ACS OMEGA

http://pubs.acs.org/journal/acsodf

Article

In Silico and Structure-Based Assessment of Similar Variants Discovered in Tandem Repeats of BRCT Domains of BRCA1 and BARD1 To Characterize the Folding Pattern

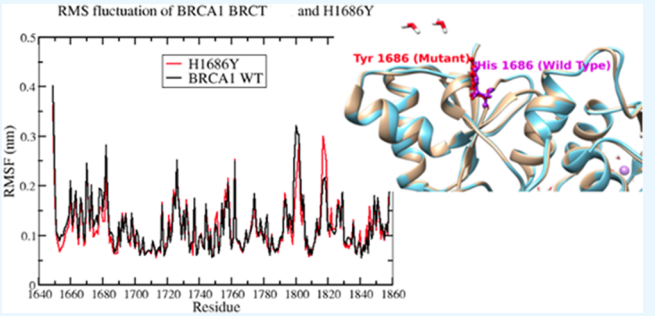
Siddhartha A. Barua, Nabajyoti Goswami, Neha Mishra, Ulka U. Sawant, and Ashok K. Varma*

Cite This: *ACS Omega* 2022, 7, 44772−44785

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: BRCA1 and BARD1 are important proteins in the homologous DNA damage repair pathways. Different genetic variants identified in these proteins have been clinically correlated with the occurrence of hereditary breast and ovarian cancer (HBOC). Variants of unknown significance (VUS) reported in the BRCT domains of BRCA1 and BARD1 substantiate the importance of BRCT domain-containing proteins for genomic integrity. To classify the pathogenicity of variants, *in silico*, structural and molecular dynamics (MD)-based approaches were explored. Different variants reported in the BRCT region were retrieved from cBioPortal, LOVD3, BRCA Exchange, and COSMIC databases to evaluate the pathogenicity. Multiple sequence alignment and superimposition of the structures of BRCA1 BRCT and BARD1 BRCT domains were performed to compare alterations in folding patterns. From 11 *in silico* predictions servers, variants reported to be pathogenic by 70% of the servers were considered for structural analysis. To our observations, four residue pairs of both the proteins were reported, harboring 11 variants, H1686Y, W1718L, P1749L, P1749S, and W1837L variants for BRCA1 BRCT and H606D, H606N, W635L, P657L, P657S, and W762F for BARD1 BRCT. MD simulations of the BRCT repeat regions of these variants and wild-type proteins were performed to evaluate the differences of folding patterns. Root mean square deviation (RMSD), R_g, solvent-accessible surface area (SASA), and root mean square fluctuation (RMSF) of variants showed slight differences in the folding patterns from the wild-type proteins. Furthermore, principal components analysis of H1686Y, P1749S, and W1718L variants of BRCA1 showed less flexibility than the wild type, whereas that of H606D, W635L, and W762F of BARD1 showed more flexibility than the wild type. Normal mode analysis of the energy minima from the simulation trajectories revealed that most of the variants do not show much differences in the flexibility compared to the wild-type proteins, except for the discrete regions in the BRCT repeats, most prominently in the 1798−1801 amino acid region of BRCA1 and at the residue 744 in BARD1.



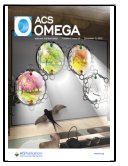
1. INTRODUCTION

Breast cancer is one of the most common types of cancer in women across the world. As per GLOBOCAN 2020 report, ∼11.7% of new cancer cases have been identified, whereas 6.9% of deaths have been reported.[1] A report published by the Centers for Disease Control and Prevention (CDC), USA, unraveled that the most common types of breast cancers are invasive ductal and lobular carcinomas which originate from the milk ducts and lactiferous lobules of the breast.[2] Furthermore, breast cancers can metastasize to other parts of the body such as the brain and lungs.

Hereditary breast and ovarian cancers (HBOC) originate due to pathogenic mutations identified in the genes involved in tumor suppressor functions. Pathogenic variants identified in breast cancer susceptibility gene 1 (BRCA1) and breast cancer susceptibility gene 2 (BRCA2) predispose the carriers to a high risk of breast and ovarian cancers.[3] HBOC cases (66% and 34%) are associated with sequence variants identified in BRCA1 and BRCA2 genes, respectively. The prospective occurrences of BRCA1 and BRCA2-associated HBOC may be suspected in individuals if the first-, second-, or third-degree relatives are likely to have pathogenic sequence variants. The lifetime risk for a woman to develop breast cancer is 12%; however, it increases to 46−87% in individuals who inherited pathogenic variants of BRCA1.[4]

The C-terminal region of BRCA1 contains a dual tandem repeat of a globular domain commonly known as the BRCT domain. The BRCT domain has human homologs in several proteins involved in DNA damage repair and cell-cycle

Received: July 28, 2022
Accepted: November 1, 2022
Published: November 28, 2022



checkpoint control.[5] The dual tandem BRCT domain repeats of BRCA1 comprise approximately 200 amino acids and bind to the proteins containing the phosphoserine-specific pS-X-X-F (pS: phosphoserine) consensus sequence motifs such as ABRAXAS1 (BRCA1 A-complex subunit 1), CtIP (C-terminal binding protein-interacting protein)/RBBP8 (RB binding protein 8), BRIP1/BACH1 (BRCA1-interacting protein and C-terminal helicase 1/BRCA1-associated C-terminal helicase 1), and ACACA (acetyl-CoA carboxylase 1).[6] The interaction with phosphorylated ABRAXAS is required for the recruitment of BRCA1 to sites of DNA damage, particularly to DNA double-strand breaks (DSBs). However, the interaction of the BRCA1 BRCT repeat with phosphorylated CtIP is required for ubiquitination of CtIP.[7]

The RING-RING complex located at the N-terminus of BRCA1 and BARD1 forms a heterodimeric complex which increases the E3 ubiquitin ligase activity of the complex. However, the C-terminus of the BRCT repeat and the ARD (ankyrin repeat domain) of BARD1 have been reported to bind to the CsTF-50 protein.[8] Poly-ADP-ribose (PAR) binding has also been reported for the BARD1 BRCT region as a means of recruitment of the BRCA1-BARD1 complex to the site of DNA damage.[9] No crystal structure of phosphoproteins bound to the BARD1 BRCT repeat has been reported; however, in vitro results predicted binding with phosphopeptides other than the sequences reported for BRCA1.[10] A sulfate ion is located at the phosphoserine binding position in the crystal structure of BARD1 BRCT; however, one of the monomeric structures is located at the phosphospecific binding motif which inhibits binding of BARD1 BRCT with other binding partners. Therefore, either the active dimer structure of BARD1 BRCT does not have reported phosphospecific binding partners or it may adopt different binding consensus sequences.

Cancer-predisposing variants have also been reported in the BRCT repeats of BRCA1 and BARD1 in families with histories of hereditary breast cancer.[11,12] In an effort to functionally annotate the pathogenicity of VUS, multidisciplinary, multifactorial likelihood predictions, in vitro transcriptional assays, as well as epidemiological case−control studies, fluorescence polarization assays, and in-silico and structure-based approaches to evaluate the differences in folding patterns between the wild type and mutants have been explored by different groups.[13−16] Therefore, noting the structural similarity and discrete differences in phosphopeptide-binding grooves, it was decided to evaluate the pathogenicity of VUS identified at sequentially and structurally identical positions of BRCA1 and BARD1.[17] Furthermore, because of the availability of crystal structures of the BRCT repeats, *in-silico*- based pathogenicity predictions and further correlations with molecular dynamics and structure-based approaches were applied for the BRCT domains of BRCA1 and BARD1.

2. MATERIALS AND METHODS

2.1. Multiple Sequence Alignment of the BRCA1 BRCT and BARD1 BRCT. Consensus amino acid sequences of BRCA1 BRCT and BARD1 BRCT repeat regions were retrieved from the National Center for Biotechnology Information (NCBI) and UniProt Knowledge Base (UniProtKB).[18] The three-dimensional coordinates for the structures of the BRCA1 BRCT and BARD1 BRCT were downloaded from RCSB PDB (Research Collaboratory for Structural Bioinformatics) using the PDB IDs 1Y98 and 2NTE, respectively.[19−21]

The amino acid sequence of the BRCT repeat region of human BRCA1 was considered as a query sequence in NCBI's Protein-BLAST to search for homologous sequences from *Pan troglodytes*, *Canis lupus*, *Bos taurus*, *Mus musculus*, *Rattus norvegicus*, *Gallus gallus*, *Xenopus laevis*, *Tetraodon nigroviridis*, *Strongylocentrotus purpuratus*, and *Arabidopsis thaliana* organisms. The multiple sequence alignment (MSA) has been used to analyze the conserved residues from the orthologues of the BRCA1 BRCT repeat.[22] The best matching sequences were selected from the lowest *E*-value and sequences of the corresponding regions to the human BRCA1 BRCT were submitted for MSA using CLUSTALW.[23] The MSA of the human BARD1 BRCT repeat region with its orthologues, in the same set of species, was also performed using CLUSTALW.

2.2. *In Silico* Pathogenicity Prediction Tools. Different missense variants identified on the BRCT regions of BRCA1 and BARD1 were retrieved from the cBioPortal database for cancer genomics[24] and the putative pathogenicity of these variants of unknown significance (VUS) was assessed using 11 different in silico prediction tools. Furthermore, clinical reports and correlation with cancer were also explored, for both the BRCT proteins from the LOVD3,[25] BRCA Exchange,[26] and COSMIC[27] databases. The *in silico* tools used to predict pathogenicity of mutations are as follows:

2.2.1. Polyphen-2 (Polymorphism Phenotyping v2). From the amino acid sequences of the protein, Polyphen-2 extracts a set of gene transcripts from transcriptome databases and also extracts structural annotations. These data are used to build conservation profiles to predict the pathogenicity of the variants. The reported classification for pathogenicity is either benign, possibly damaging, or probably damaging.[28]

2.2.2. MutPred2 (http://mutpred.mutdb.org/). MutPred2 uses amino acid sequence-based prediction methods to predict more than fifty structural and functional properties of the protein containing the variants and uses the MutPred2 score to report the posterior probabilities of pathogenicity for each variant. A cut-off threshold of 0.8 was used to classify mutants as probably pathogenic.[29]

2.2.3. PANTHER-PSEP (Protein Analysis Through Evolutionary Relationship: Position-Specific Evolutionary Preservation). The online pathogenicity prediction tool PANTHER version 9.0 database classifies the nonsynonymous single nucleotide polymorphisms (nsSNPs) as damaging or not. It compares the amino acid sequence of the mutant protein from ancestral sequences to predict the probability of pathogenicity of the missense mutations.[30]

2.2.4. PMut. PMut performs an MSA of the given protein sequence and uses comparative parameters to generate a probability score ranging from 0.0 to 1.0. Scores less than or equal to 0.5 correspond to mutations predicted to be benign, while scores greater than 0.5 correspond to mutations predicted to be pathogenic.[31]

2.2.5. PROVEAN (Protein Variation Effect Analyzer). PROVEAN generates MSAs of the given protein's sequence by using a database of precomputed PROVEAN scores to classify missense mutation as pathogenic or benign.[32]

2.2.6. SIFT (Sorting Intolerant From Tolerant). The SIFT tool performs MSA of the given query protein's amino acid sequences and uses the alignment to generate a SIFT score. Those with a score equal to or lower than 0.05 are considered

https://doi.org/10.1021/acsomega.2c04782
ACS Omega 2022, 7, 44772−44785

pathogenic, while those with a score higher than 0.05 are not considered pathogenic.[33]

2.2.7. PhD-SNP (Predictor of Human Deleterious Single Nucleotide Polymorphisms). PhD-SNP generates a sequence profile based on MSA and support vector machines (SVMs) to generate the pathogenicity prediction score and classify the variants as pathogenic or not pathogenic. Scores at or below 0.5 correspond to benign mutations, while those above 0.5 are considered to be associated with pathogenicity.[34]

2.2.8. SNAP (Screening for Nonacceptable Polymorphisms). SNAP predictions can be accessed from the Meta-SNP predictor tool.

