ORIGINAL PAPER

In silico discrimination of nsSNPs in *hTERT* gene by means of local DNA sequence context and regularity

**C. George Priya Doss · Chiranjib Chakraborty ·
B. Rajith · N. Nagasundaram**

Abstract Understanding and predicting the significance of novel genetic variants revealed by DNA sequencing is a major challenge to integrate and interpret in medical genetics with medical practice. Recent studies have afforded significant advances in characterization and predicting the association of single nucleotide polymorphisms in human *TERT* with various disorders, but the results remain inconclusive. In this context, a comparative study between disease causing and novel mutations in *hTERT* gene was performed computationally. Out of 59 missense mutations, five variants were predicted to be less stable with the most deleterious effect on *hTERT* gene by in silico tools, in which two mutations (L584W and M970T) were not previously reported to be involved in any of the human disorders. To get insight into the structural and functional impact due to the mutation, docking study and interaction analysis was performed followed by 6 ns molecular dynamics simulation. These results may provide new perspectives for the targeted drug discovery in the coming future.

Keywords Docking · hTERT · Molecular dynamics
simulation · SNPs

C. G. P. Doss (✉) · B. Rajith · N. Nagasundaram
Medical Biotechnology Division, School of Biosciences
and Technology, VIT University,
Vellore 632014 Tamil Nadu, India
e-mail: georgecp77@yahoo.co.in

C. G. P. Doss
e-mail: georgepriyadoss@vit.ac.in

C. Chakraborty
Department of Bio-informatics, School of Computer and
Information Sciences, Galgotias University, Greater Noida, India

Introduction

DNA sequencing technology is becoming the method of choice for medical genetic diagnostics. However, the important challenge in the DNA sequencing technology involves the difficulty of interpreting novel sequence variants. Most geneticists use a combination of traditional genetic methods relying on segregation with the disease in families, frequency in controls, biochemical characterization, and evolutionary conservation at the variant position [1]. It is often a time consuming and laborious task to study the molecular basis of diseases like cancer by these methods. Associations with polymorphisms in candidate genes have been confirmed in many diseases, and genome-wide association studies (GWAS) are identifying many novel associations in genes that had not been strong a priori candidates for the disease under test [2]. However, the modest increase in risk implies that large well-designed and analyzed studies exist that incorporate robust computational methods to classify novel variants accurately. The massive capacity of computational application can be harnessed for effective screening and validation of genetic variants, which could be a valuable resource for the pharmacogenomics approach.

Single nucleotide polymorphisms (SNPs) are the most common type of genetic variation in the human genome [3, 4]. However, not all the SNPs can correlate with human diseases. nsSNPs that occur in a coding region can cause an amino acid substitution, thereby impart structural and functional changes on protein [5] are termed as "deleterious" and, those nsSNPs which do not have any impact on protein functions are termed as "tolerated". Hence, it is necessary to differentiate deleterious from tolerated nsSNPs. This will definitely assist in better understanding the genetic basis of human diseases, and also help in identifying the molecular and potential therapeutic targets.

Over expression experiments in human cells have shown that TERT exhibits activities in cellular transformation, proliferation, cell survival and chromatin regulation [6, 7]. Telomerase is a specialized ribonucleoprotein complex that plays a crucial role in maintaining the integrity of telomeric DNA [8]. Telomerase consists of a protein component with reverse transcriptase activity (TERT), and an RNA component (TERC) which provides the template for the telomere repeat [9]. *hTERT* gene is located within a locus at chromosome 5p13.33. It encodes a mature protein of 1132 amino acids arranged within four domains namely, N-terminal, C-terminal, RNA-binding and reverse transcriptase (RT) domain. Telomerase associated proteins such as dyskerin, nucleolar protein-10 (NOP10), non-histone protein-2 (NHP2) and glycine arginine rich-1 (GAR1) are required for the assembly of a functional telomerase holoenzyme complex [10]. Telomerase is active in some epithelial, haemopoietic and germ line cells. The mutations in protein components of TERT are linked to certain inherited human disorders of the haemopoietic system, such as dyskeratosis congenita (DC) aplastic anemia (AA) and idiopathic pulmonary fibrosis (IPF) [11–20]. Deleterious nsSNPs in hTERT and its impact on protein structure and function have not been predicted so far using in silico approach, although they have received great attention from experimental biologists. In view of this, we carried out fine-mapping, followed by functional analysis of associated SNPs identified within the coding region of *hTERT* gene using SIFT, PolyPhen and I-Mutant 2.0 [21–23]. However, the lack of a structural framework posed serious challenges in rationalizing results from polymorphic studies to characterize the impact on protein function. While, the pursuit of a high resolution experimental structure is underway, we decided to generate a three dimensional (3D) model based on homology modeling using two *TERT* domain structures: 3KYL and 2R4G using SWISS MODEL work space [24]. Consequently, in order to understand the molecular mechanism underlying the impact of mutation, docking analysis and binding analysis were undertaken. Curcumin, a well known inhibitor for *hTERT* inhibition was used to determine the binding affinity toward *hTERT* [25]. An atomic-level look at the protein dynamics through molecular dynamics simulations helped in better understanding the effects of these mutations on the protein structure, which allows for investigating how an amino acid variation can create a ripple effect throughout the protein structure and ultimately affect function. This finding is likely to have major consequences in understanding of telomerase biology and the molecular details of telomerase activities due to polymorphisms.

Materials and methods

Retrieval of SNPs

SNP related information of *hTERT* gene was retrieved from the NCBI dbSNP [26], SwissProt [27] and Telomerase database [28].

Defining the functional context of missense mutation

The pathogenic effects of missense mutations were analyzed using SIFT, PolyPhen and I-Mutant 2.0. The default parameters of all programs were applied, and only the protein sequence and missense variant were given as input information for each program.

SIFT

SIFT is a sequence homology-based tool that predicts the variants as "neutral" or "deleterious" using normalized probabilities calculated from the input multiple sequence alignment. It uses relevant multiple sequence alignments (MSAs) from pre-computed BLAST searches from the NCBI. Variants at a position with normalized probability scores less than 0.05 or 0 to 0.05 are predicted as deleterious and scores greater than 0.05 are predicted to be neutral.

PolyPhen

PolyPhen predicts the possible impact of amino acid substitutions on protein structure and function using straight forward physical and evolutionary comparative considerations. This prediction is based on straightforward empirical rules that are applied to the sequence, phylogenetic and structural information characterizing the substitution. The input of PolyPhen is an amino acid sequence (FASTA) or corresponding IDs with the position of the amino acid variant. PolyPhen searches for the 3D protein structures, multiple alignments of homologous sequences and amino acid contact information in several protein structure databases. Subsequently, it calculates PSIC scores for each of two variants, and computes the difference of the PSIC scores of these variants. The higher a PSIC score difference the higher the functional impact a particular amino acid substitution is likely to have. A PSIC score difference of 1.5 and above is considered to be damaging and less than 1.5 considered as neutral.

