



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](https://doi.org/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: 40

 View related articles [↗](#)

 View Crossmark data [↗](#)

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*}

^aTata Memorial Centre, Advanced Centre for Treatment, Research and Education in Cancer, Kharghar, Navi Mumbai, Maharashtra 410 210, India; ^bMolecular Biology Division, Bhabha Atomic Research Centre, Mumbai 400 085, India; ^cDepartment 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)

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 dynamics 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 $62.7 \pm .25$, $56.91 \pm .16$ and $44.9 \pm .56$ 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

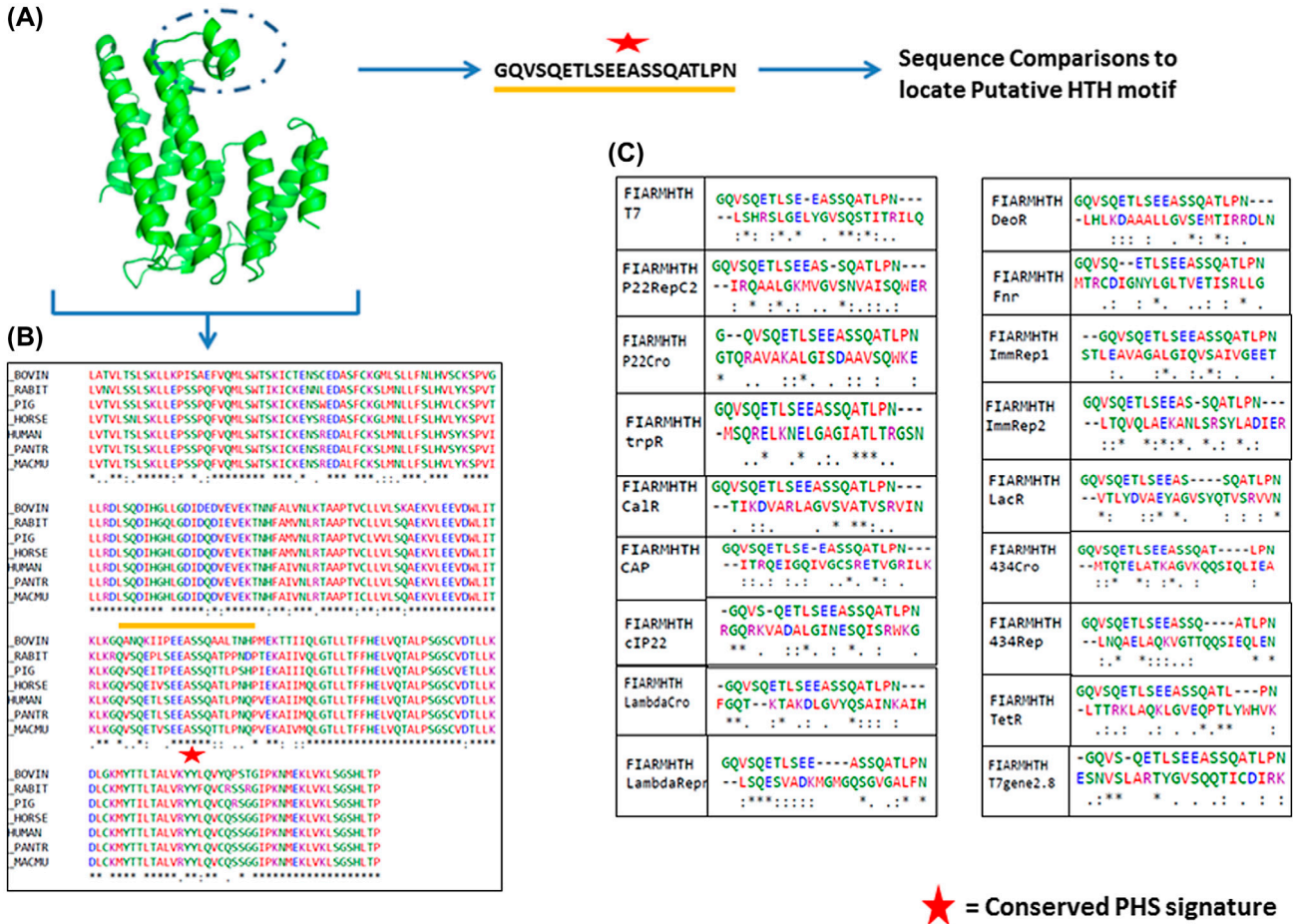


Figure 1. (A) Modelled structure of FANCI ARM repeat protein, (B) Multiple sequence alignment of FANCI ARM repeat, (C) MUSCLE alignment of ARM repeat dynamic region with different known HTH motif present in “Master Sets”.

