International Journal of ChemTech Research
CODEN(USA): IJCRGG ISSN : 0974-4290
Vol.6, No.2, pp 1251-1267, April-June 2014

Exploring the missense mutation impact on STK11 which cause Peutz-Jeghers Syndrome by computational analysis

Merlin Lopus[1]*, R. Rajasekaran[1]

[1]Bioinformatics Division, School of Biosciences and Technology, Vellore Institute of Technology,Vellore 632014, Tamil Nadu, India.

***Corres.author: rrajasekaran@vit.ac.in;**
Phone No: +91 9443491789

Abstract: In this work, the most detrimental missense mutations of STK11 that cause Peutz-Jeghers Syndrome and other types of cancer were identified computationally and the binding efficiencies of those missense mutations with STRAD were analyzed. Out of 39 missense mutations, I-Mutant 2.0, SIFT and PolyPhen programs identified 26 variants that were less stable, deleterious and damaging respectively. Subsequently, modeling of these 26 variants was performed to understand the change in their conformations with respect to the native STK11 by computing their root mean squared deviation (RMSD). Furthermore, the native protein and the 26 mutants were docked with the STRAD to explain the binding efficiencies of those detrimental missense mutations. Among the 26 mutants, 12 mutants were identified as deleterious based on the results of docking studies, RMSD scores and stability analysis. Finally, normal mode analysis determined that the loss of binding affinity of these 12 mutants was caused by altered flexibility in the amino acids that bind to STRAD compared with the native protein. Thus, the present study showed that all the active site amino acids in those 12 mutants displayed loss of flexibility, which could be the theoretical explanation of decreased binding affinity between the mutant STK11 and STRAD.
Key words: Peutz-Jeghers Syndrome, cancer, STK11 and STRAD.

INTRODUCTION

The germline mutations in STK11 gene cause a rare dominantly inherited disease , Peutz –Jeghers syndrome (PJS) [1,2] characterized by presence of gastrointestinal (GI) hamartomatous polyps and mucocutaneous melanin pigmentation which is frequently expressed in lips and oral area [3, 4, 5]. It is found that the hamartomatous polyps are commonly develop in small bowel, colon and in stomach, causing abdominal pain, bowel obstruction and severe gastrointestinal bleeding [10]. The risk factor for cancer in PJS patients has been estimated to be 15 - fold higher than the general population [6]. An increased risk for gastrointestinal (GI) and non gastrointestinal (non-GI) cancer is observed in PJS patients. Frequently observed GI targeted cancers include stomach, small intestine, colon and pancrease cancer, whereas cancers in breast, endometrial, ovary and lung tumours are frequent in non-GI targeted cancer [10].

Mutations in the tumor suppressor gene STK11 on chromosome 19p13.3 is the only known cause for PJS [11]. The gene encodes for serine-threonine kinase which has a critical role in regulating cell growth and apoptosis [7,8]. Many of the germline mutations in STK11 are small deletions or point/missense mutations and which are present in the STK11 catalytic kinase domain and minority occur within its COOH- terminal non catalytic region and this will result in STK11 protein reduction, loss or inactivation [12,13].STK11 associates with the pseudokinase STRAD (STe20-Related ADaptor) and the scaffolding MO25 (MOuse protein 25) in a 1:1:1 heterotrimeric complex in cell [14]. The catalytic kinase domain of STK11 is activated when pseudokinase domain of STRAD binds to it which leads to transport of STK11 to the cytoplasm [15]. Contrasting to the majority of the protein kinases which are regulated by phosphorilation, STK11 is activated by binding to STRAD and MO25 through an unknown, phosphorylation-independent, molecular mechanism [16,17]. Mutations in STK11 can lead to its inactivation without affecting this complex assembly.

We investigated the detrimental missense mutations of STK11 since there were 44 reported mutations for this protein and also many of the mutations were reported in PJS and other types of cancer [10, 11]. These missense mutations fall in to the category of non-synonymous SNPs (nsSNP) which cause change in amino acid residue. The computational protocol was used to identify and analyze them and model structures were proposed for the mutants. The binding partner STRAD was docked with both the native and mutants of STK11 to determine the binding effect and the nature of the flexibility in the binding pockets, which explained the decreased binding efficiency of these missense mutations.

MATERIALS AND METHODS

Datasets

The protein sequence and variants (single amino acid polymorphisms/missense mutations/point mutations) of STK11 were obtained from the Swissprot database available at http://www.expasy.ch/sprot/. The subsection of each Swissprot entry provides information on polymorphic variants. Some of the polymorphic variants may be disease(s) - associated by causing defects in a given protein. Most of these polymorphic variants were nsSNPs (non-synonymous SNPs) in the gene sequence and SAPs (single amino acid polymorphisms) in the protein sequence [19, 20]. The 3D Cartesian coordinates of the protein STK11 was obtained from Protein Data Bank with PDB ID 2WTK [21] for *in silico* mutation modeling and docking studies based on detrimental point mutants.

Predicting stability changes caused by SAPs using support vector machine (I-Mutant 2.0)

We used the program I-Mutant2.0 (http: //gpcr. biocomp.unibo.it/cgi/predictors/IMutant2.0/ I-Mutant2.0.cgi) for this study. I-Mutant2.0 is a support vector machine (SMV) based tool for the automatic prediction of protein stability changes caused by single point mutations. I-Mutant2.0 predictions were performed starting either from the protein structure or, more importantly, from the protein sequence [22]. This program was trained and tested on a dataset derived from ProTherm [23], which is the most comprehensive available database of thermodynamic experimental data of free energy changes of protein stability caused by mutations under different conditions. The output files show the predicted free energy change value or sign ($\Delta\Delta G$), which was calculated from the unfolding Gibbs free energy value of the mutated protein minus the unfolding Gibbs free energy value of the native protein (KJ/mol). Positive $\Delta\Delta G$ values meant that the mutated protein has higher stability and negative values indicate lower stability.

