Prediction of functionally significant single nucleotide polymorphisms in *PTEN* tumor suppressor gene: An *in silico* approach

Imran Khan[1]
Irfan A. Ansari[1*]
Pratichi Singh[2]
Febin Prabhu Dass J[2]

[1] *Department of Biosciences, Integral University, Lucknow, India*

[2] *School of Biosciences and Technology, Vellore Institute of Technology, Vellore, Tamilnadu, India*

Abstract

The phosphatase and tensin homolog (*PTEN*) gene plays a crucial role in signal transduction by negatively regulating the PI3K signaling pathway. It is the most frequent mutated gene in many human-related cancers. Considering its critical role, a functional analysis of missense mutations of *PTEN* gene was undertaken in this study. Thirty five nonsynonymous single nucleotide polymorphisms (nsSNPs) within the coding region of the *PTEN* gene were selected for our *in silico* investigation, and five nsSNPs (G129E, C124R, D252G, H61D, and R130G) were found to be deleterious based on combinatorial predictions of different computational tools. Moreover, molecular dynamics (MD) simulation was performed to investigate the conformational variation between native and all the five mutant *PTEN* proteins having predicted deleterious nsSNPs. The results of MD simulation of all mutant models illustrated variation in structural attributes such as root-mean-square deviation, root-mean-square fluctuation, radius of gyration, and total energy; which depicts the structural stability of *PTEN* protein. Furthermore, mutant *PTEN* protein structures also showed a significant variation in the solvent accessible surface area and hydrogen bond frequencies from the native *PTEN* structure. In conclusion, results of this study have established the deleterious effect of the all the five predicted nsSNPs on the *PTEN* protein structure. Thus, results of the current study can pave a new platform to sort out nsSNPs that can be undertaken for the confirmation of their phenotype and their correlation with diseased status in case of control studies. Volume 64, Number 5, Pages 657–666, 2017

Keywords: *deleterious, in silico, PTEN, RMSD, signaling pathways, SNP*

1. Introduction

The phosphatase and tensin homolog (*PTEN*/MMAC1/TEP1) gene located on a human chromosome 10q23.3 was recently identified as a putative tumor suppressor gene [1, 2]. The *PTEN* gene has been reported as one of the most frequently mutated tumor suppressor gene in several cancers and primarily functions as a cytoplasmic phosphatase involved in regulation of critical signal transduction pathways such as adhesion, growth, invasion, migration, and apoptosis [3]. The *PTEN* gene has nine exons that encode a 403-amino acid cytoplasmic protein [4]. It consists of two distinct domains: The C-terminal region has the C2 domain that binds *PTEN* protein to the plasma membrane, and the N-terminal region has a phosphatase domain with an active site that is responsible for binding of phosphatidyl inositol [PI] substrate [5]. *PTEN* is a primary negative regulator of the PI3K signaling pathway [6]. Loss of *PTEN* leads to overactivaton of PI3K signaling, which in turn is associated with decreased apoptosis, uncontrolled cell proliferation, and increased tumor angiogenesis [7]. Early reports suggest that the point mutations in the *PTEN* gene loci lead to inactivation of the *PTEN* protein function [8]. Mutagenesis studies have shown that single-point mutations in the protein alter the enzymatic properties and also result in loss of function *PTEN* protein. Thus knowledge of individual genetic variations plays a key role in prognosis [9]. The well-established importance of *PTEN* in various human-related cancers makes its functional analysis of missense mutation a significant approach in designing better diagnostic and therapeutic approaches.

Over past few years, various *in silico* studies have attempted to screen missense/nonsynonymous single nucleotide

Abbreviations: *PTEN, phosphatase and tensin homolog; RMSD, root mean square deviation; SNP, single nucleotide polymorphism.*

**Address for correspondence: Irfan A. Ansari, Ph.D., Assistant Professor, Department of Biosciences, Integral University, Lucknow 226 026, India. Tel.: +91-9456241184; Fax: +91-522-2890809; e-mail: iaansari@iul.ac.in, ahmadirfan.amu@gmail.com.*

Received 25 April 2015; accepted 16 January 2016

DOI: 10.1002/bab.1483

Published online 23 August 2017 in Wiley Online Library (wileyonlinelibrary.com)

polymorphisms (nsSNPs) within the protein coding region of a gene, using sequence-based information and structural attributes. Computational methods classify these SNPs as deleterious through considering various aspects such as sequence conservation among various species, structural features, and physiochemical properties of protein [10–14]. Structural variations in protein can occur due to amino acid substitutions and their biochemical properties (basic, acidic, or hydrophobic) and also by the location of the substitution in the protein sequence [15]. So, we employ two diverse approaches for the analysis of deleterious nsSNPs through empirical-based methods and support vector-based approaches to profile deleterious nsSNPs of the *PTEN* gene. Thus, the goal of the study is to identify nsSNPs for the *PTEN* gene that are likely to alter the structural and functional aspects of the *PTEN* protein.

2. Materials and Methods

2.1. SNP data retrieval

The SNPs were retrieved from the SNP database of National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/snp) using various limits of *Homo sapiens* coding nonsynonymous, stop gained, coding synonymous, mRNA UTR (5′ and 3′), and intronic regions [16].

2.2. SIFT

SIFT (Sorting Intolerant From Tolerant) (http://sift.jcvi.org/www/SIFT_BLink_submit.html) is a sequence homology-based algorithm that classifies amino acid substitutions. We performed updated version of SIFT called SIFT-Blink by submitting a query in the form of gene identification number and the amino acid substitutions obtained from NCBI. The SIFT predictions are provided as a normalized probability score chart for all 20 amino acids. SIFT scores less than 0.05 is predicted as deleterious [17, 18].

2.3. PolyPhen-2

PolyPhen-2 (Polymorphism Phenotyping v2), (http://genetics.bwh.harvard.edu/pph2/) is an automatic structure homology–based tool for the prediction of functional and structural impact of amino acid substitution in protein. PolyPhen-2 extracts information on sequence, structural, and phylogenetic features to characterize an amino acid substitution [19, 20]. We submitted the query in the form a protein FASTA sequence with mutational positions each with two amino acids variants. PolyPhen searches for three-dimensional (3D) protein structures, multiple alignments of homologous sequences, and amino acid contact information in several protein structure databases. A position-specific independent counts (PSIC) score and their difference are generated for two variants. The PSIC score difference of 1.5 or above is characterized as damaging. The PolyPhen scores are classified as probably damaging (2.00), possibly damaging (1.50–1.99), potentially damaging (1.25–1.49), or benign (0.00–0.99).

