J Mol Model (2013) 19:5587–5599
DOI 10.1007/s00894-013-2044-0

ORIGINAL PAPER

In-silico screening of cancer associated mutation on PLK1 protein and its structural consequences

Balu Kamaraj · Vidya Rajendran · Rao Sethumadhavan · Rituraj Purohit

Received: 14 September 2013 / Accepted: 21 October 2013 / Published online: 23 November 2013

Abstract The Polo-like kinases (Plks) are a conserved subfamily of serine-threonine protein kinases that have significant roles in cell proliferation. The serine/threonine protein kinases or polo-like kinase 1 (PLK1) exist in centrosome during interphase and is an important regulatory enzyme in cell cycle progression during M phase. Mutations in mammalian PLK1 were found to be over expressed in various human cancers and it is disrupting the binding ability of polo box domain with target peptide. In this analysis we implemented a computational approach to filter the most deleterious and cancer associated mutation on PLK1 protein. We found W414F as the most deleterious and cancer associated by Polyphen 2.0, SIFT, I-mutant 3.0, PANTHER, PhD-SNP, SNP&GO, Mutpred and Dr Cancer tools. Molecular docking and molecular dynamics simulation (MDS) approach was used to investigate the structural and functional behavior of PLK1 protein upon mutation. MDS and docking results showed stability loss in mutant PLK1 protein. Due to mutation, PLK1 protein became more flexible and alters the dynamic property of protein which might affect the interaction with target peptide and leads to cell proliferation. Our study provided a well designed computational methodology to examine the cancer associated nsSNPs and their molecular mechanism. It further helps scientists to develop a drug therapy against PLK1 cancer-associated diseases.

Keywords Flexibility · Molecular dynamics simulations · PLK-1 protein

Balu Kamaraj and Vidya Rajendran equally contributed to this paper.

B. Kamaraj · V. Rajendran · R. Sethumadhavan · R. Purohit (✉)
School of Bio Sciences and Technology (SBST), Bioinformatics Division, Vellore Institute of Technology University,
Vellore 632014, Tamil Nadu, India
e-mail: riturajpurohit@gmail.com

R. Purohit
Human Genetics Foundation, Torino,
Via Nizza 52, 10126 Torino, Italy

Introduction

The polo-like kinases (Plks) are a conserved subfamily of serine-threonine protein kinases that have significant roles in cell proliferation [1, 2]. In mammals, the multiple forms of plks are designated as PLK1, PLK-2/Snk, PLK-2/Prk/Fnk, and PLK/Sak. Among the Plks, PLK1 has been studied most extensively because of its ability to override cellular checkpoints and induce genetic instability, which leads to oncogenic transformation of human cells [3, 4]. PLK1 protein is composed of a common N-terminal catalytic domain and C-terminal regulatory domain with highly conserved sequences named polo boxes (PB). The polo box domain (PBD) is observed only in the PLK and contains a characteristic sequence, which is the trademark of this protein family. This domain is supposed to be involved in an auto regulatory mechanism or in targeting the kinase to its substrates [5]. The serine/threonine protein kinases or polo-like kinase 1 (PLK1) exist in centrosome during interphase and is an important regulatory enzyme in cell cycle progression during M phase [1, 6]. The polo kinases regulate diverse cellular events at various stages of M phase [7, 8] such as centrosome maturation and bipolar spindle formation [9–11]. Polo kinases also appear to regulate important biochemical steps in G2/M phase, such as activation of Cdc2 through Cdc25C phosphatase [12], DNA damage checkpoint adaptation [13], and regulation of the anaphase-promoting complex [14, 15]. One additional function attributed to polo kinases is the induction of cytokinesis-associated septal structures [10, 16].

Mutations in the polo box domain (PBD) show that the PBD is critical for localization and mitotic functions of various Plks in their native organisms [17–19]. Mutations

specifically disrupt phosphodependent interactions abolish cell-cycle-dependent localization and provide compelling phenotypic evidence that PBD-phospholigand binding is necessary for proper mitotic progression. In addition, phosphopeptide binding to the PBD stimulates kinase activity in full-length Plk1, suggesting a conformational switching mechanism for Plk regulation and a dual functionality for the PBD [20]. Consistent with these observations, a single W414F mutation is sufficient to disrupt the ability of PBD to bind with target peptide [5, 21, 22]. Mutations in mammalian PLK1 was found to be over expressed in various human cancers including head and neck squamous cell carcinoma, gastric, thyroid or B-cell lymphomas, oropharyngeal carcinomas, non-small cell lung cancer, melanomas, ovarian, and endometrial carcinomas [23].

Non-synonymous SNPs (nsSNPs) occurring in the coding regions result in single amino acid polymorphisms (SAPs) that may affect protein conformation and function which lead to cause pathogenic phenotypes. nsSNPs have the prospective to alter the function of their corresponding protein, either directly or via disruption of structure. Hence they are of particular concern for further experimental assessment. Computational algorithms were implemented to filter the non-significant SNPs from the ones that might produce major disease associated consequences. The phenotypic changes integrated with classical computational SNP prediction techniques, will ultimately provide a high accuracy prediction level and thus will help in easy classification of SNPs on the basis of their specific disease-associated consequences. In-silico studies have previously provided an efficient platform for evaluation and analysis of genetic mutations for their pathological consequence and in determining their underlying molecular mechanism [24–28]. In this study we analyzed the deleterious effect of nsSNPs reported in PLK1 protein coding region. In our study we used Polyphen 2.0 [29], SIFT [30], I-mutant 3.0 [31], PANTHER [32], PhD-SNP [33], SNP&GO [34], MutPred [35] and Dr Cancer [36] to prioritize the most deleterious and disease-associated nsSNPs from the available SNP datasets obtained from swissprot/Uniprot database. Conformational changes in the 3D structure of the protein accounts for its time dependent physiological affinities and various biochemical pathway alterations [37–42]. Molecular docking and molecular dynamics simulation (MDS) approach was applied to examine, how the predicted mutation (W414F) affect the interaction between Plk1 PBD and their target peptides which is sufficient to inhibit the over expression of plk1. The overall workflow implemented in this work was shown in Fig. 1.

Materials and methodology

Dataset collection

The protein sequence and SNP information for our in-silico analysis were obtained from swissprot/Uniprot database [43–45]. The structure of PLK1 (PDB ID: 3FVH_A) and two peptide molecules (PDB ID: 3FVH_B, 3HIK_B) were obtained from Brookhaven Protein Data Bank [46]. The mutant (W414F) structure was built by induced point mutation in the position of 414 of PLK1 protein using SPDB viewer package [47]. This structure was energetically optimized by applying the all atom OPLS force field available in GROMACS package 4.5.3 [48].