I-Mutant 2.0

I-Mutant 2.0 is a support vector machine (SVM)-based tool for the automatic prediction of protein stability changes upon single point mutations. The predictions are performed

starting either from the protein structure or, more importantly, from the protein sequence. I-Mutant 2.0 can be used both as a classifier for predicting the sign of the protein stability change upon mutation and as a regression estimator for predicting the related DeltaDeltaG values. Input for I-Mutant 2.0 is either a protein structure or a sequence. We used the sequence-based version of I-Mutant2.0 which classifies the prediction in two classes: (i) DDG<0: decrease stability (ii) DDG>0: increase stability. The output file shows the predicted free energy change (DDG) which is calculated from the unfolding Gibbs free energy change of the mutated protein minus the unfolding free energy value of the native protein (kcal mol^{-1}) [23]. DDG.0 means that the mutated protein has high stability and vice versa.

Modeling the effect of deleterious nsSNPs

Homology modeling and structural validation of human TERT was carried out on the basis of two TERT domain structures: 3KYL and 2R4G using SWISS MODEL workspace and RAMPAGE [24, 29]. Mutation analysis was performed based on the results obtained from various in silico tools as mentioned above. SwissPDB viewer was used to perform mutations at their respective coordinates [30] and hydrogen atoms were added to the structures using MolProbity [31]. MolProbity also adds all atom contacts into the structures and flips asparagine and glutamine side chains when necessary. By visualizing the position of the mutated amino acid residues, it is possible to suggest a physiochemical rationale for the effect on protein activity. The constructed models were subjected to energy minimization by steepest descent, using GROMOS96 53a6 force field. The ligand structure was downloaded from PubChem [32] in SMILE string format and converted to protein data bank (PDB) format using CORINA [33].

Docking and interaction analysis of hTERT

We used PatchDock for docking native and mutants of *hTERT* with the drug curcumin. Patchdock performs docking based on molecular shape representation, surface patch matching plus filtering and scoring [34]. 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 ligand molecule were given as input for performing the docking experiments with default root-mean-square deviation (RMSD) value (4.00 Å). It generated several complex structures based on docking scores. The complex structure file, with the best docking score was

selected for further analysis. For a better dynamic stability of the ligand-receptor complex, electrostatic energy, van der Waals interaction and hydrogen bond which mainly contribute the total interaction energy play a major role [35]. The total interaction energy of the hTERT-curcumin complex was calculated by PEARLS web server [36]. The negative value of total interaction energy enables better interaction and vice-versa.

Molecular dynamics simulation

Molecular dynamics simulations were performed using the GROMACS 4.5.5 software package [37] with the GROMOS96 53a6 force field. The systems were solvated using the 0.9 nm simple point charge (SPC) water embedded in the simulation boxes, and sufficient potassium and chloride ions were added to neutralize the charge of the systems. The system was energy-minimized using the steepest descent algorithm for 5000 steps with no constraints. The energy minimized system was equilibrated using the position restrained simulation under an NVT ensemble (constant number of particles, volume and temperature) for 1000 ps to stabilize the temperature at 300 K with Berendsen thermostat followed by an NPT ensemble (constant number of particles, pressure and temperature) for 1000 ps to stabilize the pressure at 1.0 bar with Parrinello-Rahman pressure coupling factor. Finally, unrestrained MD simulation was performed for 6 ns with Berendsen thermostat of 300 K and the pressure at 1.0 bar with Parrinello-Rahman pressure coupling factor. The trjconv, g_rms, g_sasa, g_rmsf and g_hbond [38] utilities of GROMACS 4.5.5 were used to analyze the MD results. In order to generate the three-dimensional backbone RMSD, RMSF of carbon-alpha, number of hydrogen bonds, SASA analysis and motion projection of the protein in phase space of the system were plotted for all the simulations using Graphing, Advanced Computation and Exploration (GRACE) program.

Results

Analysis of deleterious nsSNPs using SIFT

SIFT predicts whether an amino acid substitution affects the protein function based on sequence homology and the physical properties of amino acid. SIFT program focuses more on sequence conservation over evolutionary time and the nature of amino acids in predicting the effect of residue substitutions on function. About 14 % of nsSNPs were predicted as highly deleterious, exhibited a SIFT score of 0.00, 15 % of nsSNPs exhibited a score ranging from 0.01 to 0.05 were predicted as deleterious, and the remaining 71 % of nsSNPs were predicted as benign (Supplementary

Table 1). Thus, 29 % of nsSNPs were predicted to be intolerant, that could bring about a change in protein function.

Analysis of deleterious nsSNPs using PolyPhen

PolyPhen evaluates the location of the amino acid replacement within identified functional domains and 3D structures. All protein sequences submitted to SIFT were also submitted to PolyPhen. Unlike SIFT, it does not solely depend on sequence homology alone to make SNP functional prediction, but also on structural information. By PolyPhen 25 % of the nsSNPs were predicted to be "probably damaging", 27 % of the nsSNPs to be "possibly damaging", and the remaining 48 % were characterized as benign. Most of the mutations predicted to be deleterious were also predicted to be damaging by PolyPhen (Supplementary Table 1).

Identification of functional nsSNPs using I-Mutant 2.0

All the nsSNPs submitted to SIFT and PolyPhen were also submitted as input to the I-Mutant 2.0. Based on the difference in Gibbs free energy value of mutated and wild type proteins, 83 % of nsSNPs are found to destabilize the protein (DDG < 0 Kcal mol^{-1}) (Supplemental Table 1).

Homology modeling

To date, no experimentally determined 3D structure is available for hTERT. Hence, hTERT protein 3D structure analysis using homology modeling further provides structural

Table 1 Summary of nsSNP predicted to be deleterious by SIFT, PolyPhen and I-Mutant 2.0

SNP	Variant	Disease type	Reference
VAR_062535	L55Q	IPF	[13]
rs140261940	T564S	-NA-	-NA-
VAR_062536	K570N	AA	[16]
rs143457728	L584W	-NA-	-NA-
VAR_062783	R631Q	AA	[17]
rs121918662	V694M	AA	[15, 17]
VAR_062538	P721R	DKC1	[18]
VAR_062540	R811C	DKCB4	[19]
rs121918666	R865H	IPF	[20]
VAR_062541	R901W	DKCB4	[19]
rs121918665	K902N	AA/DKCA2	[13]
rs149439946	M970T	-NA-	-NA-

NA not available, *IPF* Idiopathic pulmonary fibrosis, *DKCA2* Dyskeratosis congenita autosomal domi nant type 2 *DKCB4* Dyskeratosis congenita autosomal recessive type 4; *AA* Aplastic anemia

representation of nsSNPs. BLAST search identifies two suitable templates (3KYL and 2R4G) for modeling of TERT. These matches span the second half of the protein (starting from residue 526 to 1121). The model with the tightest geometrical restraints on the Cα positions was obtained by the SWISS MODEL workspace. RAMPAGE software was used to validate the modeled structure. Ramachandran plot for the model shows 98.6 % of the residues in either the core region or allowed region and remaining 1.4 % of the residues in the generously allowed region with no residue in the disallowed region as plotted in Fig. 1.