Analysis of functional consequences of point mutations by a sequence homology-based method (SIFT)

The program, SIFT available at http://blocks.fhcrc.org/sift/SIFT.html [25], used specifically to detect deleterious single amino acid polymorphisms. SIFT is a sequence homology-based tool, which presumes that important amino acids will be conserved in a protein family; therefore, changes at well-conserved positions tend to be predicted as deleterious [24]. Queries are submitted in the form of protein sequences. SIFT takes a query sequence and uses multiple alignment information to predict tolerated and deleterious substitutions for every position of the query sequence. SIFT is a multistep procedure that, for given a protein sequence, (i) searches for similar sequences, (ii) chooses closely related sequences that may share similar function, (iii) obtains the multiple alignment of these chosen sequences, and (iv) calculates normalized probabilities for all possible substitutions at each position from

the alignment. Substitutions at each position with normalized probabilities less than a chosen cutoff are predicted to be deleterious and those greater than or equal to the cutoff are predicted to be tolerated [25]. The cutoff value in SIFT program was tolerance index of ≥0.05. The higher the tolerance index, the less functional impact a particular amino acid substitution would be likely to have.

Simulation for functional change in a point mutant by structure homology-based method (PolyPhen)

Analyzing the damage caused by point mutations at the structural level is considered very important to understand the functional activity of the protein. The server PolyPhen [26] available at http://coot.embl.de/ PolyPhen/ was used for this purpose. Input options for the PolyPhen server are protein sequence, SWALL database ID or accession number, together with the sequence position of two amino acid variants. The query is submitted in the form of a protein sequence with a mutational position and two amino acid variants. Sequence-based characterization of the substitution site, profile analysis of homologous sequences, and mapping of the substitution site to known protein 3D structures are the parameters taken into account by PolyPhen server to calculate the score. It calculates position-specific independent counts (PSIC) scores for each of the two variants and then computes the PSIC scores difference between them. The higher the PSIC score difference, the higher the functional impact a particular amino acid substitution would be likely to have.

Modeling point mutation on protein structures to compute the RMSD

Structure analysis was performed to evaluate the structural deviation between native proteins and mutant proteins by means of root mean square deviation (RMSD). The web resource Protein Data Bank [21] was used to identify the 3D structure of STK11 (PDB ID: 2WTK) and also confirmed the mutation position and the mutation residue in PDB ID 2WTK. In order to calculate the RMSD for native and mutant STK11 with STRAD, we used SWISSPDB viewer for performing mutation, and NOMAD-Ref server performed the energy minimization for 3D structures [27]. This server uses Gromacs as the default force field for energy minimization, based on the methods of steepest descent, conjugate gradient and Limited-memory Broyden-Fletcher-Goldfarb- Shanno (L-BFGS) methods [28]. The conjugate gradient method was used here to minimize the energy of the 3D structure of STK11. Divergence of the mutant structure from the native structure could be caused by substitutions, deletions and insertion [29] and the deviation between the two structures could alter the functional activity [30] with respect to binding efficiency of the binding partner, which was evaluated by their RMSD values. We used the server SRide [36] for identifying the stabilizing residues in native protein and in mutant model. Stabilizing residues were computed using the parameters such as surrounding hydrophobicity, long-range order, stabilization center and conservation score as described by Magyar *et al*. [36].

Identification of binding residues in STK11-STRAD interaction and computing atomic contact energy (ACE)

In order to understand the functional activity of STK11 with its binding partner STRAD we selected the PDB ID: 2WTK and performed the mutations by using SWISSPDB viewer and the energy minimization was done by NOMAD-Ref. Finally, the program PatchDock was used for the docking of the native and mutant STK11 with STRAD to compute the atomic contact energy. The underlying principle of this server is based on molecular shape representation, surface patch matching plus filtering and scoring [31]. It finds docking transformations that yield good molecular shape complementarity. Such transformations, when applied, induce both wide interface areas and small amounts of steric clashes. A wide interface ensured that include several matched local features of the docked molecules that have complementary characteristics were included. The PatchDock algorithm divides the Connolly dot surface representation [32, 33] of the molecules into concave, convex and flat patches. Then, complementary patches are matched to generate candidate transformations. Each candidate transformation is further evaluated by a scoring function that considers both geometric fit and atomic desolvation energy [33, 34].

To identify the binding residues between STK11 and STRAD, we submitted the PDB ID: 2WTK, to the protein interface recognition server SPPIDER [35]. This server integrates enhanced relative solvent accessibility (RSA) predictions with high resolution structural data. So that it is used to predict residues to be at the putative protein interface(s) by considering single protein chain with resolved 3D structure, to analyse protein- protein complex with given 3D structural information and to identify residues that are being inter chain contact [35].

Exploring the flexibility of binding pocket by normal mode analysis

A quantitative measure of the atomic motions in proteins could be obtained from the mean square fluctuations of the atoms relative to their average positions. Protein flexibility is important for protein function [37]. In addition, the flexibility of certain amino acids in a protein is useful for various types of interactions. It is found that the increase in flexibility upon mutation is tend to be localized, whereas increase in rigidity are likely to be coupled to remote structural sites. When a mutation occur the structure subtly rearranges itself to maximize the enthalpy-entropy compensation [18]. Hence the flexibility of amino acids in the binding site was computed from the mean-square displacement <R2> of the lowest-frequency normal mode using ElNémo server [38].