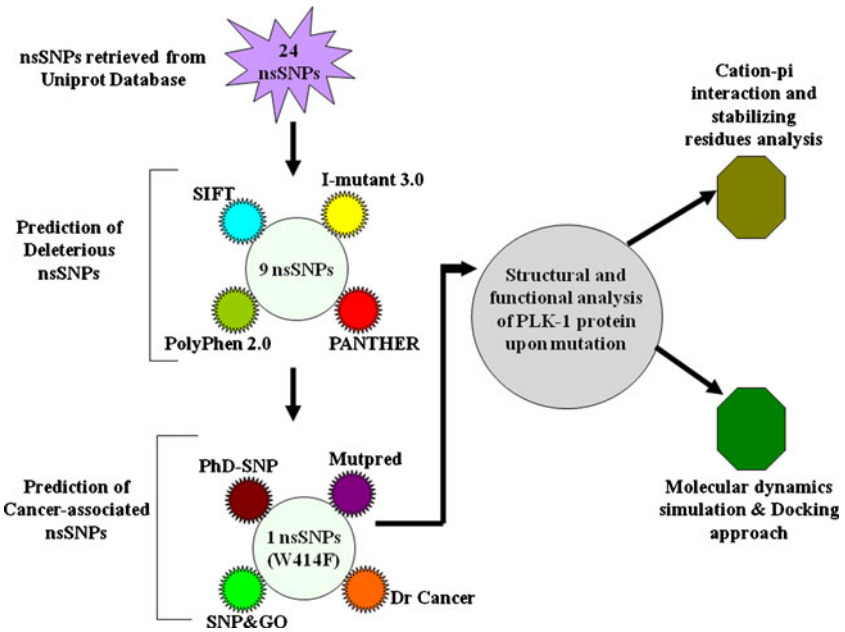

Fig. 1 Work flow model implemented in this study

Disease associated SNP prediction

The single nucleotide polymorphism occurring in the protein coding region may lead to the deleterious consequences and might disturb its 3D structure. Here we applied Polyphen 2.0 [29], SIFT [30], I-Mutant 3.0 [31], PANTHER [32], PhD-SNP [33], SNP&GO [34], MutPred [35], and Dr Cancer [36] tools in order to examine the cancer-associated nsSNP occurring in the PLK1 protein coding region. PolyPhen 2.0 is based on combination of sequence and structure based attributes and uses naive Bayesian classifier for the identification of an amino acid substitution and the impact of mutation. The output levels of probably damaging and possibly damaging were classified as functionally significant (≤0.5) and the benign level being classified as tolerated (≥0.51) [29]. SIFT prediction is based on the sequence homology and the physicochemical properties of amino acids which are dictated by the substituted amino acid. SIFT score ≥0.05 indicates the amino acid substitution is intolerant or deleterious, whereas the score ≤0.05 is predicted as tolerant [30]. I-Mutant 3.0 is a support vector machine (SVM)-based tool. We used the sequence based version of I-Mutant 3.0 that classifies the prediction into three classes: neutral mutation (−0.5 ≤ DDG ≥ 0.5 kcal mol^{-1}), large decrease (<−0.5 kcal mol^{-1}), and a large increase (>0.5 kcal mol^{-1}). The free energy change (DDG) predicted by I-Mutant 3.0 is based on the difference between unfolding Gibbs free energy change of mutant and native protein (kcal mol^{-1}) [31]. PANTHER program which is a protein family and subfamily database which predicts the frequency of occurrence of amino acid at a particular position in evolutionary related protein sequences. The threshold subPSEC score of −3 has been assigned below which the predictions are considered as deleterious [32]. We filtered the missense mutations that were combinedly predicted to be deleterious and damaging from these four servers. Further we used PhD-SNP, SNP&GO, MutPred and Dr Cancer tools to examine the cancer-associated SNPs. PhD-SNP is SVM based classifier, trained over the million amino acid polymorphism datasets using supervised training algorithm [33]. It predicts whether the given amino acid substitution leads to disease associated or neutral along with the reliability index score [37]. SNP&GO retrieves data from protein sequence, evolutionary information, and functions as encoded in the gene ontology terms [34]. MutPred is a web based tool, used to predict the molecular cause of disease related amino acid substitution [35]. It utilizes several attributes related to protein structure, function, and evolution. Thus by combining the scores of all three servers, the accuracy of prediction rises to a greater extent. Further we used Dr Cancer [36] to identify cancer associated SNPs from the available dataset.

Cation- π sites and stabilizing residues

Cation-π interactions were computed using CaPTURE program [49]. The CaPTURE program identifies energetically significant cation–π interactions within proteins in the Protein Data Bank (PDB). Cation–π interactions in protein structures are identified and evaluated by using an energy-based criterion for selecting significant side chain pairs. The percentage composition of a specific amino acid residue contributing to cation–π interactions is obtained from the equation:

$$compcat{-}\pi(i) = ncat{-}\pi(i) \times [100/n(i)],$$

where "i" stands for the five residues (Lys, Arg, Phe, Trp, and Tyr), *n*cat–π is the number of residues involved in cation–π interactions, and *n*(i) is the number of residues of type "i" in the considered protein structures.

Further the cation-π interaction is increasingly recognized as an important non-covalent binding interaction relevant to structural biology. It uses a variant of the optimized potentials for liquid simulations (OPLS) force field to provide an energetic evaluation of all potential cation-π interactions in a protein. The electrostatic energy (*E*es) is calculated using the equation:

$$Eel = \sum qiqje2/rij,$$

where *q*i and *q*j are the charges for the atoms *i* and *j*, respectively, and *r*ij is the distance between them. The van der Waals energy is given by:

$$Evw = 4\varepsilon ij[(\sigma ij12/rij12){-}(\sigma ij6/rij6)],$$

where $\sigma ij = (\sigma ii\sigma jj)1/2$ and $\varepsilon ij = (\varepsilon ii\varepsilon jj)1/2$; σ and ε are the van der Waals radius and well depth, respectively.

If Ees ≤ −2.0 kcal mol^{-1}, the pair is counted as a cation-π interaction. If Ees > −1.0 kcal mol^{-1}, the structure is rejected. If −2.0 < Ees ≤ −1.0 kcal mol^{-1}, the structure is retained only if EvdW ≤ −1.0 kcal mol^{-1}.

Identifying stabilizing residues

Sride tool [50] was used to identify the stabilizing residues in PLK1 protein. The prediction is done by calculating four essential criteria that involves surrounding hydrophobicity of a residue, the long-range order, stabilization center defined by considering the contact map of a protein and the conservation scores of residues. Two residues are in contact if there is at least one pair of heavy atoms with a distance less than the sum of the van der Waals radii of the two atoms plus 1.0 Å. A contact is considered long-range if it is between residues that are separated by at least ten residues in the amino acid

sequence. Two residues are SC elements if they are involved in long-range contacts and if at least one supporting residue can be found in each of the flanking tetra-peptides of these residues, in such a way that at least seven out of the possible nine interactions are formed between the two triplets. Combined all these criteria are compiled to report the stabilizing residues.

Molecular dynamics simulation

Molecular dynamics simulation was performed by using GROMACS 4.5.3 package [48] running on a single Intel Core2Duo machine with 3 GB RAM and running Ubuntu 11.10 Linux package. Structure of native and mutant PLK1 protein was used as the starting point for MD simulations. Systems were solvated in a cubic box with simple point charge (SPC) water molecules at 10 Å marginal radius. At physiological pH the structures were found to be positively charged, thus in order to make the simulation system electrically neutral, we added two chloride ions (Cl-) to the simulation box using the "genion" tool that accompanies with GROMACS package. Initially the solvent molecules were relaxed while all the solute atoms were harmonically restrained to their original positions with a force constant of 100 kcal mol^{-1} for 5000 steps. After this, the whole molecular system was subjected to energy minimization for 5000 iterations by steepest descent algorithm implementing GROMOS96 43a1 force field. Berendsen temperature coupling method [51] was used to regulate the temperature inside the box. Electrostatic interactions were computed using the particle mesh Ewald method [52]. The ionization states of the residues were set appropriate to pH 7 with all histidines assumed neutral. The pressure was maintained at 1 atm with the allowed compressibility range of 4.5e−5 atm. SHAKE algorithm was used to constrain bond lengths involving hydrogen, permitting a time step of 2 fs. Van der Waals and coulomb interactions were truncated at 1.0 nm. The nonbonded pair list was updated every 10 steps and conformations were stored every 0.5 ps. Position restraint simulation for 500 ps was implemented to allow solvent molecules to enter the cavity region of structure. Finally, systems were subjected to MD simulation for 40 ns. We then computed the comparative analysis of structural deviations in native and mutant structure. RMSD, RMSF, SASA, Rg, DSSP, distance and density plot analysis were carried out by using g_rms, g_rmsf, g_sas, g_gyrate, do_dssp, g_dist and g_density tool respectively. Number of distinct hydrogen bonds formed by specific residues to other amino acids within the protein during the simulation (NH bond) was calculated using g_hbond. NH bond determined on the basis of donor–acceptor distance smaller than 0.35 nm and of donor–hydrogen–acceptor. All the graphs were plotted using XMGRACE [53] program.