2.2.9. Meta-SNP (Meta-Predictor of Disease-Causing Single Nucleotide Polymorphisms). Meta-SNP is a pathogenicity prediction tool that reports PANTHER-PSEP, PhD-SNP, SNAP, and SIFT predictions and combines them to assign meta-prediction scores.[35]

2.2.10. WS-SNPs and GO [Web Server Implementation of SNPs (Single Nucleotide Polymorphisms) and GO (Gene Ontology terms)]. WS-SNPs and gene ontology (GO) accept solely the wild-type amino acid sequences of the query protein and the list of missense mutations to be assayed, along with GO terms, if desired. WS-SNPs and GO use an SVM and sequence profiles of the variants to classify them as disease (-associated) or neutral.[36]

2.2.11. WS-SNPs and GO 3D. The WS-SNPs and the GO 3D (or WS-SNPs and GO^{3d}) tool accepts the structure of the wild-type protein in addition to the list of missense variants to be assayed, along with GO terms (optionally). WS-SNPs and GO 3D use an SVM and structural features of the variants to classify them as disease (-associated) or neutral. For both WS-SNPs and GO and WS-SNPs and GO 3D, prediction scores greater than 0.5 are considered to be associated with pathogenic missense mutations, while scores equal to or less than 0.5 are not considered to be associated with pathogenicity.[36]

The *in silico*-predicted score can be used to interpret the probability of pathogenicity. None of the *in silico* tools are 100% precise; however, the accuracies of the Polyphen-2, SNPs and GO, SIFT, PROVEAN, SNAP, and Meta-SNP are ~81.4, 84.3, 81.1, 82.1, 80.1, and 74.4%, respectively.[37] The PANTHER and MutPred have 70 and 83% accuracy, respectively, which were tested using 76 sequence variants from a different study.[38] From 151 missense variants reported in BRCA1 and 134 in BRCA2, the accuracies of the PMut and PhD-SNP servers were found to be 85.43 and 68.87%, respectively, for the BRCA1 dataset and 93.28 and 78.36%, respectively, for the BRCA2 dataset.[39] Thus, the accuracies of individual *in silico* pathogenicity predictors vary from server to server and across algorithms used for predictions. Therefore, we have used 11 different servers for better reliability and variants predicted to be pathogenic by ≥70% of servers were considered for structure-based correlations.

2.3. Structural Superimposition of BRCA1 BRCT and BARD1 BRCT Repeat Regions. Structural superimpositions of BRCA1 BRCT and BARD1 BRCT were performed using the coordinates obtained from the PDB IDs 1Y98 and 2NTE, respectively. Considering the availability of full-length coordinates of the nonligand bound structure from PDB ID: 1Y98 was selected for BRCA1 BRCT, whereas chain A, which is the longest chain in PDB ID: 2NTE, was selected for BARD1 BRCT. Hence, structural coordinates of chain A were selected for energy minimization using GROMACS, for both the proteins.[40] A steepest descent minimization integrator was used to perform energy minimization of the systems for 50,000 steps so that the structures representing the minimum energy conformations could be fairly compared. The superimposition of the energy-minimized structures was performed using the MatchMaker tool of UCSF Chimera.[41] The chain pairing of the structures was selected from the best-aligned chains using the Needleman−Wunsch algorithm from the BLOSUM-62 matrix. Furthermore, the set of missense variants corresponding to wild-type residues which were closely superposed in the two proteins and predicted to be pathogenic by ≥70% *in silico* predictors was explored for alterations in folding patterns.

2.4. MD Simulations. MD simulations were performed for variants and wild types of BRCA1 and BARD1 BRCT, using PDB IDs 1Y98 and 2NTE. The missense mutations were generated with the help of the swapaa command of UCSF Chimera. A total of 13 MD simulations, for the five variants H1686Y, W1718L, P1749L, P1749S, and W1837L of the BRCA1 BRCT repeat and the six variants H606D, H606N, W635L, P657L, P657S, and W762F of the BARD1 BRCT, along with two corresponding wild type BRCT proteins, were executed using the GROMACS 2022 package of AMBER 99SB ILDN all-atomic force field at 310 K.[42] Simulations were performed on a workstation comprising 16 physical processor cores within two AMD EPYC processors accelerated by an NVIDIA RTX A4000 graphic processor unit (GPU). Each protein molecule was placed inside a cubic box in such a way that the maximum distance between the outermost atom of the protein and the nearest side of the box was 10 Å. An explicit solvent TIP3P water model was used to saturate the systems with moisture within the simulation boxes. The systems were neutralized by adding counter-ions, followed by which, energy minimizations were performed using the steepest descent minimization integrator, for 50,000 steps. Thereafter, the solvent molecules and counter-ions were allowed to move for two consecutive 500 ps position by restraining the positions of the protein backbones. Temperature and pressure were coupled using V-rescale and Berendsen's coupling algorithms during position restraint equilibration to maintain the temperature (310 K) and pressure (1 bar) constant. All simulations were performed under periodic boundary conditions using a reciprocal grid of 60 × 60 × 60 cells with the fourth-order B-spline interpolation. Finally, production MD runs were executed using a classical Newtonian leap-frog MD integrator, for 100 ns of simulation time, allowing all the molecules to move in all directions. The Parrinello−Rahman barostat was used to maintain a pressure of 1 bar, by employing isotropic pressure coupling with the time constant τ = 2.0 ps and a compressibility of 4.5×10^{-5} bar^{-1} in the x, y, and z components. The particle mesh Ewald (PME) algorithm was used to calculate the electrostatic interactions with a Coulomb cut-off of 12 Å and an interpolation order of 4, within a grid spacing of 1.6 Å. The van der Waals (vdw) forces were also analyzed using a distance cut-off of 12 Å. The integration time step for the simulations was set to 2 fs and the coordinates were stored every 10 ps. All bond lengths were constrained using the LINCS algorithm and analyses were carried out using the GROMACS package.[40,43−48] The plots were generated using GRACE (1991−1995, Paul J Turner, Portland, 1996−2007, Grace Development Team).

2.5. PCA and Normal Mode Analysis. PCA was performed on a 3 N × 3 N (N represents the number of atoms in the backbone) covariance matrix of the 3D

https://doi.org/10.1021/acsomega.2c04782
ACS Omega 2022, 7, 44772−44785

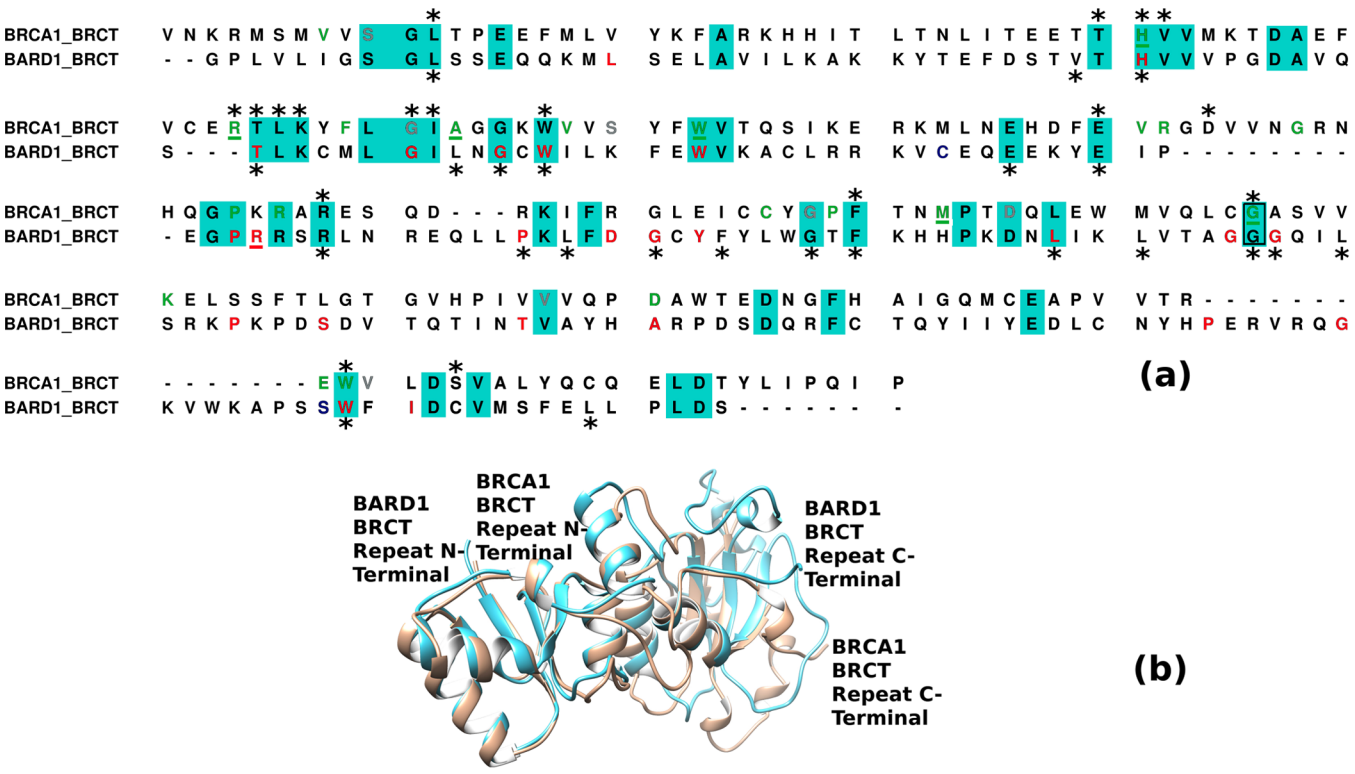


Figure 1. (a) Annotated MSA of human BRCA1 and BARD1 BRCT repeats. **Blue highlight:** Identical residues. **Red font:** *In silico*-predicted pathogenic missense mutations in BARD1. **Green font:** *In silico*-predicted pathogenic missense mutations in BRCA1. **Dark blue font:** Clinically reported cancer-associated missense variants in BARD1. **Gray font:** Positions hit by clinically reported cancer-correlated missense variants in BRCA1. **Underlined:** Both *in silico* prediction and clinical reports are applicable. **Asterisks:** Evolutionarily conserved residues. **Boxed-in resides:** Conserved residues with other dual tandem BRCT repeat-containing human proteins. (b) Structural superposition of the processed versions of the PDB files 1Y98 and 2NTE depicting the structures of the BRCT repeats of BRCA1 (in tan) and BARD1 (in sky blue), respectively. The figure was made using UCSF Chimera and GNU Image Manipulation Program (GIMP).