Table 1. Molecular weight estimation of purified protein.

Theoretical Mol. Wt. (kDa) ^a	Experimentally derived Mol. Wt. (kDa)		
	V_e/V_0^b	Size exclusion chromatography	Mass spectrometry
28.98	1.40 ± .01	26.6 ± 1	29.08 ± .05

Note: V_e/V_0 : Elution volume/Void volume ratio in gel filtration chromatography (superdex 75 16/60).

^aDetermined from ProtParam, ExPASy.

^bDetermined from standards chromatate, aprotinin, lysozyme, carbonic anhydrase, ovalbumin, albumin, ferritin, dextran.

(PDB ID: 3S4W) 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

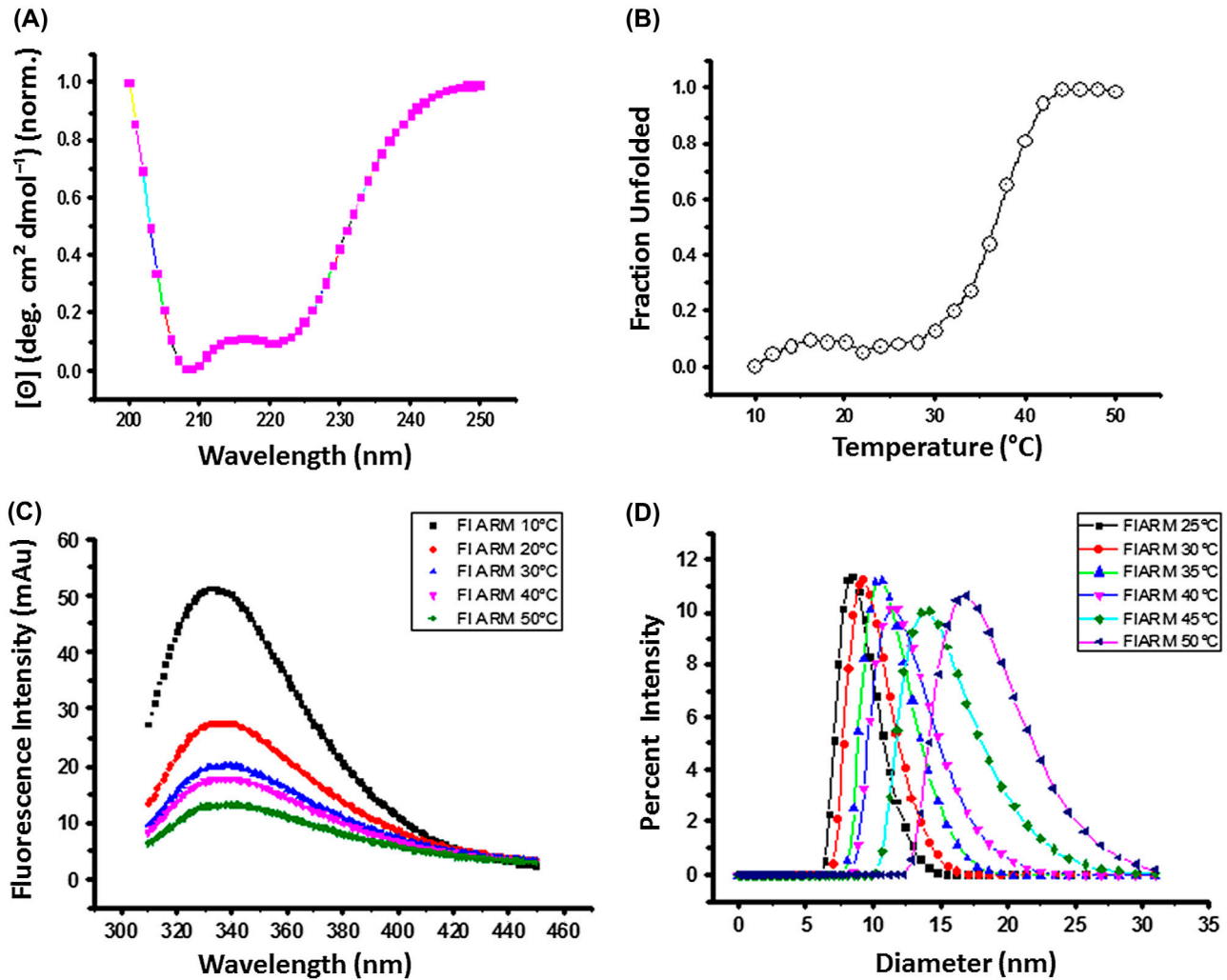


Figure 2. (A) Far-UV, CD spectra of FANCI-ARM repeat, indicating the α -helical nature of protein, (B) Thermal denaturation profile by CD showing protein unfolds at 50°C, (C) Thermal denaturation profile using Fluorescence spectroscopy showing steep decrease in intensity beyond 30°C, (D) Temperature-dependent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