Principle component analysis

The calculation of the eigenvectors and eigenvalues, and their projection along the first two principal components, was carried out using essential dynamics (ED) method according to protocol [54] within the GROMACS software package. The principle component analysis or ED is a technique that reduces the complexity of the data and extracts the concerted motion in simulations that are essentially correlated and presumably meaningful for biological function [54]. In the ED analysis, a variance/covariance matrix was constructed from the trajectories after removal of the rotational and translational movements. A set of eigenvectors and eigenvalues was identified by diagonalizing the matrix. The eigenvalues represents the amplitude of the eigenvector along the multidimensional space, and the displacement of atoms along each eigenvector shows the concerted motions of protein along each direction. The movements of structures in the essential subspace were identified by projecting the Cartesian trajectory coordinates along the most important eigenvectors from the analysis. Backbone C-alpha bonds trajectories were obtained using g_covar and g_anaeig of GROMACS utilities.

Molecular docking approach

To validate MDS results, we used PatchDock for docking native and mutants of PLK1 with two target peptides. In our paper we called, Acetyl-Leu-His-Ser-phosphoThr-Ala-NH2 peptide (PDB ID: 3FVH_B) as peptide 1 and pentamer phosphopeptide (PDB ID: 3HIK_B) as peptide 2. PatchDock performs docking based on molecular shape representation, surface patch matching plus filtering and scoring [55, 56]. PatchDock is more reliable because of its fast transformational search, which is driven by local feature matching rather than brute force searching of the six dimensional transformation space. It further speeds up the computational processing time by utilizing advanced data structures and spatial pattern detection techniques, such as geometric hashing and pose clustering. Protein and the peptide molecule were given as input for performing the docking experiments with default root-mean-square deviation (RMSD) value (4.00 Å). Binding site residues 416D, 489H, 491L, 516R, 538H, and 540K of PLK1 were given as one of the additional inputs to the server. It generated several complex structures based on docking scores. The complex structure file, with the best docking score was selected for further analysis. The global binding energy of the PLK1-peptide complex was calculated by FireDock web server [57, 58]. FireDock is an efficient method for refinement and rescoring of rigid-body protein-protein docking solutions. The higher negative value of total interaction energy enables better interaction and vice-versa.

Results and discussion

Prediction of deleterious and cancer-associated nsSNPs

Out of 24 input missense SNP dataset, 18 nsSNPs were found to be "damaging" (0.5 to 1.000) to protein structure and function and the remaining six nsSNPs were characterized as benign by Polyphen 2.0. Among these 18 deleterious nsSNPs, eight SNPs L130G, S137A, D176N, D194A, T210A, T210D, T210V, and W414F were report to be highly deleterious with Polyphen score of 1.000 (Table 1). In SIFT, 18 mutations were predicted to be deleterious with tolerance index ≥ 0.05 (Table 1). Among these 18 mutations, 15 SNPs R12L, R518H, S595L, C67V, K82M, L130G, S137A, S137D, D176N, D194A, T210A, T210D, T210V, R337A, W414F, and H538A were reported to be highly deleterious with SIFT score of 0.00 (Table 1). Furthermore, 17 mutations were identified as deleterious and damaging in SIFT and Polyphen2.0 server (Table 1) which also shows a strong correlation between the prediction methodologies implemented by these two servers. All the nsSNPs submitted to Polyphen 2.0 and SIFT were also submitted as input to the I-Mutant 3.0 server. Nineteen mutations were predicted to affect the stability of the protein structure by I-Mutant 3.0. To further validate these results we implemented a HMM based statistical prediction method to identify the functionally significant point mutations using PANTHER server. The mutations with subPSEC score less than −3 has been reported to be probably deleterious. Eleven mutations with subPSEC score less than equal to −3 were characterized to be deleterious. We filtered nine mutations (R518H, L130G, S137A, D176N, D194A, T210A, T210D, T210V, and W414F) which were commonly predicted to be deleterious and damaging by Polyphen2.0, SIFT, I-Mutant 3.0 and PANTHER servers (Table 1). Further we applied PhD-SNP, SNP&GO, MutPred, Dr Cancer tools to observe the cancer-associated mutation on PLK1 protein. Out of nine mutations, six of them (R518H, D194A, T210A, T210D, T210V, and W414F) were predicted to be disease associated in PhD-SNP server (Table 2). In SNP&GO, eight nsSNPs (R518H, L130G, D176N, D194A, T210A, T210D, T210V, and W414F) were predicted to be disease associated (Table 2). To verify this prediction we further employed Mutpred and Dr Cancer tools. We found W414F to be highly cancer-associated with general probability (g) scores of 0.813

Table 1 Deleterious and damaging nsSNPs prioritized by four computational methods PolyPhen 2.0, SIFT, I-mutant 3.0, and PANTHER in PLK-1 gene

Mutation	PolyPhen 2.0		SIFT		I-MUTANT 3.0		PANTHER	
	PSIC	Prediction	Score	Prediction	DDG	Stability	subPSEC	Prediction
R12L	0.054	Benign	0.00	Deleterious	0.13	Increase	NA	NA
L261F	0.084	Benign	0.14	Tolerated	−0.95	Decrease	−3.20686	Deleterious
N297D	0.000	Benign	0.64	Tolerated	−0.41	Decrease	−2.65669	Tolerated
L332V	0.001	Benign	0.57	Tolerated	−1.50	Decrease	−2.50672	Tolerated
I463H	0.001	Benign	0.50	Tolerated	−2.27	Decrease	NA	NA
R518H	0.967	Damaging	0.02	Deleterious	−1.50	Decrease	−3.04726	Deleterious
S595L	0.933	Damaging	0.08	Tolerated	0.16	Increase	NA	NA
R599H	0.001	Benign	0.18	Tolerated	−1.24	Decrease	NA	NA
C67V	0.961	Damaging	0.00	Deleterious	−0.05	Decrease	NA	NA
K82M	0.999	Damaging	0.00	Deleterious	0.14	Increase	NA	NA
L130G	1.000	Damaging	0.00	Deleterious	−3.17	Decrease	−6.73879	Deleterious
S137A	1.000	Damaging	0.00	Deleterious	−0.56	Decrease	−3.90559	Deleterious
S137D	1.000	Damaging	0.00	Deleterious	0.17	Increase	−5.55171	Deleterious
D176N	1.000	Damaging	0.00	Deleterious	−1.14	Decrease	−5.67212	Deleterious
D194A	1.000	Damaging	0.00	Deleterious	−0.64	Decrease	−5.98769	Deleterious
T210A	0.962	Damaging	0.00	Deleterious	−1.39	Decrease	−4.88553	Deleterious
T210D	0.992	Damaging	0.00	Deleterious	−0.99	Decrease	−6.2594	Deleterious
T210V	1.000	Damaging	0.00	Deleterious	−0.78	Decrease	−5.77384	Deleterious
R337A	0.901	Damaging	0.00	Deleterious	−1.10	Decrease	−1.91497	Tolerated
L340A	0.768	Damaging	0.01	Deleterious	−2.97	Decrease	−1.93366	Tolerated
W414F	1.000	Damaging	0.00	Deleterious	−1.26	Decrease	−4.07587	Deleterious
K492R	0.999	Damaging	0.03	Deleterious	−0.28	Decrease	−2.74906	Tolerated
H538A	0.997	Damaging	0.00	Deleterious	−0.35	Decrease	−2.19467	Tolerated
K540M	1.000	Damaging	0.00	Deleterious	0.29	Increase	NA	NA

SNPs highlighted in bold are predicted to be deleterious

Table 2 The disease associated nsSNPs are predicted from PhD-SNP, SNP&GO, Mutpred, and Dr Cancer servers

Mutation	PhD-SNP	SNP&GO	Mutpred			Dr Cancer	Inference
			g score	P score	Molecular changes		
R518H	Disease	Disease	0.575	0.0181	Loos of sheet	Neutral	Actionable hypothesis
L130G	Neutral	Disease	0.683	0.0603	Loss of loop	Cancer	No reliable inference
S137A	Neutral	Neutral	0.546	0.0062	Loss of phosphorylation at S137	Neutral	Actionable hypothesis
D176N	Neutral	Disease	0.933	0.0381	Loss of catalytic residue at D176	Cancer	Confident hypothesis
D194A	Disease	Disease	0.932	0.0535	Gain of methylation at K191	Cancer	No reliable Inference
T210A	Disease	Disease	0.837	0.028	Loss of phosphorylation at T210	Cancer	Actionable hypothesis
T210D	Disease	Disease	0.811	0.028	Loss of phosphorylation at T210	Cancer	Actionable hypothesis
T210V	Disease	Disease	0.842	0.028	Loss of phosphorylation at T210	Cancer	Actionable hypothesis
W414F	**Disease**	**Disease**	**0.813**	**0.0349**	**Loss of stability**	**Cancer**	**Confident hypothesis**

Mutations highlighted in bold has been predicted to show pathogenic property

and were predicted to induce the loss of stability with (p) score of 0.0349, showing confident hypothesis (Table 2). This prediction could be endorsed with the observed experimental data [21].