2.4. I-Mutant2.0

I-Mutant2.0 (http://folding.biofold.org/i-mutant/i-mutant2.0.html) is an automatic support vector machine (SVM)-based prediction tool for the automatic prediction of protein stability changes upon single-point mutations. I-Mutant2.0, that is a predictor of protein stability changes upon single-point mutation at a homepage, consist of protein structure and protein sequence. I-Mutant predicts the free energy change (DDG) calculated by subtracting the free energy change of the mutant protein from the free energy change of the native protein (kcal/mol) [21].We provided query protein FASTA sequence along with position and amino acid substitutions, and free energy change (DDG) were obtained. A zero value of DDG predicts high stability of mutant protein and more negative the value lesser will be the stability of the mutant protein.

2.5. PANTHER

PANTHER (Protein Analysis Through Evolutionary Relationship) is a database with systematic arrangement of protein families and their subfamilies. PANTHER database is generated via Hidden Markov Models (HMMs), a mathematical model. On the basis of evolutionary relationship between related proteins, it calculates the Substitution Position-Specific Evolutionary Conservation (subPSEC) scores. All the nsSNPs were analyzed using PANTHER for validating its impact on the protein function upon single-point mutation. Thus, the FASTA format of the protein sequence itself and substitution changes were provided as input. The functional impact of SNPs on the proteins is predicted on the basis of subPSEC scores. The protein sequence with subPSEC scores ≤–3 is predicted to be deleterious, and protein sequences with subPSEC scores 0 are predicted to be neutral [22].

2.6. 3D structure modeling

To evaluate the structural stability of the protein upon substitution, the X-ray crystallographic 3D structure of the *PTEN* protein (pdb id- 1DR5) was downloaded from the UniProt [23].The structure was validated using PROCHECK (http://www.ebi.ac.uk/thornton-srv/software/PROCHECK/) [24, 25]. The mutant protein models were generated using SWISS PDB Viewer [26].

2.7. Molecular dynamics simulation

Molecular dynamics simulation studies were carried out by GROMACS 4.5.3 program package with GROMOS9643a1 force field for energy minimization [27]. At first, all models were solvated with the 0.9-nm simple-point charge water embedded in a cubic simulation box [28]. All the native and mutant structures were subjected to the steepest decent energy minimization to the tolerance level of 100 kJ/mol. To acquire an electrically neutralized system, random water molecules were replaced with Na^+ or Cl^- ions using the GENION procedure of GROMACS package. The number of particles, pressure, and temperature were kept constant thorough applying periodic boundary conditions. The Berendsen algorithm was employed for maintaining constant temperature with a coupling time

of 0.2 [29]. Finally, all the native and mutant structures were equilibrated for 10,000 ps each at 300 K and the equilibrated structures were subjected to molecular dynamics simulations for 5 ns each at 300 K. Long-range Coulombic interactions were treated through the particle mesh Ewald method [28], and the SANDER module was used to perform simulations. Bond lengths were constrained using the SHAKE algorithm involving hydrogen's permitting a time step of 2 fs. At regular time intervals of 1 ps, the coordinates were saved. The van der Waals force was maintained at 1.4 nm, and Coulomb interactions were truncated at 0.9 nm.

2.8. Analysis of molecular dynamics trajectories

Trajectory files of native and mutant models of *PTEN* protein generated from built-in functions of GROMACS 4.5.3 were analyzed. Trajectory files were analyzed using g˙rms, g˙rmsf, g˙sas GROMACS utilities for total energy, root mean square deviation (RMSD), radius of gyration (Rg), root mean square fluctuation (RMSF), and solvent accessibility surface area (SASA). The g˙h bond utility was used for calculation of the number of distinct hydrogen bonds formed in the protein during simulation. The number of hydrogen bonds was determined at a donor–acceptor distance less than 3.9 nm and donor—hydrogen–acceptor angle larger than 90° [30].

3. Results

3.1. Data mining

For the current study, we have selected SNPs of the *PTEN* gene, for *H. sapiens* using various limits, of coding nonsynonymous (nsSNPs) region, coding synonymous (sSNPs) region, stop-gain, mRNA untranslated regions UTR (5′ and 3′), and intronic regions. Out of 1,782 SNPs, the coding region contains 35 nsSNPs (1.96%) and 15 sSNPs (0.84%); stop-gain region contains 8 SNPs (0.45%); and noncoding regions contain 1,642 SNPs (92.14%) in the intronic region and 82 SNPs (4.59%) in the mRNA UTR region with 25 SNPs in 5′ UTR and 57 SNPs in 3′ UTR. Since, vast majority of SNPs were found to be in the intronic region, therefore SNPs within the coding nonsynonymous region or regulatory region (35 nsSNPs) of the *PTEN* gene were selected for our investigation. The functional impact of these sorted nsSNPs was further assessed by various bioinformatics tools.

3.2. Analysis of deleterious nsSNPs of the *PTEN* gene using SIFT Blink

On the basis of sequence homology and the physical properties of amino acids, SIFT predicts whether the amino acid substitution alters the protein function. The protein sequence with mutational positions and amino acid residues was submitted to the SIFT Blink server as input. The intolerant range of SIFT is ≤0.05, which indicates that nsSNP is more damaging/deleterious to the protein function and score of >0.05 predicts the tolerant range. Out of 35 nsSNPs, 18 nsSNPs (51.42%) were predicted to be intolerant with 0.00 scores,

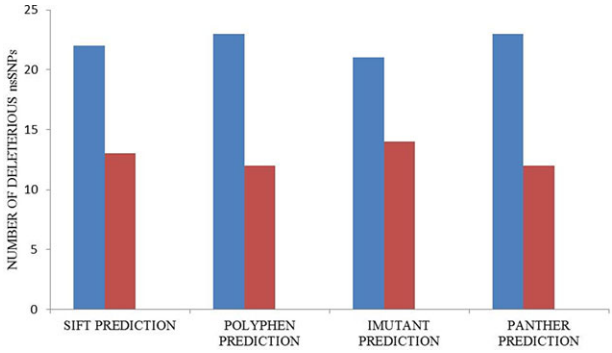

FIG. 1 *Prediction of deleterious nsSNPs of the PTEN gene by the SIFT Blink, PolyPhen-2.0, I-Mutant2.0, and PANTHER. The bar diagram indicates the number of deleterious and benign nsSNPs predicted by various tools: The blue bar indicates the deleterious nsSNPs, and the red bar indicates the benign nsSNPs.*

4 nsSNPs (11.42%) showed the score ranging from 0.01 to 0.05, and 13 nsSNPs (37.14%) were found to be tolerant with a score ranging from 0.06 to 0.10. Thus, in total 22 nsSNPs (62.84%) were predicted to be intolerant that may alter functional properties of protein, as shown in Fig. 1. Therefore, these scores enable the quantitative comparison and ranking of the nsSNPs according to their deleterious nature and allowing researchers to decide which SNP to be targeted for further investigation.