RESULTS AND DISCUSSION

Data set from Swissprot

The STK11 protein and a total of 39 variants, namely Y49D, V66M, L67P, R86G, R87K, G135R, F157S, L160P, G163D, Q170P, G171S, H174R, D176Y, I177N, N181E, D194N, D194V, D194Y, E199K, E199Q, A205T, D208N, G215D, S216F, E223V, T230P, F231L, S232P, W239C, L245R, T250P, Y272H, D277Y, P281L, L285Q, R297K, P314H, P315S and P324L were taken from Swissprot [39, 40].

Identification of functional missense mutants of STK11 by I-Mutant2.0

Of the 39 variants, 37 were observed as less stable from the I-Mutant 2.0 server as shown in the Table 1. Out of such 37 variants, 19 variants, viz., R86G, R87K, G135R, Q170P, H174R, D176Y, N181E, D194N, E199Q, A205T, D208N, S216F, E223V, T230P, S232P , T250P , D277Y, P281L and P324L had shown a $\Delta\Delta G$ value of < -1.0, the other 13 variants, viz., V66M, F157S, G163D, G171S, E199K, G215D, F231L, W239C, L245R, Y272H, R297K, P314H and P315S shown a $\Delta\Delta G$ value of > -1.0 ,and the remaining 5 variants, viz, Y49D , L67P, L160P, I177N ,and L285Q shown a $\Delta\Delta G$ value of > -2.0 were illustrated in Table 1.

Of the 37 variants that showed a negative $\Delta\Delta G$, variants Y49D and N181E changed its polar amino acid to negatively charged amino acid, variant R86G changed from positively charged to non-polar, three variants (G135R, L245R and P314H) changed from non-polar to positively charged amino acid. Six variants (F157S, G171S, I177N, A205T, L285Q, and P315S) changed from non-polar to polar, variant E199K changed from negatively charged to positively charged, another variant E223V changed from negatively charged to non-polar. Two variants (G163D, G215D) changed from non-polar to negatively charged, five variants (Q170P, S216F, T230P, S232P and T250P) changed from polar to non-polar, and the variants D176Y, D194N, E199Q, D208N and D277Y changed their polarity from negatively charged to polar. The variant Y272H changed from polar to positively charged amino acid and the remaining ten variants (V66M, L67P, R87K, L160P, H174R, F231L, W239C, P281L, R297K and P324L) retained their polarity. Indeed, by considering only amino acid substitution based on physico-chemical properties, we could not be able to identify the detrimental effect. Rather, by considering the sequence conservation along with the above said properties could have more advantages and reliable to find out the detrimental effect of missense mutations [41].

Predicting the deleterious missense mutants of STK11 by SIFT program

The degree of conservation of a particular position in a protein was determined using sequence homology based tool SIFT [25]. The protein sequences of the 39 variants were submitted to SIFT to determine their tolerance indices. As the tolerance level increases, the functional influence of the amino acid substitution decreases and vice versa. Out of 39 variants, 28 variants viz., Y49D , L67P , R86G , G135R , F157S , L160P , G163D , H174R , D176Y , I177N , N181E , D194N , D194V , D194Y , E199K , E199Q , D208N , G215D , S216F , E223V , T230P , F231L , W239C , L245R , T250P , Y272H , L285Q and R297K were found to be deleterious having the tolerance index of ≤ 0.05 (Table 1).

Among these 28 variants, 21 variants (Y49D , L67P , G135R , F157S , L160P , G163D , H174R , D176Y , I177N , N181E , D194N , D194V , D194Y , G215D , S216F , E223V , W239C , L245R , Y272H , L285Q and R297K)

showed a highly deleterious tolerance index score of 0.00, six variants (R86G, E199K, E199Q, D208N, F231L and T250P) showed a tolerance index of 0.01 and the variant T230P showed a tolerance index score of 0.04 .Interestingly, 26 deleterious variants identified by SIFT also were seen to be less stable by the I-Mutant 2.0 server.

Predicting the damaged missense mutants of STK11 by PolyPhen

Structural level alterations were determined by PolyPhen program. Protein sequence with mutational position and amino acid variants associated with the 39 single point mutants were submitted to the PolyPhen server. A PSIC score difference of 0.001 and above was considered to be damaging. Of the 39 variants, 37 were said to be damaging by PolyPhen and these variants had a PSIC score difference between 0.01 and 1.00. It was to be noted that the variants that were considered to be damaging by PolyPhen were also identified as less stable by I-Mutant 2.0 server and deleterious according to the SIFT server .

Rational consideration of detrimental point mutations

We rationally considered the 26 most potential detrimental point mutations (Y49D , L67P , R86G, G135R , F157S, L160P , G163D , H174R , D176Y , I177N , N181E , D194N , E199K , E199Q , D208N , G215D , S216F , E223V , T230P , F231L , W239C , L245R , T250P , Y272H , L285Q and R297K) for further course of investigations because they were commonly found to be less stable, deleterious, and damaging by the I-Mutant2.0, SIFT and PolyPhen servers respectively [22, 25,26]. We considered the statistical accuracy of these three programs, I-Mutant improves the quality of the prediction of the free energy change caused by single point protein mutations by adopting a hypothesis of thermodynamic reversibility of the existing experimental data. The accuracy of prediction for sequence and structure based values were 78% and 84% with correlation coefficient of 0.56 and 0.69, respectively [42]. SIFT correctly predicted 69% of the substitutions associated with the disease that affect protein function. PolyPhen-2 evaluates rare alleles at loci potentially involved in complex phenotypes, densely mapped regions identified by genome-wide association studies, and analyses natural selection from sequence data, where even mildly deleterious alleles must be treated as damaging. PolyPhen-2 was reported to achieve a rate of true positive predictions of 92% [42-45]. To obtain precise and accurate measures of the detrimental effect of our variants, comprehensive parameters of all these three programs could be more significant than individual tool parameters. Hence, we further investigated these detrimental missense mutations by structural analysis.