Structural and functional analysis of PLK1 protein upon mutation

To understand the structural and functional behavior of the prioritized cancer associated mutations, we performed molecular dynamics simulation for native and mutant PLK1 proteins. We studied RMSD, RMSF, Rg, SASA, NH bond variations, density plot and ED analysis between the native and mutant (W414F) PLK1 proteins. RMSD for all the Cα atoms were calculated from the initial structure which was considered as the central origin to measure the protein system (Fig. 2a). In Fig. 2a, native and mutant (W414F) showed a similar way of deviation till 13,640 ps from their initial structure resulting in the backbone RMSD of ~0.09 to ~0.23 nm during simulation. After, native structure showed minimum deviation till the end of simulation resulting in the backbone RMSD of ~0.14 to ~0.22 nm, where as mutant structure showed maximum deviation till the end of simulation

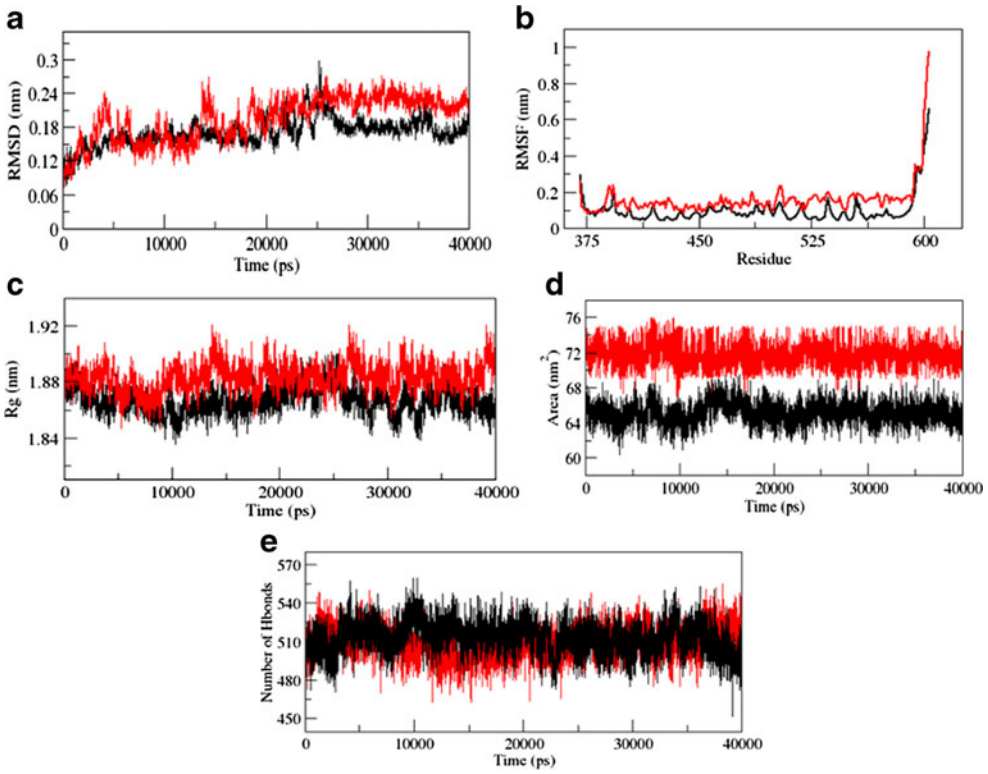

Fig. 2 **a**-**e** RMSD, RMSF, Rg, SASA, and NH bond of native and mutant PLK1 proteins versus time at 300 K. Native is shown in black, mutant (W414F) in red

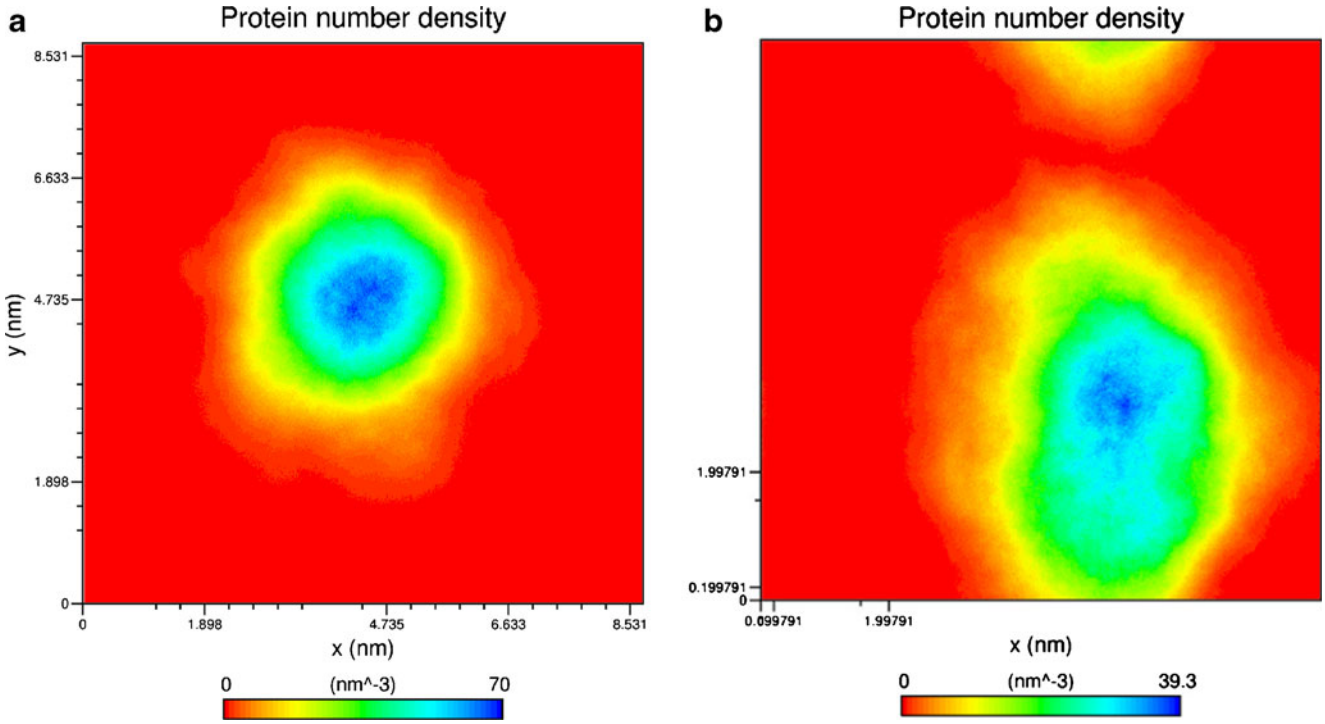

Fig. 3 Number density plot of PLK-1 protein. **a**. Native, **b**. Mutant

resulting backbone RMSD of ~0.24 to ~0.29 nm respectively. This magnitude of fluctuation together with a small difference in average RMSD value after the relaxation period led to the conclusion that the simulation produced stable trajectory, thus providing a suitable basis for further analysis. With the aim of determining whether mutation affected the dynamic behavior of residues, the RMSF values of native and mutant (W414F) structures were calculated (Fig. 2b). Higher degree of flexibility was observed in mutant when compared to native PLK1. It further indicates that mutant conformation was flexible throughout the simulation time and its structure acquired expanded conformation as compared to native.