3.3. Analysis of deleterious nsSNPs of the *PTEN* gene using PolyPhen-2

The PolyPhen prediction is based on the number of sequences, phylogenetic and structural features, characterizing the amino acid substitution. The same protein sequence with mutational positions and amino acid substitutions, submitted to SIFT as input, was also submitted to Polyphen 2 (i.e., 35 nsSNPs in total). PolyPhen scores comprise a range from zero to a positive number, where zero indicates a neutral effect on amino acid substitution and higher the positivity, higher will be the detrimental effect of substitution on the protein function. Out of 35 nSNPs, 23 nsSNPs (65.7%) were predicted to be probably damaging depicting score values ranging from 0.8 to 1, 2 nsSNPs (5.7%) were found to be possibly damaging having score values 0.6–0.8, and 10 nsSNPs (28%) were identified as benign (neutral) showing a score value of zero, as shown in Fig. 1. Thus the PolyPhen score is useful in quantitative characterization of the damaging effects of nsSNPs on the protein function.

3.4. Analysis of deleterious nsSNPs of the *PTEN* gene using I-Mutant2.0

The query protein sequence along with mutational positions and amino acid substitutions, submitted to the SIFT Blink and PolyPhen 2.0, was also submitted as input to I-Mutant2.0. I-Mutant2.0 predicts the alterations in protein stability due to the presence of single-point mutations. According to

TABLE 1 *Combined prediction of SIFT Blink, PolyPhen-2.0, I-Mutant2.0, and PANTHER for possible deleterious nsSNPs of the PTEN gene*

S. No.	rsID	Allele	Residue change	SIFT prediction	PolyPhen score	I-Mutant DDG	PANTHER subPSEC
1	rs121909236	C/G	H61D	0.01	0.998	−2.2	−6.95989
2	rs121909223	T/C	C124R	0	0.999	−2.29	−5.46188
3	rs121909218	G/A	G129E	0	1	−1.71	−3.55832
4	rs121909224	C/G	R130G	0	0.999	−2.17	−7.00463
5	rs121909239	A/G	D252G	0	0.862	−2.03	−4.22445

I-Mutant2.0, more negative the value of free energy change (DDG), lesser will be the stability of protein. According to these score, out of 35 nsSNPs, nine variants C124R, R130G, D19N, H61D, A121G, D252G, D107N, R173H, and D115N showed DDG values –2.29, –2.17, –2.31, –2.2, –3.09, –2.03, –2.07, –2.84, and –3.04, respectively, which were considered to be least stable and most deleterious nsSNPs. The other 11 variants G129E, R130Q, L112P, H93R, F241S, T167N, D153H, R173C, V119L, I135T, and Q298E, showing DDG values ranging from –0.05 to –1.9, were further identified as less stable and deleterious nsSNPs. The remaining 15 variants showing DDG values ≥0.00 were grouped as nondeleterious. Thus, in total 20 nsSNPs/variants (57.14%) were predicted to be deleterious to the protein stability, as shown in Fig. 1.

3.5. Analysis of deleterious nsSNPs of the *PTEN* gene using PANTHER

PANTHER estimates the impact of single-point mutation on the protein function by calculating the substitution position-specific evolutionary conservation score (subPSEC) based on the alignment of evolutionary-related proteins. This tool adds the next layer of complexity in refining the nsSNPs according to protein functions. subPSEC score varies from 0 [neutral] to about –10 (most likely to be deleterious), and a score ≤–3 is said to be deleterious. Out of 35 nsSNPs, nine variants S170R, C124R, R130G, L70P, R130Q, H61D, G132V, R173C, and R173H having subPSEC values –7.19905, –5.46188, –7.00463, –5.4489, –7.04173, –6.95989, –5.3268, –8.266, and –7.63074, respectively, were considered to be highly deleterious. Other 14 variants G129E, H123R, M35R, L112P, D19N, H93R, D252G, F241S, F271S, T167N, T131I, D153H, L220P, and I135T having score –3.55832, –4.62615, –4.76529, –4.17137, –4.00395, –4.62615, –4.22445, –3.24713, –3.24713, –3.09784, –4.50276, –3.07801, –3.99507, and –2.93223, respectively, were grouped under deleterious nsSNPs. The remaining 12 variants V217I, R234Q, A121G, A79T, V290L, D107N, V119L, S294R, D115N, D297Y, Q298E, and P354L showing scores ≥ –3 were classified as nondeleterious, as shown in Fig. 1.

3.6. Combinatorial prediction of functionally deleterious nsSNPs of the *PTEN* gene

As each tool works on a specific algorithm, which is based on variable biological aspect, therefore, through combining different computational tools, prediction of deleterious nsSNPs can be performed more accurately [31, 32]. Thus to reduce the false positive predictions, we opted for a combinatorial approach involving empirical and SVM-based tools and selecting only those SNPs for further trajectory analysis that were commonly predicted to be deleterious by all the tools. By comparing the result of SIFT Blink, PolyPhen-2, I-Mutant2.0, and PANTHER tools, five variants G129E, C124R, R130G, H61D, and D252G, having SNP IDs rs121909218, rs121909223, rs121909224, rs121909236, and rs121909239, respectively, with the highest predicted scores of SIFT Blink, PolyPhen 2.0, I-Mutant 2.0, and PANTHER, as shown in Table 1.

3.7. Modeling of mutant *PTEN* protein structures

Protein stability can be significantly altered by the presence of SNP. Thus, for better insight on single nucleotide change and its role in protein functional and structural stability, it is mandatory to gather knowledge about 3D structure. The 3D structure of *PTEN* protein (ID-1D5R) was obtained from Universal Protein Resource (UniProt) having 403 amino acids. Later, the PROCHECK server was used to validate the native structure; the structure had resolution of 2 Å, and 80% of residues were found in the most favored region in the Ramachandran plot. Mutant models for all the five nsSNPs sorted via combinatorial analysis were generated using SWISSPDB Viewer, as presented in Table 1. To investigate precisely the deleterious effect of these nsSNPs on the structure and function of the *PTEN* protein, further molecular dynamics studies were performed.