Table 2. Thermal stability of protein.

Circular dichroism	Fluorescence spectroscopy	Dynamic light scattering
$36.12 \pm .5^\circ\text{C}$ (T_m)	$>30^\circ\text{C}^a$	$>30^\circ\text{C}^b$

Note: T_m = Melting temperature.

^aDecrease in intensity and red shift was observed.

^bOligomeric 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 $3.969 \pm 1.712 \mu\text{M}$ (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)).

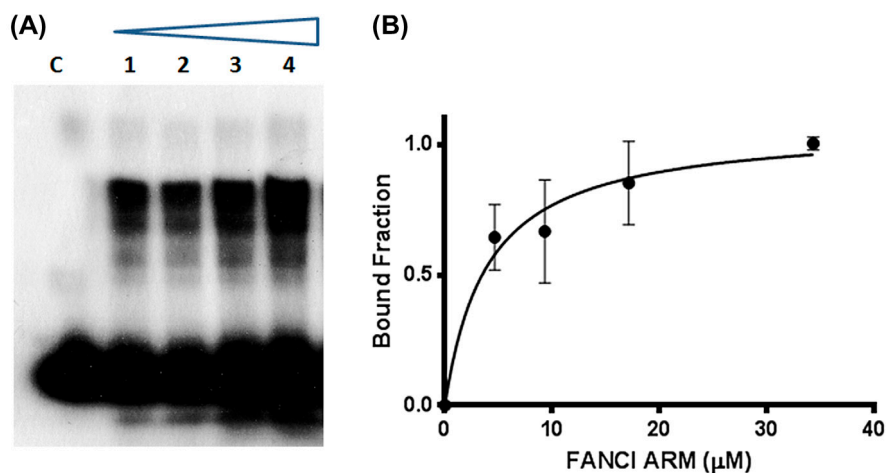


Figure 3. (A) DNA-binding analysis of FANCI ARM repeat with DNA by autoradiography (C = Control probe, 1, 2, 3, and 4 with increasing concentration of protein 4.6, 9.29, 17.14 and 34.29 μM , respectively). (B) Graph plot of bound fractions of protein to DNA.