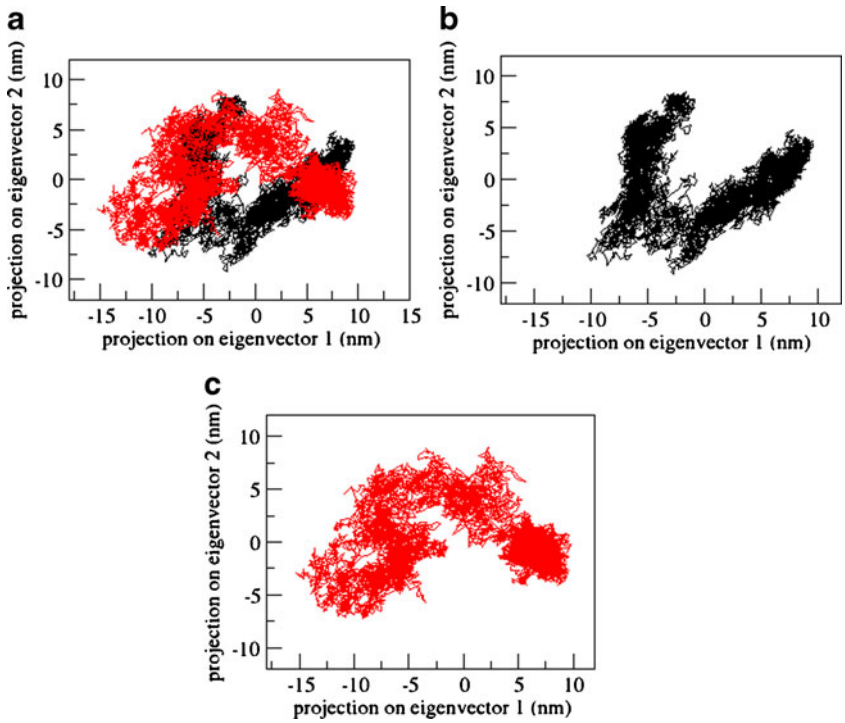

Fig. 4 Projection of the motion of the protein in phase space along the first two principal eigenvectors at 300 K: (**a**) Native is shown in black, mutant in red. For clarity's sake, each trajectory is also shown separately in **b**, **c**

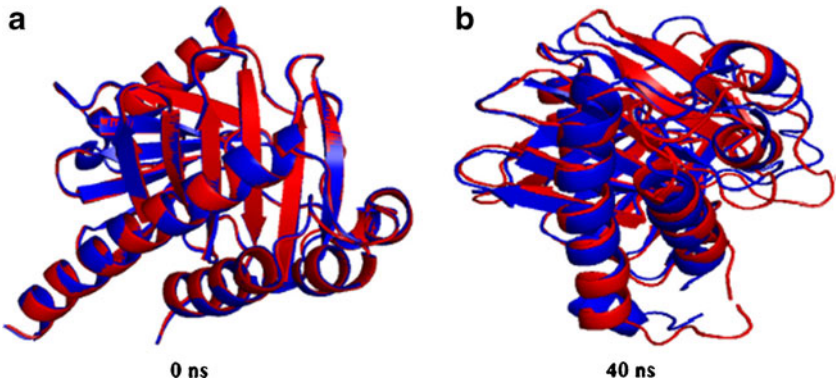

Fig. 5 Snapshots of native and mutant (W414F) PLK-1 protein conformation merged at different simulation timescale. Native is shown in blue, mutant in red

The radius of gyration (Rg) is defined as the mass-weight root mean square distance of collection of atoms from their common center of mass. It provides an insight into the overall dimension of the protein. Radius of gyration plot for Cα atoms of protein vs. time at 300 K is shown in Fig. 2c. In Fig. 2c, at the end of the simulation mutant structure showed greater Rg value than native structure. The change of SASA for native and mutant (W414F) proteins with time is shown in Fig. 2d. Mutant (W414F) structure indicated higher value of SASA with time, while native showed lower SASA value. Maximum fluctuation in Rg plot indicated that the mutant protein might be undergoing a significant structural transition. This was further supported by SASA result in which the mutant was found to exhibit higher values of SASA as compared to the native (Fig. 2d). Hydrogen bond accounts for the major factor of maintaining the stable conformation of protein. NHbond analyses of native and mutant (W414F) proteins were performed with respect to time in order to understand the relationship between flexibility and hydrogen bond formation. Mutant structure showed a significantly smaller number of NHbond formation during the simulation as compared to native (Fig. 2e).

NHbond results of native and mutant proteins are in accordance to the RMSD, RMSF, Rg, and SASA plot results. And further it was indicated that mutant (W414F) structure became more flexible upon mutation which affects the structural and

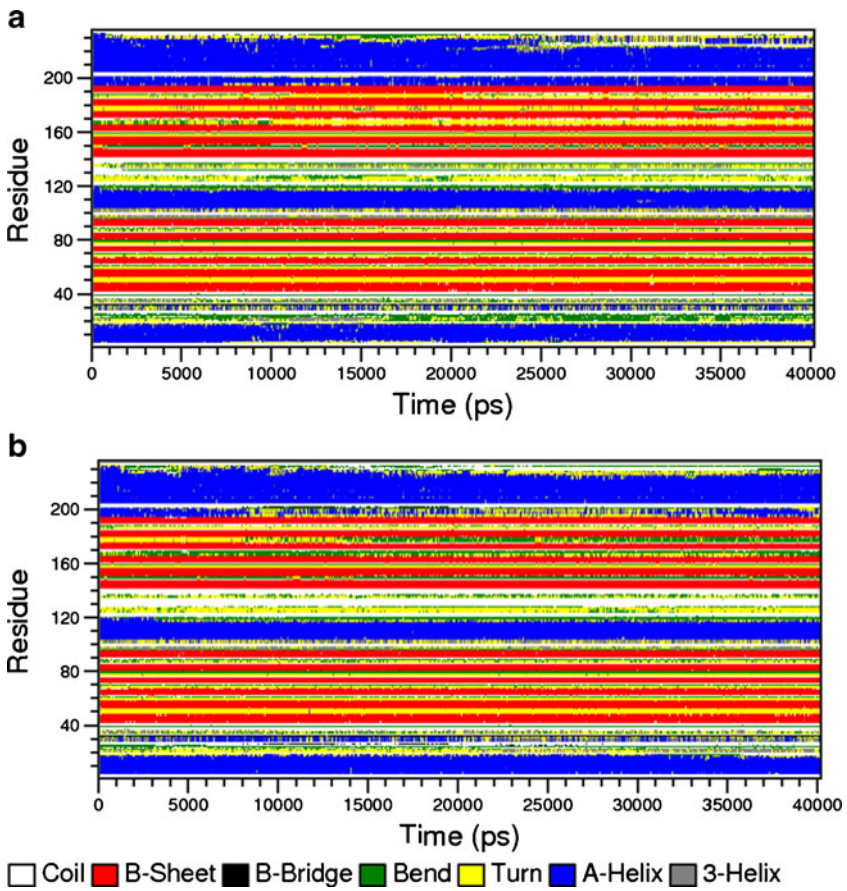

Fig. 6 Time evolution of the secondary structural elements of PLK-1 proteins at 300 K (DSSP classification). **a**. Native, **b**. Mutant

Table 3 Cation-π interactions obtained from CaPTURE program

Cation	Pi	E(es) (kcal mol^{-1})	E(vdw) (kcal mol^{-1})
Arg579	Phe534	−1.41	−1.56
Lys475	Phe436	−2.09	−0.88

functional behavior of PLK1 protein. This was further supported by the atomic density plot and PCA analysis. The consequences of these molecular changes were clearly observed in the atomic density distribution plot. Moreover the mutant structure shows highest atomic density distribution of 39.3 nm^{-3} but in native structure showed 70 nm^{-3} respectively (Fig. 3). The spectrum of the corresponding eigenvalues indicated the level of fluctuation and dynamic behavior of protein molecule in the system and was basically confined within the first two eigenvectors. Native structure showed a small region of phase space particularly along PC1 plane not in the mutant structure (Fig. 4). We obtained the following values for native and mutant (W414F) structures are 109.05 nm^2 and 124.95 nm^2 respectively and again it confirmed the overall increased flexibility in mutant (W414) than in native structure at 300 K.