Mutation structural analysis

Based on the results obtained from the above analysis (Table 1), three experimentally validated nsSNPs [K570N (AA), P721R (DKC1), R865H (IPF)], and two non-experimentally validated mutations [M970T and L584W] were chosen for the structural analysis (Fig. 2). Based on the position of amino acids in the corresponding chains of the crystallized structures, the mutation analysis were performed using SWISSPDB viewer and energy minimization were carried out using the program package GROMACS 4.5.5.

Analysis of the local environment changes

By visualizing the position and proximity of the substituted amino acid to surrounding residues, it is possible to suggest a physiochemical rationale for the effect, or the lack of any effect, of the substitution on protein activity. Hence, the interacting residual changes and polar contact points were visualized within the range of 4A° using PyMOL. The change in the local environment of the mutant residue is due to the amino acid properties such as polarity, hydrophobicity etc., thereby affecting the function of the protein through destabilization of the biological unit (Fig. 3). Superimpositions between native and mutant residues are displayed in Fig. 4.

Aplastic anemia (K570N)

AA was first identified to be associated with mutations in pseudoknot region of *hTERT*. According to the previous report, K570N mutation results in complete loss of the ability of telomerase to add hexameric repeats to telomeres, abolishing telomerase enzymatic function which in turn causes AA [14]. In K570N, lysine is smaller than the wild type residue asparagines, which may cause empty space in the core of the protein. This mutation may also lead to the change in polarity of *hTERT* from positively charged lysine
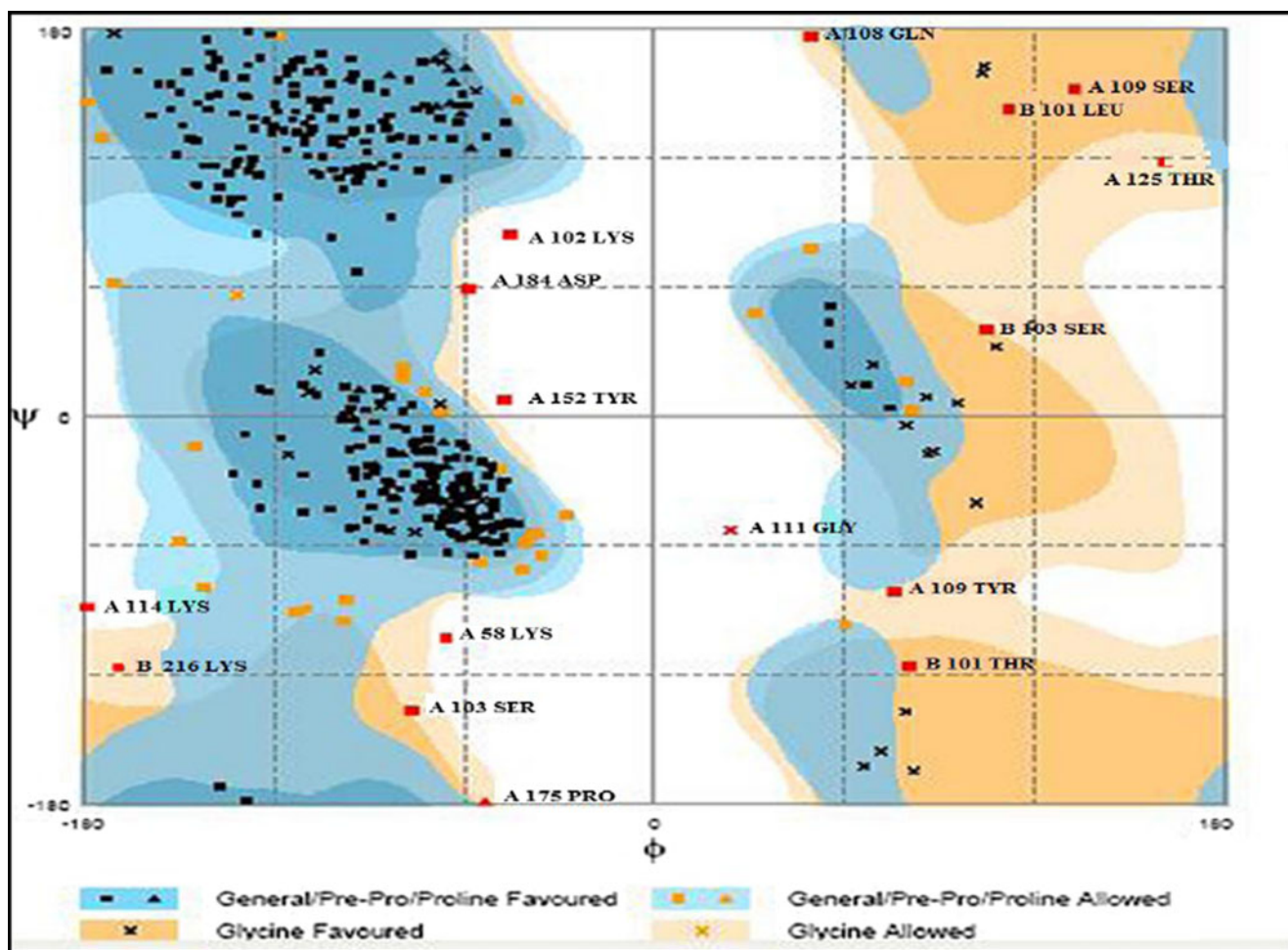
Fig. 1 Ramachandran plot of the constructed *hTERT* model. Ramachandran plot created by RAMPAGE for hTERT shows 98.6 % of residues in allowed region and 1.4 % of the residues in generously allowed region with no residues in disallowed region
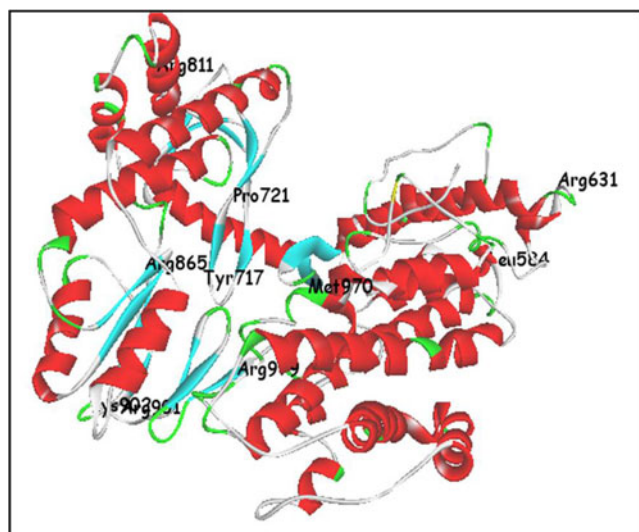

Fig. 2 Plotting of deleterious nsSNPs in modeled structure of hTERT. 3D structure of modeled hTERT with amino acid variant predicted to be deleterious by SIFT, PolyPhen and I-Mutant

to neutral asparagine. The change in the interacting residues and polar contacts due to mutation is given in Fig. 3a.