3.8. Structural stability and flexibility analysis by total energy, RMSD, RMSF, and Rg value calculation

To gain better understanding of the structural consequences of the selected functional single nucleotide polymorphisms in mutants *PTEN* protein structures, we performed trajectory analysis through molecular dynamics simulation of native and mutant *PTEN* protein structures. The trajectory analysis

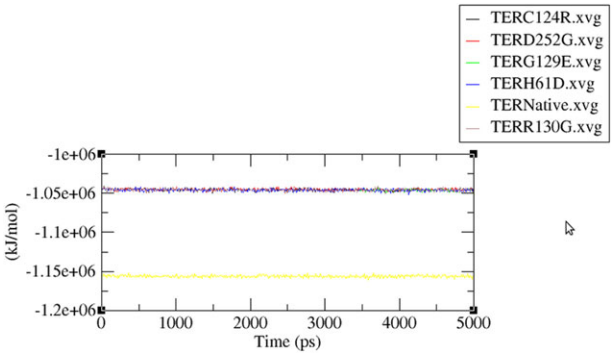

FIG. 2 *Total energies of the native and mutant PTEN protein. The ordinate is energy (kJ/mol), and the abscissa is time (ps). The yellow, black, red, green, blue, and brown lines indicate the native, C124R, D252G, G129E, H61D, and R130G PTEN structures, respectively.*

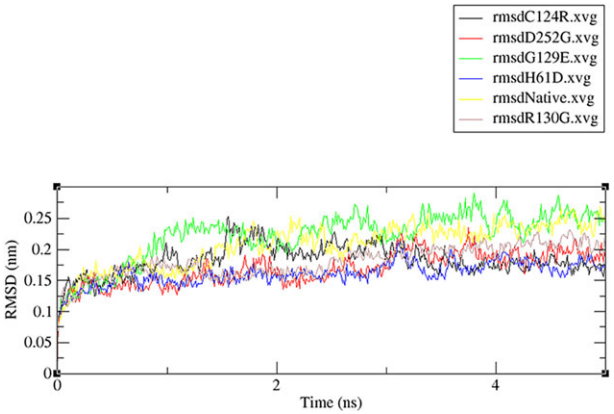

FIG. 3 *Backbone RMSD values of the native and mutant PTEN protein. The ordinate is RMSD (nm), and the abscissa is time (ps). The yellow, black, red, green, blue, and brown lines indicate the native, C124R, D252G, G129E, H61D, and R130G PTEN structures, respectively.*

involved the identification of the total energy, RMSD, RMSF, SASA, and hydrogen bond variation in selected mutant *PTEN* models (C124R, D252G, G129E, H61D, and R130G) having highly deleterious nsSNPs. The total energy of all the mutant models was compared with the native model as shown in Fig. 2. All the five mutant *PTEN* protein models G129E, C124R, D252G, H61D, and R130G demonstrated higher total energy values of ~−1.05e+06 kJ/mol compared to ~−1.15 e+06 kJ/mol total energy value of the native *PTEN* protein. The RMSD values for the Cα atom of native and mutant models are analyzed in Fig. 3. At the first 1-ns lapse, the RMSD values showed no variations. The native *PTEN* protein structure showed the RMSD values ranging from 0.1 to 0.25 nm. The mutant G129E showed variation in RMSD values from 0.1 to 0.3 nm. On the other hand, mutant structures C124R, D252G, H61D, and R130G showed RMSD values ranging from 0.1 to 0.2 nm. This

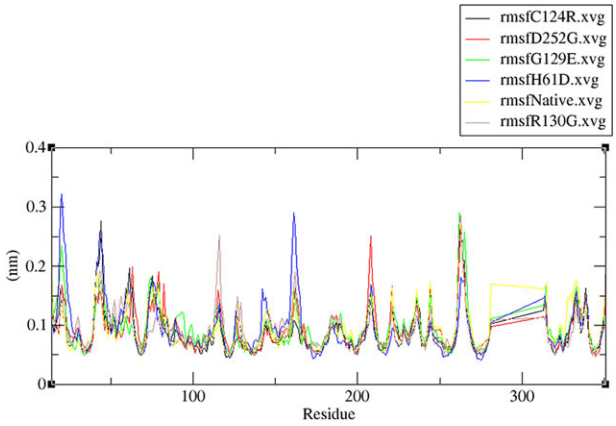

FIG. 4 *RMSF value of the native and mutant PTEN protein. The ordinate is RMSF (nm), and the abscissa is time (ps). The yellow, black, red, green, blue, and brown lines indicate the native, C124R, D252G, G129E, H61D, and R130G PTEN structures, respectively.*

deviation in the RMSD values in mutant models supports the idea of change in stability and explains the impact of these single nucleotide polymorphisms leading to substitution of an amino acid in the protein sequence.

To determine the deviation in the flexibility of the *PTEN* mutant models, the RMSF values of native and mutant models were monitored. The RMSF values of native and mutant models are shown in Fig. 4. During the entire 5-ns simulation period, mutant models H61D, R130G, and D252G showed fluctuation peaks ranging from ~0.1 to ~0.3 nm, ~0.1 to ~0.25 nm, and ~0.1 to ~0.26 nm, respectively, when compared to native RMSF values ranging from ~0.1 to ~0.25 nm. In the early ~0 to 280 ps simulation period, the mutant models C124R and G129E showed a similar pattern of RMSF values as of the native model. At ~280 to 330 ps time frame of simulation, all five mutant models H61D, R130G, C124R, G129E, and D252G showed a significant deviation from the native *PTEN* protein model. These fluctuations in the mutant models depict the flexibility changes in the mutant structures and reflect the impact of these missense mutations on the stability of the *PTEN* protein.

