Chapter 3
Biophysical characterization and
Molecular mechanism
Of BARD1 ARD-BRCT
CstF50 binding
Interface
3.1 Introduction:

Germ-line mutations in *Breast cancer susceptibility gene 1 (BRCA1)* result in early predisposition of 40–50% familial breast cancer cases and >75% cases of hereditary breast and ovarian cancer [53, 141, 234, 235]. The gene product of *BRCA1* comprises of 1863 amino acid protein that interact with different cellular partners such as BARD1, BAP1, ABRAXAS, CtIP and RAP80 [65]. BARD1 in association with BRCA1, present itself as E3 ubiquitin ligase and helps in tumor suppression function of BRCA1 [21, 236-239]. CstF50 (Cleavage Stimulating Factor 1) is key reported binding partner of BARD1 and has an important role in processing and polyadenylation of premature mRNA transcripts. BARD1 is the only binary interacting partner reported for CstF50 until now [12, 36, 131].

The polyadenylation reaction is a series of steps consisting of an endonucleolytic cleavage followed by synthesis of the poly(A) tail [130]. Cleavage stimulation factors (CstF₅₀) are the essential processing factors, responsible for processing of 3’-end of premature mRNA transcript generated by RNA polymerase II [240]. CstF50 is a 50 KDa protein implicated in direct interaction with the RNA polymerase II CTD (C-terminal domain) through its N-terminal region [130, 131]. BARD1 interacts with 7th WD40 repeat of CstF50 through the flexible linker region residues from 554-568 [12]. It has been reported that 3’-end processing is repressed after UV-induced DNA damage as a result of proteasome mediated degradation of RNAP II by the BRCA1–BARD1–CstF50 complex [6]. Depletion of CstF50 enhances sensitivity to UV irradiation and reduces the ability of BRCA1-BARD1 to ubiquitinate RNAP II [132] and further leads to cell-cycle arrest and apoptosis [241]. CstF50 is also reported to interact with the DNA replication and repair factor PCNA [12]. PCNA co-localizes with BRCA1/BARD1 at the sites of DNA damage [72, 78] and in association with
other DNA damage repair proteins [133]. Supporting these facts, studies have shown that cells lacking BRCA1 have defective transcription coupled repair mechanism [8, 242]. Since, the results of two studies link BRCA1-BARD1 to the stimulation of ubiquitination of RNAPII in non-breast cells, signifying that the BRCA1-BARD1 mediated RNA Polymerase-II CTD ubiquitination and degradation is not a breast and ovarian cell-specific phenomenon. Genomic analysis of BARD1 in 58 ovarian tumors, 50 breast tumors and 60 uterine tumors has been previously performed and mutational investigation leads to detection of seven different polymorphism [27]. Cancer predisposing mutation in BARD1 ARD-BRCT linker region Gln564His was found in the germ-line of a patient with adenocarcinoma along with loss of the wild-type BARD1 allele [27]. BARD1 Gln564His mutation abrogates the p53 and CstF50 binding to BARD1 as demonstrated by GST pull down assay [6, 95]. Loss of binding between BARD1 and CstF50 leads to the formation of premature transcript and uncontrolled mRNA regulation. The BARD1 Gln564His mutation has shown to induce apoptosis less efficiently when transfected in mammalian cells, indicating that the BARD1 linker region is necessary for the tumor-suppressor and pro-apoptotic functions [95]. These findings suggest role for BARD1 mutations in the development of sporadic and hereditary tumors.

3.2 Material and Methods:-

A comparative analysis of BARD1 ARD-BRCT wild-type and mutants were performed using different biophysical and in-silico tools.

3.2.1 In-silico analysis: Interaction analyses between BARD1 wild-type and BARD1 Gln564His mutant with CstF50 were carried out using HADDOCK server [196].
3.2.2 Gene Cloning: bard1 ard-brct (encoding 425-777aa) wild-type (cDNA of full length BARD1, a kind gift from Richard Baer, USA) was sub-cloned in pGEX-kT vector (Amersham Pharmacia). bard1 ard-brct (425-777aa) Glu564His substitution was generated using the mismatch primers considering wild-type construct as template. Different regions of CstF50 (1-431aa) and (92-431aa) (cloned in pDONR vector which was purchased from UCLA) were sub-cloned in to pGEX-kT and pET-28a vectors. The basic scheme for sub-cloning was used and oligonucleotide primers complementary to the template sequence were used in PCR reactions and their details are mentioned below in primer details.

The genes of interest were amplified using following PCR condition: 95°C denaturation (5 minutes), 95°C denaturation (45 seconds), and annealing 62°C for 35 seconds, extension 72°C at 0.5kb/min, final extension 72°C for 10 minutes, and 25 cycles. The PCR amplified products were digested with EcoR1 and BamH1 and ligated in the pGEX-kT and pET-28a vectors (pGEX-kT was a kind gift from John Ladias, BIDMC-HMS, Boston). The ligation mixture was transformed into *E.coli* DH5α cells. Colony screening were performed and positive clones were selected after the insert release from the plasmid using EcoR1 and BamH1 restriction enzymes, which was further confirmed by DNA sequencing. Sequentially correct clones were used for protein expression and purification.