Dyskeratosis congenita (P721R)

DC is an inherited disorder characterized by premature aging and also causes increase risk of cancer. In P721R mutation, the mutant Argenine (positively charged) was bigger than the wild-type residue proline (neutral). The mutated residue is located in a domain that is necessary for the main activity of the protein. In turn, the mutation will cause loss of hydrophobic interactions in the core of the protein. P721R substitution has a significant impact on telomerase structure and function because of its non-conservative nature and general importance of proline residue in protein folding [16]. The change in the interacting residues and polar contacts due to mutation is given in Fig. 3b.

Fig. 3 Changes in local environment of TERT protein. Changes in local environment of TERT protein brought about by various mutations in native (*yellow*) and mutant (*red*) **a** K570M, **b** L584W, **c** P721R, **d** R865H, and **e** M970T
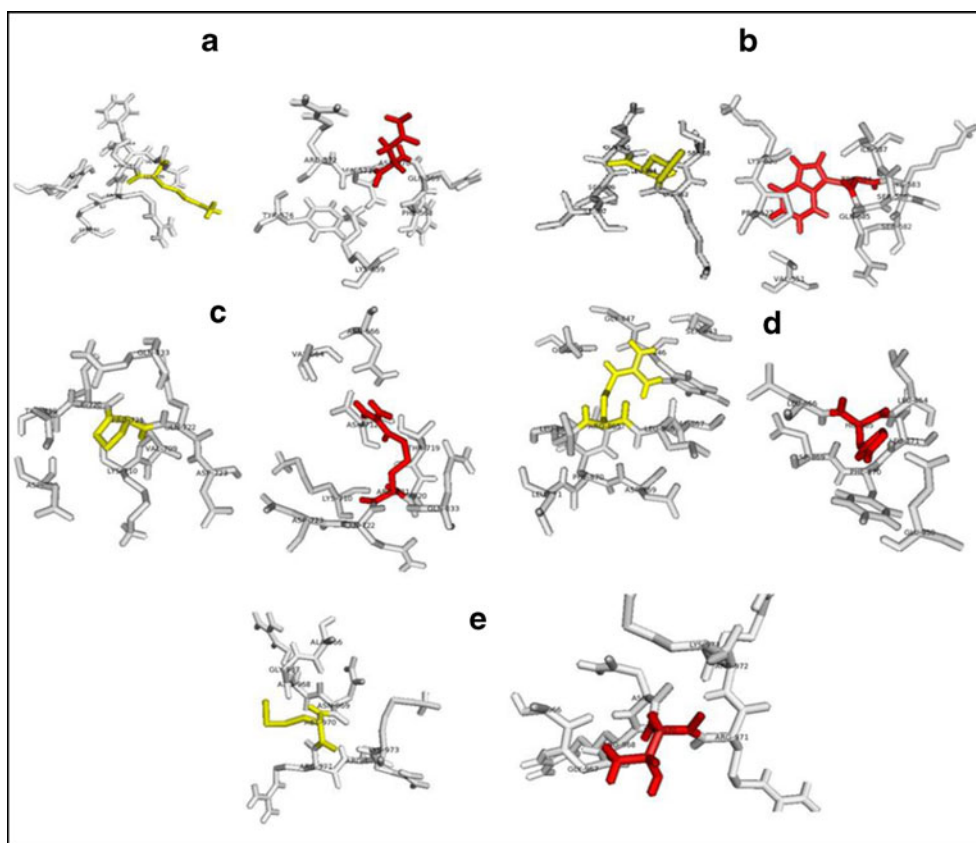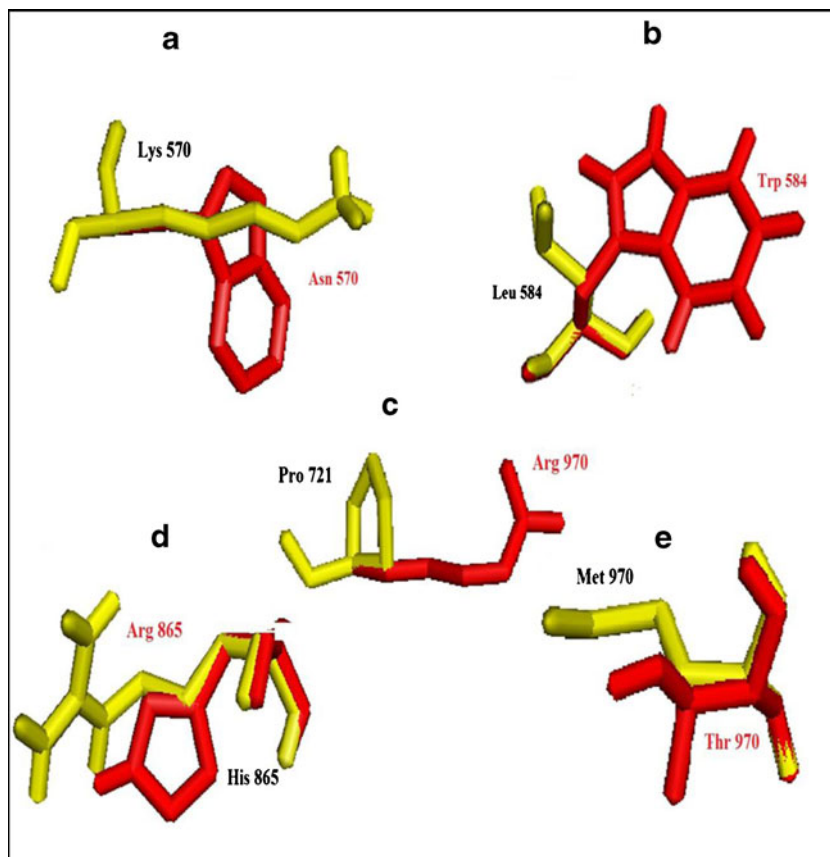

Fig. 4 Superimposition of native and mutant residues. Superimposition of the native (*yellow*) and mutant residue (*red*) in model structures. **a** K570N, **b** L584W, **c** P721R, **d** R865H, and **e** M970T

Idiopathic pulmonary fibrosis (R865H)

IPF is a form of idiopathic interstitial pneumonia characterized by progressive and chronic formation of fibrotic scar tissue in the lungs without any known causative agent [18]. The wild-type residue histidine forms a hydrogen bond with the glutamic acid on position 325. The size difference between wild-type and mutant residue may alter the hydrogen bond as the original wild-type residue did. The difference in charge will disturb the ionic interaction made by the wild-type residue. This can cause loss of interactions with other molecules and in turn leads to possible loss of external interactions. R865H mutation disrupts the nucleotide positioning in the active site and therefore, directly compromises the catalytic reaction. The change in the interacting residues and polar contacts due to mutation is given in Fig. 3c.