To gain insights into the complete 3D of the native and mutant structures, we examined the Rg. Rg is the mass-weighted root-mean-square distance of a collection of atoms in the protein from their common center of mass. Figure 5 displays the Rg value fluctuations in *PTEN* mutant structures when compared to native *PTEN* protein structures. During complete 5 ns simulation time, the native *PTEN* protein showed Rg values ranging from ~2.15 to ~2.25 nm. All the five mutant structures H61D, R130G, C124R, G129E, and D252G showed a notable decrease in the Rg values. Mutant structures D252G and G129E showed a significant decrease in Rg values ranging from ~2.15 to ~2.12 nm and ~2.15 to ~2.1 nm, respectively; whereas mutant structures H61D, R130G, and C124R showed a slight decrease in Rg values compared to the native *PTEN*

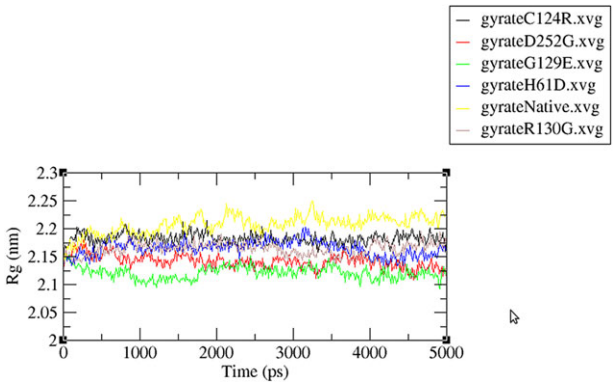

FIG. 5 *Rg of the backbone carbon alpha for the native and mutant PTEN protein. The ordinate is Rg (nm), and the abscissa is time (ps). The yellow, black, red, green, blue, and brown lines indicate the native, C124R, D252G, G129E, H61D, and R130G PTEN structures, respectively.*

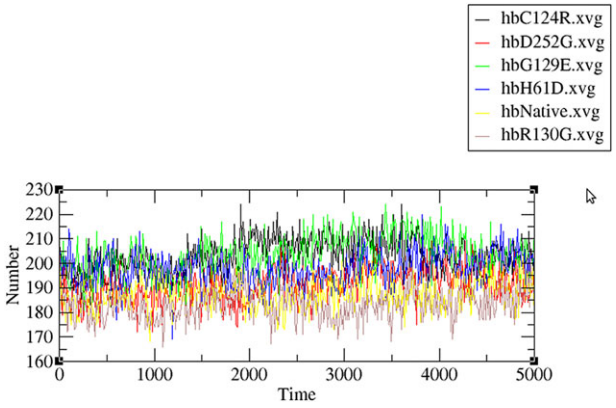

FIG. 6 *The number of hydrogen bonds formed in native and mutant PTEN protein. The ordinate is RMSD (nm), and the abscissa is time (ps). The yellow, black, red, green, blue, and brown lines indicate the native, C124R, D252G, G129E, H61D, and R130G PTEN structures, respectively.*

structure ranging from ~2.15 to ~2.2 nm, ~2.15 to ~2.2 nm, and ~2.15 to ~2.12 nm, respectively.

3.9. Prediction of intermolecular hydrogen bond frequencies

The intermolecular hydrogen bond frequency is responsible for the stability of the protein structure [33]. Thus, examining the number of hydrogen bonds in native and mutant structures was important for the better insight into the stability of residues in these structures. Figure 6 depicts the number of hydrogen bonds formed in the *PTEN* protein in its native and all the five prioritized mutant models. The native *PTEN* protein structure exhibited on average ~170 to ~210 hydrogen bonds throughout the 5-ns simulation period. Mutant structures C124R, G129E, and H61D showed an increase in the average number of

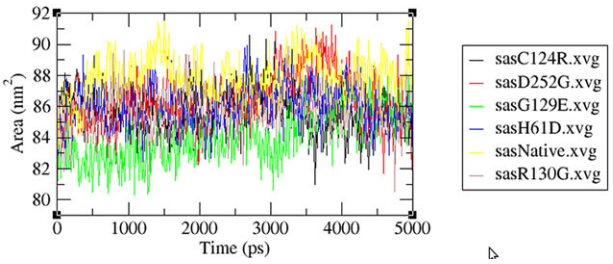

FIG. 7 *Solvent accessible surface area of the native and mutant PTEN protein. The ordinate is area (nm^2), and the abscissa is time (ps). The yellow, black, red, green, blue, and brown lines indicate the native, C124R, D252G, G129E, H61D, and R130G PTEN structures, respectively.*

hydrogen bonds to the highest value of ~225, ~220, and ~215, respectively. The mutant structure D252G showed a minor decrease in the number of hydrogen bonds to the maximum value of ~170. During the entire 5-ns simulation time period, the mutant structure R130G displayed a significant decrease in the number of hydrogen bonds ranging from ~185 to ~165. This variation in the number of hydrogen bonds in the mutant *PTEN* protein structures reflects the deleterious effect of these single nucleotide mutations on the *PTEN* protein.

3.10. Prediction of variations in solvent accessibility

The SASA of a protein is the surface of protein in contact with the solvent molecules. The solvation process plays an important role in protein stability and rearrangement of protein residues during protein folding. The SASA values were calculated for both native and mutant *PTEN* protein structures. Figure 7 displays the SASA values of the native *PTEN* protein structure ranging from ~83 to 91 nm^2 during the entire 5-ns simulation time period. All the mutant models showed notable reduction in the solvent accessible area compared to the native model. The mutant structure G129E displayed notable reduction in the SASA value during the initial 0 to 3,000 ps simulation time frame. Mutant structures H61D and D252G showed a slight reduction in solvent accessible area ranging from ~82 to 88 nm^2 during the initial 0 to 2,500 ps simulation time period. Later on, during the 2,500–4,000-ps simulation time period, mutant structures H61D and D252G displayed a hike in solvent accessible area and reaching a SASA value of ~91 nm^2. Mutant models C124R and R130G showed a slight decrease in the SASA value compared to the native *PTEN* protein structure but followed a similar pattern of the SASA value during the entire 5-ns simulation time.