coordinates of each protein's backbone, using the gmx covar tool of GROMACS. The first six eigenvectors and their corresponding eigenvalues were extracted from each trajectory, out of which, the first two eigenvectors, known as principal components (PCs) 1 and 2, were analyzed by projecting them on each other. The spread of the data points in each of these graphs depicting the projections on the first two eigenvectors, in phase space, was quantified as the area covered, in nm^2. This quantification was performed using the graphics software ImageJ, with the areas being selected using ImageJ's freehand selection tool.[49] From the projection of the first two eigenvectors (PCs 1 and 2), the Gibbs' free energy landscape (FEL) was evaluated by the Boltzmann inverting multidimensional histogram method implemented in the GROMACS' gmx sham tool. For each simulation, a three-column data file, comprising the simulation time in nanoseconds, PC1 and PC2 in the three columns, was supplied to the program to produce another three-column matrix data file of Gibbs' free energy (ΔG) for PC1 and PC2 in the X PixMap (xpm) format. The number of bins for the energy landscape was set to 100 in the *x*, *y*, and *z* directions, for the energy level of 60, for all datasets. Matrix data files (xpm) were converted to 3D contour plots with the help of an add-on utility script, xpm2txt.py, and the GNU Plot program.[50] The structures contained within the conformational basins for the Gibbs' free energy minima were extracted to perform normal mode analysis (NMA) using the ProDy plugin under the NMWiz analysis tool of the VMD software.[51,52]

NMA was performed on the MD simulation trajectories containing the structures with ΔG = 0 kJ mol^{-1} from the BRCA1 and BARD1 BRCT wild-type and mutant proteins. This yielded a plot, called a "porcupine" plot, depicting vectors in the form of global modes distributed over the backbone of the representative structure of each set of energy minima belonging to a particular wild type or mutant protein. The global modes represent reconfigurations, along the directions of the principal axes, that are most easily accessible on the multidimensional energy landscape. The magnitudes of these vectors are referred to as PC square fluctuations. In order to facilitate the comparison of the values of these PC square fluctuations between the wild-type and mutant proteins, the PC square fluctuations were normalized between 0 and 1, yielding normalized PC square fluctuation (NPSF) plots. Relative NPSF plots were obtained by subtracting the NPSF value at each residue position of the *X*-axis of a wild-type NPSF plot from the NPSF value of respective mutants.

2.6. Analysis of Time-Averaged SASA of Hydrophobic Residues. Core hydrophobic side-chain-bearing residues were selected based on 0% solvent accessibility relative to the theoretical maximum solvent accessible surface area (SASA) of the respective residue. A spherical probe, of 1.40 Å radius, was employed to gauge the solvent accessibility, using a grid density of 240 grid points per atom. A GROMACS index file was then created for these most buried residues, with each residue forming its own group, followed by the extraction of the SASA profile of each group, using the gmx sasa tool of GROMACS. The time-averaged SASA value of each residue's

https://doi.org/10.1021/acsomega.2c04782
ACS Omega 2022, 7, 44772−44785

SASA profile was calculated and the time-averages, of the most buried residues bearing hydrophobic side chains, were summed up to yield the cumulative time-averaged SASA for the BRCA1 BRCT and BARD1 BRCT repeats.

3. RESULTS AND DISCUSSION

3.1. Characterizing Frequencies of BRCA1 BRCT and BARD1 BRCT Amino Acids. BRCA1 BRCT comprises 10.3% positively charged, 12.2% negatively charged, and 41.1% hydrophobic side-chain-bearing, and 58.9% hydrophilic side-chain-bearing residues, whereas the BARD1 BRCT comprises 13.3% positively charged, 11.4% negatively charged residues, 39.5% hydrophobic side-chain-bearing residues, and 60.5% hydrophilic side-chain-bearing residues. To our observation, 46 residues were sequentially aligned and identical among BRCA1 and BARD1 BRCT proteins (Figure 1a), 26 missense variants at 20 different positions of BRCA1 and 28 missense variants at 24 different positions of BARD1 were predicted to be pathogenic by ≥70% in silico prediction tools (Table 1). Furthermore, very few clinical reports on correlation with cancers with family pedigree were found for the missense mutations of the BRCA1 BRCT at sequentially identical positions of the BARD1 BRCT. Our literature search could find seven reported family pedigrees showing the cosegregation of the variants with the occurrence of cancer in BRCA1 BRCT at positions that were sequentially identical with BARD1 BRCT. However, no family pedigrees were found for missense variants occurring at any of the identical BARD1 residues (Table 2).

Table 1. List of Missense Variants of the BRCA1BRCT and BARD1 BRCT Repeat Regions That Have Been Predicted To Be Pathogenic from ≥70% of In Silico Predictors

s. no.	residue position in BRCA1	unique BRCA1 mutants	residue position in BARD1	unique BARD1 mutants
1	1653	V1653A	585	L585P
2	1686	H1686Y	606	H606D and H606N
3	1699	R1699L and R1699W	617	T617I
4	1704	F1704S	623	G623W
5	1708	A1708E, A1708P, A1708T, and A1708V	627	G627R
6	1713	V1713A	629	W629S
7	1718	W1718L	635	W635L
8	1736	V1736A	657	P657L and P657S
9	1737	R1737T	658	R658C
10	1743	G1743V	669	P669Q
11	1749	P1749L and P1749S	673	D673Y
12	1751	R1751L and R1751Q	674	G674V
13	1768	C1768Y	676	Y676C
14	1771	P1771L	691	L691F
15	1775	M1775I	698	G698D
16	1788	G1788V	700	G700C and G700S
17	1793	K1793N	707	P707S
18	1813	D1813H	711	S711R
19	1836	E1836K	719	T719R
20	1837	W1837L	724	A724V
21			747	P747Q
22			753	G753C and G753D
23			762	W762F
24			764	I764T

3.2. BRCT Repeats of BRCA1 and BARD1 Are Sequentially Different but Have Similar Structural Folding. In order to study the evolutionary conservation of the residues between the BRCT regions of BRCA1 and BARD1 and their respective orthologues, ranging from humans to *Arabidopsis thaliana*, MSA was performed using CLUSTALW. To our observations, only ∼10% residues are conserved from humans to *A. thaliana*. However, the BRCT regions of human BRCA1 and BARD1 show ∼21.5% sequence identity (Figure 1a).

The two BRCT repeats have 29 sequentially aligned, identical and structurally superimposed residue pairs of BRCA1/BARD1 such as S1655/S575 within a β-strand/random coil, G1656/G576 and L1657/L577 within random coils, A1669/A589 within α-helices, T1685/T605 within random coils, H1686/H606, V1687/V607, and V1688/V608 within β-strands, T1700/T617 within random coils, L1701/L618, L1702/L619, L1705/L622, G1706/G623, and I1707/I624 within α-helices, G1710/G627 within random coils, W1712/W629 within β-strands, W1718/W635 and V1719/V636 within α-helices, E1731/E648 within α-helices, E1735/E652 within random coils, P1749/P657 within α-helices, F1761/F672 and G1763/G674 within random coils, F1772/F683 and P1776/P687 within random coils, G1788/G699 within random coils, and W1837/W762, D1840/D765, and V1842/V767 within α-helices (Figure 1b, S1). However, of these, only four residue pairs harbor missense mutations predicted to be pathogenic by ≥70% of *in silico* prediction tools. These residue pairs for BRCA1/BARD1 are H1686/H606, W1718/W635, P1749/P657, and W1837/W762 (Figure 2a,b). Therefore, MD simulations of the BRCT domains of the BRCA1 and BARD1 proteins and the 11 corresponding missense variants, H1686Y, W1718L, P1749L/S, and W1837L in BRCA1 and H606D, H606N, W635L, P657L/S, and W762F in BARD1, were performed to evaluate the alterations in structures.

3.3. Variants Reported in the BRCA1 BRCT Are Structurally More Dynamic Than those of the BARD1 BRCT. Root mean square deviation (RMSD) values of the BRCA1 BRCT wild type over the 100 ns of the simulation revealed that the wild type has the lowest average RMSD (0.198 nm), whereas the variant P1749L has the highest average RMSD (0.220 nm), followed by P1749S (0.219 nm), W1837L (0.219 nm), H1686Y (0.206 nm), and W1718L (0.201 nm) (Figure 3a). Furthermore, the 100 ns simulation trajectory of the BARD1 BRCT wild type shows a high average RMSD (0.223 nm) than P657L (0.218 nm), H606D (0.217 nm), P657S (0.202 nm), and H606N (0.193 nm) but low average RMSD for variants W635L (0.261 nm) and W762F (0.236 nm) (Figure 4a, Table 3).