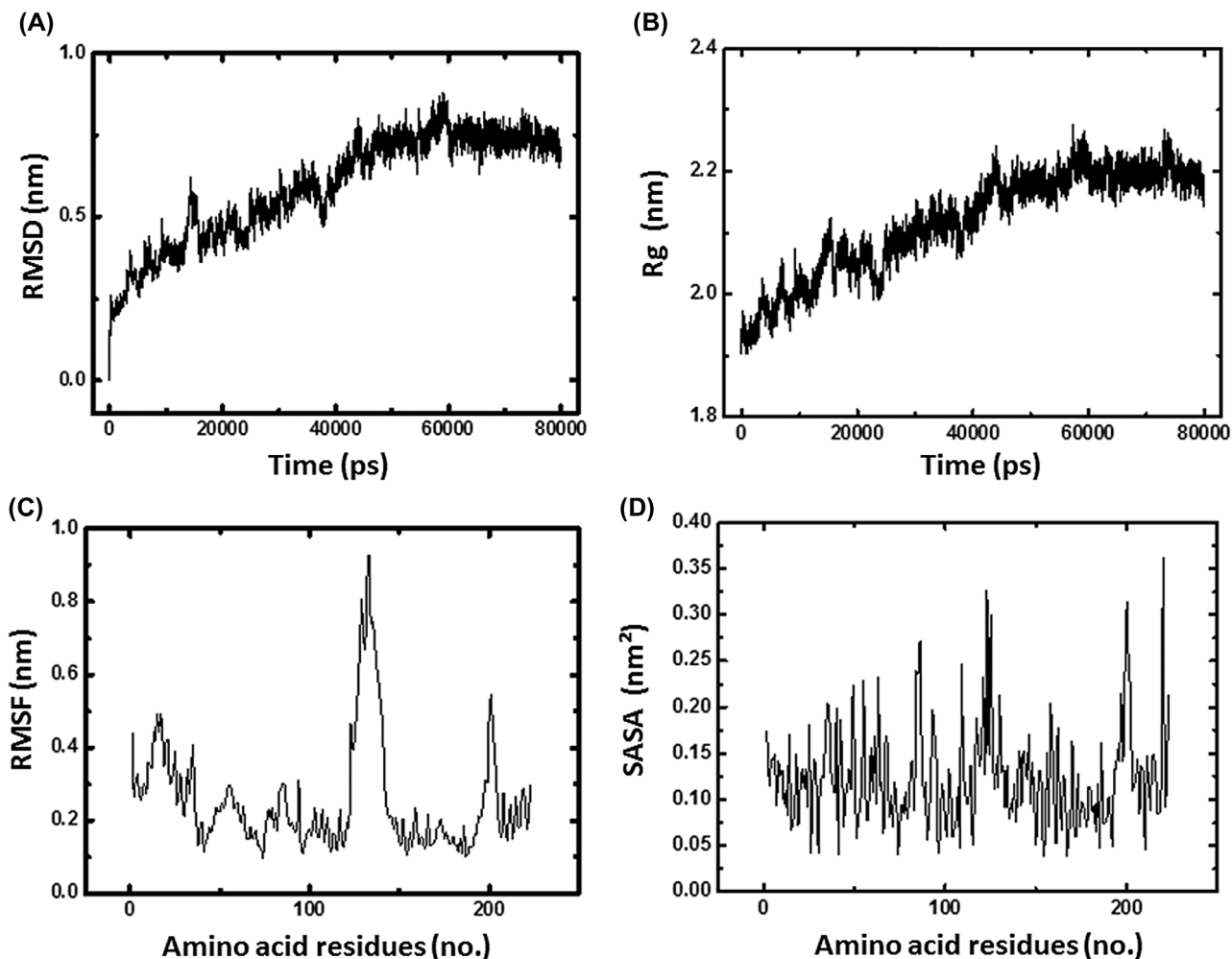


Figure 4. Molecular dynamics simulation profile of FANCI ARM repeat (A) RMSD profile and (B) Radius of gyration (R_g) profile of 80 ns showing structural transitions up to 60 ns, (C) RMSF profile showing large fluctuation of C-terminus, (D) Solvent accessible surface area (SASA) profile of h-FANCI ARM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