4. Discussion

The increasing bioinformatics approaches for *in silico* analysis of amino acid substitution have progressed considerably nowadays [34]. The current study is focused on profiling the deleterious nsSNPs of the *PTEN* tumor suppressor gene

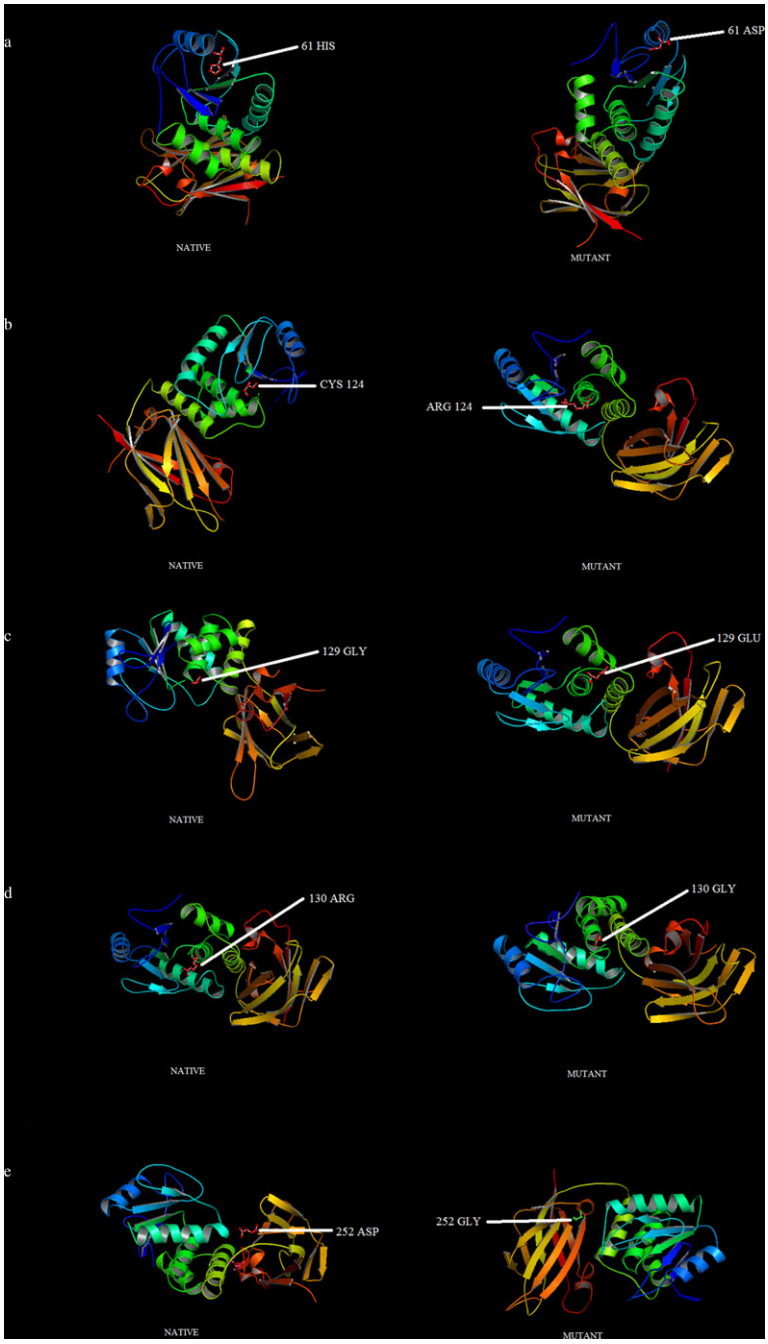

FIG. 8 *3D structure of the native and mutant PTEN protein models representing the residues changes in native and mutant models denoted in ball-and-stick forms: (a) The native-type histidine and substitution of aspartic acid at position 61, (b) the native type cysteine and substitution of arginine at position 124, (c) the native-type arginine and substitution of glutamic acid at position 129, (d) the native-type arginine and substitution of glutamine at position 130, and (e) the native-type aspartic acid and substitution of glycine at position 252.*

associated with various cancers. Initially, the SNP data were extracted from the NCBI database and nsSNPs were sorted for further computational analysis by different bioinformatics tools. Various tools like SIFT Blink, PolyPhen-2, I-Mutant2.0, and PANTHER are helpful in predicting the functional phenotypes of nsSNPs based on protein structure, protein sequence cross species conservation, and physicochemical properties [10–14, 35]. Previous studies have reported that using multiple tools and algorithms for prioritizing functional mutations enhances the accuracy of prediction [15, 36]. The SIFT algorithm calculates the score based on the sequence homology; it determines the extent of evolutionary conservation to the candidate amino acid and predicts whether a particular substitution is deleterious or tolerant to the protein stability. Of the 35 missense substitutions, 22 substitutions were predicted by SIFT to be deleterious. These 22 substitutions showed SIFT scores ≤ 0.5, and the remaining 13 substitutions were predicted as tolerant as the SIFT score was ≥ 0.5. Further analysis was performed to predict the functional impact of the substitutions on the protein 3D structure, as the predictive power of sequence information combined with structural information is crucial in studying the role of mutations in a gene and its possible association with the disease. Polyphen-2 was employed to predict the effect of substitutions on the 3D structure of the *PTEN* protein. The results from Polyphen-2 were in concordance with the SIFT results and predicted 23 probably damaging substitutions. Previous studies by Rajith and George [15] have also reported good concordance between SIFT and Polyphen results. Other studies performed to check the performance of these tools have also reported that SIFT and Polyphen as reliable tools for identifying functional nsSNPs [37]. It is well established that the protein stability is changed upon site-specific mutations; there are many tools that have been developed to calculate the free energy change for predicting the change in stability upon substitution of amino acid residues in protein structures [38, 39]. Thus, we used SVM-based tool I-Mutant2.0, which is based on a new approach of neural network proposed by Capriotti et al. [40]. This tool predicts the direction of the stability shift due to mutation in protein structures as a function of $\Delta\Delta G$. The value of free energy change ($\Delta\Delta G$) can vary from negative to positive corresponding to a decrease or increase in stability of protein structures [41]. The results predicted 21 missense substitutions as deleterious with negative $\Delta\Delta G$ values, and there was no significant variation with the previous predictions of SIFT and Polyphen. Further analysis for the functional nsSNPs was done through the HMM algorithm-based tool PANTHER. It is sequence-based software, which calculates the substitution-specific evolutionary conservation scores (SubPSEC). It predicts the functional impact of a substitution if the SubPSEC score is ≥ 3. In our analysis of 35 missense substitutions, 23 were predicted as deleterious having PANTHER score ≥ 3 and 12 substitutions were predicted to be nondeleterious. The variation in the predictions of SIFT Blink, PolyPhen-2, I-Mutant 2.0, and PANTHER was found to be very less, but we incorporated a combinatorial approach to dismiss any possible false positive results. Thus, we selected only those nsSNPs for further investigation that were predicted as deleterious by all selected tools.

For the better understanding of functional impact of nucleotide substitutions in a gene sequence, it is required to gain knowledge about the 3D structure of protein encoded by the gene. Thus, to further improve our analysis, we downloaded the template structure 1DR5 of *PTEN* protein from UniProt and checked for the stability via the PROCHECK server and the structure was found to be a good quality model having 90% residues in the most favorable region A, B, L, on the Ramachandran plot [25].