To examine how the structure got disturbed and affected the functions upon mutation, we analyzed the super imposed initial and final structure of native and mutant PLK1 protein at different time scales. It was clearly observed that there is an increase in loop content in mutant structure as compared to native structure till the end of simulation (Fig. 5). It was well supported by the DSSP analysis. The structural flexibility was

Table 4 The RMSF values of residues acting as stabilizing center in PLK-1 protein

Residues	RMSF(nm)	
	Native	Mutant
Pro407	0.0530	0.1452
Val411	0.0449	0.1233
Ser412	0.0514	0.1171
Gly571	0.0726	0.1236

observed by the analysis of time dependent secondary structure fluctuations. Figure 6 showed helix, coil, bends, sheets and turns in both native and mutant PLK1 protein during simulation period. Mutant (W414F) structure showed an increase in loop and turns content in the amino acid residual position from 15 to 30, 120 to 140, and 185 to 190 as compared to native structure. Moreover the native structure showed more helical content from the amino acid residual position from 195 to 200 when compared to mutant (W414F) structure. In general helices are mostly rigid whereas spanning loop regions are mostly flexible [59–61]. Here, native structure showed continuous loss of coils and turns than mutant structure till the end of simulation. It confirmed that due to mutation the structure became more flexible in the conformation.

The above results strongly showed stability loss in the mutant (W414F) structure. To examine the reason behind such phenotype, we further investigated the cation- π interactions and the stabilizing residues in PLK1 protein. Cation–π interactions were known to be important contributors to protein

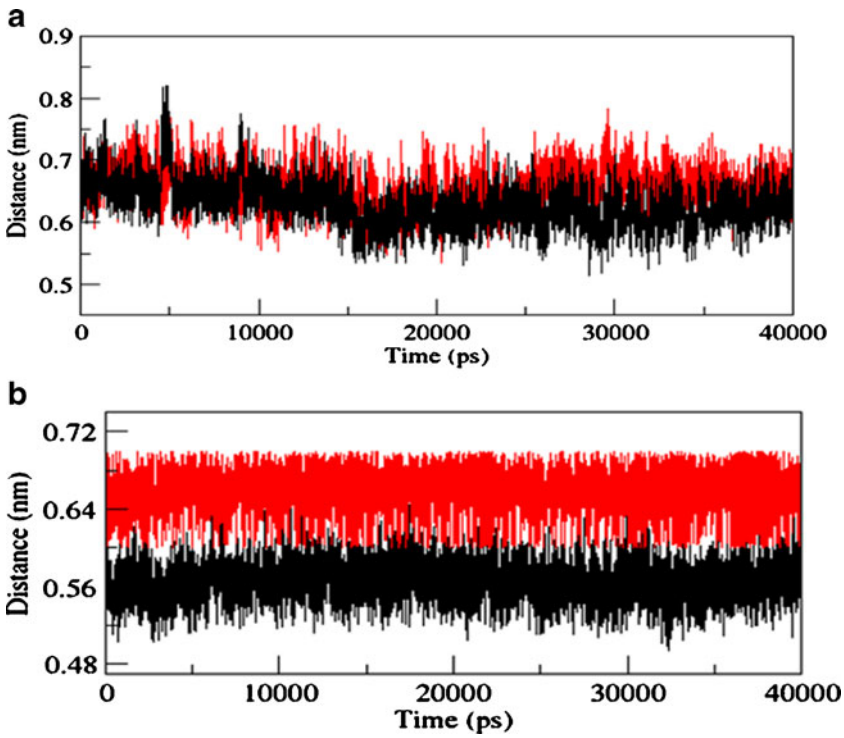

Fig. 7 Distance fluctuation between the cation-π residues (**a**) Arg579-Phe534 (**b**) Lys475-Phe436. Native is shown in black and mutant in red

Table 5 Docking analysis of initial structure (0 ns) of native and mutant PLK1 protein with peptide 1 and peptide 2

Parameters	Complex I	Complex II	Complex III	Complex IV
Binding energy (kcal mol^{-1})	−44.88	−40.47	−44.26	−41.94
Attractive vdW	−19.84	−14.85	−22.23	−17.89
Repulsive vdW	4.45	3.39	13.99	6.66
ACE	−6.35	−8.94	−7.83	−9.43
HB	−1.99	−1.59	−1.57	−0.89
Area	587.1	629.3	731.3	837.3

vdW van der Waals energy; *ACE* Atomic contact energy; *HB* Hydrogen bonds

stability. The importance of cation–π interaction has been examined in several research works for their corresponding role in maintaining the stability of proteins. We observed a total of two energetically significant cation–π interactions (Table 3). To examine if the cation–π interactions were retained in mutant structure, we further investigated the bond distance fluctuations for the distant cation–π interaction (Fig. 7a–b). The bond distance fluctuation showed that the Arg579-Phe534 and Lys475-phe436 cation–π interactions have undergone major interaction losses (Fig. 7a–b). Mutant trajectory of Arg579-Phe534 & Lys475-Phe436 bond distance was comparatively higher than native structure. These results indicate that there was a major loss in cation–π interactions in the mutant (W414F) structure which might have induced higher fluctuation and lower stability of the structure. Furthermore, the protein structure contains certain stabilizing centers (SC) that act as an essential component to maintain the stable conformation of corresponding protein. Residues Pro407, Val411, Ser412, and Gly571 were identified as stabilizing centers. When the RMSF of these residues were investigated, we observed a significant rise in their fluctuation levels (Table 4). Apart from cation-π interactions, the mutation has also introduced major deflection in the PLK1 protein stabilizing centers, which might have significantly contributed toward phenotypic changes observed in mutant W414F structure.

Protein-peptide interaction analysis

From the molecular dynamics approach we examined that, due to mutation PLK1 protein lost stability and became more flexible. And this flexibile conformation can alter the binding phenomenon of protein with ligand molecule. To validate this, further we applied molecular docking approach to evaluate the interaction between the initial (0 ns) and end (40 ns) structure of PLK1 protein with two target peptides upon mutation. In our paper we called native PLK1-peptide 1 complex as complex I, mutant PLK1-peptide 1 complex as complex II, native PLK1-peptide 2 complex as complex III and mutant PLK1-peptide 2 complex as complex IV at 0 ns and 40 ns respectively.

In-depth analysis of the docked complex reveals notable features. Calculation of binding energy is very important to understand the affinity level of biological partners. Overall binding energy of the complex mainly contributes to attractive and repulsive van der Waals interaction energy, atomic contact energy (ACE), hydrogen bond (HB) and area (interfacial surface area of the complex) between PLK1 (native and mutant) and peptide 1 and peptide 2.

Using initial structure of the PLK1 protein, complexes I and II showed binding energy of −44.8 kcal mol^{-1} and −40.47 kcal mol^{-1}, respectively. Complex III showed binding energy of −44.26 kcal mol^{-1} and complex IV exhibited binding energy of −41.94 kcal mol^{-1}, respectively (Table 5). Less interaction has been found in complexes II and IV when compared to complexes I and III and this was due to more interfacial surface area and less contribution of hydrogen bond and this is depicted in Table 5. In a similar manner using end structure of the PLK1 protein, complexes II and IV showed less interaction than complexes I and III, respectively (Table 6). At both stages (0 ns and 40 ns), a more negative value of binding energy was observed in complex I and

Table 6 Docking analysis of end structure (40 ns) of native and mutant PLK1 protein with peptide 1 and peptide 2