This implies that the BRCA1 BRCT mutants H1686Y, W1718L, P1749L, P1749S, and W1837L and the BARD1 BRCT repeat mutants W635L and W762F are more dynamic than the respective wild-type proteins, whereas the BARD1 BRCT mutants H606D, H606N, P657L, and P657S are less dynamic than the wild type. It has been reported that the variants with simulation trajectory having average RMSD values >0.3 were deleterious, and those with ≤0.3 were tolerated. The cut-off was based on the simulation trajectory

https://doi.org/10.1021/acsomega.2c04782
ACS Omega 2022, 7, 44772−44785

Table 2. Residue Positions with Sequential Identity in the BRCT Repeats of BRCA1 and BARD1[a]

name of residue	position no. of residue in BRCA1	known mutants for BRCA1 BRCT[b]	position no. in BARD1	known mutants for BARD1 BRCT[b]
S	**1655**	F (Mohammadi et al.,[53]), P, and A	575	G
G	1656	D	576	R and V
L	1657		577	
E	1660	Q and A	580	K
A	1669	T, S, and V	589	T and V
T	1685		605	
H	**1686**	Q (Giannini et al.,[22]), P, R, and L	606	D and N
V	1687	I and A	607	I and D
V	1688	L, I, G, and F	608	G
D	1692	E, V, G, A, Y, H (BRCA exchange), and N (CBIOPORTAL)	612	V (LOVD3)
A	1693	V, G, D, S, P, and T	613	
T	1700	A and I	617	I
L	1701	R, P, Q, V, M	618	
K	1702	N	619	
L	1705	P	622	F
G	**1706**	E (Osorio et al.,[54]), R, and A (Phelan et al.,[14])	623	W and E
I	1707	M, S, T, N, F, V, and L	624	V
G	1710	R	627	R
W	1712	C, L, S, G, and R	629	S and C
W	**1718**	C (Mirkovic et al.,[55]), R, G, and S	635	L and R
V	1719	A	636	I
E	1731	D, V, G, Q,and K	648	
E	1735	K	652	
G	1748	V, A, D, C, R, and S	656	
P	1749	A and R	657	L and S
R	1751	Q	659	
R	1753	T	661	G
K	1759	E and R	670	E
F	1761	S	672	
G	1763		674	
G	**1770**	V (Tudini et al.,[56]) and W	681	
F	1772	C	683	
P	1776	H and L	687	S
D	1778		689	
L	1780	P	691	F
V	1784	G, A, E, L, and I	695	I and D
G	**1788**	V (Easton et al.,[13]), D, C, and S	699	E
V	**1809**	F (Phelan et al.,[14]) and A	720	A
D	1818	G	729	H and Y
F	1821	L, C, S, Y, V, and I	732	L, S, and V
E	1829	D, V, G, A, Q, and K	740	
W	1837	L, R, G, and C	762	F and R
D	1840	E, V, G, A, Y, H, and N	765	G
V	1842	G, A, E, L, and I	767	A, M, and L
L	1850	R, P, Q, V, and M	775	P
D	1851	E, V, G, A, Y, H, and N	776	

[a]References for evidence of putative pathogenicity of variants based on family pedigree analysis are marked in bold letters. [b]The reference for a clinical report of co-occurrence with cancer based on family pedigree analysis refers only to the specific unique mutant.

average RMSD values of known/standard pathogenic sequence variants of the BRCA1 BRCT.[57]

3.4. Root Mean Square Fluctuation (RMSF) and Radius of Gyration (R_g) Determine Discrete Folding Patterns and Extents of Compactness for the BRCA1 BRCT and BARD1 BRCT. RMSF values of the BRCA1 BRCT wild type over the period of the 100 ns simulation show that the wild type has an average RMSF value (0.120 nm) higher than those of W1718L (0.119 nm) and H1686Y (0.116 nm) but smaller than those of W1837L (0.139 nm), P1749L (0.136 nm), and P1749S (0.132 nm) variants (Figure 3b). RMSF analyses revealed that the BRCA1 BRCT variants P1749L, P1749S, and W1837L are more flexible, on an average, than the wild-type protein, whereas the variants H1686Y and W1718L are less flexible than their respective wild-type proteins.

The BRCA1 BRCT wild type has an average R_g value (1.943 nm) higher than those of H1686Y (1.942 nm), P1749L (1.935 nm), and P1749S (1.933 nm), but lower than those of the variants W1718L (1.947 nm) and W1837L (1.944 nm) (Figure 3c). R_g analysis revealed that the BRCA1 BRCT repeat variants H1686Y, P1749L, and P1749S are more compact, whereas the variants W1718L and W1837L are less compact, on an average, than the wild type.

The BARD1 BRCT repeat wild type has an average RMSF value (0.131 nm) which is equal to that of the average RMSF of H606D (0.131 nm) and higher than those of P657S (0.124 nm) and H606N (0.116 nm), but less than those of W635L (0.144 nm), P657L (0.132 nm), and W762F (0.132 nm) (Figure 4b). The BARD1 BRCT variants W635L, P657L, and W762F are more flexible, on an average, than the WT, while the variants H606N and P657S are less flexible, on an average, than the WT. However, the variant H606D was found to be as flexible as the wild-type protein.

The BARD1 BRCT repeat has an average R_g value (1.955 nm) higher than those of H606D (1.951 nm), H606N (1.948 nm), W762F (1.948 nm), P657S (1.947 nm), and W635L (1.941 nm), but lower than the average R_g of P657L (1.956 nm) (Figure 4c). R_g analysis showed that the BARD1 BRCT repeat variants H606D, H606N, W635L, P657S, and W762F are more compact and the variant P657L is less compact than the wild type.

RMSF analyses revealed that the BRCA1 BRCT variants P1749L, P1749S, and W1837L and the BARD1 BRCT variants W635L, P657L, and W762F are more flexible, on an average, than their respective wild-type proteins, whereas the BRCA1 BRCT repeat variants H1686Y and W1718L and the BARD1 BRCT variants H606N and P657S are less flexible than their respective wild-type proteins. However, the BARD1 BRCT repeat variant H606D was found to be as flexible as wild-type protein.

Time-evolution profiles of RMSF for the WT and variant BRCT repeats show that the RMSF profiles of the variants are similar to those of the respective wild-type BRCT repeats, with the exceptions being P1749L and W1837L, having slightly higher RMSF values than the BRCA1 BRCT repeat at the

https://doi.org/10.1021/acsomega.2c04782
ACS Omega 2022, 7, 44772−44785

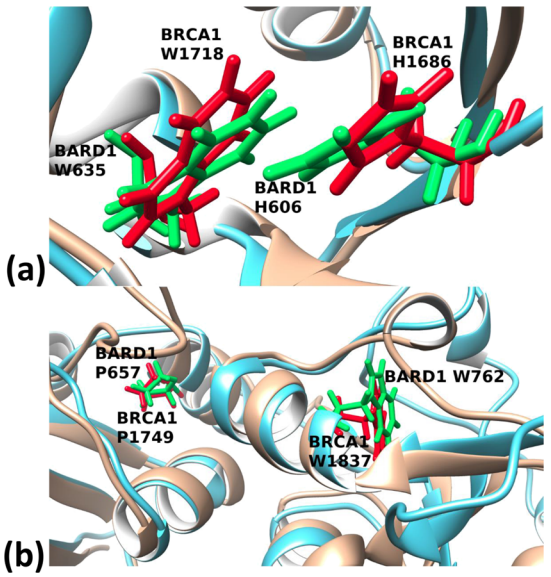


Figure 2. (a) Superposition between the longest polypeptide chains of the BRCA1 and BARD1 BRCT repeats, showing close superposition between similar residue positions, H1686/H606 and W1718/W635 of BRCA1/BARD1, respectively, associated with missense variants predicted to be pathogenic by 70% or more of the *in silico* pathogenicity predictors employed. (b) Same kind of the superposition shown for residue pairs P1749/P657 and W1837/W762 of BRCA1/BARD1. The figure was made using UCSF Chimera and GIMP.

amino acid regions ∼1690−1700 and ∼1846−1855, respectively.

Moreover, R_g analysis revealed that the BRCA1 BRCT repeat variants H1686Y, P1749L, andP1749S and the BARD1 BRCT repeat variants H606D, H606N, W635L, P657S, and W762F are more compact than their respective WT proteins, whereas the BRCA1 BRCT repeat variants W1718L and W1837L and the BARD1 BRCT repeat variant P657L are less compact than their respective wild-type proteins.

3.5. Average SASA of BRCA1 BRCT and BARD1 BRCT Wild Type and Mutants. The average SASA of the BRCA1 BRCT (114.225 nm^2) was found to be less than those of the variants P1749L (115.386 nm^2), P1749S (115.244 nm^2), W1837L (115.050 nm^2), W1718L (114.879 nm^2), and H1686Y (114.869 nm^2) The time-dependent SASA profiles of the BRCA1 BRCT repeat variants are not much different from that of the wild type, except for slight deviations in P1749S (Figure 3d).

However, the average SASA of the BARD1 BRCT repeat wild type (116.456 nm^2) was found to be more than that of H606D (116.156 nm^2) and less than those of W635L (118.907 nm^2), P657L (118.825 nm^2), W762F (117.999 nm^2), P657S (117.048 nm^2), and H606N (116.697 nm^2) (Table 3). The SASA values of the BARD1 BRCT repeat variants, over the 100 ns of simulation time, do not differ much from those of the wild type, except for slight deviations in the cases of P657L, W635L, and W762F (Figure 4d).

The time-evolution profiles of the SASA for the wild-type and BRCT variants show that the variants have very similar

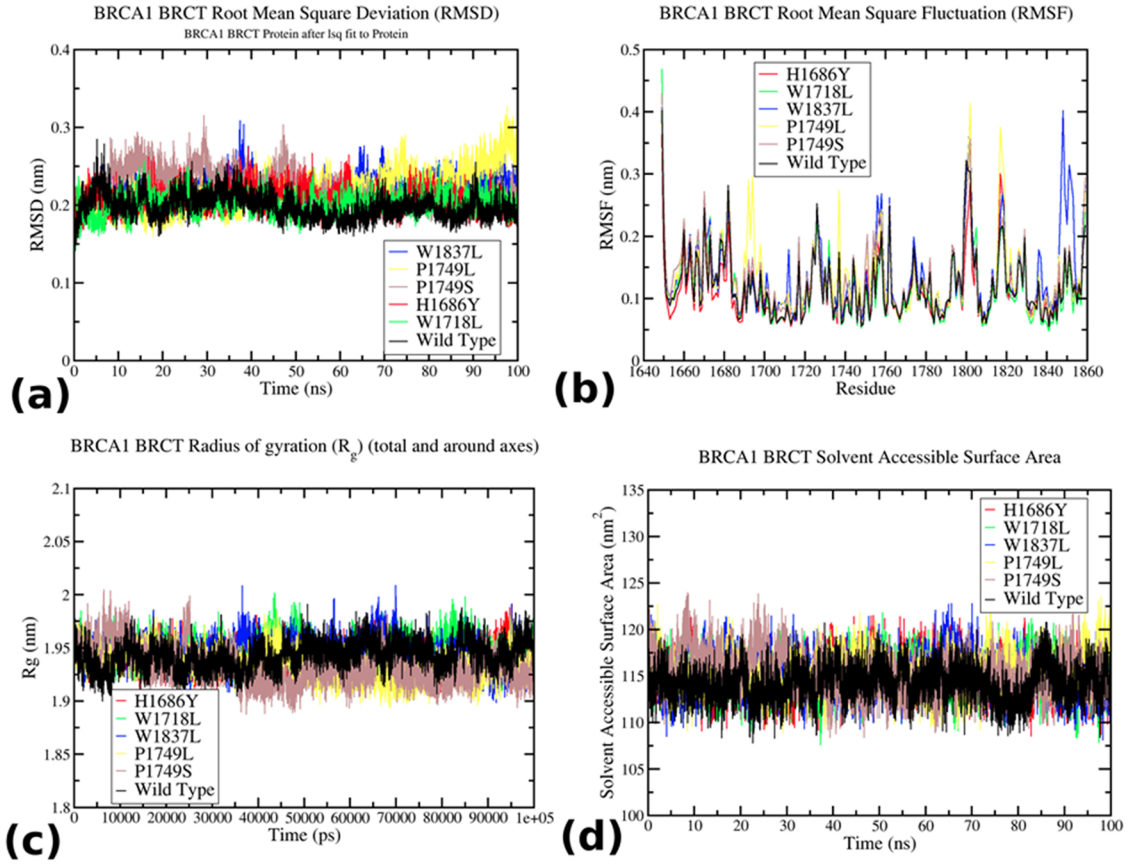


Figure 3. Time-evolution profiles of the following parameters for the BRCA1 BRCT repeat wild type and variants (a) RMSD, (b) RMSF, (c) R_g, and (d) SASA.

