V. S., & Bonner, W. M. (1998). DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. *Journal of Biological Chemistry, 273*, 5858–5868.

Rouse, J., & Jackson, S. P. (2002). Interfaces between the detection, signaling, and repair of DNA damage. *Science, 297*, 547–551.

Sartori, A. A., Lukas, C., Coates, J., Mistrik, M., Fu, S., Bartek, J., … Jackson, S. P. (2007). Human CtIP promotes DNA end resection. *Nature, 450*, 509–514.

Sato, Y., Yoshikawa, A., Mimura, H., Yamashita, M., Yamagata, A., & Fukai, S. (2009). Structural basis for specific recognition of Lys 63-linked polyubiquitin chains by tandem UIMs of RAP80. *The EMBO Journal, 28*, 2461–2468.

Shiotani, B., & Zou, L. (2009a). ATR signaling at a glance. *Journal of Cell Science, 122*, 301–304.

Shiotani, B., & Zou, L. (2009b). Single-stranded DNA orchestrates an ATM-to-ATR switch at DNA breaks. *Molecular Cell, 33*, 547–558.

Sobhian, B., Shao, G., Lilli, D. R., Culhane, A. C., Moreau, L. A., Xia, B., … Greenberg, R. A. (2007). RAP80 targets BRCA1 to specific ubiquitin structures at DNA damage sites. *Science, 316*, 1198–1202.

Solyom, S., Aressy, B., Pylkas, K., Patterson-Fortin, J., Hartikainen, J. M., Kallioniemi, A., … Winqvist, R. (2012). Breast cancer-associated Abraxas mutation disrupts nuclear localization and DNA damage response functions. *Science Translational Medicine, 4*, 122ra123.

Sreerama, N., Venyaminov, S. Y., & Woody, R. W. (2000). Estimation of protein secondary structure from circular dichroism spectra: Inclusion of denatured proteins with native proteins in the analysis. *Analytical Biochemistry, 287*, 243–251.

Stephens, P. J., McKenna, C. E., McKenna, M. C., Nguyen, H. T., & Devlin, F. (1981). Circular dichroism and magnetic circular dichroism of reduced molybdenum-iron protein of *Azotobacter vinelandii* nitrogenase. *Biochemistry, 20*, 2857–2864.

Vikrant, Kumar, R., Yadav, L. R., Nakhwa, P., Waghmare, S. K., Goyal, P., & Varma, A. K. (2013). Structural and functional implication of RAP80 ΔGlu81 mutation. *PLoS One, 8*, e72707.

Vikrant, Nakhwa, P., Badgujar, D. C., Kumar, R., Rathore, K. K., & Varma, A. K. (2013). Structural and functional characterization of the MERIT40 to understand its role in DNA repair. *Journal of Biomolecular Structure and Dynamics,* 1–16. doi:10.1080/07391102.2013.843473

Vikrant, Sawant, U. U., & Varma, A. K. (2014). Role of MERIT40 in stabilization of BRCA1 complex: A protein–protein interaction study. *Biochemical and Biophysical Research Communications, 446*, 1139–1144.

Walters, J., Milam, S. L., & Clark, A. C. (2009). Practical approaches to protein folding and assembly: Spectroscopic strategies in thermodynamics and kinetics. *Methods in Enzymology, 455*, 1–39.

Wang, B., & Elledge, S. J. (2007). Ubc13/Rnf8 ubiquitin ligases control foci formation of the Rap80/Abraxas/Brca1/Brcc36 complex in response to DNA damage. *Proceedings of the National Academy of Sciences, 104*, 20759–20763.

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.

Wang, W., & Kirschner, M. W. (2013). Emi1 preferentially inhibits ubiquitin chain elongation by the anaphase-promoting complex. *Nature Cell Biology, 15*, 797–806.

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 and Development, 14*, 927–939.

Whitmore, L., & Wallace, B. A. (2004). DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Research, 32*, W668–W673.

Whitmore, L., & Wallace, B. A. (2008). Protein secondary structure analyses from circular dichroism spectroscopy: Methods and reference databases. *Biopolymers, 89*, 392–400.

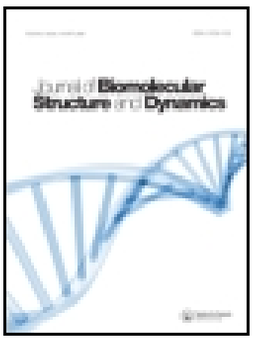
Williams, R. S., & Glover, J. N. (2003). Structural consequences of a cancer-causing BRCA1-BRCT missense mutation. *Journal of Biological Chemistry, 278*, 2630–2635.

Zarrine-Afsar, A., Larson, S. M., & Davidson, A. R. (2005). The family feud: Do proteins with similar structures fold via the same pathway? *Current Opinion in Structural Biology, 15*, 42–49.

Zerovnik, E. (2011). Oligomerization preceding amyloid fibril formation: A process in common to intrinsically disordered and globular proteins. *Network, 22*, 154–161.

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

Zhukov, I., Jaroszewski, L., & Bierzynski, A. (2000). Conservative mutation Met8 → Leu affects the folding process and structural stability of squash trypsin inhibitor CMTI-I. *Protein Science, 9*, 273–279.



Taylor & Francis
Taylor & Francis Group

Journal of Biomolecular Structure and Dynamics

ISSN: 0739-1102 (Print) 1538-0254 (Online) Journal homepage: http://www.tandfonline.com/loi/tbsd20

Multimodal approach to explore the pathogenicity of BARD1, ARG 658 CYS, and ILE 738 VAL mutants

Rajan Kumar Choudhary, Vikrant, Quadir M. Siddiqui, Pankaj S. Thapa, Sweta Raikundalia, Nikhil Gadewal, Nachimuthu Senthil Kumar, M.V. Hosur & Ashok K. Varma

To cite this article: Rajan Kumar Choudhary, Vikrant, Quadir M. Siddiqui, Pankaj S. Thapa, Sweta Raikundalia, Nikhil Gadewal, Nachimuthu Senthil Kumar, M.V. Hosur & Ashok K. Varma (2015): Multimodal approach to explore the pathogenicity of BARD1, ARG 658 CYS, and ILE 738 VAL mutants, Journal of Biomolecular Structure and Dynamics, DOI: 10.1080/07391102.2015.1082149

To link to this article: http://dx.doi.org/10.1080/07391102.2015.1082149

View supplementary material

Accepted online: 26 Aug 2015.Published online: 23 Sep 2015.

Submit your article to this journal

Article views: 10

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
http://www.tandfonline.com/action/journalInformation?journalCode=tbsd20

Download by: [14.139.127.177] **Date:** 25 September 2015, At: 23:12

Journal of Biomolecular Structure and Dynamics, 2015
http://dx.doi.org/10.1080/07391102.2015.1082149

Taylor & Francis
Taylor & Francis Group

Multimodal approach to explore the pathogenicity of BARD1, ARG 658 CYS, and ILE 738 VAL mutants

Rajan Kumar Choudhary[a], Vikrant[a,†], Quadir M. Siddiqui[a], Pankaj S. Thapa[a], Sweta Raikundalia[a], Nikhil Gadewal[a], Nachimuthu Senthil Kumar[b], M.V. Hosur[a] and Ashok K. Varma[a]*

[a]*Advanced Centre for Treatment, Research and Education in Cancer, Kharghar, Navi Mumbai 410 210, Maharashtra, India;*
[b]*Department of Biotechnology, Mizoram University (A Central University), Aizawl 796 004, Mizoram, India*

Communicated by Ramaswamy H. Sarma

(*Received 28 May 2015; accepted 7 August 2015*)

BARD1–BRCA1 complex plays an important role in DNA damage repair, apoptosis, chromatin remodeling, and other important processes required for cell survival. BRCA1 and BARD1 heterodimer possess E3 ligase activity and is involved in genome maintenance, by functioning in surveillance for DNA damage, thereby regulating multiple pathways including tumor suppression. BRCT domains are evolutionary conserved domains present in different proteins such as BRCA1, BARD1, XRCC, and MDC1 regulating damage response and cell-cycle control through protein–protein interactions. Nonetheless, the role of BARD1BRCT in the recruitment of DNA repair mechanism and structural integrity with BRCA1 complex is still implicit. To explicate the role of BARD1BRCT in the DNA repair mechanism, in silico, *in vitro*, and biophysical approach were applied to characterize BARD1 BRCT *wild-type* and Arg658Cys and Ile738Val mutants. However, no drastic secondary and tertiary structural changes in the mutant proteins were observed. Thermal and chemical denaturation studies revealed that mutants Arg658Cys and Ile738Val have a decrease in T_m and ΔG than the *wild type*. *In silico* studies of BARD1 BRCT (568-777) and mutant protein indicate loss in structural compactness on the Ile738Val mutant. Comparative studies of *wild-type* and mutants will thus be helpful in understanding the basic role of BARD1BRCT in DNA damage repair.