3.2.3 Primer details

**BARD1 ARD-BRCT domain (425-777aa) wild-type**

**FORWARD PRIMER**

5’-GTCGGATCCATATGGAGAACCTGTACTTTCAGGGTAACCATCGTGGTGAGACTTTTGGCTCCAT-3’
BARD1 ARD-BRCT (425-777aa) and Gln564His mutant were expressed in Rosetta 2(DE3) bacterial strain. For protein expression, 50 ng/µl plasmid constructs were transformed into Rosetta 2(DE3) cells and grown on LB agar plate containing ampicillin (100 µg/ml). CstF50 (1-431aa) was expressed in Rosetta 2(DE3) bacterial system. For protein expression, 50ng/µl plasmid constructs were transformed into Rosetta 2(DE3) cells and grown on LB agar plate containing kanamycin (100 µg/ml). Detailed protocol is described below.

3.2.4 Protocol for protein purification of BARD1 ARD-BRCT (425-777aa) and Gln564His mutant:

Purification buffer: 50 mM Tris (pH 6.2), 300 mM NaCl, 0.1% triton, 5% glycerol pH 6.2 (Buffer A)
**FPLC buffer:** 20 mM Tris (pH 7.5), 100 mM NaCl, 1 mM 0.1% triton (Buffer B)

1. **Re-suspension:** Re-suspended the pellet of BARD1 ARD-BRCT (425-777aa) and Gln564His mutant in 60 ml of buffer A; supplemented with 200 mM PMSF and 20µl of protease inhibitor.

2. **Ultra sonication:** Transferred the resuspended cell pellet into centrifuge tube and sonicated at 70 pulse rate and 70 power with 1 minutes of duty cycle. Repeated the sonication cycle 5 times with 1 min break.

3. **Centrifugation:** After sonication, the suspension was subjected to centrifugation at 18000 rpm for 45 minutes at 4°C to obtain cleared lysate. The soluble protein in the supernatant was collected and cell debris was discarded.

4. **Binding:** The soluble wild-type and mutant protein fractions were allowed to pass from column containing pre-equilibrated affinity resin. The sepharose beads in affinity column charged with glutathione binds specifically to proteins having GST (Glutathione-S-transferase) tag. The protein bound resin ~ 10 µl was loaded onto SDS-PAGE gel to check the binding.

5. **Washing:** After binding, the protein bound column was washed with 10 column buffer A to remove any non-specific protein.

6. **Cleavage:** Added 400 µl of TEV protease (20 units), 40 µl of protease inhibitor cocktail and 100 µl of PMSF in 20 ml of buffer A and completed the cleavage step in 3 hours by passing the TEV containing buffer solution through column at an interval of 1 hour. Nearly 10 µl of protein bound beads were loaded onto 12 % SDS-PAGE get to detect the cleavage of protein.
7. **Elution**: After TEV cleavage, the protein was eluted with 30 ml of buffer A.

8. **Equilibration of Ni-NTA resin**: Provided 2 column washes with double distilled water and then 5 to 6 column washes with buffer A.

9. **Metal Ion Chelate Affinity Chromatography**: After calibration of Ni-NTA resin, passed the eluted fractions through Ni-NTA to get rid of His-tagged TEV protease contamination.

10. **Concentrating the protein**: Transferred the eluted protein in a 10 KDa pre-equilibrated centricon and concentrate the protein up to 2 ml by centrifuging at 4500 rpm for 10 minutes at 4°C. Check the concentration on Nanodrop spectrophotometer (280 nm). Centrifuged for 10 minutes at 13000 rpm at 4°C for removal of insoluble aggregates or precipitates.

11. **Gel filtration**: 2 ml of concentrated protein was injected in AKTA- FPLC against buffer B.

12. **Fraction collection**: Collected the purified protein obtained through FPLC in 1.7 ml microfuge tube at its elution volume according to gel filtration spectra profile of the sample.

13. **Loading on 12% SDS-PAGE gel**: Loaded 20 µl of FPLC fractions on 12% SDS-PAGE, stained with coomassie dye, and then destained it to visualize the protein of interest.

14. **Concentrate the protein**: The fractions, which showed purified protein band was concentrated as per the requirement.

**Protocol for protein expression of CstF50:**

1. **Inoculation**: picked a single transformed colony of pET-28a-6HIS-CstF50 from antibiotic resistant LB agar plate and inoculated it in 100 ml LB broth preinoculum containing 100 µg/ml of kanamycin. Grown at 37°C overnight in shaking condition.
2. **Dilution:** Added 10 ml of pre-inoculum to 1000 ml (1:100) of autoclaved LB broth containing 100 µg/ml of kanamycin. Grown the medium on a shaker incubator at 37°C until it has reached mid-log phase i.e. $A_{600}$ between 0.6-0.8.

3. **Induction:** Cooled down the flasks and added 100 µl IPTG (stock 1M), and incubate on shaker incubator at 20°C for 18 hours.

4. **Harvesting:** The culture was centrifuged for 10 minutes at 6000 rpm at 4°C. The pellet was resuspended in a small volume of supernatant and centrifuged for 15 minutes at 5000 rpm, 4°C.

5. **Storage:** The bacterial pellet obtained was stored at -80°C for further use.

Proteins were purified by affinity chromatography followed by FPLC.