https://doi.org/10.1021/acsomega.2c04782
ACS Omega 2022, 7, 44772−44785

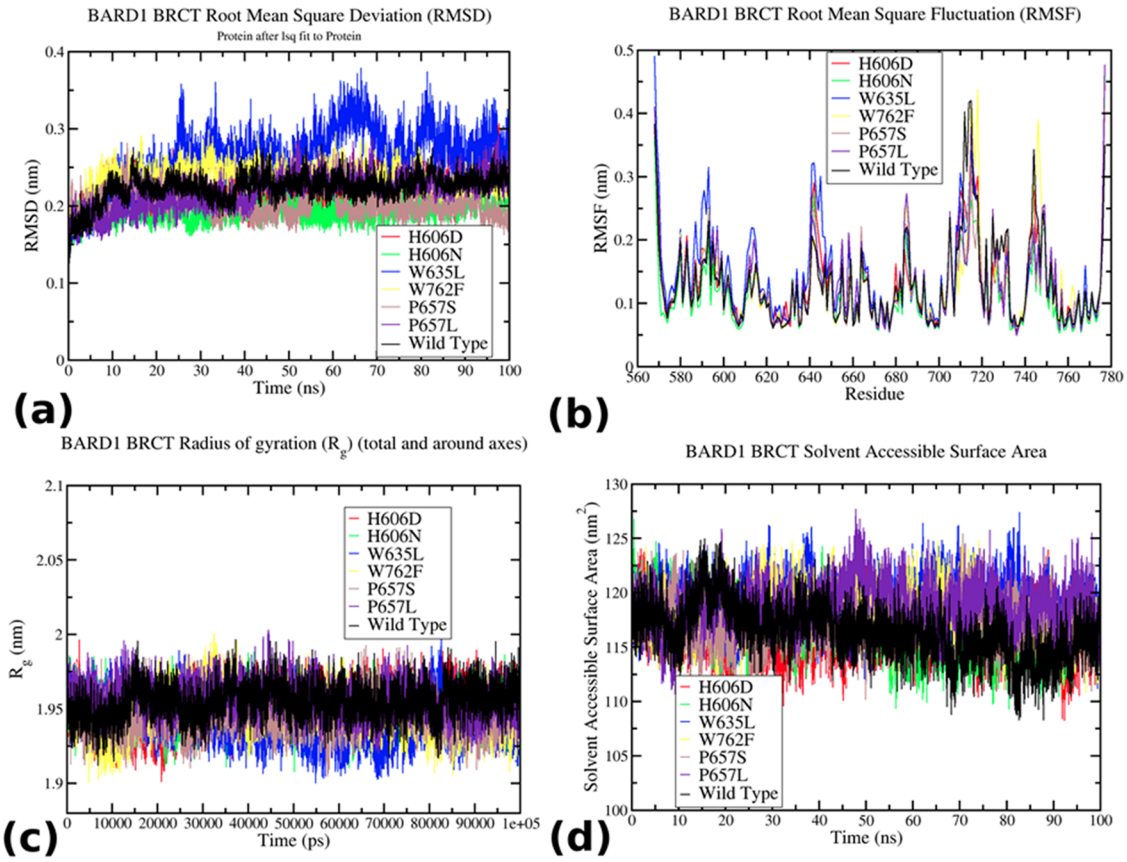


Figure 4. Time-evolution profiles of the following parameters for the BARD1 BRCT repeat wild type and variants (a) RMSD, (b) RMSF, (c) R_g, and (d) SASA.

Table 3. Average RMSD, RMSF, R_g, and SASA Values for the Simulation Trajectories of the Wild Type and Mutants of BRCT Repeats of BRCA1 and BARD1

protein	average RMSD (nm)	average RMSF (nm)	average R_g (nm)	average SASA (nm^2)
BRCA1 BRCT repeat WT	0.198	0.120	1.943	114.225
H1686Y	0.206	0.116	1.942	114.869
W1718L	0.201	0.119	1.947	114.879
P1749L	0.220	0.136	1.935	115.386
P1749S	0.219	0.132	1.933	115.244
W1837L	0.219	0.139	1.944	115.050
BARD1 BRCT repeat WT	0.223	0.131	1.955	116.456
H606D	0.217	0.131	1.951	116.156
H606N	0.193	0.116	1.948	116.697
W635L	0.261	0.144	1.941	118.907
P657L	0.218	0.132	1.956	118.825
P657S	0.202	0.124	1.947	117.048
W762F	0.236	0.132	1.948	117.999

Table 4. Cumulative Time-Averaged SASA of Hydrophobic Side-Chain-Bearing Residues, Wherein These Residues Were Initially Most Buried, Just Prior to MD Simulations

protein	cumulative time-averaged SASA of most buried hydrophobic residues (nm^2)
BRCA1 BRCT wild type	51.231
H1686Y	51.235
W1718L	41.802
P1749L	49.236
P1749S	49.234
W1837L	39.450
BARD1 BRCT wild type	48.165
H606D	48.158
H606N	48.158
W635L	26.452
P657L	48.179
P657S	50.262
W762F	44.980

SASA profiles to those of their respective wild-type proteins. The only exceptions are P1749S, with very slightly higher SASA values than the BRCA1 BRCT repeat, for the first ~30 ns of the simulation and P657L, with slightly higher SASA values than the BARD1 BRCT repeat wild-type, for the last ~60 ns of the simulation.

Furthermore, to check whether the buried residues with hydrophobic side chains contribute to SASA values of the variants, the sum of the simulation time averages of the SASA values of all the most buried, hydrophobic side chain-bearing residues were evaluated for the BRCT repeats of BRCA1 and BARD1. It has been found that most of the variants had comparatively similar cumulative average SASAs for the most buried hydrophobic residues, compared to the respective wild-type proteins except for the BARD1 BRCT variants P657S, W762F, and W635L with slightly higher, slightly lower, and very much lower cumulative time-averaged SASA, respectively. For the BRCA1 BRCT repeat, the notable exceptions to this

https://doi.org/10.1021/acsomega.2c04782
ACS Omega 2022, 7, 44772−44785

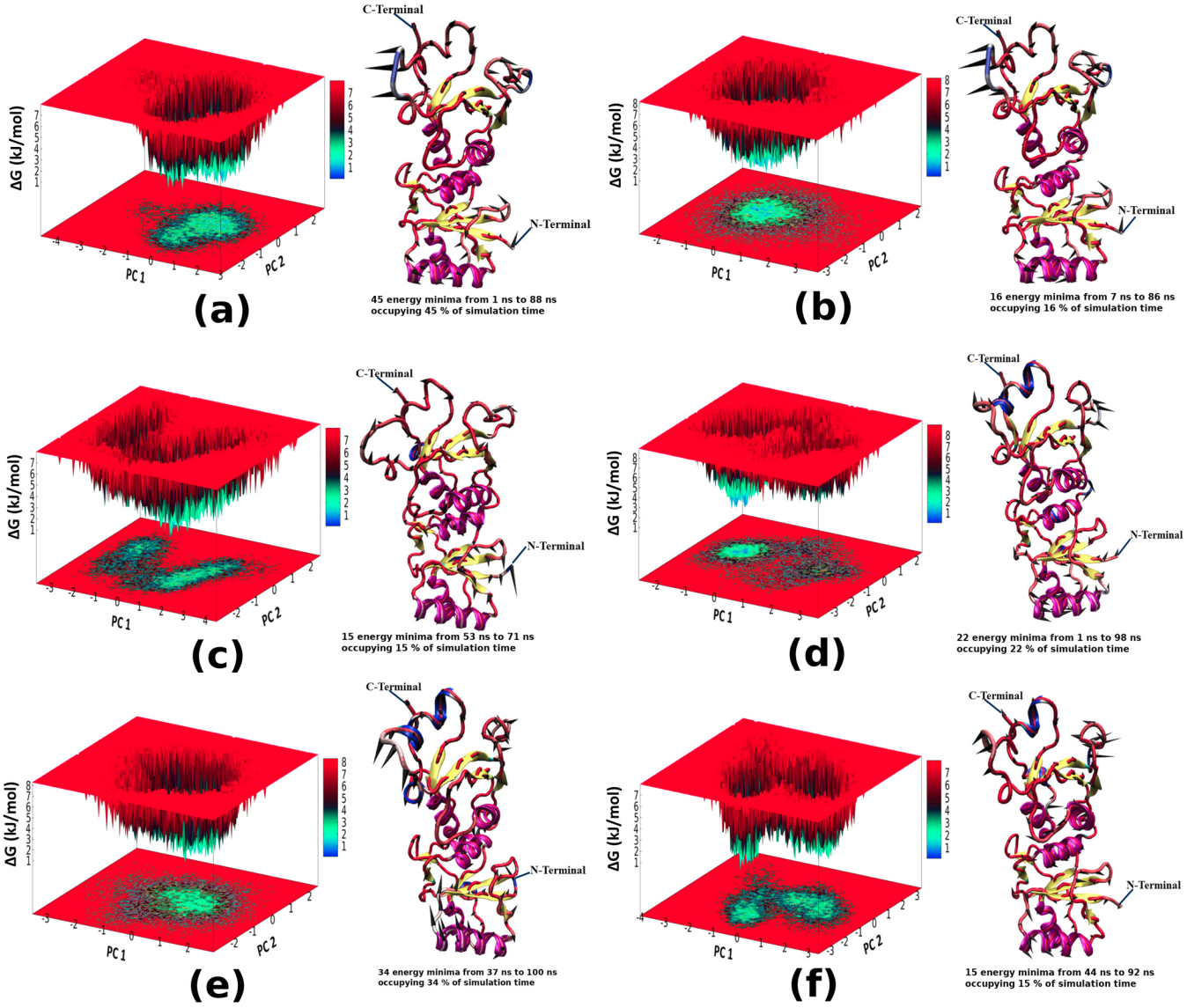


Figure 5. In each subfigure above, on the left is shown the Gibbs' FEL superimposed over the projections on the first two eigenvectors, in phase space, for the BRCA1 BRCT repeat. On the right, each subfigure depicts the representative structure computed from the NMA of the 3D coordinates of the energy minima (where Gibbs' free energy, $\Delta G = 0$) depicted in the Gibbs' FEL plot. Vectors depicting the amplitudes of the prominent 1st PCA mode, of the NMA, are shown projecting from the structure, turning it into a "porcupine" style plot. (a) BRCA1 BRCT repeat wild-type (b) H1686Y, (c) P1749L, (d) P1749S, (e) W1718L, and (f) W1837L.

trend were observed in W1718L and W1837L variants, in which each had significantly lower cumulative time-averaged SASAs of buried hydrophobic residues compared to wild type (Table 4). Hydrophobic to hydrophilic changes such as P1749S in the BRCA1 BRCT repeat and P657S in the BARD1 BRCT repeat and neutral to hydrophilic changes, such as H1686Y in the BRCA1 BRCT repeat and H606N in the BARD1 BRCT repeat, could have changed the structural stabilities of the associated variants.