Modeling the mutant structures and computing RMSD values

The available structure of STK11 is PDB ID 2WTK. The mutational position and amino acid variants were mapped onto 2WTK native structure. Mutations at a specified position were performed *in silico* by SWISSPDB viewer independently to obtain a modeled structure. NOMAD-Ref server performed the energy minimizations, for both native structure and the 26 mutant modeled structures. To determine the deviation between the native structure and the mutants, the native structure was superimposed with all 26 mutant modeled structures and calculated the RMSD. The higher the RMSD value, the more deviation there is between the native and mutant structure, which in turn changes the binding efficiency with the binding partner because of deviation in the 3D space of the binding residues of STK11. Table 2 shows the RMSD values for native structure with each mutant modeled structure. Of the 26 mutants, 16 mutants (Y49D , L67P , R86G , H174R , I177N , D208N , S216F , E223V , T230P , F231L , W239C , L245R , T250P , Y272H , L285Q and R297K) exhibited a higher RMSD >2.00 Å as illustrated in Table 2 and the other mutants exhibited an RMSD between 0.25 Å and 1.99 Å .The superimposed structure of native with mutants R86G, N181E and Y272H are also shown in Fig 1.

The total energy for the native type structure was found to be -16734.766KJ/mol. Of the 26 mutants, 18 mutants had higher total energy than native, ranging from -11049.088 to -16723.133KJ/mol as shown in Table 2. Other eight mutants (G135R, H174R, I177N, D194N, D208N, T230P, L245R and T250P) found to be with decreased total energy than native. These 8 mutants might not be contributing to the change in stability of the protein as some of the mutations give more stable but less active protein. Higher the total energy, less stable will the protein structure be.

Table 1 List of variants that were predicted to be functionally significant by I-Mutant 2.0, SIFT and PolyPhen

.

Variant	ΔΔG	Tolerant index	PSIC SD
Y49D	**-2.01**	**0.00**	**1.00**
V66M	-1.67	0.31	0.08
L67P	**-2.28**	**0.00**	**1.00**
R86G	**-0.82**	**0.01**	**0.99**
R87K	-0.65	0.20	0.05
G135R	**-0.90**	**0.00**	**1.00**
F157S	**-1.35**	**0.00**	**1.00**
L160P	**-2.02**	**0.00**	**1.00**
G163D	**-1.58**	**0.00**	**1.00**
Q170P	-0.58	0.16	0.99
G171S	-1.14	0.22	0.99
H174R	**-0.82**	**0.00**	**1.00**
D176Y	**-0.17**	**0.00**	**1.00**
I177N	**-2.04**	**0.00**	**1.00**
N181E	**-0.40**	**0.00**	**1.00**
D194N	**-0.47**	**0.00**	**1.00**
D194V	0.65	0.00	1.00
D194Y	0.34	0.00	1.00

E199K	**-1.12**	**0.01**	**1.00**
E199Q	**-0.75**	**0.01**	**1.00**
A205T	-0.52	0.19	0.69
D208N	**-0.53**	**0.01**	**1.00**
G215D	**-1.22**	**0.00**	**1.00**
S216F	**-0.24**	**0.00**	**1.00**
E223V	**-0.26**	**0.00**	**1.00**
T230P	**-0.56**	**0.04**	**0.91**
F231L	**-1.28**	**0.01**	**0.99**
S232P	-0.73	0.11	0.34
W239C	**-1.39**	**0.00**	**1.00**
L245R	**-1.61**	**0.00**	**1.00**
T250P	**-0.84**	**0.01**	**0.99**
Y272H	**-1.45**	**0.00**	**0.90**
D277Y	-0.15	1.00	0.01
P281L	-0.67	0.31	0.02
L285Q	**-2.56**	**0.00**	**1.00**
R297K	**-1.46**	**0.00**	**1.00**
P314H	-1.25	0.07	0.00
P315S	-1.16	0.81	0.00
P324L	-0.84	0.11	0.14

Letters in **bold** indicate less stable, deleterious and damaging by I-Mutant, SIFT and PolyPhen, respectively

Table 2 RMSD, total energy and stabilizing residues for native and mutants of STK11

Variant	RMSD (Å)	Total Energy (KJ/mol)	No. of SR	Stabilizing Residues	ACE Kcal/mol
Native	0.00	-16734.766	9	**Ala (30), Lys (32), Ile (33), Met (80), Cys (85), Gly (86), Leu (134), Lys (142),Ile (143)**	53.31
Y49D	2.59	-16678.254	11	**Ala(30),** Val(31),**Lys(32), Ile(33),** Lys(35), Tyr(77), **Cys(85), Gly(86), Leu(134), Lys(142), Ile(143)**	441.93
L67P	2.08	-16517.932	6	**Ala(30),** Val(31),**Lys(32), Ile(33)**, Leu(34), **Ile(143)**	168.55
R86G	2.70	-16456.994	7	**Ala(30),**Val(31), **Lys(32), Ile(33), Lys(142), Ile(143),** Thr(200)	81.96
G135R	1.91	-16980.611	5	**Ala(30),**Val(31), **Lys(32), Ile(33), Lys(142)**	58.02
F157S	1.92	-16626.852	7	**Ala(30),**Val(31), **Lys(32), Ile(33), Lys(142),** Phe(170), Ser(191)	159.49
L160P	0.25	-11190.762	3	**Ala(30),**Val(31), **Lys(32)**	455.00
G163D	0.26	-11574.894	3	**Ala(30),**Val(31), **Lys(32)**	-37.55
H174R	2.10	-16937.945	9	**Ala(30),**Val(31), **Lys(32), Ile(33),** Gly(114), Leu(141), **Lys(142),**	415.86