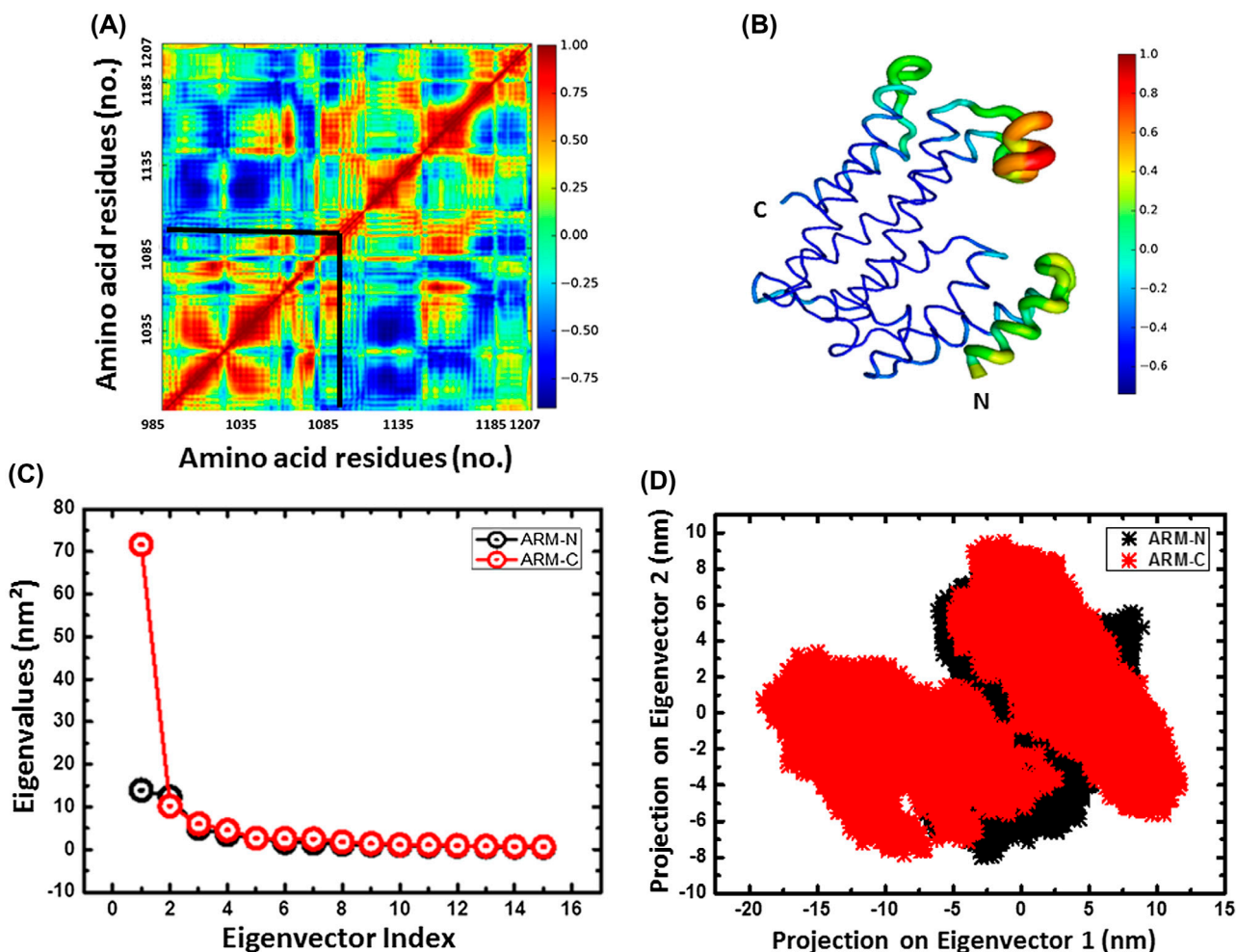


Figure 5. (A) Cross-correlation diagonal matrix of h-FANCI ARM repeat, (B) RMSF sausage profile, (C) Scree plot of N-terminus (1–100) amino acids and C-terminus (100–223) amino acids, (D) PCA profile of Eigenvector 1 and 2 of N-terminus (1–100) amino acids and C-terminus (100–223) amino acids.

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-GAACCTGTACTTTCAGGGTCTAGTCACGGTCTT-ACCAG-3' and 5'-GTCCTCGAGCTATTAGGGGGTCA-GATGAGAACCAG-3', respectively. The forward primer was designed with the TEV protease site having