3.6. Screen Plots and PCA of the MD Simulation Trajectories of the BRCA1 BRCT and BARD1 BRCT. Analysis of the screen plots of the eigenvalues plotted against their respective eigenvectors for the wild type and variants of the BRCA1 and BARD1 BRCT structures showed that most of the variance in the motions of the proteins was accounted for by the first two eigenvectors (Figure S2).

The BRCA1 BRCT WT protein's projection in phase space on eigenvectors 1 and 2 covered an area of ~24.782 nm^2, which was higher than those of H1686Y (~20.808 nm^2), P1749S (~20.214 nm^2) and W1718L (~20.889 nm^2) but less than those of P1749L (~30.497 nm^2) and W1837L (~27.170 nm^2). The corresponding value for the BARD1 BRCT repeat WT protein was ~15.557 nm^2, which was higher than that of P657S (~15.221 nm^2) but lower than those of H606D (~24.356 nm^2), H606N (~16.142 nm^2), P657L (~15.949 nm^2), W635L (~21.477 nm^2), and W762F (~21.073 nm^2). Thus, based on the area covered by the data points of the projections on the first two eigenvectors in phase space, the H1686Y, P1749S, and W1718L variants were less mobile, while the P1749L and W1837L variants were more mobile than the BRCA1 BRCT repeat wild type. Similarly, the H606D, W635L, and W762F variants were more mobile than the BARD1 BRCT repeat WT, while the H606N, P657L, and P657S variants had comparable and similar mobility to that of the BARD1 BRCT repeat WT (Figures S3 and S4).

https://doi.org/10.1021/acsomega.2c04782
ACS Omega 2022, 7, 44772−44785

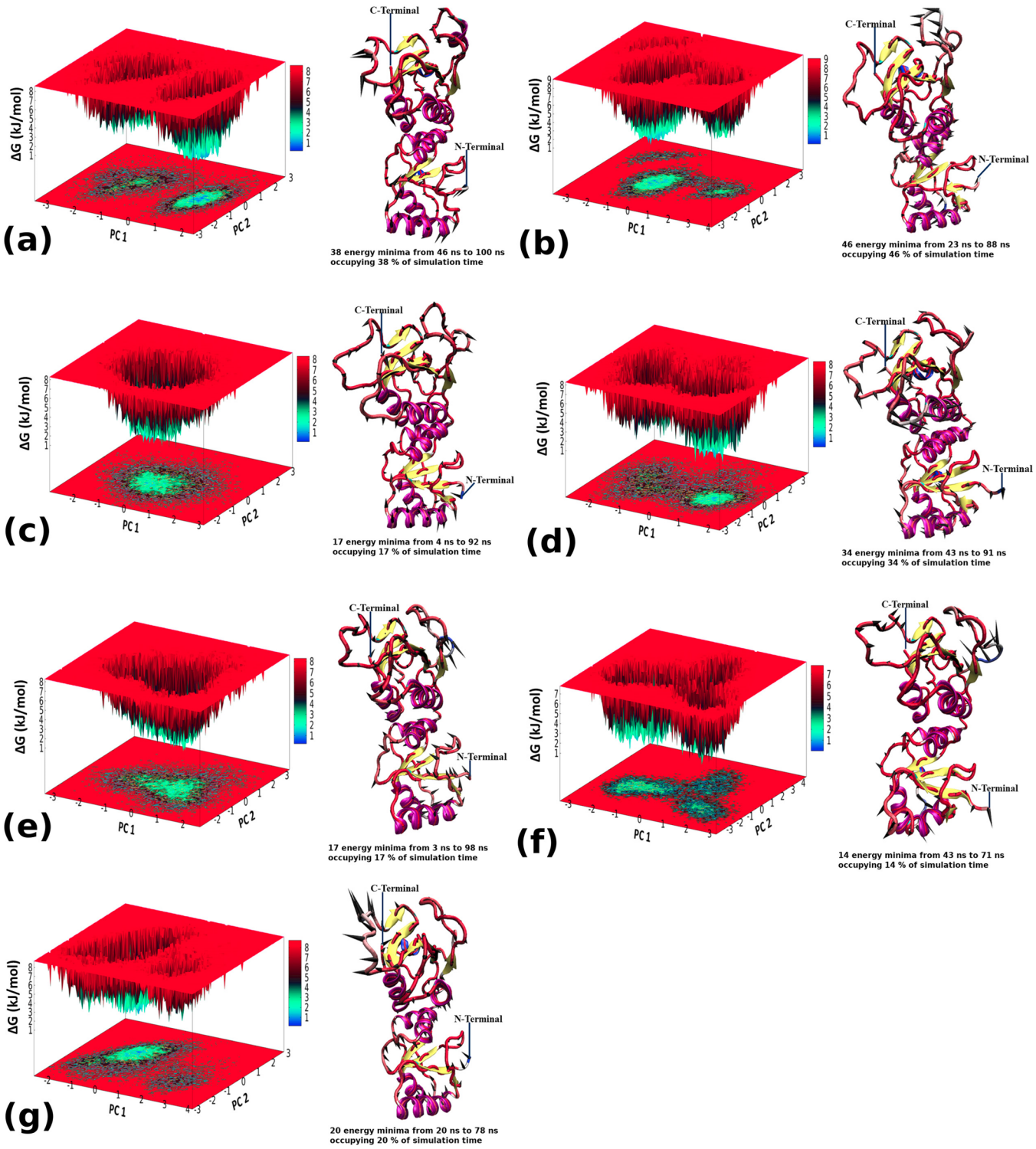


Figure 6. In each subfigure above, on the left is shown the Gibbs' FEL superimposed over the projections on the first two eigenvectors, in phase space, for the BARD1 BRCT repeat. On the right, each subfigure depicts the representative structure computed from the NMA of the 3D coordinates of the energy minima (where Gibbs' free energy, $\Delta G = 0$) depicted in the Gibbs' FEL plot. Vectors depicting the amplitudes of the prominent 1st PCA mode, of the NMA, are shown projecting from the structure, turning it into a "porcupine" style plot. (a) BARD1 BRCT repeat wild-type (b) H606D, (c) H606N, (d) P657L, (e) P657S, (f) W635L, and (g) W762F.

Based on the Gibbs' free energy landscapes derived from the projection of PC 1 and 2, the structures having $\Delta G = 0$ kJ mol^{-1} were extracted, from different time frames of the simulated trajectories, to acquire Gibbs' free energy minima during the simulations. For the BRCA1 BRCT wild type, 45 structures had acquired the energy minimum ($\Delta G = 0$ kJ mol^{-1}) within the range of 1−88 ns, occupying ∼45% of simulation time. Similarly, the BRCA1 BRCT repeat variants H1686Y, P1749L, P1749S, W1718L, and W1837L acquired 16, 15, 22, 22, and 34 energy minima within the ranges 7−86 ns, 53−71 ns, 1−98 ns, 37−100 ns, and 44−92 ns, respectively. On the other hand, the BARD1 BRCT repeat WT exhibited 38

https://doi.org/10.1021/acsomega.2c04782
ACS Omega 2022, 7, 44772−44785

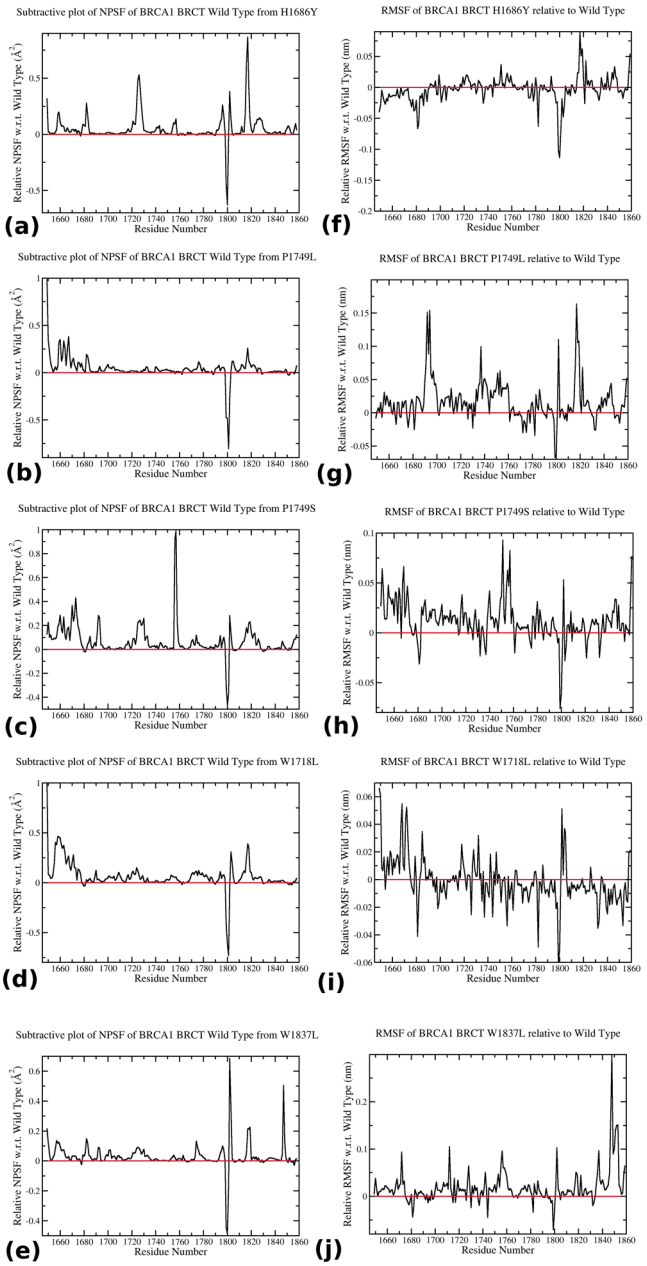


Figure 7. NPSF (Normalized PCA Square Fluctuations) relative to the wild-type protein for variants of the BRCA1 BRCT repeat, obtained by subtracting each data point's NPSF from that of the wild-type, at the corresponding residue position (a) H1686Y, (b) P1749L, (c) P1749S, (d) W1718L, and (e) W1837L. Relative RMSF of variants of the BRCA1 BRCT repeat (relative to the wild type) (f) H1686Y, (g) P1749L, (h) P1749S, (i) W1718L, and (j) W1837L.

energy minima that were spread in the range 46−100 ns, occupying ∼38% of simulation time. Likewise, the BARD1 BRCT repeat variants H606D, H606N, P657L, P657S, W635L, and W762F acquired 46, 17, 34, 17, 14, and 20 energy minima in the ranges 23−88 ns, 4−92 ns, 43−91 ns, 3−98 ns, 43−71 ns, and 20−78 ns, respectively (Figure S5).