				Ile(143), Asp(188)	
D176Y	0.25	-11049.088	3	**Ala(30),**Val(31), **Lys(32)**	12.54
I177N	2.00	-16791.643	7	**Ala(30),**Val(31), **Lys(32), Ile(33), Met(80), Lys(142),** Asp(188)	149.88
N181E	1.97	-16520.436	7	**Ala(30),**Val(31), **Lys(32), Ile(33), Cys(85), Gly(86), Ile(143)**	323.02
D194N	1.99	-16878.275	11	**Ala(30),**Val(31), **Lys(32), Ile(33), Cys(85),Gly(86),** Gly(114), **Leu(134), Lys(142), Ile(143)**	349.11
E199K	1.86	-16630.598	11	**Ala(30),**Val(31), **Lys(32), Ile(33), Met(80), Cys(85), Gly(86), Leu(134), Lys(142),Ile(143)**	26.14
E199Q	1.84	-16723.133	7	**Ala(30)**,Val(31), **Lys(32), Ile(33), Met(80), Lys(142), Ile(143)**	85.24
D208N	2.70	-17025.410	6	**Ala(30),**Val(31), **Lys(32), Ile(33), Lys(142), Ile(143)**	70.74
G215D	0.27	-11732.503	3	**Ala(30),**Val(31), **Lys(32)**	183.93
S216F	2.23	-16508.961	7	**Ala(30),**Val(31), **Lys(32**), **Ile(33), Met(80), Lys(142), Ile(143)**	365.67
E223V	2.00	-16612.900	8	**Ala(30),**Val(31), **Lys(32), Ile(33), Met(80), Cys(85), Lys(142), Ile(143)**	283.63
T230P	2.52	-16783.074	6	**Ala(30),**Val(31), **Lys(32), Ile(33), Lys(142), Ile(143)**	-73.51

F231L	2.05	-16665.414	9	**Ala(30),**Val(31), **Lys(32), Ile(33),** Leu(34), **Met(80), Cys(85), Lys(142), Ile(143)**	335.25
W239C	2.16	-16642.031	6	**Ala(30),**Val(31), **Lys(32), Ile(33), Lys(142), Ile(143)**	343.05
L245R	2.74	-16971.711	8	**Ala(30), Lys(32), Ile(33), Cys(85), Gly(86), Leu(134), Lys(142), Ile(143)**	0.09
T250P	2.58	-16805.973	11	**Ala(30),**Val(31), **Lys(32), Ile(33), Cys(85),Gly(86),**Lys(97), Arg(98), **Leu(134), Lys(142), Ile(143)**	183.49
Y272H	2.63	-16664.336	8	Val(31), **Lys(32), Ile(33), Cys(85),Gly(86), Leu(134), Lys(142),** Thr(200)	74.26
L285Q	2.05	-16699.883	6	**Ala(30)**,Val(31), **Lys(32), Ile(33), Lys(142), Ile(143)**	202.69
R297K	2.58	-16407.490	10	**Ala(30),**Val(31), **Lys(32), Ile(33), Cys(85), Gly(86), Leu(134), Lys(142), Ile(143),**Gln(171)	451.48

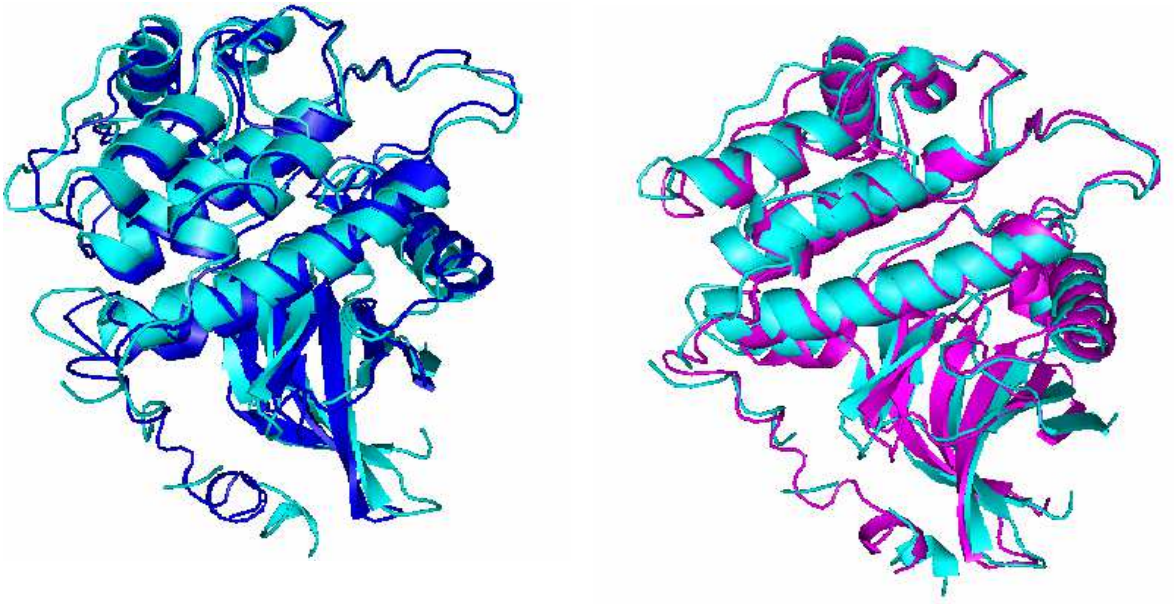
RMSD - root mean square deviation; SR-Stabilizing Residue; the common stabilizing residues are shown in bold; ACE-Atomic Contact Energy

Table 3 Comparison of normalized mean square displacement of binding residues of native protein and mutants