Keywords: BARD1 BRCT; protein–protein interactions; secondary and tertiary structure; thermal and chemical denaturation

Downloaded by [14.139.127.177] at 23:12 25 September 2015

1. Introduction

Breast cancer is one of the prominent causes for cancer deaths in the developing countries. It has been found that mutations in the *BRCA1* and *BRCA2* genes can predispose someone at the risk of developing breast and ovarian cancer (Collins et al., 1995; Miki et al., 1994; Wooster et al., 1994). The inherited germ line mutations in the genes like *RAD50, ATM, CHEK2, NBS1, 53BP1, PALB2, BRIP1 BRCA 1/2*, and *BARD1* regulate genomic integrityand cancer predisposing factors (Ahmed & Rahman, 2006; Heikkinen et al., 2006; Karppinen et al., 2006; Meijers-Heijboer et al., 2002; Rapakko, Heikkinen, Karppinen, Erkko, & Winqvist, 2007; Renwick et al., 2006). BARD1 is a structurally correlated but sequentially distinct protein from BRCA1. BARD1 and BRCA1 both harbor N-terminal RING domain and two C-terminal tandem BRCT repeat motifs (Wu et al., 1996). BARD1 contains four ankyrin repeats in the region (425–550), which is absent in BRCA1. BRCA1 and BARD1 heterodimerize by RING–RING domain interaction and colocalize to distinct nuclear assemblies at different stages of cell-cycle progression (Jin et al., 1997; Scully et al., 1997). Furthermore, *in vivo* BRCA1–BARD1 heterodimer forms active E3 ubiquitin ligase complex, and plays a significant role in tumor suppression (Brzovic, Meza, King, & Klevit, 2001; Brzovic, Rajagopal, Hoyt, King, & Klevit, 2001; Wu-Baer, Lagrazon, Yuan, & Baer, 2003). Mis-sense mutations within the BRCA1 RING domain lead to loss in heterodimerization which further reduces E3 ubiquitin ligase activity (Brzovic, Meza, et al., 2001). BARD1 BRCT (568-777) domain is conserved in many other multiple DNA damage response (DDR) proteins (Bork et al., 1997; Callebaut & Mornon, 1997).

Different DDR proteins have diverse repeats of BRCTs (Mohammad & Yaffe, 2009). It has been reported that BARD1 BRCT is required for the early recruitment of BRCA1 at the DNA damage site (Li & Yu, 2013). Noting the role of BARD1 in genomic integrity, we decided to explore functional consequences of genetic alterations discovered in the BARD1 BRCT region. Two cancer predisposing mutations, Arg658Cys in Caucasian, African, Finnish populations, and Ile738Val in Polish and

*Corresponding author. Email: avarma@actrec.gov.in
†Current address: Quintiles Research Pvt. Ltd. Andheri, Mumbai-400059, India.

© 2015 Taylor & Francis

Downloaded by [14.139.127.177] at 23:12 25 September 2015

Belgian families have been detected (Alshatwi, Hasan, Syed, Shafi, & Grace, 2012; Antoniou et al., 2010; De Brakeleer et al., 2010; Karppinen, Heikkinen, Rapakko, & Winqvist, 2004; Thai et al., 1998), but the molecular mechanism of how these mutations of BARD1 gene lead to breast cancer has not been explored. Further, in sporadic and BRCA1-associated breast cancers, expression of BARD1 is observed to be too low suggesting a significant role of BARD1 in breast carcinogenesis (Ghimenti et al., 2002; Yoshikawa et al., 2000).

In this study, we have used *in vitro* experiments and *in silico* approaches to understand the folding pattern of BARD1 BRCT (568-777) and its two cancer-disposing mutants, Arg658Cys and Ile738Val. Molecular dynamics simulations (MDS) were carried out to explore alterations in the structure at the atomic level. To our conclusion, both the mutants show loss in thermodynamic stability; however, BARD1 Ile738Val protein structure shows higher structural flexibility.

2. Material and methods

2.1. Molecular dynamics simulation

The crystal structure of BARD1 BRCT (568-777) *wild-type* protein was taken from PDB (pdb id: 2NTE), (Birrane, Varma, Soni, & Ladias, 2007). BARD1 BRCT (568-777) amino acid sequence was retrieved from the Uniprot database (Accession ID: q99728). Point mutants, Arg658Cys and Ile738Val, were introduced using SPDB viewer (Kaplan & Littlejohn, 2001). MDS was performed using GROMACS 4.5.5 package (Hess, Kutzner, van der Spoel, & Lindahl, 2008) installed on a 80-node cluster of Xeon quad-core processors. The molecular systems were solvated with TIP3P water molecules in a cubic box with periodic boundary conditions of 1.0 nm from the edge of the box. Counter ions were added to neutralize the systems. The systems were subjected to energy minimization for 5000 iterations by steepest descent algorithm implementing under OPLS-AA force field (Kaminski, Friesner, Tirado-Rives, & Jorgensen, 2001) with a tolerance of 1000 kJ/mol/nm. Electrostatic, columbic, and van der Waals interactions were calculated with a distance cut-off of 1.4 nm. Systems were equilibrated using NVT (Moles Volume Temperature) followed by NPT (Moles Pressure Temperature) ensemble for 50,000 steps each. SHAKE algorithm (Ryckaert, Ciccotti, & Berendsen, 1977) was used to constrain bond lengths. Finally, systems were subjected to MD simulation for 50 ns of production run with time-step integration of 2 fs. The trajectories were saved at every 2 ps and analyzed using GROMACS. The comparative analysis of BARD1 BRCT (568-777) *wild-type* and mutant structures, Arg658Cys, Ile738 Val, were performed to calculate the RMSD, RMSF, radius of gyration, solvent accessible surface area (SASA), Dictionary of secondary structure pattern (DSSP) (Kabsch & Sander, 1983), and hydrogen-bond interactions.

2.2. Principal component analysis

Principal component (PC) was calculated using eigenvectors and eigenvalues for the covariance matrix. Projection of first two principal components was carried out by essential dynamics method described within the GROMACS package (Amadei, Linssen, & Berendsen, 1993). After removing the rotational and translational movements, variance/covariance matrix was built and further eigenvectors and eigenvalues were calculated. The eigenvalues are representative of amplitude of eigenvectors in multidimensional space. The concerted motion of the protein can be detected by the displacement of atoms along each eigenvector.

2.3. Protein expression and purification

BARD1 BRCT (568-777) *wild-type* protein was cloned in pET-28a vector (cDNA of full-length BARD1 was kind gift from Prof. Richard Bayer, Institute of Cancer Genetics, Columbia University). Mismatch primer method of site-directed mutagenesis was used to generate the Arg658Cys and Ile738Val mutants (forward primer Arg 658Cys 5′-GAAATTCCTGAAGGTCCATGCAGAAGC-AGGCTCAACAG-3′, reverse primer 5′-CTGTTGAGCC-TGCTTCTGCATGGACCTTCAGGAATTTC-3′, forward primer Ile738Val 5′-CTGCACACAGTATATCGTCTAT-GAAGATTTGTGT-3′ reverse primer 5′-ACACAAATC-TTCATAGACGATATACTGTGTGCAG-3′). The BARD1 BRCT (568-777) *wild-type*, Arg658Cys, and Ile738V mutant proteins were expressed and purified in 6His fusion proteins using Rosetta 2(DE3) bacterial cells. Fusion proteins were first purified on Ni-NTA affinity chromatography, then further treated with TEV to remove the fusion tag. Protein was purified using AKTA FPLC size exclusion chromatography on superdex 75 column (GE Healthcare).

2.4. Dynamic light scattering experiment

Purified BARD1 BRCT (568-777), Arg658Cys, and Ile738Val mutant proteins were analyzed on Malvern particle size analyzer (Zetasizer μV) dynamic light scattering instrument to characterize the oligomeric nature of these proteins. The protein samples were filtered (0.22-μm filter) and degassed prior to scanning. 1.0 mg/ml of *wild-type* and mutant proteins were scanned at 5-min intervals for 15 min, and effective diameters were obtained from peak positions. The experiments were repeated three times and the average was considered to be the true value.

Downloaded by [14.139.127.177] at 23:12 25 September 2015

2.5. Circular dichroism spectroscopy

Circular dichroism spectroscopy was performed using different concentrations of purified proteins on the JASCO-815 spectropolarimeter. For far–UV CD spectra, the protein concentration was 10 µM and the wavelength range was taken from λ = 200–240 nm, and for near-UV spectra, the wavelength range was from λ = 350–250 nm with the protein concentration of 40 µM (buffer: 2.5 mM HEPES pH 7.5, 50 mM NaCl). Blank-corrected seven independent data scans were averaged, and the average value is presented for further comparative structural analysis. Furthermore, proteins at a concentration of 10 µM were unfolded in a temperature range from 10 to 65°C, and folding pattern was monitored by measuring ellipticity at λ = 222 nm. Near-UV range, thermal denaturation studies were performed for 40 µM of proteins at different temperatures 20, 30, 45, and 60°C. The observed values of fractions unfolded were used for curve fitting.