The data points of the projections on the first two eigenvectors were grouped into "bins", with each bin containing structures/frames, associated with the corresponding data points in the projection graph, with the same value of ΔG (in kJ mol^{-1}). This made identification of the structures located in the energy minima (where ΔG = 0 kJ mol^{-1}) easier.

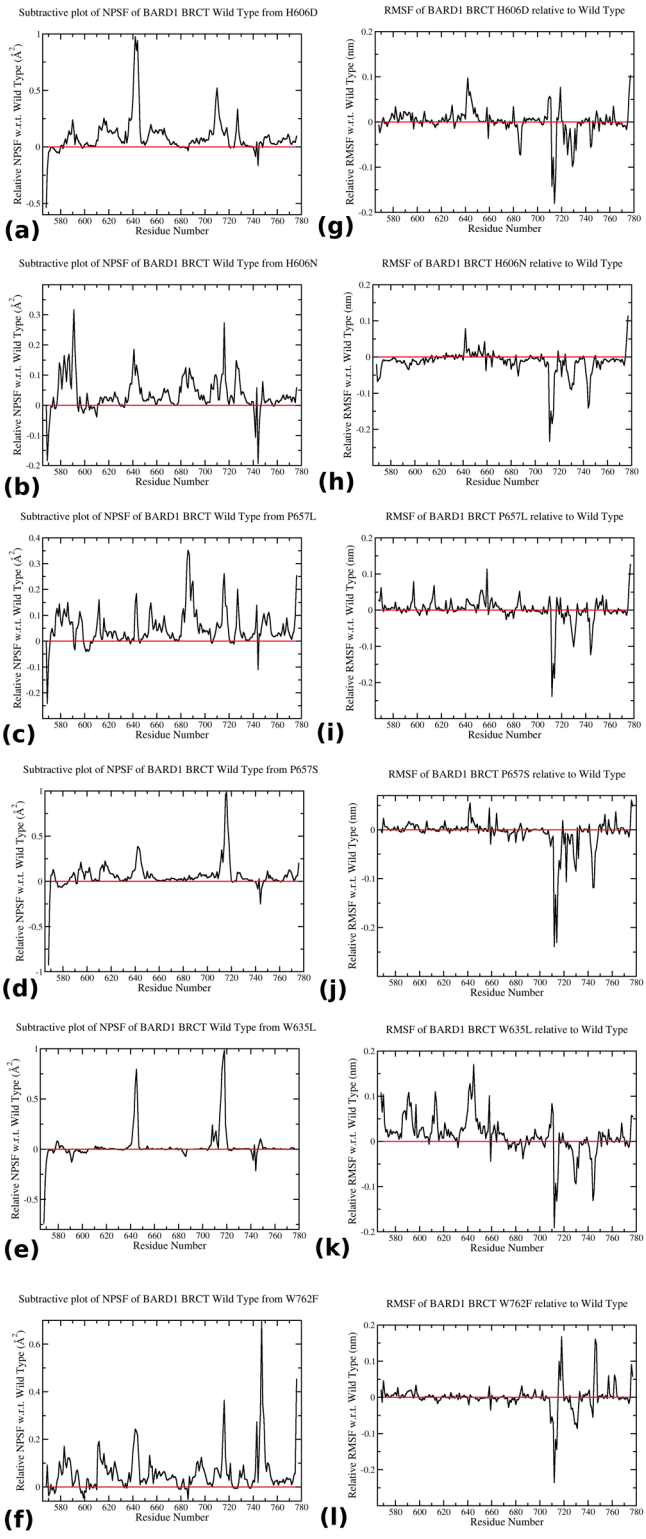


Figure 8. NPSF relative to the wild-type protein for variants of the BARD1 BRCT repeat, obtained by subtracting each data point's NPSF from that of the wild type (a) H606D, (b) H606N, (c) P657L, (d) P657S, (e) W635L, and (f) W762F. Relative RMSF of variants of the BARD1 BRCT repeat (g) H606D, (h) H606N, (i) P657L, (j) P657S, (k) W635L, and (l) W762F.

Normal mode analysis (NMA) was performed on each set of structures corresponding to the energy minima, from every simulation trajectory, to generate a representative structure

https://doi.org/10.1021/acsomega.2c04782
ACS Omega 2022, 7, 44772−44785

showing directional spikes, projecting from its backbone, which represent the PC square fluctuations or extent of deviation of the farthermost structure, from the representative one, at the base of each spike. (Figures 5 and 6).

3.7. NMA of the Energy Minima of the MD Simulation Trajectories of the BRCA1 and BARD1 BRCT Repeats and Their Variants. The NPSF profile of the BRCA1 BRCT repeat wild type shows prominent NPSF peaks in the amino acid ranges 1676−1683, 1797−1804, 1824−1834, and 1850−1858. The most prominent peak of these, by far, was the single, tall peak in the amino acid region 1797−1804. The relative NPSF profiles of the variants have positive and negative peaks; a positive peak indicates a greater range of mobility, while a negative peak indicates a smaller range of motion than the WT. The BRCA1 BRCT repeat variants had lower mobility than the wild type in the amino acid region 1798−1801 (spanning a portion of a loop and a portion of a contiguous helix) and the lowest mobility, within this dip, at residue 1800 (Figure 7). The NPSF profile of the BARD1 BRCT repeat WT shows prominent NPSF peaks in the amino acid regions 568−572, 574−605, 609−613, 636−646, 681−687, 709−713, 740−742, and 744−749. The most prominent of these peaks is the one at 568−572, followed by the one at 744−749 and then the one at 740−742. BARD1 BRCT variants had significantly lower mobility, than the WT, at residue 744 (within a long loop), except W762F, which had higher mobility at this position (Figure 8).

4. CONCLUSIONS

The structural superposition and sequential identity of BRCA1 BRCT and BARD1 BRCT are found to be highly correlated. Furthermore, *in silico* analyses, through MD simulations, revealed that the missense variants do not vastly differ from their respective wild-type proteins in terms of RMSD, RMSF, R_g, and SASA, except for the W635L variant of BARD1, which has a notably higher RMSD than the BARD1 BRCT, during the last ∼60 ns of the simulation trajectory. Furthermore, analyses of the essential motions of potential biological relevance, through NMA of the energy minima in the simulation trajectories, revealed that all of the BRCA1 BRCT variants had a significantly lower flexibility than the wild type in the residue region 1798−1801, with the mobility in this dip being the lowest at residue 1800, while all of the BARD1 BRCT variants showed significantly lower flexibility than the wild type at the residue 744, with the exception of the W762F variant, which had higher flexibility than the wild type. This suggests that the residues which are structurally superposed between the two BRCT repeats and are identical to both the proteins and the variants at these positions are worthy of functional characterization to reveal putative clinical significance. We would like to mention here that all the results are obtained using different *in silico* and bioinformatics-based structural approaches, which may not be without some drawbacks. Therefore, no statements are made on the pathogenicity of the variants vis-à-vis clinical management and genetic counseling.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.2c04782.

(1) Images of the closely superposed residue pairs from the BRCA1 BRCT and BARD1 BRCT repeat; (2) screen plots of the MD simulations, depicting the eigenvalues associated with the first 30 eigenvectors of each simulation; (3) 2D projections, in phase space, on eigenvectors 1 and 2, of each of the 13 MD simulations; and (4) histogram depicting the approximate percentage of total simulation time (100 ns) occupied by the energy minima evolved during each of the 13 MD simulations (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Ashok K. Varma − *Advanced Centre for Treatment, Research and Education in Cancer, Navi Mumbai, Maharashtra 410210, India; Homi Bhabha National Institute, Mumbai, Maharashtra 400094, India;* orcid.org/0000-0001-6091-5315; Email: avarma@actrec.gov.in

Authors

Siddhartha A. Barua − *Advanced Centre for Treatment, Research and Education in Cancer, Navi Mumbai, Maharashtra 410210, India; Homi Bhabha National Institute, Mumbai, Maharashtra 400094, India*

Nabajyoti Goswami − *Advanced Centre for Treatment, Research and Education in Cancer, Navi Mumbai, Maharashtra 410210, India;* orcid.org/0000-0002-6300-5410

Neha Mishra − *Advanced Centre for Treatment, Research and Education in Cancer, Navi Mumbai, Maharashtra 410210, India; Homi Bhabha National Institute, Mumbai, Maharashtra 400094, India*

Ulka U. Sawant − *Advanced Centre for Treatment, Research and Education in Cancer, Navi Mumbai, Maharashtra 410210, India*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.2c04782

Author Contributions

S.A.B. and N.G. performed the experiments, analyzed the data, generated the figures and tables, and wrote the first draft. N.M. and U.U.S. analyzed the data and made the figures, while A.K.V. designed the experiments and wrote the paper.

Funding

The study has been supported by *DBT (BT/PR40181/BTIS/137/15/2021)* to A.K.V.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors would like to thank the Bioinformatics Centre (BIC) at ACTREC for providing the necessary computational facility.

■ REFERENCES

(1) Ferlay, J.; et al. Cancer statistics for the year 2020: An overview. *Int. J. Cancer* **2021**, *149*, 778−789.

(2) Centres for Disease Control and Prevention (the CDC, USA). *Centres for Disease Control and Prevention* (2021).

(3) Roy, R.; Chun, J.; Powell, S. N. BRCA1 and BRCA2: Different roles in a common pathway of genome protection. *Nat. Rev. Cancer* **2012**, *12*, 68−78.