3.2.5 **Protocol for purification and refolding of CstF50:**

- **Cell lysis buffer:** 50 mM Tris (pH 6.2), 500 mM NaCl, 0.1% triton, 5% glycerol pH 7.2, 8M urea (Buffer A)

- **Pellet washing buffer:** 50 mM Tris (pH 6.2), 500 mM NaCl, 0.1% triton, 5% glycerol pH 7.2, 1% sodium do-decyl sulphate and 1% lauryl sulphate. (Buffer B)

- **CstF50 solubilization buffer:** 50 mM Tris (pH 6.2), 500 mM NaCl, 0.1% triton, 5% glycerol pH 7.2, 8M urea (Buffer C)

- **Refolding buffer:** 50 mM Tris (pH 6.2), 500 mM NaCl, 0.1% triton, 10% glycerol pH 7.2, 2mM EDTA, 700mM arginine, 100mM KCl, 50mM MgCl$_2$, 100mM CaCl$_2$, 5mM DTT and 0.2% triton (Buffer D).
FPLC buffer: 20 mM Tris (pH 7.5), 100 mM NaCl, 1 mM 0.1% triton, 200mM arginine (Buffer E).

1) **Re-suspension**: Re-suspended the pellet of CstF50 (1-431aa) in 100 ml of Buffer A supplemented with 200 mM PMSF and 20µl of protease inhibitor.

2) **Ultra sonication**: Transferred the resuspended cell pellet into centrifuge tube and sonicated at 70-pulse rate and 50 power with 1 minutes of duty cycle. Repeated the cycle 5 times with 1 min break.

3) **Centrifugation**: After sonication, the suspension was subjected to centrifugation at 18000 rpm for 45 minutes at 4°C to obtain cleared lysate. Collected the supernatant and discarded the cell debris.

4) **Pellet washing**: The soluble fraction was discarded (CstF50 is insoluble and forms inclusion bodies) and pellet was washed with pellet buffer B (composition described above) by resuspending in the buffer B thoroughly and centrifugation at 18000r.p.m. Discarded the supernatant and washed the pellet again. Repeated the process thrice.

5) **Solubilisation in 8M urea**: Resuspended the pellet after washing in Buffer C. Resuspended the pellet in buffer C in such a way that no pellet clump should left over. Solubilized all the protein present in pellet by incubating the washed pellet in 8M urea solubilisation for four hours.

6) **Metal Ion Chelate Affinity Chromatography**: The soluble 6HIS-CstF50 in 8M urea was allowed to bind fraction obtained is brought at room temperature and then mixed with pre-equilibrated affinity resin (with Buffer C) and incubated at room temperature for 1 hour.
The Ni\(^{2+}\) in the affinity column binds specifically to those proteins having Hexa-histidine tag. Take out 40-μl bead for binding check.

7) Washing: After binding, the protein bound column was washed with 4-column wash buffer c to remove any non-specific protein bound to the affinity column.

8) Elution: Eluted the bound protein after passing buffer containing increasing concentration of imidazole (100-600 mM).

9) Dialysis: Eluted protein was dialyzed by using 10KDa dialysis filter bag in presence of the buffer D. After every 4 hrs buffer changes were provided to refold the protein and at every change, the concentration of arginine was reduced 100mM and five serial changes were provided.

10) Concentrating the protein: Transferred the eluted protein in a 10 KDa pre-equilibrated centricon and concentrated the protein up to 2 ml by centrifuging at 4500 rpm for 10 minutes at 4\(^{\circ}\)C. Check the concentration on nanodrop spectrophotometer (280 nm). Centrifuged for 10 minutes at 13000 rpm at 4\(^{\circ}\)C for removal of insoluble aggregates or precipitates.

11) Gel filtration: Injected 2 ml of concentrated protein in AKTA- FPLC against FPLC buffer.

1. Fraction collection: Collected the purified protein obtained through FPLC in 1.7 ml microfuge tube at its elution volume according to gel filtration spectra profile of the sample.
12) **Loading on SDS-PAGE 10% gel:** Loaded 20 µl of FPLC fractions on SDS-PAGE, stained with coomassie dye, and then destained it to visualize the protein of interest.

13) **Concentrate the protein:** The fractions, which showed purified protein band was, concentrated as per the requirement.

### 3.3 Results and discussion

#### 3.3.1 Cloning, expression and purification of BARD1 ARD-BRCT domain (425-777aa), BARD1 ARD-BRCT Gln564His mutant and functional domains of CstF50:-

Selected potential clones when digested with the EcoR1 and BamH1 restriction enzymes showed the insert release of correct size ([Figure: 3.3.1](#)).

**Figure 3.3.1:** (A) PCR and (B) Cloning of BARD1 ARD-BRCT wild-type

DNA sequencing results confirmed the presence of ligated gene of interest in the vector with desired frame of codon sequence. Purified BARD1 ARD-BRCT (425-777aa) *wild-type* and Gln564His mutant showed similar solubility in the identical buffer and pH conditions.
(Figure: 3.3.2). This indicates that Gln564His mutation does not change the solubility of the wild-type protein.

**Figure 3.3.2:** Expression and purification profile of (A) BARD1 ARD-BRCT wild-type and (B) cloning and purification of Gln564His mutant.