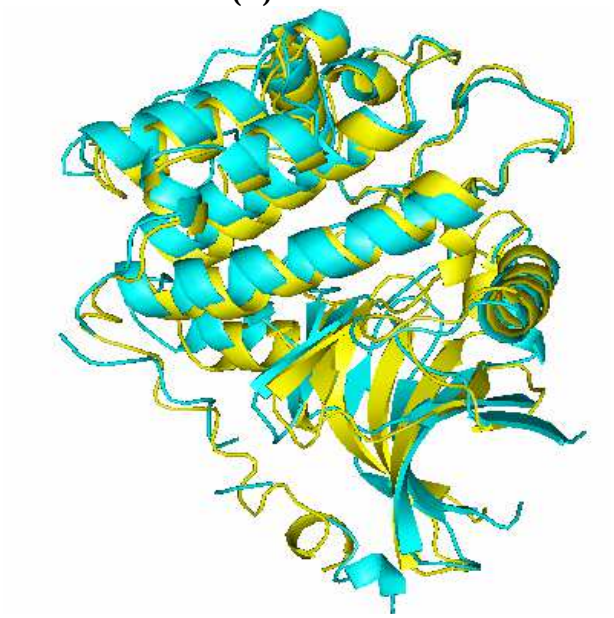
BR	Native	L67P	R86G	F157S	L160P	N181E	E199Q	G215D	S216F	E223V	W239C	Y272H	L285Q
Normalized mean square displacement $\langle R^2 \rangle$													
Leu50	0.0011	*0.001	*0.0009	**0.0015**	**0.0232**	**0.0024**	**0.0021**	**0.0241**	**0.0015**	*0.001	*0.001	*0.0002	**0.0016**
Leu67	0.0030	**0.0034**	*0.0029	**0.0034**	**0.0135**	*0.0023	**0.0042**	**0.0143**	*0.0022	*0.0029	*0.0029	*0.0023	**0.0035**
Ser69	0.0043	*0.0036	*0.0035	**0.0048**	**0.0276**	*0.0042	**0.0059**	**0.0284**	*0.0041	*0.0039	*0.0038	*0.003	**0.0048**
Glu70	0.0060	*0.0047	*0.0041	**0.0064**	**0.0331**	**0.0069**	**0.0075**	**0.0337**	**0.0062**	*0.0053	*0.0052	*0.0045	**0.0062**
Thr71	0.0071	*0.0063	*0.0058	**0.0075**	**0.0298**	**0.0075**	**0.0087**	**0.0304**	*0.007	*0.0065	*0.0063	*0.0058	**0.0074**
Leu72	0.0060	*0.0058	*0.0055	**0.0065**	**0.0219**	*0.0056	**0.0078**	**0.0225**	*0.0054	*0.0057	*0.0056	*0.005	**0.0066**
Cys73	0.0056	*0.0054	*0.0051	**0.0059**	**0.0165**	*0.0054	**0.0068**	**0.0167**	*0.0052	*0.0052	*0.0052	*0.0048	**0.0059**
Arg74	0.0045	**0.0051**	*0.0044	**0.0047**	**0.0085**	*0.0039	**0.0055**	**0.0085**	*0.0039	*0.0042	0.0042	*0.0041	**0.0048**
Arg106	0.0060	0.006	**0.0064**	*0.0045	**0.0105**	*0.0046	**0.0061**	**0.0081**	*0.0052	**0.0064**	*0.0049	**0.0062**	*0.0048
Gln112	0.0051	***0.0049	**0.0054**	*0.0050	**0.0081**	*0.0048	**0.0053**	**0.0082**	*0.005	*0.005	*0.005	**0.0053**	**0.0052**
Val133	0.0030	**0.0037**	**0.0035**	**0.0033**	**0.0141**	**0.0033**	**0.0035**	**0.0139**	**0.0032**	0.003	**0.0031**	**0.0036**	**0.0033**
Thr186	0.0050	**0.0055**	**0.0053**	**0.0051**	**0.0142**	**0.0052**	**0.0055**	**0.0139**	**0.0051**	*0.0049	*0.0048	**0.0053**	**0.0052**
Gly187	0.0055	**0.0059**	*0.0054	0.0055	**0.0139**	**0.0059**	**0.0058**	**0.0138**	**0.0058**	0.0055	*0.0054	**0.0059**	**0.0057**
Lys312	0.0064	*0.0061	**0.0067**	**0.0065**	**0.0142**	**0.0081**	*0.0058	**0.0147**	**0.0075**	**0.0066**	**0.0065**	**0.0069**	*0.0058
Pro321	0.0065	*0.0063	*0.0063	*0.0064	**0.0173**	**0.0083**	*0.0062	**0.0179**	**0.0074**	*0.0062	*0.006	**0.0067**	*0.0061
Ile322	0.0053	0.0053	0.0053	0.0053	**0.0175**	**0.0069**	*0.0052	**0.0178**	**0.006**	*0.0052	*0.0049	**0.0057**	*0.0052
Arg331	0.0073	*0.0079	*0.0072	0.0073	**0.0193**	*0.0072	**0.0084**	**0.0183**	0.0073	*0.0069	*0.0068	*0.0068	0.0073
Val337	0.0045	**0.0053**	**0.0051**	**0.0047**	**0.012**	*0.0031	**0.006**	**0.013**	*0.0036	**0.0046**	**0.0046**	*0.0042	**0.0051**
Val338	0.0060	**0.0067**	**0.0065**	**0.0062**	**0.0181**	*0.0044	**0.0077**	**0.02**	*0.0051	**0.0061**	0.006	*0.0057	**0.0067**
Leu341	0.0042	**0.0047**	**0.0046**	**0.0046**	**0.0322**	*0.0017	**0.0061**	**0.0342**	*0.0027	**0.0046**	**0.0046**	*0.0041	***0.0053**

BR-Binding Residues; Bold numbers indicate amino acids with increased flexibility in the mutant compared with the native protein; * mark indicates amino acids with decreased flexibility in the mutants compared with the native protein