Parameters	Complex I	Complex II	Complex III	Complex IV
Binding energy (kcal mol^{-1})	−56.64	−39.89	−55.18	−46.31
Attractive vdW	−23.11	−20.75	−23.49	−22.42
Repulsive vdW	16.86	6.18	7.83	6.28
ACE	−7.82	−10.74	−7.74	−13.04
HB	−1.18	−2.49	−4.15	−0.71
Area	661.9	679.2	690.9	726

vdW van der Waals energy; *ACE* Atomic contact energy; *HB* Hydrogen bonds

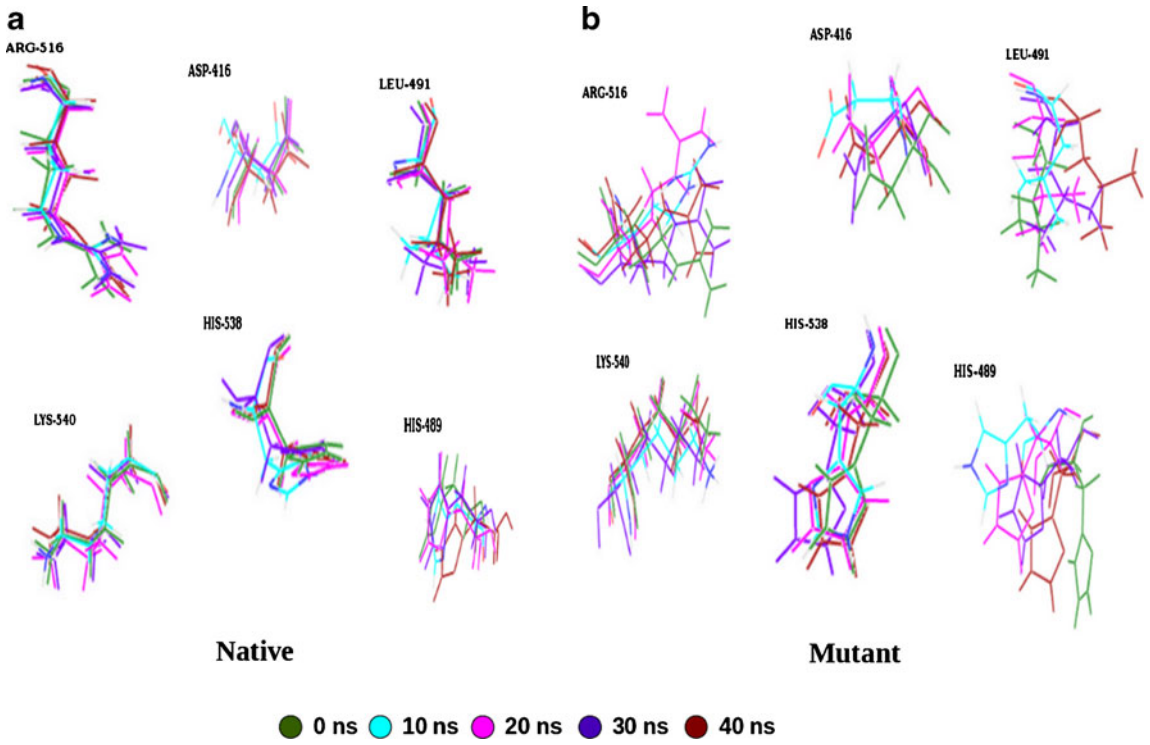

Fig. 8 Movements of PLK1 binding residues (416D, 489H, 491L, 516R, 538H, and 540K) of native and mutant protein during simulations. Average structure of PLK1 binding residues at 0 ns (*green*), 10 ns (*cyan*), 20 ns *(magenta*), 30 ns (*purple*), and 40 ns (*red*) has been superimposed

complex III which indicates high binding affinity toward their biological partners (peptide 1 and 2) when compared to complex II and IV. Major differences between binding energy, ACE and area were observed in all the complexes (Tables 5 and 6).

The binding energy of complex II and IV is significantly higher than complex I and complex III respectively (Tables 5 and 6). This clearly indicates that the effect of mutation in PLK1 loss and the interaction with peptides. And this was well supported by alteration in binding residues of PLK1 protein upon mutation. Superimposed conformers of binding residues of native and mutant (W414F) PLK1 protein at different time scales during simulation were shown in Fig. 8. The significant conformational changes in binding residues are clearly shown. In mutant (W414F) PLK1, the binding residues of 416D, 489H, 491L, 516R, 538H, and 540K exhibited more flexibility than native. Due to flexible conformation the mutant PLK1 residues have less participation in H-bonding with other amino acids, while native PLK1 residues were rigid and have more H-bonds.

MD simulation analysis has collectively suggested that the W414F mutation has strong evidence of inducing phenotypic damages in PLK1. And this damage can alter the binding phenomenon between protein and ligand molecule. In docking analysis we clearly observed that, mutation in PLK1 protein might disturb the interaction with two target peptides which is sufficient to inhibit the over expression of PLK1 protein. These results can be further implemented for drug designing process and develop a potent drug target for PLK1 associated diseases.

Conclusions

In-silico analysis has now become a significant way to characterize a standard disease specific SNP at atomic level. Using multiple computational methods and molecular dynamics simulation approaches, we identified the PLK1 W414F mutation as highly deleterious as well as their molecular mechanisms. The stability loss is clearly observed in RMSD, RMSF, Rg, and SASA analysis in mutant (W414F) structure. Due to mutation, PLK1 protein became more flexible in nature, which is well supported by NHbond, density plot, PCA, and DSSP analysis. This might disturb the structural orientation and binding phenomenon of the PLK1 protein. From the docking approach we clearly observed that, mutation in PLK1 protein might disturb the interaction with target peptide which is sufficient to inhibit the over expression of PLK1 protein. Our analysis suggests an important roadmap to identify the cancer associated mutation and its molecular mechanism of native and mutant PLK1 protein. The insight of mutation (W414F) in PLK1 protein significantly helps researchers to develop drug therapies for PLK1 cancer-associated diseases.

Acknowledgments Authors gratefully acknowledge the management of Vellore Institute of Technology for providing the facilities to carry out this work.