2.6. Limited proteolysis

BARD1 BRCT *wild-type* and mutant proteins (2 mg/ml) were treated with trypsin (37°C) and chymotrypsin (25°C), 10 ρg/µl each in the digestion cocktail. At the time points of 0, 5, 10, 30, 60, 120, and 180 min, the reaction was stopped by adding protease inhibitor PMSF (1 mM). The experiment was repeated thrice considering untreated *wild-type* and mutant samples as controls.

2.7. Fluorescence spectroscopy

2.7.1. Chemical denaturation

Micro-environment of tryptophan residue in the proteins were investigated using fluorescence spectrophotometer (Horiba, USA).While the excitation was at λ = 295 nm, the individual fluorescence emission spectra were recorded from λ = 310–400 nm. In chemical denaturation, BARD1 BRCT proteins (2 µM) were treated with increasing concentrations of GuHCl (0–6 M) keeping the final volume constant, and emission spectra were recorded between λ = 310 and 400 nm. Averaged blank-corrected data were analyzed by fitting to a three-state transition model (Pace, 1986; Pace & Shaw, 2000; Vikrant et al., 2013).

2.7.2. Data analysis and curve fitting

Thermodynamic parameters were determined during the GuHcl unfolding process by detecting the folded (*N*) to intermediate (*I*) and intermediate to unfolded (*U*) transitions of the curve. Further, thermodynamic constraints analysis and curve fitting were performed (Pace & Shaw, 2000; Vikrant, Nakhwa, Badgujar, Kumar, & Rathore, 2014).

2.8. Chemical cross-linking assay

BARD1 BRCT (568-777) *wild-type* and mutant proteins were incubated with 0.1% glutaraldehyde, and the reaction was terminated by addition of 5 µl of 1 M Tris pH-8.0. Reaction was stopped in a time-dependent manner (0, 2.5, 5, 10, 15, 30, and 60 min, respectively) and analyzed on 12% SDS–PAGE. Untreated protein sample was taken as control.

3. Results and discussion

Deleterious non-synonymous single nucleotide polymorphisms (nsSNPs) are involved in inducing disease-associated phenomena, because of structural alterations and consequent loss of molecular functions. The modern bioinformatics-based advanced methods have now enabled us to determine the structural effects due to deleterious nsSNPs in the target candidate genes. In the current study, multidisciplinary, *in vitro,* biophysical, and *in silico* approaches were employed to understand the pathogenicity of BARD1 BRCT, Arg658Cys, and Ile738Val mutants (Figure 1(F)).

3.1. Folding pattern of BARD1 BRCT (568-777) **wild-type** *and mutants*

BARD1 BRCT, Arg658Cys, and Ile738Val mutations have been discovered from the affected families of breast and ovarian cancer. Considering the importance of mutations in predisposing the cancer risk, we decided to characterized abrogative effect of mutations on the structure and function of the protein. It has been well documented that different classes of proteins with exposed hydrophobic residues become highly prone to non-specific interactions (Kolchanov, Soloviov, & Zharkikh, 1983; Zhukov, Jaroszewski, & Bierzynski, 2000), which in turn lead to the formation of high molecular weight aggregates. The residual Arg658Cys mutation is the substitution of a basic amino acid to a small non-polar residue, cysteine. However, Ile738Val substitution is a change from a large to small hydrophobic residue. Ile to Val mutation leads to removal of –CH_2–group from Ile, and potentially create a void volume of 25 $Å^3$ in the mutant protein (Harpaz, Gerstein, & Chothia, 1994). Such mutations may lead to the loss of hydrophobic interactions and van der Waals interactions among the substituted and neighboring residues. To explore this prospect, we have used gel filtration chromatography (AKTA FPLC, superdex 75 column) to probe the oligomeric status of the mutants Arg658Cys and Ile738Val. The non-synonymous amino acid substitution in protein tends to show differential solubility and elution profiles in gel filtration chromatography. There was no substantial alteration in solubility and elution profiles of Arg658Cys and

Downloaded by [14.139.127.177] at 23:12 25 September 2015

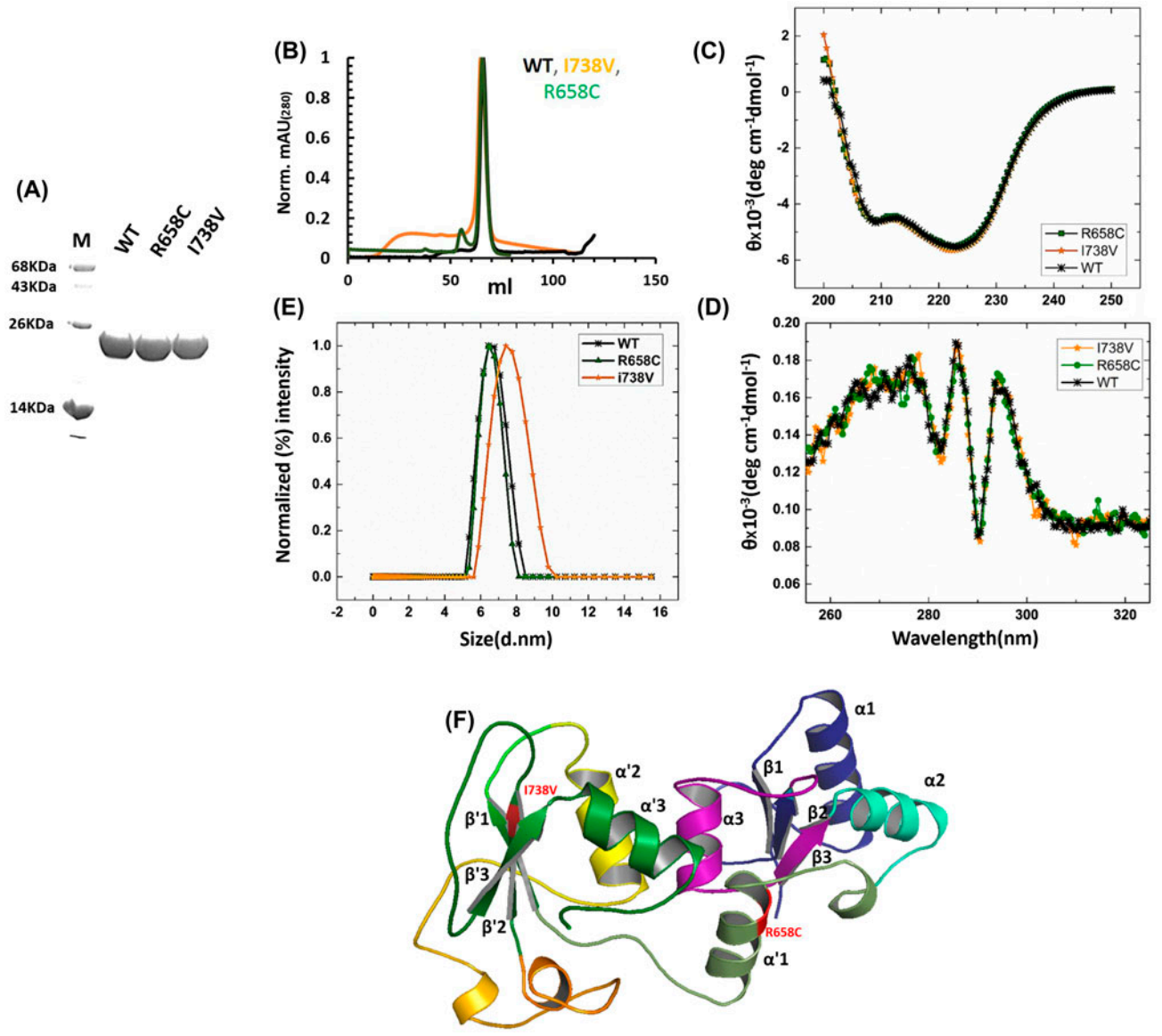


Figure 1. Expression and purification profile of BARD1 BRCT (568-777) *wild-type*, Arg658Cys, and Ile738Val proteins.
Notes: (A) Purified protein after gel filtration chromatography on SDS–PAGE. FPLC purified protein were heated with Laemmli buffer and loaded on SDS–PAGE BARD1 BRCT (568-777) *wild-type* (lane 1) , Arg658Cys (lane 2), and Ile738Val (lane 3). (B) Comparative gel filtration spectra of BARD1 BRCT (568-777) wild-type, Arg658Cys, and Ile738Val (Superdex-75, GE). Elution profile of BARD1 BRCT (568-777) *wild-type* Arg658Cys, and Ile738Val were similar and showed the presence of single peak correspond to monomer. (C) BARD1 BRCT (568-777), Arg658Cys, and Ile738Val have well-defined α/β characteristics and mutants does not show any drastic alteration in the secondary structure. (D) Overlay of near-UV spectra of BARD1 BRCT (568-777) *wild-type*, Arg658Cys, and Ile738Val mutant proteins indicating well compact overall structure. (E) Dynamic light scattering profile of BARD1 BRCT (568-777) *wild-type*, Arg658Cys, and Ile738Val showed monomeric population of all the proteins. (F) Crystal structure of BARD1 BRCT (PDB ID: 2NTE) with the sites for cancer-predisposing mutations. Mutant residues are indicated in red.