Variant L584W and M970T

L584W and M970T mutations have not been previously identified, and their effect in structural level is not known. To address this, we generated and compared the structure of native and mutant models. Leucine, which is smaller than tryptophan is buried in the core of the protein. Hence, the mutant residue probably will not fit in the protein structure and may in turn destabilize the protein. Whereas in variant M970T, mutant threonine is smaller that the native methionine which causes an empty space in the core of the protein. This mutation will lead to loss of hydrophobicity. All these nsSNPs might lead to decrease in the stability of protein and therefore a proper validation is needed to know the conformational as well as functional implications. The change in the interacting residues and polar contacts due to mutation is given in Fig. 3d and e.

Docking and interaction analysis

In order to understand the binding affinity of hTERT with its drug molecule, docking analysis was performed using PatchDock. According to the literature review Val867, Ala868 and Asp869 were identified as the active residues [39]. A comparative analysis between disease causing mutations (K570N, R721P, H865D and L902N) and, the novel mutations (L584W and T970R) were carried out in this analysis. Docking score and atomic contact analysis (ACE) were calculated for both the native and mutant models. The docking score of native-curcumin complex was found to be 7282, whereas for the mutant model docking score ranges from 6992 to 7126. Similarly, ACE for native mutation was found to be −304.28, whereas, for the mutant it was found to be in the range of −164.17 to −298.22 (Table 2). Furthermore, the total interaction energy between hTERT (native and mutant) and curcumin complex, which is mainly

contributed by electrostatic energy and van der Waals interaction energy were calculated for validating the binding efficiency. In the native-curcumin complex, the total interaction energy was found to be −8.81 kcal mol^{-1}, whereas, for mutant models, the range is from −4.01 to −6.89 kcal mol^{-1}. From this analysis, it was clear that all mutant hTERT models exhibited less binding affinity with curcumin than the native model indicating the effect of mutation on the drug binding efficiency. In the above analysis, the two novel mutations (L584W and T970R) showed the least ligand protein interaction (−4.01, −5.00 kcal mol^{-1}) and docking score (6992, 7014) compared to the other mutants, which confirmed the impact of these mutations on the binding affinity of hTERT. This finding inspired us to investigate the alteration in structural-dynamics properties due to mutation in hTERT.

Molecular dynamics simulation

A comparative MD analysis of disease associated mutants K570N, P721R, R865H and the two novel mutations L584W, K970N with the native was carried out. In the 6 ns simulation trajectory, different parameters such as root mean square deviation (RMSD), root mean square fluctuation (RMSF), hydrogen bond formation, and solvent accessibility surface area (SASA) were applied to analyze the level of structural changes. The backbone RMSD was calculated from the trajectory value of native and mutant models. Mutant L584W showed maximum deviation resulting in backbone RMSD of ~0.8 nm at 2600 ps, while, the native type structure obtained maximum deviation of 0.46 nm at 4200 ps (Fig. 5a). Mutant model K570N and K970N showed almost consistent deviation throughout the simulation period with a maximum deviation of ~0.62 nm attained at ~2500 ps (Fig. 5a, c). R865H showed less deviation till 2800 ps from their starting structure when compared to the native structure. Between 2800 ps till the end of simulation R865H showed higher deviation when compared to the native structure (Fig. 5c). Whereas, P721R showed frequent decrement in backbone RMSD till 4600 ps when compared to the native structure, after which a marginal increase in backbone RMSD of about 0.1 nm was observed till the end of the simulation (Fig. 5c). The RMSD of L584W was higher throughout the process, and a maximum deviation of 0.4 nm was observed at around 2.6 ns between the native and mutant model (Fig. 5a). The RMSF profile was determined for the mutants, and it was observed that L584W mutation affects neighboring residues at the maximum of around 0.7 nm fluctuation indicating a gain of flexibility due to mutation (Fig. 5b). Further, the RMSF values of K570N, P721R, R865H and K970N observed a similar fluctuation throughout the process indicating an increase of flexibility as compared to the native protein

Table 2 Comparison of docking score, atomic contact energy (ACE) and ligand-receptor

hTERT	Score	ACE	Ligand receptor electrostatic energy (kcal mol⁻¹)	Ligand receptor van der Waals energy (kcal mol⁻¹)	Ligand receptor total interaction energy (kcal mol⁻¹)
Native	7282	−304.28	−0.81	−7.80	−8.81
K570N	7114	−182.54	−0.62	−5.06	−5.90
L584W	6992	−164.17	−0.54	−3.41	−4.01
R721P	7126	−298.22.	−0.77	−5.89	−6.89
H865N	7034	−201.32	−0.64	−5.49	−6.37
T970N	7014	−169.24	−0.57	−4.12	−5.00