### 3.3.2 Structural insights into BARD1 ARD-BRCT domain:

To comprehend the thermodynamic and biophysical parameters of BARD1 ARD-BRCT wild-type and Gln564His mutant, gel filtration chromatography, thermal denaturation, and
chemical denaturation were performed. Molecular exclusion analytical chromatogram of BARD1 ARD-BRCT and mutant shows that BARD1 ARD-BRCT wild-type and mutant exist mostly in the monomeric form, and mutation has not affected the monomeric property of the wild-type protein. DLS performed for wild-type and mutant did not show any large change in the hydrodynamic size of the proteins providing evidence that mutation is not changing the oligomeric status of the wild-type protein (Figure: 3.3.3).

Figure 3.3.3: - In-solution oligomeric characterization of BARD1 ARD-BRCT wild-type and Gln564His mutant. (A) Wild-type elution profile, (B) Elution profile of standard proteins (C) Stokes radii calculation, (D) DLS profile of wild-type and mutant, (E) Elution profile of mutant.
Glutaraldehyde crosslinking experiment with BARD1 ARD-BRCT wild-type also yielded similar results as, there was no higher molecular weight aggregates conjugates were visible on the coomassie brilliant blue counterstained gel (Figure: 3.3.4). Secondary structural components of BARD1 ARD-BRCT and mutant were characterized with far-UV Circular-Dichroism (Figure: - 3.3.5 A). To determine independent behavior of ARD domain of BARD1 from its BRCT domain, we have performed limited proteolysis in a time-dependent manner using trypsin. After treating the wild-type and mutant for limited time period with equal concentration of enzyme, same domain stability of wild-type and mutant was observed and mutation had not brought any significant changes on the structural compactness (Figure:- 3.3.5).

**Figure 3.3.4:** (A, B) Limited proteolysis and (C, D) Chemical crosslinking of BARD1 ARD-BRCT wild-type and Gln564His respectively.
Comparative secondary structure characterization of BARD1 ARD-BRCT wild-type and Gln564His mutant protein using CD spectroscopy show similar spectra having a mixed profile of α-helices, β-sheets, turns, and random coil component in the structure. (Figure: - 3.3.5).

**Figure:-3.3.5-** (A) CD spectra and (B) thermal denaturation of BARD1 ARD-BRCT wild-type and Gln564His mutant protein respectively.

Thermal denaturation of wild-type and mutant proteins was performed using circular dichroism spectroscopy and fraction unfolded for wild-type and mutant proteins were calculated as function of temperature. It has been observed that wild-type and mutant proteins unfold via a two state pathway and both the proteins unfold completely at 55°C. The T_m calculated for wild-type and mutant were 46.2±.25°C and 45.6°C±.45°C respectively. Insignificant change in the T_m indicates that the mutation does not affect the thermodynamic and unfolding pattern of the wild-type protein (Figure: - 3.3.5 B). Chemical denaturation of wild-type and mutant proteins with guanidium hydrochloride (GuHcl) indicates that wild-type and mutant protein unfolds via intermediate formation. This observation is also supported by
blue-shift of the emission maximum in the fluorescence spectra consistent with the presence of an intermediate molten globule (Figure: - 3.3.6).

Figure 3.3.6:- Chemical denaturation of (A) BARD1 ARD-BRCT wild-type (WT) and (B) Gln564His mutant.

The emission maxima shifted from ($\lambda_{\text{max}}$) shifted from 341nm to 335nm in wild-type and from 341nm to 336nm for mutant at 1.8M GuHCl. However, the $\lambda_{\text{max}}$ at 6M concentration of GuHCl was observed to be close to 350nm, thus signifying that proteins unfold via molten globule formation.

CstF50 was expressed in bacterial system using pGEX-kT and pET-28a vector (Figure: - 3.3.6). However, CstF50 (1-431aa) and WD40 domain (92-431aa) shows that most of the protein is insoluble in nature and forms inclusion bodies (Figure: - 3.3.7). CstF50 full length was cloned in pET-28a vector (Figure: - 3.3.7) and purified as 6HIS-CstF50 from inclusion bodies by solubilizing in 8M urea (Figure: - 3.3.8) and further refolded in native condition. However, no good quality CD spectra were obtained due to presence of arginine in the
refolding buffer, this interfered with the measurement. Efforts to remove arginine hydrochloride led to precipitation of the refolded protein.

Figure 3.3.7: (A) Cloning of CstF50 (B) pGEX-KT and pET-28a vectors

Figure 3.3.8: (A) Gradient PCR of CstF50 full length (B) CstF50 8M urea purification

Figure 3.3.8: Expression and purification of CstF50.
To further confirm CstF50 protein folding in native condition, we have performed fluorescence spectroscopy, and the unfolded protein shows loss of fluorescence intensity and red-shift in the emission maximum ($\lambda = 348$nm) (Figure: - 3.3.9). Fluorescence spectra of refolded protein shows an increase in the fluorescence intensity and blue-shift ($\lambda = 335$nm) of the emission maximum which indicates that tryptophan residues are buried in the hydrophobic core of the protein (Figure: - 3.3.9).