Figure 1: Pymol view of **(A**) superimposed structure of the mutant R86G (blue) with native (cyan) (**B)** superimposed structure of the mutant N181E (magenta) with native (cyan) **(C)** superimposed structure of the mutant Y272H (yellow) with native (cyan)



(A) **(B)**



(C)

Computation of stabilizing residues

In order to understand the stability of the native and mutant structures furthermore, we used the SRide server [36]. It can be seen from Table 2 that 9 stabilizing residues were identified in native structure. The variants Y49D, D194N, T250P and E199K found more number of stabilizing residues than in native, i.e. 11 stabilizing residues. Two mutants H174R and F231L were identified with the same number of stabilizing residues of native and the mutant R297K identified with 10 stabilizing residues. The remaining 19 mutants **(**L67P, R86G, G135R, F157S,

L160P, G163D, D176Y, I177N, N181E, E199Q, D208N, G215D, S216F, E223V, T230P, W239C, L245R, Y272H and L285Q) were found to be less in number of stabilizing residues than the native structure. Only 3 stabilizing residues were present in the mutants L160P, G163D, D176Y and G215D and which is found to be the lowest number of stabilizing residues among the other mutants. Two stabilizing residues namely Ala(30) and Lys(32) were found to be common in native and these mutants and all the other stabilizing residues were missing.

Rationale of binding efficiency for native and mutant structures of STK11 with STRAD

To determine the binding efficiency of STK11 with STRAD, the PDB ID 2WTK structure was selected, and the program SPPIDER was used to calculate contacts between the binding residues of STK11 and STRAD. In this analysis we found that, 20 amino acids viz., Leu (50), Leu (67), Ser (69), Glu (70), Thr (71), Leu (72), Cys (73), Arg (28), Arg (62), Gln (112), Val (133), Thr (186), Gly (187), Lys (312), Pro (321), Ile (322), Arg (331) , Val (337), Val (338) and Leu (341) act as binding residues in STK11- STRAD interaction (Table 3).

Docking was performed using the PatchDock server between STRAD and the native and mutant modeled structures of STK11 to determine the binding efficiency in the form of the Atomic Contact Energy (ACE).The ACE between STRAD and native STK11 was found to be 53.31kcal/mol. Five of the mutants namely G163D, D176Y, E199K, T230P and L245R were found to be with increased binding affinity in terms of ACE. This might be a result of the 3D conformation of STRAD, which had a comfortable fit into the 3D space of the binding residues of these mutants as compared with the native. All the other mutants were found less binding affinity with STRAD than the native STK11 in terms of ACE (Table 2) ranging from 58.02 to 455kcal/mol.

The mutants with increased total energy, RMSD, decreased number of stabilizing residues and low binding affinity only chosen for the further work. Thus, 12 mutants (L67P, R86G, F157S, L160P, N181E, E199Q, G215D, S216F, E223V, W239C, Y272H and L285Q) were selected as they were satisfying all the parameters above mentioned. Moreover these mutants were commonly found to be less stable, deleterious and damaging by the I-mutant 2.0, SIFT and PolyPhen servers respectively. These 12 mutants were also confirmed as detrimental by experimental and clinical observations performed elsewhere. These studies did not use structural and binding analysis but rather used molecular genetic analysis, population prevalence and epidemiology survey [10-14 and 47-51]. Of the 12 mutational residues, 8 residues namely L67, R86, N181, E199, G215, S216, E223, Y272 were also identified as surface accessible residues [15] which contributes more to their candidature of deleterious mutation as many of the disease related mutations lie in the solvent accessible region of the protein [45,46]. Hence we further investigated the 12 detrimental point mutations by normal mode analysis to understand the flexibility of the active site region for the native and mutant structures.

The majority of amino acids in active site show loss of flexibility

To understand the variation of STRAD binding efficiency of the 12 detrimental missense mutations, the program ElNémo was used to compare the flexibility of amino acids that are involved in binding with STRAD of both the native protein and the mutants. Table 3 depicts the flexibility of the amino acids in the active site of both the native and mutant proteins of STK11 by means of the normalized mean square displacement $\langle R^2 \rangle$. These data were further sorted into three different categories of flexibility as shown in Fig 2. First one is where the $\langle R^2 \rangle$ of the amino acids in the active site of the mutant was the same as that of the native protein (termed identical flexibility). The second category was where the $\langle R^2 \rangle$ of the amino acids in the active site of the mutant was higher than that of the native protein (termed increased flexibility). The last category was where the $\langle R^2 \rangle$ of the amino acids in the active site of a mutant was lower than that of the native protein (termed decreased flexibility). In this analysis, we found that fewer active sites of these 12 mutants have identical flexibility than have increased and decreased flexibility (Fig 2). However, some active site amino acids have increased flexibility than have decreased flexibility (Fig 2). Decrease in the flexibility creates an unfavorable reduction of conformational entropy and increase in flexibility creates a large loss in enthalpy (weakened native contacts) that is also unfavorable [18]. Thus the majority of the amino acids participating in STRAD binding of these mutants lost their flexibility, leading to a loss of binding efficiency with the STRAD.