Ile738Val mutant. In size exclusion chromatography, *wild-type* as well as mutants elute at the same position in superdex 75 column (GE) (Figure 1(A) and (B)). Hence, it is concluded that Arg658Cys and Ile 738Val mutations are not affecting the monomeric nature of BARD1 BRCT (568-777) *wild-type* protein in solution. This monomeric behavior of BARD1 BRCT (568-777) *wild-type*, Arg658Cys, and Ile 738Val mutant proteins were further supported by chemical cross-linking experiments (Supplementary Figure 1A). No higher molecular aggregate formations were observed using cross-linking assay of BARD1 BRCT (568-777) *wild-type*, Arg658Cys, and Ile738Val proteins. Furthermore, dynamic light scattering of BARD1 BRCT (568-777) *wild- type*, Arg658Cys, and

Downloaded by [14.139.127.177] at 23:12 25 September 2015

Ile738Val mutant samples also indicated their monomeric nature. Albeit, BARD1 BRCT (568-777) *wild- type* and Arg658Cys did not show any change in hydrodynamic size, a slight increment was observed in Ile738Val mutant protein (Figure 1(E)). The hydrodynamic diameter of BARD1 BRCT (568-777) *wild- type* and Arg658Cys proteins were found to be 6.44 ± 0.45 nm, whereas Ile738Val mutant protein showed a diameter of 7.40 ± .23 nm. The relative increase in effective hydrodynamic size indicates molecular expansion in Ile738Val mutant. Relative increase in the hydrodynamic radii in Ile738Val mutant may be due to the void volume generated in the hydrophobic core of the protein. This may further increase the conformational entropy of neighboring residues making mutant protein structurally more flexible and leading to the increase in effective hydrodynamic diameter of the protein.

3.2. *Structural organization*

Far-UV CD spectra show predominant α-helix and β-strand characteristic, and no significant differences were observed in the mean residual ellipticities of *wild-type* and mutant proteins (Figure 1(C)). Near-UV CD spectra were obtained to dissect the effect of mutations on overall packing of the protein. Comparative near-UV spectra of BARD1 BRCT (568-777) *wild-type*, Arg658Cys, and Ile738Val mutant proteins do not reveal change in ellipticity, hence mutations are not abrogating the tertiary structure of BARD1 BRCT (568-777) *wild-type* protein (Figure 1(D)). Treating with trypsin (10 pg/mol) and chymotrypsin (10 pg/mol) for three hours, the pattern on 12% SDS–PAGE for the *wild-type* and both the mutant proteins was very similar (Figure 2). This observation suggests that the mutants also have compact tertiary structure similar to that of the *wild-type* protein. However, *wild-type* and mutants were different when treated with the chemical denaturant GuHCl. To determine the unfolding pattern of BARD1 BRCT (568-777) *wild-type*, Arg658Cys, and Ile738Val mutant proteins, all proteins with 2 μM concentration were incubated with increasing concentrations of GuHCl (0–6 M) for 24 h until equilibrium was achieved. Emission maxima were recorded between $\lambda = 310$ and 390 nm. In the absence of GuHCl, while *wild-type* protein shows emission maximum at $\lambda = 333$ nm, the emission maxima for BARD1 BRCT (568-777), Arg658Cys, and Ile738Val mutant proteins were at $\lambda = 336$ and 337 nm, respectively. This red shifting in the emission maximum for the mutants is due to the local change in the environment, resulting in exposure of buried tryptophan to polar solvent interface. BARD1 BRCT *wild-type* and mutant proteins denatured completely in 6 M GuHCl with an emission maximum of $\lambda = 345$ nm, which indicates that tryptophans are exposed to polar interface (Figure 3(A)). Thus, collective observations from CD spectroscopy,

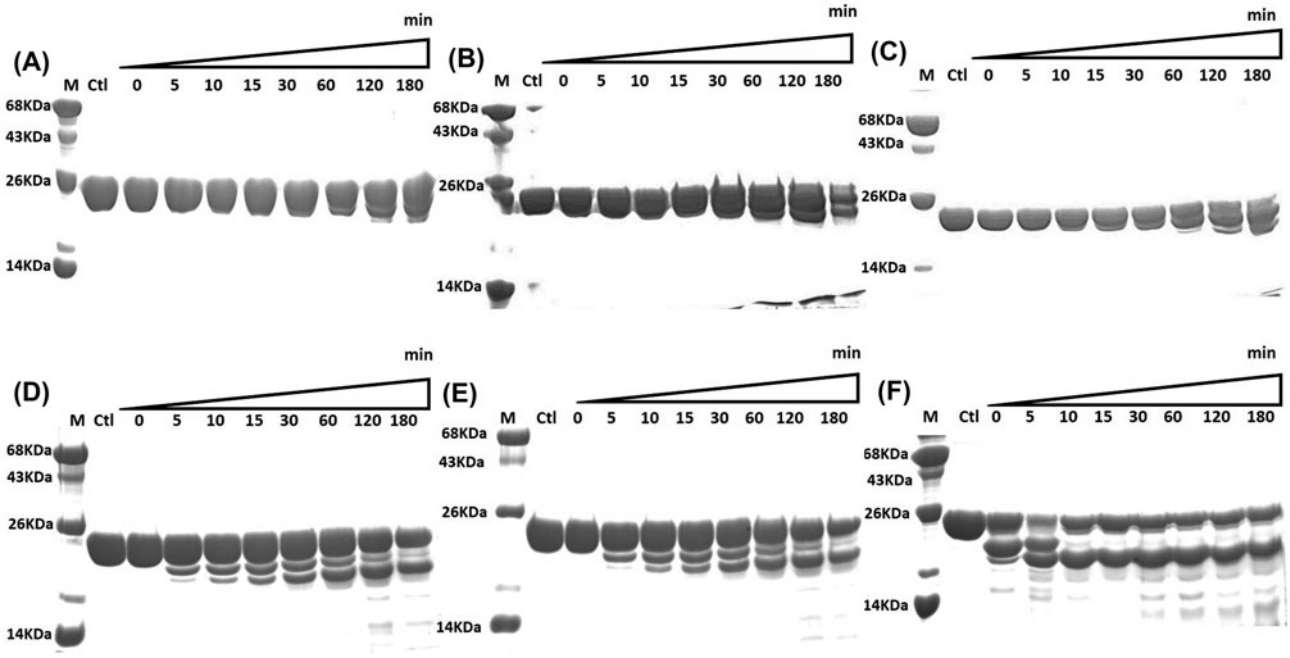


Figure 2. Resistivity profile of BARD1 BRCT (568-777) *wild-type*, Arg658Cys, and Ile738Val toward protease digestion.
Notes: (A) Limited proteolysis of BARD1 BRCT (568-777) *wild-type*, (B) Arg658Cys, and (C) Ile738Val using trypsin and (D) BARD1 BRCT (568-777) *wild-type*, (E) Arg658Cys, and (F) Ile738Val using chymotrypsin as proteases.

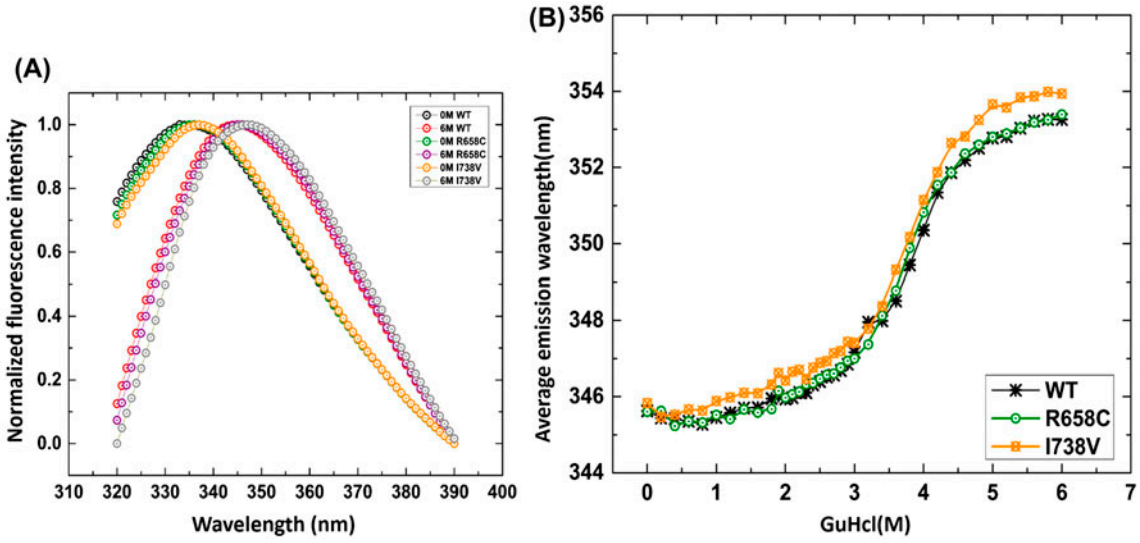


Figure 3. Chemical denaturation profile of BARD1 BRCT (568-777) *wild-type*, Arg658Cys, and Ile738Val.
Notes: (A) Overlay of normalized fluorescent intensity pattern of BARD1 BRCT (568-777) *wild-type*, Arg658Cys, and Ile738Val mutants. Normalized fluorescence intensity of BARD1 BRCT (568-777) *wild-type* native, Arg658Cys, and Ile738Val (0 M GuHcl) unfolded (6 M GuHcl), respectively. (B) Non-linear least-squares fit of a three-state equilibrium unfolding model of BARD1 BRCT (568-777) *wild-type*, Arg658Cys, and Ile738Val on the GuHcl-induced denaturation plot measured at 10°C.