3.3.3 Structural insight into the binding interface of BARD1 ARD-BRCT and CstF50:- BARD1 ARD-BRCT region (425-777aa) was modeled using Robetta server [37] and mutation Gln564His was incorporated using the swiss PDB viewer [40]. Furthermore, CstF50 was also modeled using Robetta server. Both the models were validated by saves (Metaserver for analyzing and validating protein structures (https://services.mbi.ucla.edu/SAVES/) [243] and Molprobity server [244, 245]. Further model was refined by Modloop server [246] and based on stereo chemistry and Ramachandran plot, best model was selected for structural analysis [247] (Figure:- 3.3.10).
Figure 3.3.10: - (A) Ramachandran plot and (B) structural details of validated model for BARD1 ARD-BRCT (425-777aa).

Table 3.1:- Stereochemistry details BARD1 ARD-BRCT *wild-type*

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Poor rotamers</td>
<td>2</td>
<td>0.64%</td>
<td>Goal: &lt;1%</td>
</tr>
<tr>
<td>Ramachandran outliers</td>
<td>0</td>
<td>0.00%</td>
<td>Goal: &lt;0.05%</td>
</tr>
<tr>
<td>Ramachandran favored</td>
<td>344</td>
<td>98.01%</td>
<td>Goal: &gt;98%</td>
</tr>
<tr>
<td>Cβ deviations &gt;0.25Å</td>
<td>1</td>
<td>0.30%</td>
<td>Goal: 0</td>
</tr>
<tr>
<td>Bad backbone bonds:</td>
<td>0 / 1411</td>
<td>0.00%</td>
<td>Goal: 0%</td>
</tr>
<tr>
<td>Bad backbone angles:</td>
<td>1 / 1762</td>
<td>0.06%</td>
<td>Goal: &lt;0.1%</td>
</tr>
</tbody>
</table>
In BARD1 ARD-BRCT structure (425–777aa), ARD domain is connected to BARD1BRCT domain through a flexible linker of 14 amino acids in a head-to-tail manner. The three-dimensional model structure of ARD domain is predominantly α-helical in nature and BRCT domain shows a combination of two repeats consisting of three α–helices and β-strand connected by linkers.

Modeling BARD1 ARD-BRCT region also serves the purpose to study the extent of domain motion alone and also in complex with the CstF50. Errat evaluation [248] submits that model has an average overall quality factor of 92.151, and may have good structural resolution. 3D-1D analysis [249] indicates that 82.77% of the residues had an averaged 3D-1D score > 0.2 and never dipped below zero. MolProbity stereochemical parameters were also found satisfactory for the BARD1 ARD-BRCT (425–777aa) model and indicate that model structure parameters are in normal range (Table 3.1).

However, CstF50 model is mainly α–helical at the N-terminal region and is connected to WD40 repeat region with large flexible loop. The C-terminal 7 WD40 repeats domain comprised of ~300 amino acids. WD40 repeat domain is an evolutionarily conserved domain, and it is composed of 5 to 7 repeats majorly consisting of β-strand acquiring a β propeller shape (Figure: - 3.3.11). In the structural model, majority of CstF50 residues (90%) are in the most favored regions of Ramachandran plot indicating that the corresponding coordinates for the folded amino acids are satisfactory.
Figure: - 3.3.11: (A) Model structure of CstF50, (B) Structural details of validated model for CstF50 and (C). Ramachandran plot of cstF50.

The Verify 3D-1D profile analysis of the predicted protein model of CstF50 revealed that the average score stayed above 0.2 and never decreased below zero.
Table 3.2:- Stereochemistry details of CstF50 (1-431aa)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Count</th>
<th>Percentage</th>
<th>Goal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poor rotamers</td>
<td>3</td>
<td>0.79%</td>
<td>&lt;1%</td>
</tr>
<tr>
<td>Ramachandran outliers</td>
<td>0</td>
<td>0.00%</td>
<td>&lt;0.05%</td>
</tr>
<tr>
<td>Ramachandran favored</td>
<td>400</td>
<td>93.24%</td>
<td>&gt;98%</td>
</tr>
<tr>
<td>Cβ deviations &gt;0.25Å</td>
<td>0</td>
<td>0%</td>
<td>0</td>
</tr>
<tr>
<td>Bad backbone bonds:</td>
<td>0 / 1723</td>
<td>0.00%</td>
<td>0%</td>
</tr>
<tr>
<td>Bad backbone angles:</td>
<td>2 / 2152</td>
<td>0.09%</td>
<td>&lt;0.1%</td>
</tr>
</tbody>
</table>

The overall structure calculation from PROVE program [250] gave Z-score RMS of 1.53. The Z-score obtained from SAVES [251] is significant as the Z-score is not above 4.0 and below-4.0, which indicates that the hydrophobic amino acids buried inside the core are well packed. A good average overall quality factor (Errat score) is around 87% for the predicted model which is satisfactory [248]. Furthermore, for evaluation of the structural geometry of the modeled CstF50 protein, we have submitted the structure to the Molprobity server [244, 245]. As the calculated structural parameters of the CstF50 model were within the allowed range (Table 2). Therefore, we have used this model for docking and molecular dynamics simulation studies. Similarly, being a member of WD40 family of proteins, CstF50 contains seven WD40 repeat domains spanning C-terminal region (92-431aa) comprising of β-sheets. N-terminal domain (1-91aa) is homodimerization domain, which is required for interactions with RNA pol-II CTD, is predominantly comprised of the α-helices. Inter-secondary structure segments/loops were identified by DSSP and STRIDE [252, 253]. These interactions can then
form an anchor for assembly of larger protein complexes. Such interactions were first structurally characterized for heterotrimeric G-proteins where the β-subunit is a WD40 protein [59, 60].