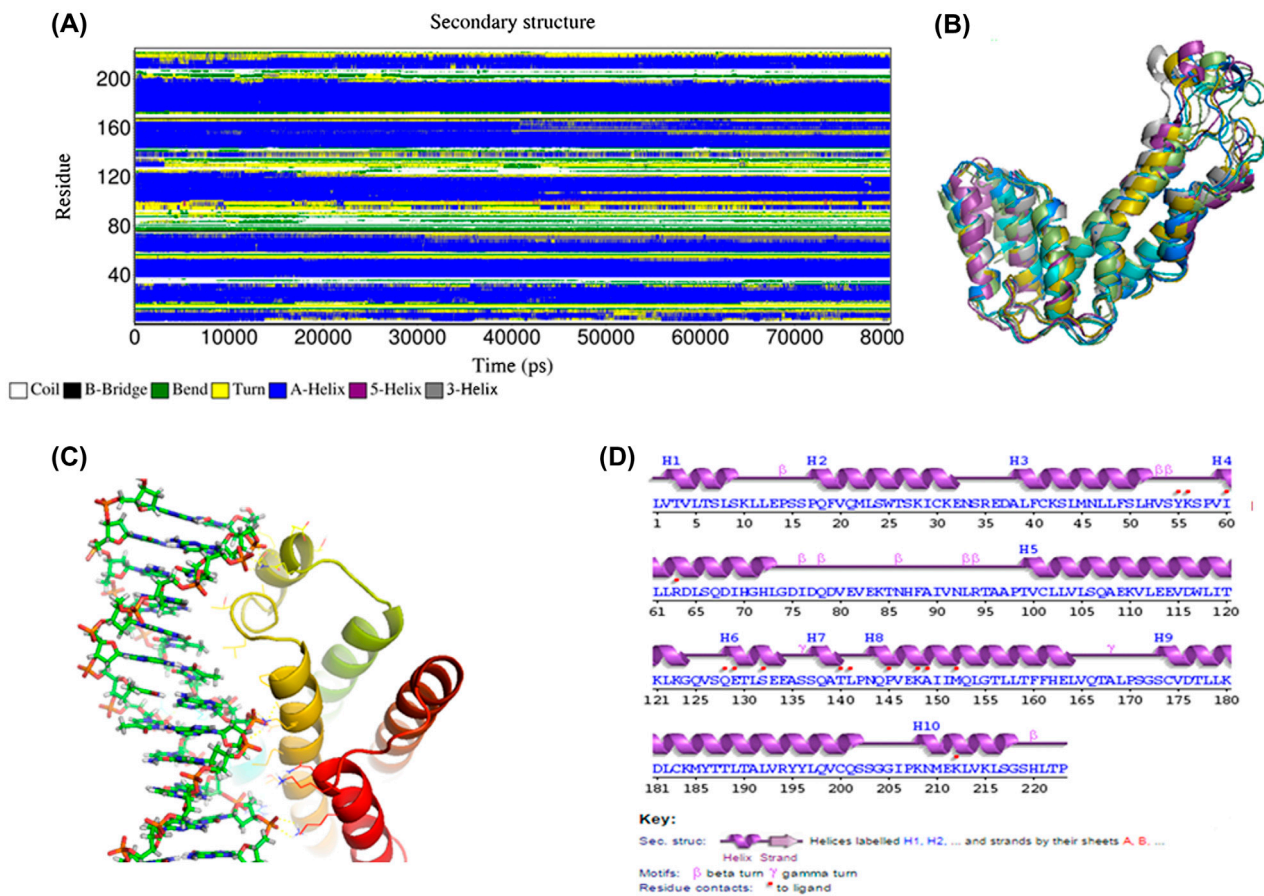


Figure 6. (A) DSSP profile of FANCI ARM repeat protein, (B) Superimposed structures in different time points (10–40 ns) of trajectories showing large structural rearrangement in HTH type motif, (C) Docking profile of FANCI ARM repeat with DNA do-decamer shows major groove binding mode of putative HTH 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 μ l 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 μ g/ μ l 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 μ l 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 μ m) 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 ($\lambda = 190$ –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 $\lambda = 295$ nm. The emission spectra were recorded from $\lambda = 310$ to 450 nm. For thermal and chemical unfolding experiments, 10 μ M 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

[³²P] γ ATP 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 μ l of the reaction containing buffer 35 mM Tris-HCl, pH 7.5, 2.5 mM MgCl₂, 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: Q9NV11) and submitted to Robetta server for 3D model building (Kim, Chivian, & Baker, 2004). Stereochemical 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'GAATTCGGTGCGCATAATGTATATTATGTAAATCATGTCC- CTGCCCAATATAACCAAGCGTATGCAGTAAGCTTCGATC3'	EMSA
2.	82R	5' GATCGAAGCTTACTGCATACGCTTGTTTTATATTGGGGCAGG- 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., Zamboni, M., Polverino de Lauro, 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., Matsuo, 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.