Interaction energies of native and mutant *hTERT* with curcumin

(Fig. 5b, d). Native model of hTERT protein obtained an average number of ~410 hydrogen bonds throughout the simulation period (Fig. 6a). The mutant models K570N, L584W, P721R, R865H and M970T showed fewer number of hydrogen bonds, when compared to the native protein (Fig. 6a, c). Since the numbers of hydrogen bonds were less in the mutant models, the reduced number of hydrogen bonds may affect the protein stability. It is noted that
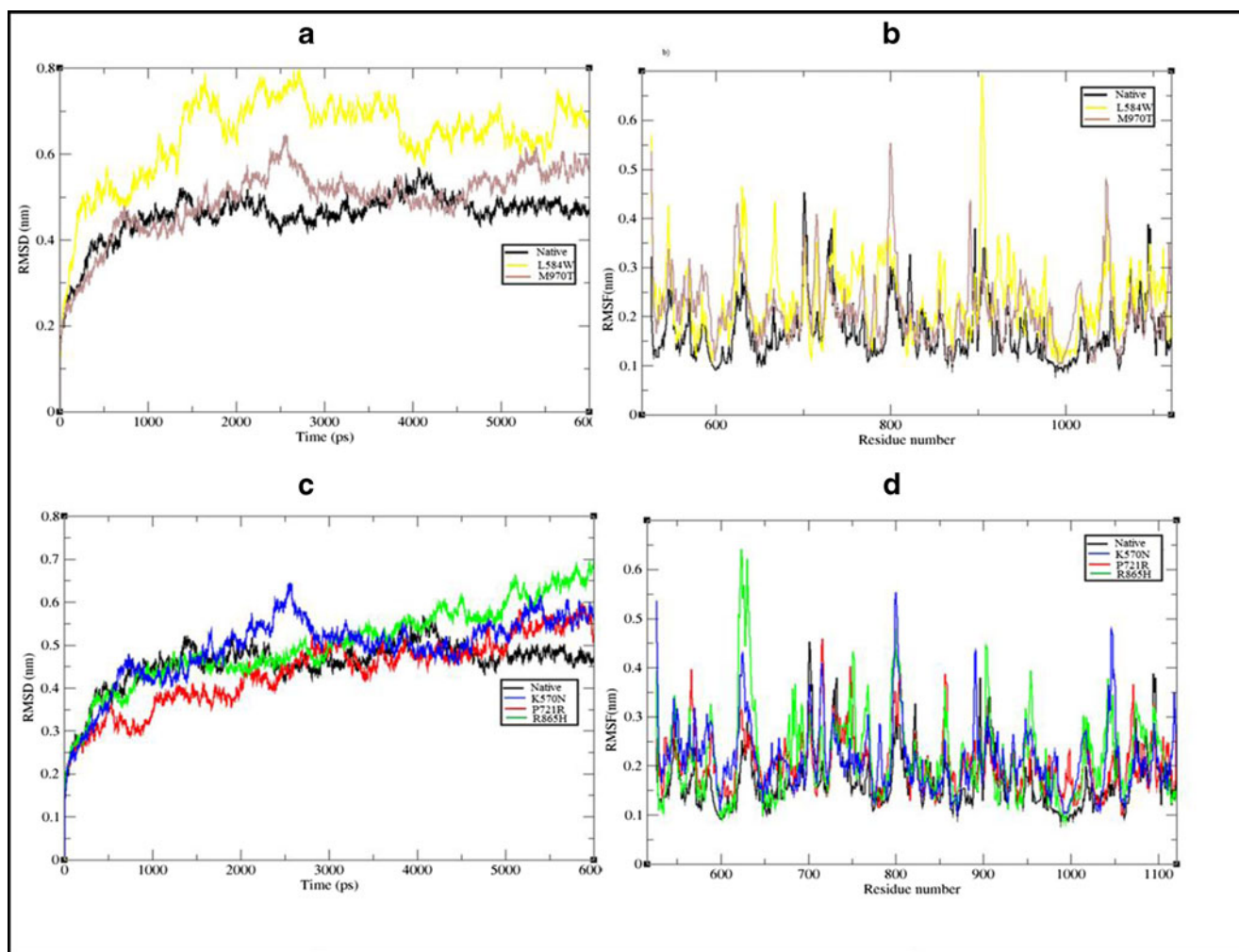

Fig. 5 RMSD and RMSF of native and mutant model of hTERT. A. RMSD of native, non experimentally verified mutant L584W and M970T B. RMSF of native, non experimentally verified mutant L584W and M970T C. RMSD of native and experimentally verified mutant L570W, P721R, R865H D. RMSF of native and experimentally verified mutant L570W, P721R, R865H. Coloring scheme - native (*black*), L570W (*blue*), P721R (*red*), R865H (*green*), L584W (*yellow*), and M970T (*purple*)
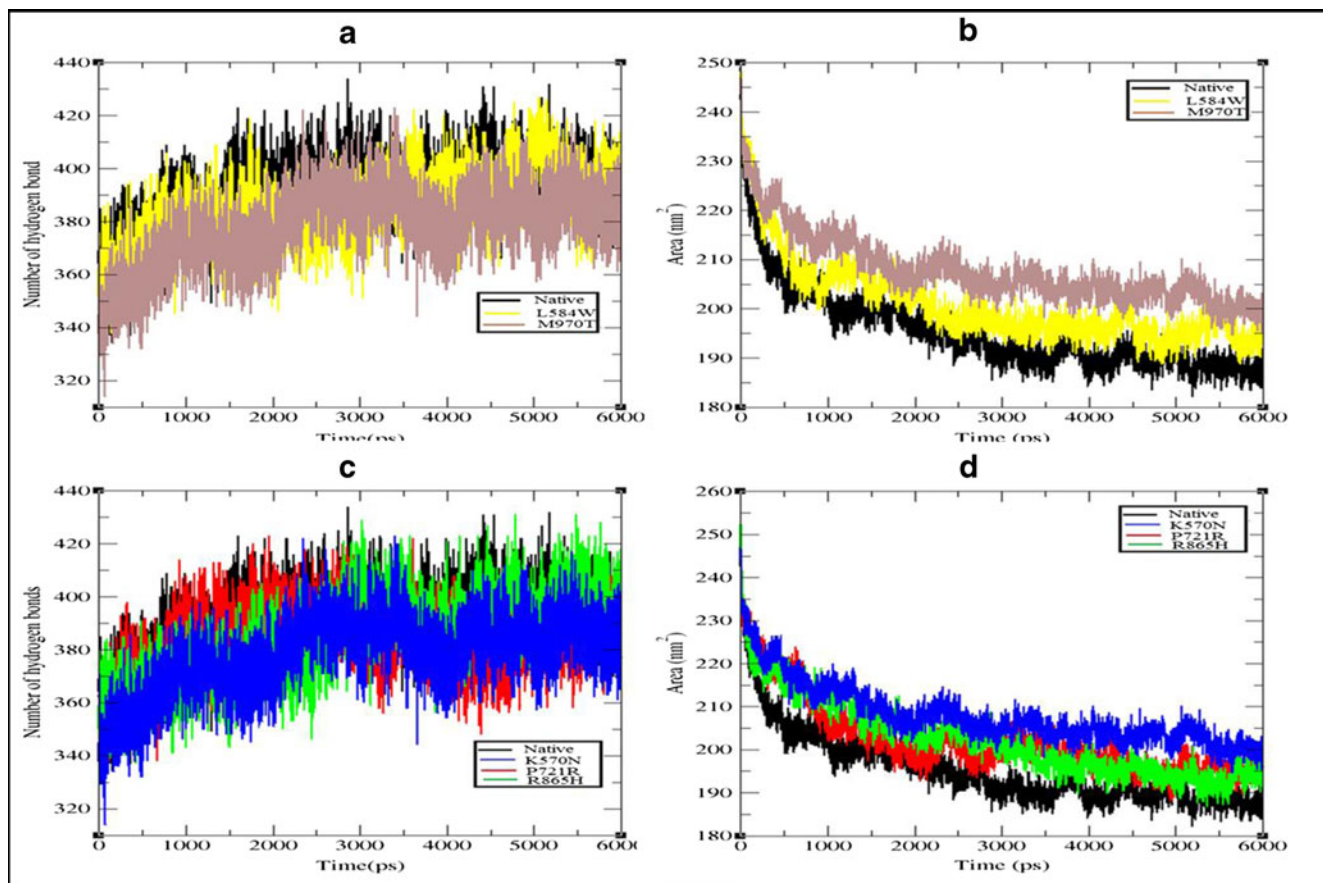
Fig. 6 SASA and H bond of native and mutant model of hTERT. A. SASA of native, non experimentally verified mutant L584W and M970T. B. H bond of native (*black*), non experimentally verified mutant L584W and M970T. C. SASA of native and experimentally verified mutant L570W, P721R, R865H. D. H-bond of native and experimentally verified mutant L570W, P721R, R865H. Coloring scheme - native (*black*), L570W (*blue*), P721R (*red*), R865H (*green*), L584W (*yellow*) and M970T (*purple*)