3.3.4 Domain flexibility of ARD with respect to BRCT domain:-

*In-silico* Normal Mode Analysis (*NMA*) is used to study the protein dynamics of BARD1 ARD-BRCT domain and mutant, which is precisely well suited for demonstrating inter-domain motions, embedded in the structure.

![Comparative NMA first mode profile of BARD1 ARD-BRCT](image)

**Figure 3.3.12:** Comparative NMA first mode profile of BARD1 ARD-BRCT (A) *wild-type* and Gln564His mutant (B) Comparative Deformation analysis of *wild-type* and mutant showing hinge bending.
To study the extent of domain motion in BARD1 ARD-BRCT, we have performed NMA for BARD1 ARD-BRCT wild-type and mutant protein structure. First few modes are the low frequency modes and represent the large motions within the protein structure (Figure: - 3.3.12 A). NMA of BARD1 ARD-BRCT wild-type and Gln564His indicates large domain motion between the ARD and BRCT domain, and the nature of concerted motion indicates that residues in the ARD domain and BRCT domain are positively co-related while linker region shows anti-correlated motion to the ARD1 and BRCT domain respectively. First five modes show that ARD and BRCT domains fluctuate between open and closed states because of bending of hinge at the linker region. Deformation analysis indicates that linker region of wild-type and mutant protein has high value of deformation as the RMSF of the linker region is recorded to be the highest.

Large-scale conformational transitions in proteins can occur because of relative movement between domains. Flexibility in short segments of the protein backbone allows corresponding domain motions to occur, with only minor structural perturbations within domains [254, 255]. The collective motion of a simulated protein can be identified by analyzing both the covariance and time correlation in the positional fluctuation of its atoms [256]. It has been observed that full-length protein always has a rigid conformation than its individual domains. Here, MD simulation has been applied to investigate the conformational changes induced in the BARD1 ARD-BRCT domain. Comparative analysis of the RMSD data was performed after fitting the trajectory input of wild-type and mutant protein structures (Figure: - 3.3.13).
Figure 3.3.13: - Comparative (A) RMSD and (B) Rg profile of BARD1 ARD-BRCT (Black) and Gln564His (green) mutant and (C) Structures sampled at every 10ns for wild-type.

The simulation for 100ns reveals that BARD1 ARD-BRCT wild-type structure is comparatively more stable than the mutant. After 5ns of simulation, RMSD of both mutant and wild type protein increased up to 5Å but after 10ns, the wild-type RMSD decreases to 3.5Å, whereas mutant RMSD increases to 6.5Å. Sharp increase in the RMSD has been observed at 16ns in wild-type structure. Nevertheless, looking at the mutant RMSD profile, instead of sharp increase gradual change in RMSD at 18ns was observed. Sharp increase in the RMSD for mutant structure at 95ns was observed in contrast to the wild-type protein.
Figure 3.3.14: - (A) Comparative aligned structure sampled at every 10ns for mutant Gln564His, (B) highest RMSD and input structure aligned for wild-type and (C) mutant (Gln564His).

We have analyzed the highest RMSD extracted structure for wild-type and mutant protein from the trajectory. After comparing the structures of wild-type at 0ns and 16ns (highest RMSD frame) and for mutant 0ns and 98ns, a large shift in the ARD domain orientation was observed.
Figure 3.3.15: - Comparative RMSD representation of ARD and BRCT domain. ARD wild-type (Black), mutant (Green) and BRCT domain wild-type (yellow), mutant (Blue) respectively.

Structural alignment with input structure and highest RMSD attained structure for wild-type and mutant protein shows large rotation about 70° (RMSD=8.3 Å) in the ARD domain from the axis for wild-type, and 84.4° (RMSD=9.27 Å) for mutant protein respectively. This indicates that BRCT domain of BARD1 ARD-BRCT wild-type and mutant protein structures are comparatively rigid but hinge bending at the linker leads to the domain motion (Figure: - 3.3.14, 3.3.15).

Furthermore, domain wise comparative study of RMSD for wild-type and mutant protein structures shows that ARD domain is more flexible than the BRCT domain (Figure: - 3.3.15). Decrease in the RMSD in both wild-type and mutant up to 4Å is indicative of stabilization in structural change, but the RMSD of mutant remained high as compared to wild-type protein. After 35ns wild-type structure is completely stabilized, as there was only small fluctuation in the RMSD in the wild-type structure. On the other hand, BARD1 ARD-
BRCT Gln564His mutant structure shows a sudden increase in the RMSD profile at 90ns, indicating another change in the conformation leading to high average RMSD in mutant structure than the wild-type structure.

Figure 3.3.16: (A) RMSD, (B) $R_g$, and (C) RMSF attained by CstF50 during 100ns MD production.

Moreover, CstF50 structure did not show any large RMSD, $R_g$ deviation during 100ns simulation and residual RMSF fluctuations that contributed either from the linkers between
N-ter and WD40 domain or loops between beta strands of 1st WD40 repeat (Figure: - 3.3.16). The comparative radius of gyration for BARD1 ARD-BRCT, Gln564His mutant structure was also analyzed by molecular dynamics simulation.