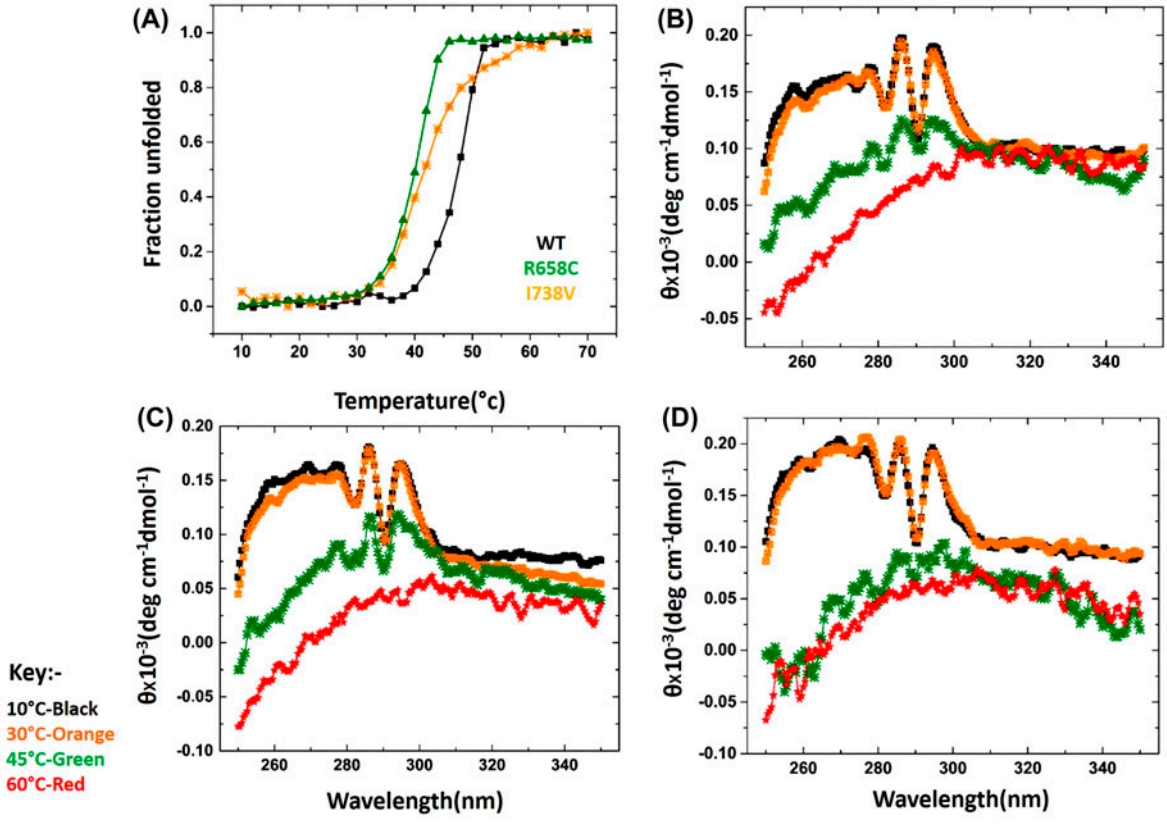


Figure 4. Structure and stability analysis of BARD1 BRCT (568-777) *wild-type*, Arg658Cys, and Ile738Val mutant proteins.
Notes: (A) Overlay of fraction unfolded in far-UV CD spectra of BARD1 BRCT (568-777) *wild-type*, Arg658Cys, and Ile738Val. Thermal stability assessment of BARD1 BRCT (568-777) *wild-type*, Arg658Cys, and Ile738Val in near-UV range. (B) Overlay of mean residual ellipticity obtained after thermal denaturation of BARD1 BRCT (568-777) *wild-type*, (D) Arg658Cys, and (C) Ile738-Val in CD spectroscopy.

Downloaded by [14.139.127.177] at 23:12 25 September 2015

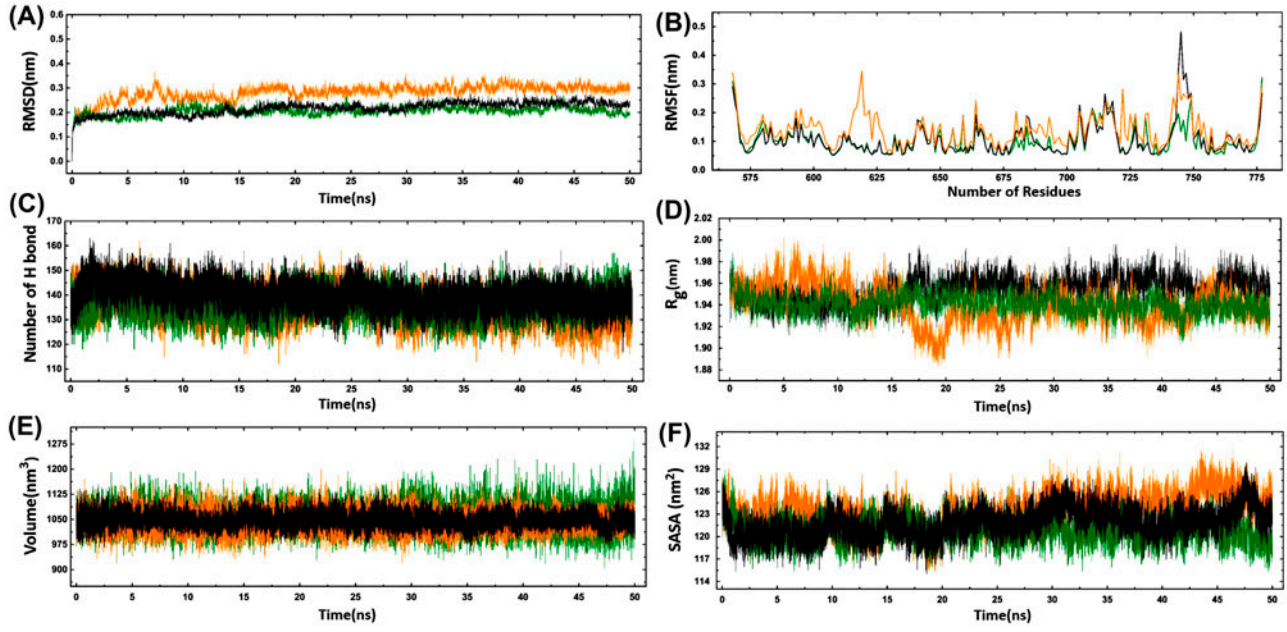


Figure 5. Structural analysis of BARD1 BRCT (568-777) *wild-type* and mutant proteins using molecular dynamics simulation.
Notes: Comparative overlay of (A) RMSD, (B) RMSF, (C) N–H bond, (D) R_g, (E) SASA, and (F) volume overlay of BARD1 BRCT (568-777) *wild-type* (black), Arg658Cys (green), and Ile738Val (orange) with respect to time (ns).

limited proteolysis, and fluorescence spectroscopy suggested that Arg658Cys and Ile738Val substitutions might be affecting secondary and tertiary structures only locally, and the effect of mutation is collectively responsible for tertiary structural alterations.