The best possible approach to study the variation in the protein structure and stability upon mutation *in silico* is to simulate *PTEN* mutant models and analyze how these residue changes affect the protein structure at its functional and structural attributes compared to the native *PTEN* models. To achieve this objective, we incorporated the SWISSPDB viewer tool to construct mutant *PTEN* models [26] and employed a molecular dynamics approach to compare the native and mutant *PTEN* protein models.

Five basic parameters RMSD, RMSF, SASA, hydrogen bonds, and Rg were analyzed for 5 ns of simulation trajectory. To detect the degree of change in the arrangement of atoms arising due to point mutations, the RMSD was calculated, which illustrated the stability of the mutant model compared to the native model. The results of *PTEN* mutant models G129E, C124R, D252G, H61D, and R130G showed notable fluctuations in the RMSD values. A higher RMSD value denotes higher stability, which further leads to higher protein rigidity and vice versa [42]. The RMSF analysis indicated the fluctuations in molecular flexibility of all the five *PTEN* mutant models, but concomitantly the analysis directed the greater fluctuation in G129E, D252G, and H61D mutant models. Consistent with RMSD and RMSF results, the Rg and total energy values also demonstrated similar results. All the five mutant models showed a notable increase in the total energy. Greater fluctuation in the Rg value was observed in G129E, D252G, and H61D mutant models.

It is well established that stability to protein structures is provided through several noncovalent interactions, for example, hydrogen bonding, hydrophobic, van der Waals, and electrostatic interactions [43, 44]. Alterations in the hydrogen bond pattern can be important characteristics helpful in understanding the effect of deleterious mutations on the protein structure [45]. The results of hydrogen bond analysis showed significant variation in the intermolecular hydrogen frequency of mutant *PTEN* structures compared to the native *PTEN* structure. The native *PTEN* model displayed hydrogen frequency on average ranging from ~170 to ~210, whereas all five mutant structures D252G, R130G, C124R, G129E, and H61D demonstrated a variable number of hydrogen bonds ranging from ~170, ~185 to ~165, ~225, ~220, and ~215, respectively. The driving forces for the protein folding are thought to be the hydrophobic interactions [46]. The tendency to resist the folding in the structure is driven by cooperative interaction

between the residues in protein structures [47, 48]. Thus, variation in hydrogen bond frequencies of the mutant *PTEN* models demonstrated the deleterious nature of predicted nsSNPs in *PTEN* gene resulting into the residue changes. The SASA values illustrated the solvent accessible area of the protein that interacts with the surrounding solvent molecules. Furthermore, the SASA analysis of the mutant *PTEN* protein displayed notable fluctuations from the native *PTEN* protein structure. The increased solvent-accessible area in the mutant protein structures indicates that the single nucleotide polymorphism causing residues changes in the protein structures may increase the probability of their interaction with surrounding molecules and vice versa [49]. Thus, the trajectory analysis for the mutant structures G129E, C124R, D252G, H61D, and R130G revealed the extent of deleterious effects of single nucleotide polymorphisms influencing the functional and structural attributes of the protein.

The 3D structures of the native and the mutant *PTEN* protein were drawn through PyMol molecular graphics system providing a clear image of the residues changes as shown in Fig. 8 [50]. The *PTEN* protein structure consists of two active domains: the N-terminal domain spanning from residues 7–185 and the C-terminal domain spanning from residues 186–351. The residue change R130G lies in the N-terminal phosphatase activity domain, whereas residue change D252G lies under the C-terminal domain. Studies have reported the mutation in the phosphatase domain that leads to loss of *PTEN* activity [51]. Furthermore, it is also well documented that specific amino acid residue phosphorylation within the C-terminal domain plays an important role in stabilizing the *PTEN* protein [52, 53]. As the matter of fact, several studies have shown that mutations in these domains are related to diseased status, but still majority of mutation in these two domains and outside these domains remain to be analyzed. In summary, the results of this study suggested two highly deleterious nsSNPs of the *PTEN* gene that can be undertaken for epidemiological studies to assess their phenotypic association and correlation with different human cancers.

5. Conclusion

In conclusion, the investigation of the *PTEN* tumor suppressor gene, having 35 SNPs in the coding region, leads to profiling of five highly deleterious SNPs rs121909218, rs121909223, rs121909224, rs121909236, and rs121909239 using different bioinformatics tools in combination with the molecular dynamics approach. Since there is an enormous amount of SNP data available for the *PTEN* gene, it might not be feasible for a researcher to carry out wet laboratory studies on every single SNP to determine their biological impact on structural and functional stability of protein encoded. The methods involved will be difficult, time consuming, and expensive to design experiments for characterizing the impact of each nsSNPs on the protein function. Thus, this study paves an alternative approach, which is inexpensive and less time consuming to prioritize amino acid substitutions and short-list candidate SNPs of the *PTEN* gene for further wet laboratory analysis.

6. Acknowledgements

The authors thank the management of the Integral University for providing the facilities to carry out this study. The authors declare no conflict of interest.