SASA value for both native and mutant protein stabilizes after 3000 ps (Fig. 6b, d). The average SASA value of disease mutations K570N, P721R and R865H was found to be higher in comparison with the native protein, thereby validating the experimental results indicating loss of hTERT function due to mutations (Fig. 6d). In the case of novel mutations L584W and M970T, the fluctuation in SASA indicated greater value when compared to native protein (Fig. 6b). This indicates that the mutant proteins might be undergoing a significant structural transition, when compared to native protein.

Discussion

In this comprehensive analysis, we provide functional evidence for the disease-associated point mutations of the protein component of human telomerase. We have presented a list of nsSNPs that could constitute as relevant genetic markers and also useful for disease association and linkage disequilibrium studies. The selection of nsSNPs likely to

cause the most severe effects on the function of the protein and on the phenotype could be facilitated considering several criteria. Some amino acid variations are more likely to alter 3D structure of the candidate proteins than others. The possible impact of amino acid allelic variants on protein activity is thus a function of both the structural locations of nsSNPs and phylogenetic conservation [40]. The basic criteria for these computational methods are sequence homology, physicochemical properties of the substituted residues and structural information. To study the functional consequences of nsSNPs in relation to the molecular basis of diseases at the structural level requires the integration of heterogeneous information such as protein sequence, protein structure (3D), and their associated variants. Mapping of deleterious nsSNPs to protein 3D structures and, analyzing at the structural level will reveal the full extent to which they can alter the activity of protein. Proteins with mutations do not always have 3D structures that are analyzed and submitted in Protein data bank (PDB). Therefore, it is necessary to model 3D structure of protein by locating the mutation in 3D structures. This is a simple way of detecting what kind of

adverse effects that a mutation can have on a protein. In silico approaches such as homology modeling and molecular dynamics approach will aid in elucidating the structural impact of deleterious nsSNPs at the molecular level. To determine the functional nsSNPs in *hTERT* gene, in silico tools with diverse approaches like SIFT, PolyPhen and I-Mutant 2.0 were used. From the results obtained, SIFT, PolyPhen and I-Mutant 2.0 predicted 29 %, 52 % and 83 %, nsSNPs to be deleterious and 71 %, 48 % and 17 %, to be tolerated respectively. The variation in the prediction score of SIFT and PolyPhen is mainly due to the difference in protein sequence alignment, and the scores used to classify the variants [41]. A recent analysis by Flanagan et al. 2010 confirmed the accuracy of SIFT and PolyPhen in predicting the effect of nsSNPs on protein function [42]. Our group also tried to evaluate the accuracy of SIFT, PolyPhen and I-Mutant 2.0 based predictions on *ATM, G6PD, F8* and *F9* genes [43–45]. In order to improve our efficiency and rationality for validating the deleterious nsSNPs in *hTERT*, the 3D homology structure was constructed. The modeled structure was verified using RAMPAGE server, to be of good quality and thus was used for docking analysis followed by MD approach. We first considered the functional impact of mutations in hTERT that have recently been identified in association with diseases like AA, DKC and IPF. Our findings successfully identified the following mutations K570N, P721R and R865H that lead to drastic change in protein stability and showed a good concordance with experimentally proved data. It was noteworthy that, K570N mutation effectively abolished telomerase enzymatic function, even though it is highly divergent among the telomerase of the different species [14]. Similarly, P721R and R865H mutations drastically reduced telomerase enzymatic activity, suggesting that these seemingly non-conserved residues may be involved in either the structural formation or functional property of telomerase [16, 18]. The precise function of L584W and M970T is still not elucidated, but we investigated computationally using homology modeling and molecular dynamics approach. Calculating the interaction energy is extremely crucial to understand the biological activity of most protein interacting with its partner. All mutant models exhibited low docking score and high interacting energy indicating loss of interaction of curcumin with hTERT when compared with native protein. From these studies, it can be concluded that mutation had altered the residues surrounding the binding residues thereby disturbing the normal biological process. This information might provide molecular insights into the impact of mutations on protein stability, folding and function. Furthermore, one of the novel findings in this study was the identification of two deleterious mutations L584W and M970T, for which there is no information regarding the biological role in telomerase database and literature search.

Conclusions

In conclusion, we have addressed the problem faced by the experimental biologist in identifying novel mutations. The main aim of this analysis is to suggest the impact of several important nsSNPs, both disease causing and novel that could impart structural and functional alteration in *hTERT* gene. A comparative analysis between the disease associated, and novel mutations, ascertains that L584W and M970T could play a major role in affecting the telomerase activity. To the best of our knowledge this is the first ever reported study that incorporates in silico tools in combination with docking study and interaction analysis followed by molecular dynamics approach for prioritizing of deleterious nsSNPs in *hTERT* gene. The set of in silico SNPs we have identified provides information necessary for investigating its mechanism further for polymorphism analysis in addition to the available resources assembled in telomerase database [28].

References

1. Richards CS, Bale S, Bellissimo DB (2006) ACMG recommendations for standards for interpretation and reporting of sequence variations: revisions 2007. Genet Med 10:294–300
2. Wellcome Trust Case Control Consortium (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature. 447:661–78
3. Cargill M, Altshuler D, Ireland J, Sklar P, Ardlie K, Patil N, Shaw N, Lane CR, Lim EP, Kalyanaraman N, Nemesh J, Ziaugra L, Friedland L, Rolfe A, Warrington J, Lipshutz R, Daley GQ, Lander ES (1999) Characterization of single-nucleotide polymorphisms in coding regions of human genes. Nat Genet 22:231–238
4. Nadeau JH (2002) Single nucleotide polymorphisms: tackling complexity. Nature 420:517–518
5. Ramensky V, Bork P, Sunyaev S (2002) Human non-synonymous SNPs: server and survey. Nucleic Acids Res 30:3894–3900
6. Smith LL, Coller HA, Roberts JM (2003) Telomerase modulates expression of growth-controlling genes and enhances cell proliferation. Nat Cell Biol 5:474–479
7. Lee J, Sung YH, Cheong C, Choi YS, Jeon HK (2008) TERT promotes cellular and organismal survival independently of telomerase activity. Oncogene 27:3754–3760
8. Wong JMY, Collins K (2003) Telomere maintenance and disease. Lancet 362:983–988
9. Hug N, Lingner J (2006) Telomere length homeostasis. Chromosoma 115:413–425
10. Chen JL, Greider CW (2003) Determinants in mammalian telomerase RNA that mediate enzyme processivity and cross-species incompatibility. EMBO J 22:304–314
11. Ly H (2009) Genetic and environmental factors influencing human diseases with telomere dysfunction. Int J Clin Exp Med 2:114–130