**Figure 3.3.17:** - (A and B) RMSF Structure representing for BARD1 ARD-BRCT wild-type and mutant. (C) RMSF for wild-type (green) and mutant(black) (D) RMSF structure and profile of CstF50.
Change in radius of gyration during simulation indicates that the structural changes are time dependent and determines the protein structure compactness. Comparative evaluation of radius of gyration throughout the simulation of wild-type and mutant during the early phase of simulation doesn’t show any sharp changes in the structural compactness. After 16ns, there is steep rise in the $R_{\text{gyr}}$ in wild-type which is a large structural juxtapositioning of the ARD domain (Figure: - 3.3.13B). However, in mutant structure, there is sharp increase in the $R_{\text{gyr}}$ at 98ns. This may be due to bulky R-group attached to glutamine.

### 3.3.5 Principal Component Analysis:-

PCA was performed over the 100ns trajectory of BARD1 ARD-BRCT wild-type, Gln564His mutant and CstF50 to analyze the dynamics of proteins in the essential subspace. First three eigenvector was deduced by diagonalizing the covariance matrix and further projected on each other. Projection of Eigenvector 3 on 1 and 2 on 1 indicates large tertiary structural fluctuation in the conformational spaces which indicates that protein is highly dynamic. Furthermore projection of eigenvector 2 on 1 indicates that wild-type and mutant shows large tertiary structural changes during the simulation (Figure: - 3.3.18 A, B).

Similarly, CstF50 also shows conformational fluctuations when projected for eigenvector 3 on 1 and 2 on 1. Observed large tertiary structural changes may be because of the two large interconnecting loops between N-terminal domain and WD40 domain and second in loop between 1st $\beta$-strands of WD40 repeat (Figure: - 3.3.18 C). The Cross-correlation for PCA of wild-type and mutant indicates that differential residual correlation in both the domains. Compared to mutant wild-type shows more positive correlation within the residues of ARD domain and BRCT domain but mutant show comparatively reduced positive correlation in both the domains (Figure:- 3.3.19A, B).
Figure 3.3.18: - Projection of first three eigenvectors for the structure (A, B, C) BARD1 ARD-BRCT wild-type, (D, E, F) mutant and (G, H, I) CstF50.

CstF50 structure shows high value of positive correlation in WD40 repeat domain and comparatively low positively correlated motion within the N-ter domain (Figure:- 3.3.19C).

To understand comparative residual displacements in the subspaces spanned by the first two eigenvectors were performed to understand the fluctuation at the residual level (Figure:- 3.3.18).
Figure 3.3.19: - Cross-correlation for the structure (A) BARD1 ARD-BRCT wild-type (A), (B) mutant and (C) CstF50.

Projection of eigenvector 1 on residue of wild-type, mutant protein N-terminal, and linker region shows comparatively high fluctuations as compared to the BRCT domain (Figure: 3.3.20). Similarly, projection of eigenvector 2 on residues shows a similar pattern of fluctuation in second eigenvector values, which shows the concerted character of residual
fluctuation in the ARD domain and linker region of the wild-type and mutant protein, substantiating results obtained from the NMA and RMSF from MD simulations (Figure: 3.3.20 and 3.3.17).

**Figure 3.3.20:** - Residual fluctuation against first two eigenvectors for the structure (A, B) BARD1 ARD-BRCT wild-type, (C, D) mutant and (E, F) CstF50.
Moreover, CstF50 shows large fluctuation in eigenvector 1 against the residual region related to the N-terminal connecting linker and 1\textsuperscript{st} WD40 interconnecting loop. Eigenvector 2 projection on residues indicates flexible N-terminal domain may be due to the flexible linker and stable WD40 domain (Figure: - 3.3.20 E and F).

3.3.6 Binding mechanism of BARD1 ARD-BRCT and CstF50:

We have used HADDOCK server to understand the binding interactions between CstF50 and BARD1. It is well established that 7\textsuperscript{th} WD40 repeat domain from 395-431 amino acids of CstF50 is required to establish the functional complex with BARD1 [6]. PDBsum software was used to analyze the interactions between BARD1 ARD-BRCT and CstF50 complex [257].

![Figure 3.3.21: - Free-energy calculation for the structure (A) BARD1 ARD-BRCT wild-type and (B) CstF50.](image)

As Gln564His mutation leads to loss in CstF50 binding to BARD1 [258]. Best docked complex was selected, and PDB sum was used to understand the residues involved in the
binding interface. Proteins were simulated for 100ns and minimum energy structure was selected for docking studies (Figure: 3.3.21).

**Figure 3.3.22:** - Binding interface of (A) BARD1 ARD-BRCT wild-type and CstF50 complex, (B) BARD1Gln564His and cstF50 complex.
Docking was performed between minimum energy structure of BARD1 ARD-BRCT, mutant and CstF50 (Figure: - 3.3.22). Similar, docking parameters were used for BARD1 ARD-BRCT mutant and CstF50 docking, and results were analyzed using PDB sum. BARD1 Gln at 564 position forms hydrogen bond with Asn409, Tyr425 and participates in non-hydrogen bonding interaction with Asn409, Tyr425, Pro410, and Thr 408. While BARD1 Thr 562 facilitates other major interactions between BARD1 and CstF50, and forms hydrogen bonding interaction with Tyr425 and non-hydrogen bonding interaction with Arg427, Ser428. Thr430 of CstF50 supports the interaction by forming hydrogen and non-hydrogen bonding interactions with linker residues (Figure: - 3.3.22 A).