(4) Petrucelli, N.; Daly, M. B.; Pal, T. BRCA1- and BRCA2-Associated Hereditary Breast and Ovarian Cancer Summary Genetic counseling Suggestive Findings. *Gene Rev.* **1998**, 1−37.
(5) Yu, X.; Chini, C. C. S.; He, M.; Mer, G.; Chen, J. The BRCT Domain Is a Phospho-Protein Binding Domain. *Science* **2003**, *302*, 639−642.
(6) Wu, Q.; Jubb, H.; Blundell, T. L. Phosphopeptide interactions with BRCA1 BRCT domains: More than just a motif. *Prog. Biophys. Mol. Biol.* **2015**, *117*, 143−148.
(7) Yu, X.; Fu, S.; Lai, M.; Baer, R.; Chen, J. BRCA1 ubiquitinates its phosphorylation-dependent binding partner CtIP. *Genes Dev.* **2006**, *20*, 1721−1726.
(8) Edwards, R. A.; et al. The BARD1 C-terminal domain structure and interactions with polyadenylation factor CstF-50. *Biochemistry* **2008**, *47*, 11446−11456.
(9) Li, M.; Yu, X. Function of BRCA1 in the DNA damage response is mediated by ADP-ribosylation. *Cancer Cell* **2013**, *23*, 693−704.
(10) Billing, D.; et al. The BRCT Domains of the BRCA1 and BARD1 Tumor Suppressors Differentially Regulate Homology-Directed Repair and Stalled Fork Protection. *Mol. Cell* **2018**, *72*, 127−139.e8.
(11) Glover, J. N. M. Insights into the molecular basis of human hereditary breast cancer from studies of the BRCA1 BRCT domain. *Fam. Cancer* **2006**, *5*, 89−93.
(12) De Brakeleer, S.; et al. Cancer predisposing missense and protein truncating BARD1 mutations in non-BRCA1 or BRCA2 breast cancer families. *Hum. Mutat.* **2010**, *31*, E1175−E1185.
(13) Easton, D. F.; et al. A systematic genetic assessment of 1,433 sequence variants of unknown clinical significance in the BRCA1 and BRCA2 breast cancer-predisposition genes. *Am. J. Hum. Genet.* **2007**, *81*, 873−883.
(14) Phelan, C. M.; et al. Classification of BRCA1 missense variants of unknown clinical significance. *J. Med. Genet.* **2005**, *42*, 138−146.
(15) Fernandes, V. C.; et al. Impact of amino acid substitutions at secondary structures in the BRCT domains of the tumor suppressor BRCA1: Implications for clinical annotation. *J. Biol. Chem.* **2019**, *294*, 5980−5992.
(16) Coquelle, N.; Green, R.; Glover, J. N. M. Impact of BRCA1 BRCT domain missense substitutions on phosphopeptide recognition. *Biochemistry* **2011**, *50*, 4579−4589.
(17) Sheng, Z. Z.; Zhao, Y. Q.; Huang, J. F. Functional evolution of BRCT domains from binding DNA to protein. *Evol. Bioinform.* **2011**, *7*, 87−97.
(18) Bateman, A.; et al. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.* **2021**, *49*, D480−D489.
(19) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.
(20) Varma, A. K.; Brown, R. S.; Birrane, G.; Ladias, J. A. A. Structural basis for cell cycle checkpoint control by the BRCA1-CtIP complex. *Biochemistry* **2005**, *44*, 10941−10946.
(21) Birrane, G.; Varma, A. K.; Soni, A.; Ladias, J. A. A. Crystal structure of the BARD1 BRCT domains. *Biochemistry* **2007**, *46*, 7706−7712.
(22) Giannini, G.; et al. Clinical classification of BRCA1 DNA missense variants: H1686Q is a novel pathogenic mutation occurring in the ontogenetically invariant THV motif of the N-terminal BRCT domain. *J. Clin. Oncol.* **2008**, *26*, 4212−4214.
(23) Thompson, J. D.; Higgins, D. G.; Gibson, T. J. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **1994**, *22*, 4673−4680.
(24) Cerami, E.; et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* **2012**, *2*, 401−404.
(25) Fokkema, I. F. A. C.; et al. The LOVD3 platform : efficient genome-wide sharing of genetic variants. *Eur. J. Hum. Genet.* **2021**, *29*, 1796−1803.
(26) Cline, M. S.; Liao, R. G.; Parsons, M. T.; Paten, B.; Alquaddoomi, F.; Antoniou, A.; Baxter, S.; Brody, L.; Cook-Deegan, R.; Coffin, A.; Couch, F. J.; Craft, B.; Currie, R.; Dlott, C. C.; Dolman, L.; den Dunnen, J. T.; Dyke, S. O. M.; Domchek, S. M.; Easton, D.; Fischmann, Z.; Foulkes, W. D.; Garber, J.; Goldgar, D.; Goldman, M. J.; Goodhand, P.; Harrison, S.; Haussler, D.; Kato, K.; Knoppers, B.; Markello, C.; Nussbaum, R.; Offit, K.; Plon, S. E.; Rashbass, J.; Rehm, H. L.; Robson, M.; Rubinstein, W. S.; Stoppa-Lyonnet, D.; Tavtigian, S.; Thorogood, A.; Zhang, C.; Zimmermann, M.; BRCA Challenge Authors; Burn, J.; Chanock, S.; Rätsch, G.; Spurdle, A. B. BRCA Challenge: BRCA Exchange as a global resource for variants in BRCA1 and BRCA2. *PLoS Genet.* **2018**, *14*, No. e1007752.
(27) Tate, J. G.; Bamford, S.; Jubb, H. C.; Sondka, Z.; Beare, D. M.; Bindal, N.; Boutselakis, H.; Cole, C. G.; Creatore, C.; Dawson, E.; Fish, P.; Harsha, B.; Hathaway, C.; Jupe, S. C.; Kok, C. Y.; Noble, K.; Ponting, L.; Ramshaw, C. C.; Rye, C. E.; Speedy, H. E.; Stefancsik, R.; Thompson, S. L.; Wang, S.; Ward, S.; Campbell, P. J.; Forbes, S. A. COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res.* **2019**, *47*, D941−D947.
(28) Adzhubei, I.; Jordan, D. M.; Sunyaev, S. R. Predicting functional effect of human missense mutations using PolyPhen-2. in *Current Protocols in Human Genetics 2*; (Wiley, 2013).
(29) Pejaver, V.; et al. Inferring the molecular and phenotypic impact of amino acid variants with MutPred2. *Nat. Commun.* **2020**, *11*, 5918.
(30) Tang, H.; Thomas, P. D. PANTHER-PSEP: Predicting disease-causing genetic variants using position-specific evolutionary preservation. *Bioinformatics* **2016**, *32*, 2230−2232.
(31) López-Ferrando, V.; Gazzo, A.; De La Cruz, X.; Orozco, M.; Gelpí, J. L. PMut: A web-based tool for the annotation of pathological variants on proteins, 2017 update. *Nucleic Acids Res.* **2017**, *45*, W222−W228.
(32) Choi, Y.; Chan, A. P. PROVEAN web server: A tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics* **2015**, *31*, 2745−2747.
(33) Sim, N. L.; Kumar, P.; Hu, J.; Henikoff, S.; Schneider, G.; Ng, P. C. SIFT web server: Predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res.* **2012**, *40*, W452−W457.
(34) Capriotti, E.; Calabrese, R.; Casadio, R. Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. *Bioinformatics* **2006**, *22*, 2729−2734.
(35) Capriotti, E.; Altman, R. B.; Bromberg, Y. Collective judgment predicts disease-associated single nucleotide variants. *BMC Genomics* **2013**, *14*, S2.
(36) Capriotti, E.; et al. WS-SNPs&GO: a web server for predicting the deleterious effect of human protein variants using functional annotation. *BMC Genomics* **2013**, *14*, S6.
(37) Leong, I. U. S.; Stuckey, A.; Lai, D.; Skinner, J. R.; Love, D. R. Assessment of the predictive accuracy of five in silico prediction tools, alone or in combination, and two metaservers to classify long QT syndrome gene mutations. *BMC Med. Genet.* **2015**, *16*, 34.
(38) Montenegro, L. R.; Lerário, A. M.; Nishi, M. Y.; Jorge, A. A. L.; Mendonca, B. B. Performance of mutation pathogenicity prediction tools on missense variants associated with 46,xy differences of sex development. *Clinics* **2021**, *76*, No. e2052.
(39) Qorri, E.; et al. A Comprehensive Evaluation of the Performance of Prediction Algorithms on Clinically Relevant Missense Variants. *Int. J. Mol. Sci.* **2022**, *23*, 7946.
(40) Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. GROMACS: A message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* **1995**, *91*, 43−56.
(41) Pettersen, E. F.; et al. UCSF Chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605−1612.
(42) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved side-chain torsion potentials for theAmber ff99SB protein force field. *Proteins* **2010**, *78*, 1950−1958.

https://doi.org/10.1021/acsomega.2c04782
ACS Omega 2022, 7, 44772−44785

(43) Lindahl, E.; Hess, B.; van der Spoel, D. GROMACS 3.0: A package for molecular simulation and trajectory analysis. *J. Mol. Model.* **2001**, *7*, 306−317.
(44) Van Der Spoel, D.; et al. GROMACS: Fast, flexible, and free. *J. Comput. Chem.* **2005**, *26*, 1701−1718.
(45) Hess, B.; Kutzner, C.; Van Der Spoel, D.; Lindahl, E. GRGMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435−447.
(46) Pronk, S.; et al. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **2013**, *29*, 845−854.
(47) Páll, S.; Abraham, M.J.; Kutzner, C.; Hess, B.; Lindahl, E. Tackling Exascale Software Challenges in Molecular Dynamics Simulations with GROMACS. (ed. Markidis, S., Laure, E.). in *Solving Software Challenges for Exascale*; (Springer: Cham., 2015), 3−27.
(48) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1-2*, 19−25.
(49) Schneider, C. A.; Rasband, W. S.; Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **2012**, *9*, 671−675.
(50) Janert, P. K. *Gnuplot in action: understanding data with graphs*; Manning Publications Co.: (Shelter Island, NY, 2016).
(51) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **1996**, *14*, 33−38.
(52) Bakan, A.; Meireles, L. M.; Bahar, I. ProDy: Protein Dynamics Inferred from Theory and Experiments. *Bioinformatics* **2011**, *27*, 1575−1577.
(53) Mohammadi, L.; et al. A simple method for co-segregation analysis to evaluate the pathogenicity of unclassified variants; BRCA1 and BRCA2 as an example. *BMC Cancer* **2009**, *9*, 211.
(54) Osorio, A.; et al. Classification of missense variants of unknown significance in BRCA1 based on clinical and tumor information. *Hum. Mutat.* **2007**, *28*, 477−485.
(55) Mirkovic, N.; Marti-Renom, M. A.; Weber, B. L.; Sali, A.; Monteiro, A. N. A. Structure-based assessment of missense mutations in human BRCA1: Implications for breast and ovarian cancer predisposition. *Cancer Res.* **2004**, *64*, 3790−3797.
(56) Tudini, E.; et al. Substantial evidence for the clinical significance of missense variant BRCA1 c. 5309G>T p.(Gly1770Val). *Breast Cancer Res. Treat.* **2018**, *172*, 497−503.
(57) Sinha, S.; Wang, S. M. Classification of VUS and unclassified variants in BRCA1 BRCT repeats by molecular dynamics simulation. *Comput. Struct. Biotechnol. J.* **2020**, *18*, 723−736.

https://doi.org/10.1021/acsomega.2c04782
ACS Omega 2022, 7, 44772−44785