References

1. Barr FA, Sillje HH, Nigg EA (2004) Polo-like kinases and the orchestration of cell division. Nat Rev Mol Cell Biol 5:429–440
2. Van de Weerdt BC, Medema RH (2006) Polo-like kinases: a team in control of the division. Cell Cycle 5:853–864
3. Eckerdt F, Yuan J, Strebhardt K (2005) Polo-like kinases and oncogenesis. Oncogene 24:267–276
4. Strebhardt K, Ullrich A (2006) Targeting polo-like kinase 1 for cancer therapy. Nat Rev Cancer 6:321–330
5. Lee KS, Grenfell TZ, Yarm FR, Erikson RL (1998) Mutation of the polo-box disrupts localization and mitotic functions of the mammalian polo kinase plk. Proc Natl Acad Sci U S A 95:9301–9306
6. Nigg EA (1998) Polo-like kinases: positive regulators of cell division from start to finish. Curr Opin Cell Biol 10:776–783
7. Glover DM, Ohkura H, Tavares A (1996) Polo kinase: the choreographer of mitotic stage. J Cell Biol 135:1681–1684
8. Lane H, Nigg EA (1997) Cell-cycle control: POLO-like kinases join the outer circle. Trends Cell Biol 7:63–68
9. Llamazares S, Moreira A, Tavares A et al (1991) Polo encodes a protein kinase homolog required for mitosis in Drosophila. Genes Dev 5:2153–2165
10. Ohkura H, Hagan IM, Glover DM (1995) The conserved chizosaccharomyces pombe kinase plo1, required to form a bipolar spindle, the actin ring, and septum, can drive septum formation in G1 and G2 cells. Genes Dev 9:1059–1073
11. Lane HA, Nigg EA (1996) Antibody microinjection reveals an essential role for human polo-like kinase 1 (Plk1) in the functional maturation of mitotic centrosomes. J Cell Biol 135:1701–1713
12. Kumagai A, Dunphy WG (1996) Purification and molecular cloning of Plx1, a Cdc25-regulatory kinase from Xenopus egg extracts. Science 273:1377–1380
13. Toczyski DP, Galgoczy DJ, Hartwell LH (1997) CDC5 and CKII control adaptation to the yeast DNA damage checkpoint. Cell 90: 1097–1106
14. Shirayama M, Zachariae W, Ciosk R (1998) The Polo-like kinase Cdc5p and the WD-repeat protein Cdc20p/fizzy are regulators and substrates of the anaphase promoting complex in Saccharomyces cerevisiae. EMBO J 17:1336–1349
15. Descombes P, Nigg EA (1998) The polo-like kinase Plx1 is required for M phase exit and destruction of mitotic regulators inXenopus egg extracts. EMBO J 17:1328–1335
16. Lee KS, Erikson RL (1997) Plk is a functional homolog of Saccharomyces cerevisiae Cdc5, and elevated Plk. Mol Cell Biol 17(6):3408-3417.
17. Seong YS, Kamijo K, Lee JS et al (2002) A spindle checkpoint arrest and a cytokinesis failure by the dominant-negative polo-box domain of Plk1 in U-2 OS cells. J Biol Chem 277:32282–32293
18. Song S, Grenfell TZ, Garfield S et al (2000) Essential function of the polo box of Cdc5 in subcellular localization and induction of cytokinetic structures. Mol Cell Biol 20:286–298
19. Liu J, Lewellyn AL, Chen LG, Maller JL (2004) The polo box is required for multiple functions of Plx1 in mitosis. J Biol Chem 279: 21367–21373
20. Jeong K, Jeong JY, Lee HO (2010) Inhibition of Plk1 induces mitotic infidelity and embryonic growth defects in developing zebrafish embryos. Dev Biol 345:34–48
21. García-Alvarez B, de Cárcer G, Ibañez S et al (2007) Molecular and structural basis of polo-like kinase 1 substrate recognition: implications in centrosomal localization. Proc Natl Acad Sci U S A 104:3107–3112
22. Yun SM, Moulaei T, Lim D et al (2009) Structural and functional analyses of minimal phosphopeptides targeting the polo-box domain of polo-like kinase 1. Nat Struct Mol Biol 16:876–882
23. Strebhardt K (2001) In: Creighton TE (ed) PLK (Polo-like kinase): encyclopedia of molecular medicine. Wiley, New York, pp 2530–2532
24. Balu K, Purohit R (2013) Mutational analysis of TYR gene and its structural consequences in OCA1A. Gene 513:184–195
25. Kamaraj B, Purohit R (2013) In silico screening and molecular dynamics simulation of disease-associated nsSNP in TYRP1 gene and its structural consequences in OCA3. BioMed Res Int. doi:10.1155/2013/697051
26. Kamaraj B, Purohit R (2013) Computational screening of disease-associated mutations in oca2 gene. Cell Biochem Biophys. doi:10.1007/s12013-013-9697-2
27. Kumar A, Rajendran V, Sethumadhavan R et al (2013) Evidence of colorectal cancer-associated mutation in MCAK: a computational report. Cell Biochem Biophys. doi:10.1007/s12013-013-9572-1
28. Kumar A, Rajendran V, Sethumadhavan R et al (2013) Roadmap to determine the point mutations involved in cardiomyopathy disorder: a Bayesian approach. Gene 519:34–40
29. Adzhubei IA, Schmidt S, Peshkin L et al (2010) A method and server for predicting damaging missense mutations. Nat Methods 7:248–249
30. Kumar P, Henikoff S, Ng PC (2009) Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc 4:1073–1081
31. Capriotti E, Fariselli P, Rossi I, Casadio R (2008) A three-state prediction of single point mutations on protein stability changes. BMC Bioinforma 9S6
32. Thomas PD, Campbell MJ, Kejariwal A et al (2003) PANTHER: a library of protein families and subfamilies indexed by function. Genome Res 13:2129–2141
33. Capriotti E, Calabrese R, Casadio R (2006) Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. Bioinformatics 22:2729–2734
34. Calabrese R, Capriotti E, Fariselli P et al (2009) Functional annotations improve the predictive score of human disease-related mutations in proteins. Hum Mutat 30:1237–1244
35. Li B, Krishnan VG, Mort ME et al (2009) Automated inference of molecular mechanisms of disease from amino acid substitutions. Bioinformatics 25:2744–2750
36. Capriotti E, Altman RB (2011) A new disease-specific machine learning approach for the prediction of cancer-causing missense variants. Genomics 98:310–317
37. Purohit R, Rajendran V, Sethumadhavan R (2011) Relationship between mutation of serine residue at 315th position in M. tuberculosis catalase-peroxidase enzyme and isoniazid susceptibility: an in silico analysis. J Mol Model 17:869–877
38. Purohit R, Rajendran V, Sethumadhavan R (2011) Studies on adaptability of binding residues and flap region of TMC-114 resistance HIV-1 protease mutants. J Biomol Struct Dyn 29:137–152
39. Rajendran V, Sethumadhavan R (2013) Drug resistance mechanism of PncA in Mycobacterium tuberculosis. J Biomol Struct Dyn. doi:10.1080/07391102.2012.759885
40. Rajendran V, Purohit R, Sethumadhavan R (2012) In silico investigation of molecular mechanism of laminopathy cause by a pointmutation (R482W) in lamin A/C protein. Amino Acids 43: 603–615
41. Purohit R (2013) Role of ELA region in auto-activation of mutant KIT receptor; a molecular dynamics simulation insight. J Biomol Struct Dyn. doi:10.1080/07391102.2013.803264

42. Balu K, Rajendran V, Sethumadhavan R et al (2013) Investigation of binding phenomenon of NSP3 and p130Cas mutants and their effect on cell signalling. Cell Biochem Biophys: 1–11
43. Yip YL, Scheib H, Diemand AV et al (2004) The Swiss-Prot variant page and the ModSNP database: a resource for sequence and structure information on human protein variants. Hum Mutat 23:464–470
44. Yip YL, Famiglietti M, Gos A et al (2008) Annotating single amino acid polymorphisms in the UniProt/Swiss-Prot knowledgebase. Hum Mutat 29:361–366
45. Boeckmann B, Bairoch A, Apweiler R et al (2003) The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. Nucleic Acids Res 31:365–370
46. Berman HM, Westbrook J, Feng Z et al (2000) The protein data bank. Nucleic Acids Res 28:235–242
47. Kaplan W, Littlejohn TG (2001) Swiss-PDB viewer (deep view). Brief Bioinform 2:195–197
48. Hess B, Kutzner C, Spoel DVD, Lindahl E. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. J Chem Theory Comput 4:435–447
49. Gallivan JP, Dougherty DA (1999) Cation-pi interactions in structural biology. Proc Natl Acad Sci U S A 96:9459–9464
50. Magyar C, Gromiha MM, Pujadas G et al (2005) SRide: a server for identifying stabilizing residues in proteins. Nucleic Acids Res 33: W303–W305
51. Berendsen HJC, Postma JPM, van Gunsteren WF et al (1984) Molecular dynamics with coupling to an external bath. J Chem Phys 8:3684–3690
52. Cheatham TE III, Miller JL, Fox T et al (1995) Molecular dynamics simulations on solvated biomolecular systems: the particle mesh Ewald method leads to stable trajectories of DNA, RNA, and proteins. J Am Chem Soc 117:4193–4194
53. Turner PJ (2005) XMGRACE, version 5.1.19. Center for Coastal and Land-Margin Research, Oregon Graduate Institute of Science and Technology, Beaverton
54. Amadei A, Linssen AB, Berendsen HJ (1993) Essential dynamics of proteins. Proteins 17:412–425
55. Duhovny D, Nussinov R, Wolfson HJ (2002) Efficient unbound docking of rigid molecules. In Gusfield et al. (ed) Proceedings of the 2'nd Workshop on Algorithms in Bioinformatics(WABI) Rome, Italy, lecture notes in computer science 2452. Springer Verlag, pp 185–200
56. Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ (2005) PatchDock and SymmDock: servers for rigid and symmetric docking. Nucleic Acids Res 33:W363–W367
57. Andrusier N, Nussinov R, Wolfson HJ (2007) FireDock: fast interaction refinement in molecular docking. Proteins 69(1):139–159
58. Mashiach E, Schneidman-Duhovny D, Andrusier N et al (2008) FireDock: a web server for fast interaction refinement in molecular docking. Nucleic Acids Res 36:W229–W232
59. Verma D, Jacobs DJ, Livesay DR (2012) Changes in lysozyme flexibility upon mutation are frequent, large and long-ranged. PLoS Comput Biol 8:e1002409
60. Ribeiro AA, de Alencastro RB (2013) Mixed Monte Carlo/molecular dynamics simulations of the prion protein. J Mol Graph Model 42:1–6
61. Kabsch W, Sander C (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22:2577–2637