CstF50 Thr 430 forms non–hydrogen bonding interactions with BARD1 linker residues like Asn561. From the mutant’s structure, it has been observed that BARD1 His564 has lost all hydrogen and non-hydrogen bonding interactions with the CstF50, 7th WD40 repeat region residues and form non hydrogen bonding interaction with CstF50 Glu100, Asp394 and Leu394 (Figure:- 3.3.22 B). To determine the stability of the docked complex and interaction we simulated the complex for 100 ns (Figure: - 3.3.23). Further simulation of complex structure shows no large changes in RMSD, R_g and RMSF (Figure: - 3.3.23). Hence, it can be concluded that binding of CstF50 binding to BARD1 ARD-BRCT stabilizes the ARD domain flexibility. Hydrogen bonding analysis between the BARD1 ARD-BRCT domain and CstF50 shows that linker region has intact hydrogen bonding interactions with CstF50 C-terminal throughout the simulation (Figure: - 3.3.23 C).
Figure 3.3.23: - (A and B) Comparative RMSD and $R_g$ of BARD1 ARD-BRCT *wild-type* (Black), Gln564His (green), CstF50 (blue) and complex (yellow) respectively, and (C) Hydrogen bonding profile for BARD1 ARD-BRCT-CstF50 complex.

Furthermore, cross-correlation of PCA for BARD1 ARD-BRCT, CstF50 and complex indicates that positive correlation has increased within the N-terminal region of CstF50; it shows increased positive residual correlation within the WD40 repeat and N-terminal residues in CstF50 in the complex.
Figure 3.3.24: - Comparative RMSF (A, B and C) of BARD1 ARD-BRCT *wild-type* (Black), Gln564His (green), CstF50 (blue) and BARD1 ARD-BRCT *wild-type-CstF50 complex* (Red) respectively, (D) RMSF structure for CstF50, (E) complex, and (F) BARD1 ARD-BRCT.
Figure 3.3.25: Comparative cross-correlation for PCA of (A) BARD1 ARD-BARD1BRCT-CstF50 complex, (B) BARD1 ARD-BRCT wild-type and (C) CstF50.
Figure 3.3.26: Free-energy calculation for the ARD-BRCT-CstF50 complex. (A) Free-energy landscape of BARD1 ARD-BRCT wild-type CstF50 complex and (B) Model of structure of the BARD1 ARD-BRCT-CstF50 complex.

Moreover, cross-correlation of PCA for BARD1 ARD-BRCT in complex also shows significant increase in the positively correlated motion within the ARD-domain and BRCT domain (Figure: - 3.3.23). The representative complex structure derived after free-energy calculation shows CstF50 binds to the BARD1 ARD-BRCT domain in such a way that it prevents the motion of ARD domain and keep it in a compact conformation (Figure: - 3.3.26). Further distance calculated between ARD domain and BRCT domain for wild-type and complex structure indicates that CstF50 binding has reduced the domain dynamics of BARD1 ARD-BRCT wild-type protein structure (Figure: - 3.3.27).
Figure 3.3.27: - Distance calculation between ARD domain and BRCT domain for the structure BARD1 ARD-BRCT wild-type- CstF50 complex

In BARD1 ARD-BRCT wild-type ARD and BRCT domain shows a fluctuation of 5.5nm, when alone. However, distance fluctuation calculated for ARD and BRCT domain in complex has drastically reduced demonstrating that CstF50 binding restricts the domain dynamics of ARD domain.

3.4 Crystallization of BARD1 ARD-BRCT

An attempt was made to crystallize the BARD1 ARD-BRCT protein from 7 mg/ml of protein concentration using vapor diffusion method. The protein and mother liquor solution was mixed in 1:1 ratio (1 µl+1µl) and allowed to crystallize at 22°C with 500-µl reservoir solution in vibration free incubator. A clear drop or light precipitation was detected in most of the crystallization conditions. No attempt was made to crystallize CstF50 as refolding process
yielded very a low concentration of protein and furthermore, it precipitated during concentration.

3.5 Conclusions:-

It has been observed that BARD1 ARD-BRCT wild-type and mutant Gln564His are monomeric in nature and mutation does not affect the biophysical property of wild-type protein. CstF50 was found insoluble in bacterial system, therefore efforts were made to purify from inclusion body. NMA analysis and deformation analysis predicted large domain fluctuation between ARD and BRCT domain of BARD1. Molecular dynamics simulation data shows high correlation with small angle X-ray scattering (SAXS) of BARD1 ARD-BRCT domain (425-777aa) [147]. Docking studies indicate that linker region residue Gln564 plays important role in complex formation between BARD1 ARD-BRCT domain and CstF50. Binding of CstF50 to BARD1 ARD-BRCT domain provides rigidity to the structure and moderates the inter-domain motion between ARD and BRCT domain (Figure: 3.3.28).

**Figure 3.3.28:** - Model for BARD1 ARD-BRCT and CstF50 complex formation.