7. References

[1] Li, D. M., and Sun, H. (1997) Cancer Res. 57, 2124–2129.
[2] Steck, P. A., Pershouse, M. A., Jasser, S. A., Yung, W. K., Lin, H., Ligon, A. H., Langford, L. A., Baumgard, M. L., Hattier, T., Davis, T., Frye, C., Hu, R., Swedlund, B., Teng, D. H., and Tavtigian, S. V. (1997) Nat. Genet. 15, 356–362.
[3] Li, J., Yen, C., Liaw, D., Podsypanina, K., Bose, S., Wang, S. I., Puc, J., Miliaresis, C., Rodgers, L., McCombie, R., Bigner, S. H., Giovanella, B. C., Ittmann, M., Tycko, B., Hibshoosh, H., Wigler, M. H., and Parsons, R. (1997) Science 275, 1943–1947.
[4] Chia, J. Y., Gajewski, J. E., Xiao, Y., Zhu, H. J., and Cheng, H. C. (2010) Biochem. Biophys. Acta. 1804, 1785–1795.
[5] Yamada, K. M., and Araki, M. (2001) J. Cell Sci. 114, 2375–2382.
[6] Keniry, M., and Parsons, R. (2008) Oncogene 27, 5477–5485.
[7] Carnero, A., Blanco-Aparicio, C., Renner, O., Link, W., and Leal, J. F. (2008) Curr. Cancer Drug Targets 8, 187–198.
[8] Dong, J. T. (2006) J. Cell Biochem. 97, 433–447.
[9] Zaghloul, N. A., and Katsanis, N. (2010) Trends Genet. 26, 168–176.
[10] Burke, D. F., Worth, C. L., Priego, E. M., Cheng, T., Smink, L. J., Todd, J. A., and Blundell, T. L. (2007) BMC Bioinformatics 8, 301–315.
[11] Chasman, D., and Adams, R. M. (2001) J. Mol. Biol. 307, 683–706.
[12] Sunyaev, S., Ramensky, V., Koch, I., Lathe, W., Kondrashov, A. S., and Bork, P. (2001) Hum. Mol. Genet. 10, 591–597.
[13] Wang, Z., and Moult, J. (2001) Hum. Mutat. 17, 263–270.
[14] Grantham, R. (1974) Science 185, 862–864.
[15] Rajith, B., and George, P. D. C. (2011) PLos One 6(9), e24607.
[16] Sherry, S. T., Ward, M. H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M., and Sirotkin, K. (2001) Nucleic Acids Res. 29, 308–311.
[17] Kumar, P., Henikoff, S., and Ng, P. C. (2009) Nat. Protoc. 4, 1073–1081.
[18] Ng, P. C., and Henikoff, S. (2001) Genome Res. 11, 863–874.
[19] Ramensky, V., Bork, P., and Sunyaev, S. (2002) Nucleic Acids Res. 30, 3894–3900.
[20] Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., Kondrashov, A. S., and Sunyaev, S. R. (2010) Nat. Methods 7, 248–249.
[21] Bava, K. A., Gromiha, M. M., Uedaira, H., Kitajima, K., and Sarai, A. (2004) Nucleic Acids Res. 32, D120–D121.
[22] Mi, H., Lazareva-Ulitsky, B., Loo, R., Kejariwal, A., Vandergriff, J., Rabkin, S., Guo, N., Muruganujan, A., Doremieux, O., Campbell, M. J., Kitano, H., and Thomas, P. D. (2005) Nucleic Acids Res. 33, D284–D288.
[23] Lee, J. O., Yang, H., Georgescu, M. M., Di Cristofano, A., Maehama, T., Shi, Y., Dixon, J. E., Pandolfi, P., and Pavletich, N. P. (1999) Cell 99(3), 323–334.
[24] Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) J. Appl. Cryst. 26, 283–291.
[25] Laskowski, R. A., Rullmannn, J. A., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996) J. Biomol. NMR 8, 477–486.
[26] Guex, N., and Peitsch, M. C. (1997) Electrophoresis 18, 2714–2723.
[27] Jorgensen, W. L. (1983) J. Chem. Phys. 79, 926.
[28] Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., and Pedersen, L. G. (1995) J. Chem. Phys. 103, 8577–8593.
[29] Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., and Haak, J. R. (1984) J. Chem. Phys. 81, 3684–3690.
[30] Baker, E. N., and Hubbard, R. E. (1984) Prog. Biophys. Mol. Biol. 44, 97–179

[31] Yuan, H. Y., Chiou, J. J., Tseng, W. H., Liu, C. H., Liu, C. K., Lin, Y. J., Wang, H. H., Yao, A., Chen, Y. T., and Hsu, C. N. (2006) Nucleic Acids Res. 34, 635–641.
[32] Grillo, G., Turi, A., Licciulli, F., Mignone, F., Liuni, S., Banfi, S., Gennarino, V. A., Horner, D. S., Pavesi, G., Picardi, E., and Pesole, G. (2010) Nucleic Acids Res. 38, D75–D80.
[33] Eisenberg, D., and McClachlan, A. (1986) Nature 319, 199–203.
[34] Frédéric, M. Y., Lalande, M., Boileau, C., Hamroun, D., Claustres, M., Béroud, C., and Collod-Béroud, G. (2009) Hum. Mutat. 30, 952–959.
[35] Chen, J., and Shen, B. (2009) Curr. Proteomics 6, 228–234.
[36] Ramesh, A. S., Khan, I., Farhan, M., and Thiagarajan, P. (2013) Cell Biochem. Biophys. 67(3), 1391–1396.
[37] Ramesh, A. S., Sethumadhavan, R., and Thiagarajan, P. (2013) Protein J. 32, 657–665.
[38] Daggette, V., and Fersht, A. R. (2003) Trends Biochem. Sci. 28, 18–25.
[39] Prevost, M., Wodak, S. J., Tidor, B., and Karplus, M. (1991) Proc. Natl. Acad. Sci. U S A 88, 10880–10884.
[40] Capriotti, E., Fariselli, P., and Casadio, R. (2004) Bioinformatics 20, I63–I68.
[41] Capriotti, E., Fariselli, P., Calabrese, R., and Casadio, R. (2006) Proteins 62(4), 1125–1132.
[42] Vihinen, M. (1987) Protein Eng. 1, 477–480.
[43] Han, J. H., Kerrison, N., Chothia, C., and Teichmann, S. A. (2006) Structure 14, 935–945.
[44] Ponnuswamy, P. K., and Gromiha, M. M. (1994) J. Theor. Biol. 166, 63–74.
[45] Ahmad, S., Gromiha, M. M., and Sarai, A. (2004) Bioinformatics 20, 477–486.
[46] Abkevich, V. I., Gutin, A. M., and Shakhnovich, E. I. (1995) J. Mol. Biol. 252, 460–471.
[47] Gromiha, M. M., and Selvaraj, S. (2004) Prog. Biophys. Mol. Biol. 86, 235–277.
[48] Voet, D., and Voet, J., (1990) Biochemistry, Wiley, New York.
[49] George, D. C. P., Chakraborty, C., Haneef, S. S., NagaSundaram, N., Chen, L., and Zhu, H. (2014) Theranostics 4(4), 366–385.
[50] Delano, W. L. (2010) The PyMOL molecular graphics system Version 1.3r1., DeLano Scientific, South San Carlos, CA.
[51] Waite, K. A., and Eng, C. (2002) Am. J. Hum. Genet. 70, 829–844.
[52] Tolkacheva, T., Boddapati, M., Tsuchida, S. A. K., Kimmelman, A. C., and Chan, A. M.-L. (2001) Cancer Res. 61, 4985–4989.
[53] Vazquez, F. (2000) Mol. Cell. Bio. 20, 5010–5018.