3.3. *Thermodynamic stability of BARD1 BRCT (568-777)* wild-type *and mutants*

To evaluate unfolding phenomenon and thermodynamic stability, thermal denaturation studies of proteins were performed using CD spectroscopy. The change in ellipticity due to gradual increase in the temperature from 10 to 65°C was recorded. Ellipticity at $\lambda = 222$ nm was used to calculate fractions unfolded (Figure 4(A)). The data obtained from CD at $\lambda = 222$ nm could fit satisfactorily into a two-state denaturation model for *wild-type* and all mutants (Pace, 1986; Pace & Shaw, 2000). The transition mid-point (T_m) calculated for BARD1 BRCT (568-777) *wild-type*, Arg658Cys, and Ile738Val mutant proteins were 40.6, 41.7, and 47.8°C, respectively. The changes in free energies corresponding to the denaturation were: $\Delta G°H_2O$ 9.6 ± 0.32, 7.3 ± 0.12, and 7.6 ± .42 kcal/mol, respectively. For tertiary structure stability assessment, near-UV CD spectra of BARD1 BRCT(568-777) *wild-type*, Arg658Cys, and Ile738Val mutant proteins were monitored at four different temperatures: 10, 30, 45, and 60°C corresponding to $\lambda = 280$ and $\lambda = 295$. At the temperature ranges from 10 to 30°C, BARD1 BRCT (568-777) *wild-type*, Arg658Cys, and Ile738Val, peaks at $\lambda = 295$ nm and $\lambda = 280$ nm were observed, which is characteristic of well folded 3-D structures. Furthermore, at 45°C, BARD1 BRCT (568-777) *wild-type* and Ile738-Val mutant proteins show 40% loss in the mean residual ellipticity, but a complete loss in ellipticity was observed in Arg658Cys mutant. It is therefore concluded that BARD1 BRCT (568-777) *wild-type* and Ile738Val mutant proteins have higher thermal stability than Arg658Cys (Figure 4(B)–(D)). Fluorescence spectroscopy was used to monitor unfolding at 283 K induced by GuHCl. For chemical denaturation of BARD1 BRCT (568-777) *wild-type*, Arg658Cys, and Ile738Val mutant proteins, thermodynamic parameters were calculated by plotting average fluorescence emission wavelength against GuHcl concentration (Figure 3(B)). $\Delta G°$ H_2O for BARD1 BRCT, Arg658Cys, and Ile738Val were 7.19 ± .36, 6.29 ± .41, and 6.8 ± .21 kcal/mol, respectively. This value of the BARD1 BRCT *wild-type* protein is in agreement with the earlier reported results (Thanassoulas et al., 2010). ΔG for folded to intermediate and then from intermediate to unfolded state for *wild-type* proteins was calculated to be 2.68 and 4.5 kcal/mol, for Ile738Val mutant to be 1.8 and 4.8658 kcal/mol, for Arg658Cys mutant to be 2.2 and 4.09 kcal/mol, respectively. Thus, the folded mutants are thermodynamically less stable than the *wild type*.

Downloaded by [14.139.127.177] at 23:12 25 September 2015

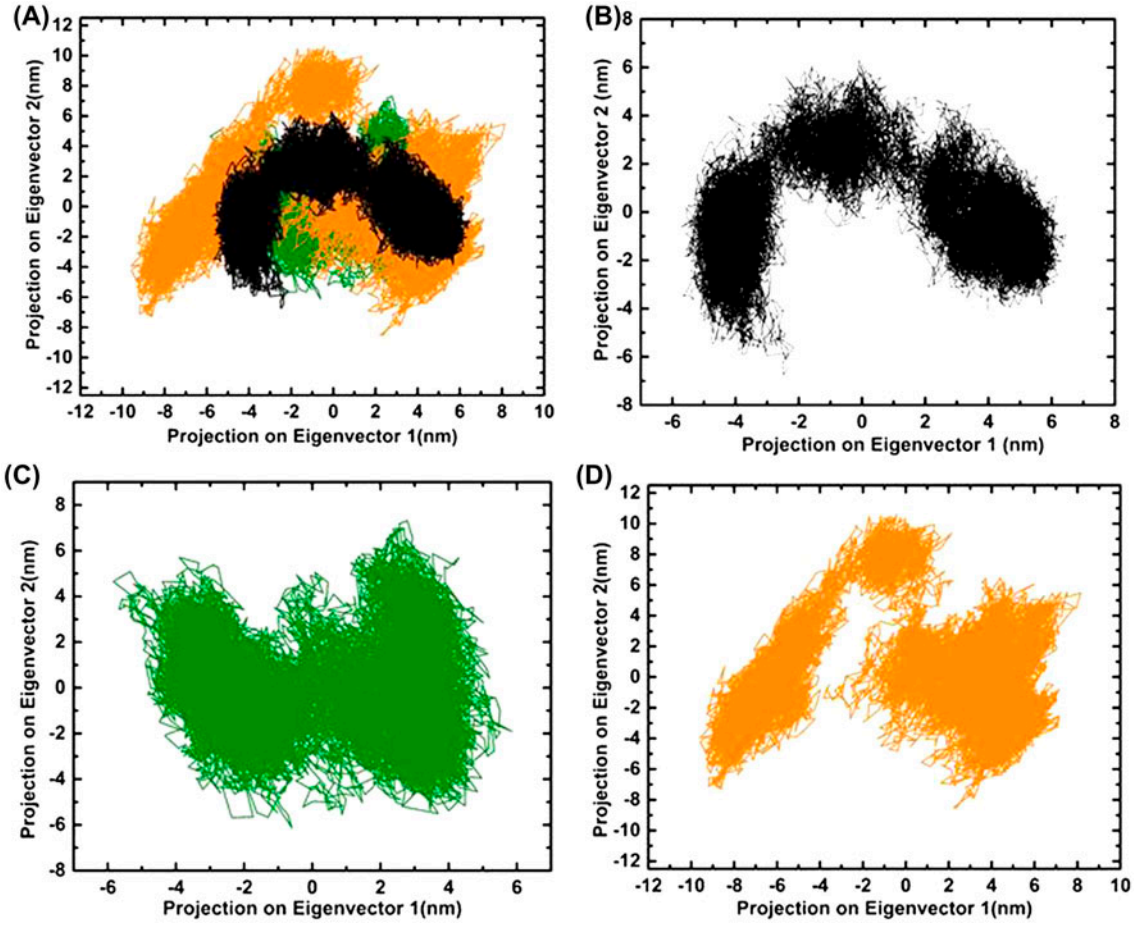


Figure 6. Principal component analysis of BARD1 BRCT (568-777) *wild-type*, Arg658Cys, and Ile738Val proteins.
Notes: (A) BARD1 BRCT (568-777) *wild-type*, Arg658Cys, and Ile738Val are shown in black, green, and yellow, respectively. For unblemished perception, individual protein has been represented independently, (B) BARD1 BRCT, (C) Arg658Cys, and (D) Ile738-Val. Dynamics of the proteins are projected with two eigenvectors.

3.4. MDS of BARD1 BRCT, ARG658CYS, and ILE 738VAL mutations

Molecular dynamics calculations were performed to visualize the structural alterations in BARD1 BRCT (568-777) wild-type, Arg658Cys, and Ile738Val mutant proteins. C*α,* RMSD values from the input structure were plotted against time (50 ns) (Figure 5(A)). It has been observed that *wild-type* and mutant structures showed similar deviations till 1500 ps, and at 1750 ps, the structures behaved differently. However, after 1.75 ns, Ile738-Val mutant structure RMSD increases to about 3 Å, whereas *wild-type* and Ile738Val structures remained with a RMSD value of ~1.75 to 2 Å and ~1.75 to 3 Å, respectively. In BARD1 BRCT (568-777) *wild-type* and Arg658Cys and Ile738Val mutant protein structures, the magnitude of Cα RMSD fluctuations, after relaxation period produces stable trajectories throughout the simulation (Figure 5(A)). The RMSF values of BARD1 BRCT (568-777) *wild-type*, Arg658Cys, and Ile738Val mutant protein structures were obtained, and higher degree of residual flexibility were observed for Ile738Val mutant protein (Figure 5(B)). The BARD1 BRCT (568-777) *wild-type* structure showed radius of gyration (R_g) value of ~1.924 nm at 0 ns, ~1.946 nm at 2 ns, ~1.92 nm at 5 ns, ~1.94 nm at 10 ns, ~1.97 nm at 15 ns, ~1.93 nm at 20 ns, ~1.96 nm at 25 ns, and 1.95 nm at 30 ns. Ile738-Val mutant structure showed R_g value of ~1.92 nm at 0 ns, ~1.954 nm at 2 ns, ~1.97 nm at 5 ns, ~1.94 nm at 10 ns, ~1.95 nm at 15 ns, ~1.92 nm at 20 ns, ~1.91 nm at 25 ns, and 1.94 nm at 30 ns. Arg658Cys mutant structure showed R_g value of ~1.925 nm at 0 ns, ~1.94 nm at 2 ns, ~1.95 nm at 5 ns, ~1.94 nm at 10 ns, ~1.95 nm at 15 ns, ~1.93 nm at 20 ns, and ~1.93 nm at 25 ns, and 1.94 nm at 30 ns. Radius of gyration of Arg658Cys mutant protein structure does not show much fluctuation, but Ile738Val mutant protein structure shows great variation throughout 50 ns simulation. Random change in radius of gyration at 20 ns of Ile738Val mutant protein