12. Chen JL, Greider CW (2004) Telomerase RNA structure and function: implications for dyskeratosis congenita. Trends Biochem Sci 29:183–192

13. Armanios MY, Chen JJ, Cogan JD, Alder JK, Ingersoll RG, Markin C, Lawson WE, Xie M, Vulto I, Phillips JA 3rd, Lansdorp PM, Greider CW, Loyd JE (2007) Telomerase mutations in families with idiopathic pulmonary fibrosis. N Engl J Med 356:1317–1326

14. Calado RT, Regal JA, Hills M, Yewdell WT, Dalmazzo LF, Zago MA, Lansdorp PM, Hogge D, Chanock SJ, Estey EH, Falcão RP, Young NS (2009) Constitutional hypomorphic telomerase mutations in patients with acute myeloid leukemia. Proc Natl Acad Sci U S A 6:1187–1192

15. Yamaguchi H, Calado RT, Ly H (2005) Mutations in TERT, the gene for telomerase reverse transcriptase, in aplastic anemia. N Engl J Med 352:1413–1424

16. Xin ZT, Beauchamp AD, Calado RT (2007) Functional characterization of natural telomerase mutations found in patients with hematologic disorders. Blood 109:524–532

17. Kirwan M, Vulliamy T, Marrone A (2009) Defining the pathogenic role of telomerase mutations in myelodysplastic syndrome and acute myeloid leukemia. Hum Mutat 30:1567–1573

18. Vulliamy TJ, Marrone A, Knight SW (2006) Mutations in dyskeratosis congenita: their impact on telomere length and the diversity of clinical presentation. Blood 107:2680–2685

19. Marrone A, Walne A, Tamary H (2007) Telomerase reverse-transcriptase homozygous mutations in autosomal recessive dyskeratosis congenita and Hoyeraal-Hreidarsson syndrome. Blood 110:4198–4205

20. Tsakiri KD, Cronkhite JT, Kuan PJ, Xing C, Raghu G, Weissler JC, Rosenblatt RL, Shay JW, Garcia CK (2007) Adult-onset pulmonary fibrosis caused by mutations in telomerase. Proc Natl Acad Sci U S A 104:7552–7557

21. Kumar P, Henikoff AS, Ng PC (2009) Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc 4:1073–1081

22. Adzhubei IA, Schmidt S, Peshkin L (2010) A method and server for predicting damaging missense mutations. Nat Methods 7:248–249

23. Capriotti E, Fariselli P, Casadio R (2005) I-Mutant 2.0: predicting stability changes upon mutation from the protein sequence or structure. Nucleic Acids Res 33:306–310

24. Bordoli L, Kiefer F, Arnold K (2009) Protein structure homology modeling using SWISS-MODEL workspace. Nat Protoc 4:1–13

25. Lee JH, Chung IK (2010) Curcumin inhibits nuclear localization of telomerase by dissociating the Hsp90 co-chaperone p23 from hTERT. Cancer Lett 290:76–86

26. Sherry ST, Ward M, Sirotkin K (1999) dbSNP-Database for Single Nucleotide Polymorphisms and other classes of minor genetic variation. Genome Res 9:677–679

27. Bairoch A, Apweiler R (1996) The SWISS-PROT protein sequence data bank and its new supplement TREMBL. Nucleic Acids Res 24:21–25

28. Podlevsky JD, Bley CJ, Omana RV, Qi X, Chen JJ (2007) The telomerase database. Nucleic Acids Res 36:339–D343

29. Lovell SC, Davis IW, Arendall WB 3rd, de Bakker PI, Word JM, Prisant MG, Richardson JS, Richardson DC (2002) Structure validation by Calpha geometry: phi, psi and Cbeta deviation. Proteins 50:437–450

30. Guex N, Peitsch MC (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis 18:2714–2723

31. Chen VB, Arendall WB III, Headd JJ (2010) MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D: Biol Crystallogr 66:12–21

32. Feldman J, Snyder KA, Ticoll A (2006) A complete small molecule dataset from the protein data bank. FEBS Lett 580:1649–1653

33. Gasteiger J, Rudolph C, Sadowski J (1990) Automatic generation of 3D-atomic coordinates for organic molecules. Tetrahedron Comput Meth 3:537–547

34. Schneidman D, Inbar Y, Nussinov R (2005) PatchDock and SymmDock: servers for rigid and symmetric docking. Nucleic Acids Res 33:363–367

35. Chothia C, Janin J (1975) Principles of protein-protein recognition. Nature 256:705–708

36. Han LY, Lin HH, Li ZR, Zheng CJ, Cao ZW, Xie B, Chen YZ (2006) PEARLS: program for energetic analysis of receptor ligand system. J Chem Inf Model 46:445–450

37. Hess B, Kutzner Z, van der Spoel D (2008) GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. J Chem Theory Comput 4:435–447

38. Baker EN, Hubbard RE (1984) Hydrogen bonding in globular proteins. Prog Biophys Mol Biol 44:97–179

39. Drosopoulos WC, Prasad VR (2007) The active site residue Valine 867 in human telomerase reverse transcriptase influences nucleotide incorporation and fidelity. Nucleic Acids Res 35:1155–1168

40. Chasman D, Adams RM (2001) Predicting the functional consequences of non-synonymous single nucleotide polymorphisms: structure-based assessment of amino acid. J Mol Biol 307:683–706

41. Karchin R, Diekhans M, Kelly L (2005) LS-SNP: large-scale annotation of coding non-synonymous SNPs based on multiple information sources. Bioinformatics 21:2814–2820

42. Flanagan SE, Patch AM, Ellard S (2010) Using SIFT and PolyPhen to predict loss-of-function and gain-of-function mutations. Genet Test Mol Biomarkers 14:533–7

43. Rajith B, Doss CGP (2011) Path to facilitate the prediction of functional amino acid substitutions in red blood cell disorders—a computational approach. PLoS One 6:e24607

44. Doss CGP, Rajith B (2012) Computational refinement of functional single nucleotide polymorphisms associated with ATM gene. PLoS One 7:e34573

45. George Priya Doss C (2012) In Silico profiling of deleterious amino acid substi-tutions of potential pathological importance in haemophlia A and haemophlia B. J Biomed Sci 19:30