Downloaded by [14.139.127.177] at 23:12 25 September 2015

Downloaded by [14.139.127.177] at 23:12 25 September 2015

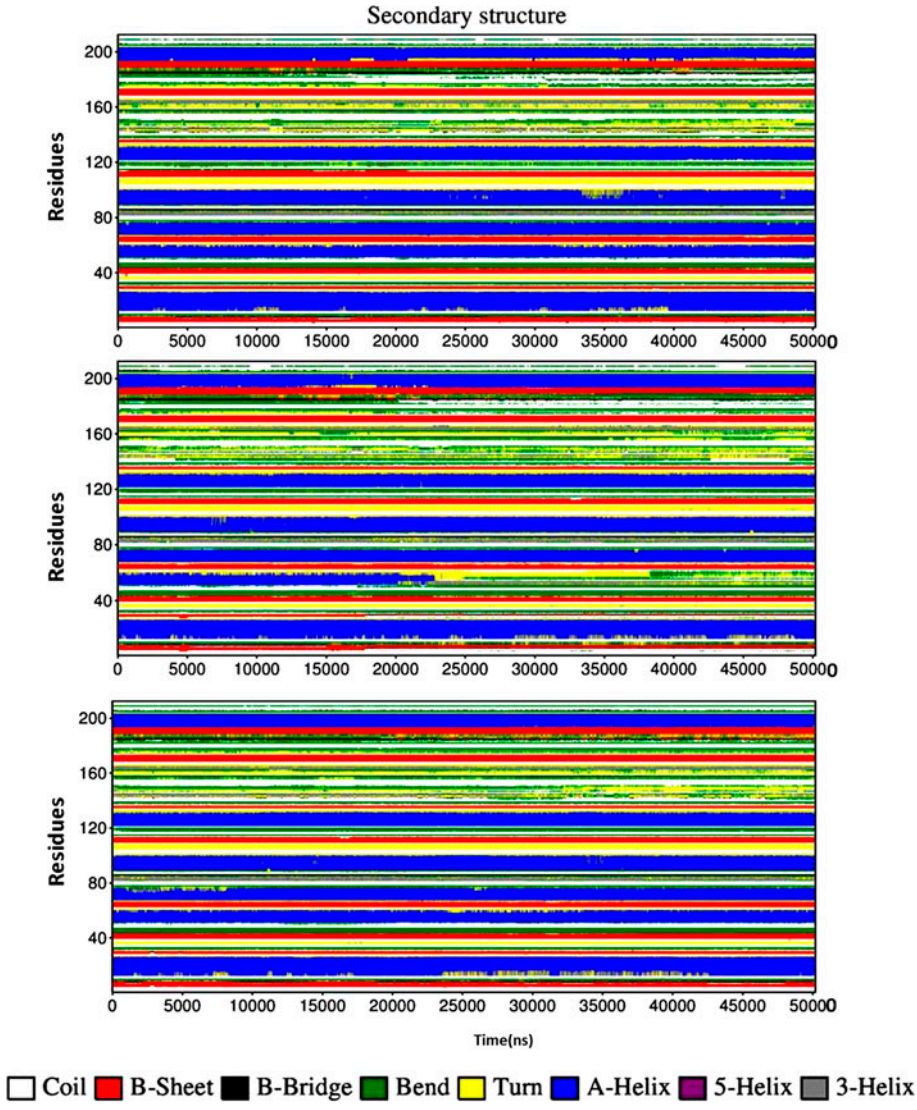


Figure 7. DSSP analysis of BARD1 BRCT (568-777) *wild-type*, Arg658Cys, and Ile738Val proteins with respect to time.
Notes: (A) BARD1 BRCT (568-777) *wild-type*, (B) Ile738Val, and (C) Arg658Cys.

structure indicates that structure has lost rigidity (Figure 5(D)). BARD1 BRCT (568-777) *wild-type* and Arg658Cys mutant protein structures do not show any drastic change in the SASA values, in sharp contrast to Ile738Val mutant protein, which shows increase in the solvent accessible surface area during the initial phase of the simulation (Figure 5(E)). We have also observed significant differences in the hydrogen bonding pattern during the simulation. BARD1 BRCT (568-777) *wild-type* protein has shown better retention in hydrogen bonds over the period of stimulation compared to Ile738Val and Arg658Cys mutant protein structures (Figure 5(C)). Hydrogen bond sampling at every 10 ns further indicated that *wild-type* protein shows higher retention of hydrogen bond than the mutant proteins (Supplementary Figure 3). Loss of hydrogen bond formation leads to more flexible conformation, and might result in loss of function of the mutant protein.

To understand the overall dynamics and evaluate the degree of flexibility, we have performed PCA and plotted eigenvector 1 v/s eigenvector 2 (Figure 6(B)–(D)). The area spanned/dispersion by this plot is indicative of the nature of motion. It has been observed that BARD1 BRCT (568-777) *wild-type* protein has the compact structure. Mutant Ile738Val is more flexible than BARD1 BRCT (568-777) *wild-type* and Arg658Cys proteins (Figure 6(A)). The DSSP plug-in of GROMACS monitored the secondary structures at different times during simulation (Kabsch & Sander, 1983). BARD1 BRCT Ile738Val mutant started losing the α-helix occurring between 50th and 60th residues and lost completely at about 25 ns (Figure 7(A) and (B)). However, no drastic

change in the secondary structure of *wild-type* and Arg658Cys mutant were observed (Figure 7(C)). It can be concluded that loss in α helical content in Ile738Val mutant protein may lead to structural flexibility. This was further supported by the change in volume recorded in DLS and fluorescence spectroscopic studies (Figure 5(F)).

3.5. Weak intramolecular interactions

Minimum energy structures of BARD1 BRCT (568-777) *wild-type*, Arg658Cys, and Ile738Val mutant proteins were used to study the weak intramolecular interactions. In the *wild-type* structure, Arg 658 forms hydrogen bond with the Glu 652, 649, and 655, and non-covalent interactions with Ser 660, Gly 656, Leu 662, and Arg 661. In Cys 658 mutant structure, there is a reduction in the number of hydrogen bonds. Furthermore, Cys 658 forms non-covalent interactions with Gly 656, Glu 652, Pro 654, Ile 653, and Ser 660 (Supplementary Figure 2 (A) and (B)). Mutant Ile738Val is involved in the hydrogen bonding with the Tyr 678 and non-hydrogen bonding interaction with Tyr 736, Phe 736, and Ala 758. Furthermore, Val 738 forms hydrogen bonds with Tyr 678 and a new hydrogen bond is established with Trp680 and non-hydrogen bonding interactions with Phe 683, Leu 679, Phe 677, Ala 758, and Tyr 736. Mutant structure Arg658Cys comprehensively loses the hydrogen bonds as compared to *wild-type* structure, further destabilizing the structure and surface charge of the protein molecule (Supplementary Figure 2 (C) and (D)).

4. Conclusions

To explore the structural effect of Arg658Cys and Ile738Val mutations on BARD1 BRCT, we have carried out biophysical experiments and MD simulations. In the crystal structure of BARD1 BRCT, Arg658 is a surface exposed residue, whereas Ile 738 is buried inside the structure. Expression and purification profile of BARD1 BRCT *wild-type* and mutant proteins do not show any drastic change in solubility and monomeric behavior. Secondary structures characterized by CD spectroscopy suggest that mutation is not inducing immense global changes, as the change in the ellipticity values is almost identical for all proteins. Far-UV spectra of proteins show insignificant change in the secondary structure, but slight local changes in the structure are also irrefutable. Near-UV CD spectroscopy, fluorescence spectroscopy, and limited proteolysis experiments suggested that the mutations are not affecting the overall compactness. The T_m value has reduced by 7 and 6.2°C for the mutants. Interestingly, the same concentration of GuHCl appears to be required to denature the *wild-type* and mutant proteins. Results from molecular dynamics simulation indicated that Ile738Val structure has gain flexibility compared to *wild-type* and Arg658Cys mutant proteins. The gain of flexibility has been detected in RMSD, RMSF, and R_g plot for Ile738Val mutant, which is further supported by red shift in native fluorescence maximum, decrease in the number of hydrogen bonds, increase in volume, SASA, and PCA. Loss of T_m and ΔG of BARD1 BRCT and Arg658Cys and Ile738Val mutants with respect to *wild-type* suggests that substitutions are physiologically detrimental. It has been observed collectively that *wild-type* and mutant proteins exhibited analogous secondary structural constituents, whereas fluorescence spectroscopy suggested that the comparative positioning of Trp and Tyr were significantly hindered. MDS of BARD1 BRCT (568-777) *wild-type*, Arg658Cys, and Ile738Val mutant protein structures has been comprehensively analyzed to understand the structural alterations. Our results from the biophysical, *in vitro*, and in silico approaches predict that Arg658Cys and Ile738Val mutants are destabilizing the BARD1 BRCT domain which may lead to the loss of thermodynamic stability and functional activity.

Supplementary material

The supplementary material for this paper is available online at http://dx.doi.org/10.1080/07391102.2015.1082149.

Acknowledgments

The authors would like to acknowledge Mass Spectrometry, BTIS facility of TMC-ACTREC. MVH thanks DAE for the award of Raja Ramanna fellowship.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the Department of Biotechnology, Ministry of Science and Technology [grant number BT/PR10765/BRB/664/2008], [grant number BT/PR12565/BID/07/303/2009], [grant number DBT BT/513/NE/TBP/2013]; Seed in Air grant from TMC.

References

Ahmed, M., & Rahman, N. (2006). ATM and breast cancer susceptibility. *Oncogene, 25*, 5906–5911.

Alshatwi, A. A., Hasan, T. N., Syed, N. A., Shafi, G., & Grace, B. L. (2012). Identification of functional SNPs in BARD1 gene and in silico analysis of damaging SNPs: Based on data procured from dbSNP database. *PLoS One, 7*, e43939.

Amadei, A., Linssen, A. B., & Berendsen, H. J. (1993). Essential dynamics of proteins. *Proteins: Structure, Function, and Genetics, 17*, 412–425.

Downloaded by [14.139.127.177] at 23:12 25 September 2015

Downloaded by [14.139.127.177] at 23:12 25 September 2015

Antoniou, A. C., Beesley, J., McGuffog, L., Sinilnikova, O. M., Healey, S., Neuhausen, S. L., … Easton, D. F. (2010). Common breast cancer susceptibility alleles and the risk of breast cancer for BRCA1 and BRCA2 mutation carriers: Implications for risk prediction. *Cancer Research, 70*, 9742–9754.

Birrane, G., Varma, A. K., Soni, A., & Ladias, J. A. (2007). Crystal structure of the BARD1 BRCT domains. *Biochemistry, 46*, 7706–7712.

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

Brzovic, P. S., Meza, J. E., King, M. C., & Klevit, R. E. (2001). BRCA1 RING domain cancer-predisposing mutations: Structural consequences and effects on protein–protein interactions. *Journal of Biological Chemistry, 276*, 41399–41406.

Brzovic, P. S., Rajagopal, P., Hoyt, D. W., King, M. C., & Klevit, R. E. (2001). Structure of a BRCA1-BARD1 heterodimeric RING-RING complex. *Nature Structural Biology, 8*, 833–837.

Callebaut, I., & Mornon, J. P. (1997). From BRCA1 to RAP1: A widespread BRCT module closely associated with DNA repair. *FEBS Letters, 400*, 25–30.

Collins, N., McManus, R., Wooster, R., Mangion, J., Seal, S., Lakhani, S. R., … Easton, D. F. (1995). Consistent loss of the wild type allele in breast cancers from a family linked to the BRCA2 gene on chromosome 13q12-13. *Oncogene, 10*, 1673–1675.

De Brakeleer, S., De Grève, J., Loris, R., Janin, N., Lissens, W., Sermijn, E., & Teugels, E. (2010). Cancer predisposing missense and protein truncating BARD1 mutations in non-BRCA1 or BRCA2 breast cancer families. *Human Mutation, 31*, E1175–E1185.

Ghimenti, C., Sensi, E., Presciuttini, S., Brunetti, I. M., Conte, P., Bevilacqua, G., & Caligo, M. A. (2002). Germline mutations of the BRCA1-associated ring domain (BARD1) gene in breast and breast/ovarian families negative for BRCA1 and BRCA2 alterations. *Genes, Chromosomes & Cancer, 33*, 235–242.

Harpaz, Y., Gerstein, M., & Chothia, C. (1994). Volume changes on protein folding. *Structure* 2: 641–649.

Heikkinen, K., Rapakko, K., Karppinen, S. M., Erkko, H., Knuutila, S., Lundan, T., … Winqvist, R. (2006). RAD50 and NBS1 are breast cancer susceptibility genes associated with genomic instability. *Carcinogenesis, 27*, 1593–1599.

Hess, B., Kutzner, C., van der Spoel, D., & Lindahl, E. (2008). GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of Chemical Theory and Computation, 4*, 435–447.

Jin, Y., Xu, X. L., Yang, M. C., Wei, F., Ayi, T. C., Bowcock, A. M., & Baer, R. (1997). Cell cycle-dependent colocalization of BARD1 and BRCA1 proteins in discrete nuclear domains. *Proceedings of the National Academy of Sciences, 94*, 12075–12080.

Kabsch, W., & Sander, C. (1983). Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers, 22*, 2577–2637.

Kaminski, G. A., Friesner, R. A., Tirado-Rives, J., & Jorgensen, W. L. (2001). Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *The Journal of Physical Chemistry B, 105*, 6474–6487.

Kaplan, W., & Littlejohn, T. G. (2001). Swiss-PDB Viewer (Deep View). *Briefings in Bioinformatics, 2*, 195–197.

Karppinen, S. M., Barkardottir, R. B., Backenhorn, K., Sydenham, T., Syrjakoski, K., Schleutker, J., … Winqvist, R. (2006). Nordic collaborative study of the BARD1 Cys557Ser allele in 3956 patients with cancer: enrichment in familial BRCA1/BRCA2 mutation-negative breast cancer but not in other malignancies. *Journal of Medical Genetics, 43*, 856–862.

Karppinen, S.-M., Heikkinen, K., Rapakko, K., & Winqvist, R. (2004). Mutation screening of the BARD1 gene: Evidence for involvement of the Cys557Ser allele in hereditary susceptibility to breast cancer. *Journal of Medical Genetics, 41*, 856–862.

Kolchanov, N. A., Soloviov, V. V., & Zharkikh, A. A. (1983). The effects of mutations, deletions and insertions of single amino acids on the three-dimensional structure of globins. *FEBS Letters, 161*, 65–70.

Li, M., & Yu, X. (2013). Function of BRCA1 in the DNA damage response is mediated by ADP-ribosylation. *Cancer Cell, 23*, 693–704.

Meijers-Heijboer, H., van den Ouweland, A., Klijn, J., Wasielewski, M., de Snoo, A., Oldenburg, R., … Stratton, M. R. (2002). Low-penetrance susceptibility to breast cancer due to CHEK2*1100delC in noncarriers of BRCA1 or BRCA2 mutations. *Nature Genetics, 31*, 55–59.

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.

Mohammad, D. H., & Yaffe, M. B. (2009). 14-3-3 proteins, FHA domains and BRCT domains in the DNA damage response. *DNA Repair, 8*, 1009–1017.

Pace, C. N. (1986). Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods in Enzymology, 131*, 266–280.

Pace, C. N., & Shaw, K. L. (2000). Linear extrapolation method of analyzing solvent denaturation curves. *Proteins: Structure, Function, and Genetics, 41*, 1–7.

Rapakko, K., Heikkinen, K., Karppinen, S. M., Erkko, H., & Winqvist, R. (2007). Germline alterations in the 53BP1 gene in breast and ovarian cancer families. *Cancer Letters, 245*, 337–340.

Renwick, A., Thompson, D., Seal, S., Kelly, P., Chagtai, T., Ahmed, M., … Rahman, N. (2006). ATM mutations that cause ataxia-telangiectasia are breast cancer susceptibility alleles. *Nature Genetics, 38*, 873–875.

Ryckaert, J.-P., Ciccotti, G., & Berendsen, H. J. (1977). Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics, 23*, 327–341.

Scully, R., Chen, J., Ochs, R. L., Keegan, K., Hoekstra, M., Feunteun, J., & Livingston, D. M. (1997). Dynamic changes of BRCA1 subnuclear location and phosphorylation state are initiated by DNA damage. *Cell, 90*, 425–435.

Thai, T. H., Du, F., Tsan, J. T., Jin, Y., Phung, A., Spillman, M. A., … Bowcock, A. M. (1998). Mutations in the BRCA1-associated RING domain (BARD1) gene in primary breast, ovarian and uterine cancers. *Human Molecular Genetics, 7*, 195–202.

Thanassoulas, A., Nomikos, M., Theodoridou, M., Yannoukakos, D., Mastellos, D., & Nounesis, G. (2010). Thermodynamic study of the BRCT domain of BARD1 and its interaction with the -pSER-X-X-Phe- motif-containing BRIP1 peptide. *Biochimica et Biophysica Acta (BBA) – Proteins and Proteomics, 1804*, 1908–1916.

Vikrant, Kumar, R., Yadav, L. R., Nakhwa, P., Waghmare, S. K., Goyal, P., & Varma, K. (2013). Structural and functional implication of RAP80 ΔGlu81 mutation. *PLoS One, 8*, e72707.

Vikrant, Nakhwa, P., Badgujar, D. C., Kumar, R., Rathore, K. K. S., & Varma A. K. (2014). Structural and functional characterization of the MERIT40 to understand its role in DNA repair. *Journal of Biomolecular Structure and Dynamics, 32*, 2017–2032.

Wooster, R., Neuhausen, S. L., Mangion, J., Quirk, Y., Ford, D., Collins, N., ... Averill, D. (1994). Localization of a breast cancer susceptibility gene, BRCA2, to chromosome 13q12-13. *Science, 265*, 2088–2090.

Wu, L. C., Wang, Z. W., Tsan, J. T., Spillman, M. A., Phung, A., Xu, X. L., … Baer, R. (1996). Identification of a RING protein that can interact *in vivo* with the BRCA1 gene product. *Nature Genetics, 14*, 430–440.

Wu-Baer, F., Lagrazon, K., Yuan, W., & Baer, R. (2003). The BRCA1/BARD1 heterodimer assembles polyubiquitin chains through an unconventional Linkage involving lysine residue K6 of ubiquitin. *Journal of Biological Chemistry, 278*, 34743–34746.

Yoshikawa, K., Ogawa, T., Baer, R., Hemmi, H., Honda, K., Yamauchi, A., … Takahashi, R. (2000). Abnormal expression of BRCA1 and BRCA1-interactive DNA-repair proteins in breast carcinomas. *International Journal of Cancer, 88*, 28–36.

Zhukov, I., Jaroszewski, L., & Bierzynski, A. (2000). Conservative mutation Met8 –> Leu affects the folding process and structural stability of squash trypsin inhibitor CMTI-I. *Protein Science, 9*, 273–279.

Downloaded by [14.139.127.177] at 23:12 25 September 2015