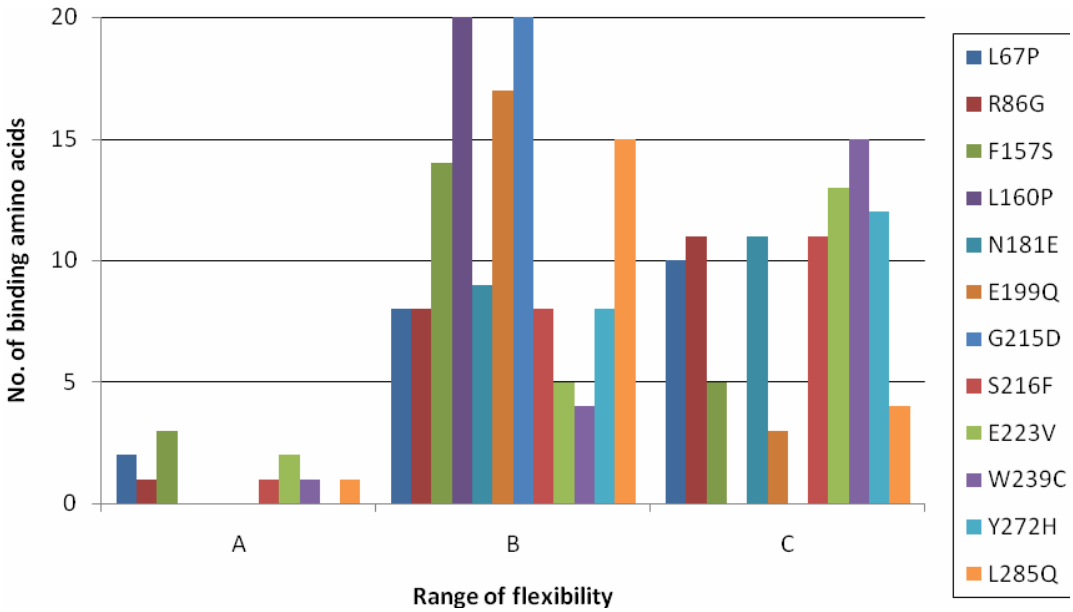


Figure 2: Normalized mean square displacement <R2> of active site amino acids; Where the letter 'A' denotes amino acids with identical flexibility of both native and mutants; 'B' denotes amino acids with increased flexibility of mutants than native; 'C' denotes amino acids with decreased flexibility of mutants than native

CONCLUSION

Of the 39 variants that were retrieved from Swissprot, 37 variants were found less stable by I-Mutant2.0, 28 variants were found to be deleterious by SIFT and 37 variants were considered damaging by PolyPhen. 26 variants were selected as potentially detrimental point mutations because they were commonly found to be less stable, deleterious and damaging by the I-Mutant 2.0, SIFT and PolyPhen servers, respectively. The structures of these 26 variants were modeled and the RMSD between the mutants and native structures ranged from 0.25Å to 2.74Å. Docking analysis between STRAD and the native and mutant modeled structures generated ACE scores between -73.51 and 455kcal/mol. Protein stability analysis of native and mutants exposed the variation in total energy from -11049.088 to -17025.410KJ/mol, and stabilizing residues ranging from 3 to 11. Finally, we concluded that the lower binding affinity of 12 mutants (L67P, R86G, F157S, L160P, N181E, E199Q, G215D, S216F, E223V, W239C, Y272H and L285Q) with STRAD compared with native in terms of their total energy, ACE, stabilizing residue and RMSD scores identified them as deleterious mutations. Normalized mean square displacement $\langle R^2 \rangle$ by normal mode analysis allows us to conclude that majority of the active site amino acids in the mutants bind to STRAD had lost their flexibility which could be the cause for their decreased binding affinity. Thus the results indicate that our approach successfully allowed us to (i) consider computationally a suitable protocol on the basis of stability and structural information to predict the impact of a missense mutation (point mutation/ single amino acid polymorphism) on the protein function before wet lab experimentation (ii) provided an optimal path for further clinical and experimental studies to characterize STK11 mutants in PJS and other types of cancer in depth.

Acknowledgment

The authors thank the management of Vellore Institute of Technology University for providing the facilities to carry out this work.

REFERENCES

1. Hemminki A, Markie D, Tomlinson I, Avizienyte E, Roth S, Loukola A, Bignell G,Warren W, Aminoff M, Höglund P, Järvinen H, Kristo P, Pelin K, Ridanpää M, Salovaara R, Toro T, Bodmer W, Olschwang S, Olsen AS, Stratton MR, de la Chapelle A, Aaltonen LA. A serine/threonine kinase gene defective in Peutz-Jeghers syndrome. Nature. 1998; 391:184-7.
2. Jenne DE, Reimann H, Nezu J, Friedel W, Loff S, Jeschke R, Müller O, Back W,Zimmer M. Peutz-Jeghers syndrome is caused by mutations in a novel serine threonine kinase. Nat Genet. 1998; 18:38-43.
3. Calva D, Howe JR. Hamartomatous polyposis syndromes. Surg Clin North Am. 2008; 88:779-817.
4. Schreibman IR, Baker M, Amos C, McGarrity TJ. The hamartomatous polyposis syndromes: a clinical and molecular review. Am J Gastroenterol. 2005; 100:476-90.
5. Giardiello FM, Trimbath JD. Peutz-Jeghers syndrome and management recommendations. Clin Gastroenterol Hepatol. 2006; 4:408-15.
6. Brosens LA, van Hattem WA, Jansen M, de Leng WW, Giardiello FM, Offerhaus GJ. Gastrointestinal polyposis syndromes. Curr Mol Med. 2007; 7:29-46.
7. Alessi DR, Sakamoto K, Bayascas JR. LKB1-dependent signaling pathways. Annu Rev Biochem. 2006;75:137-63.
8. Karuman P, Gozani O, Odze RD, Zhou XC, Zhu H, Shaw R, Brien TP, Bozzuto CD,Ooi D, Cantley LC, Yuan J. The Peutz-Jegher gene product LKB1 is a mediator of p53-dependent cell death. Mol Cell. 2001; 7:1307-19.
9. Spigelman AD, Arese P, Phillips RK. Polyposis: the Peutz-Jeghers syndrome. Br J Surg. 1995 ; 82:1311-4.
10. Resta N, Pierannunzio D, Lenato GM, Stella A, Capocaccia R, Bagnulo R, Lastella P, Susca FC, Bozzao C, Loconte DC, Sabbà C, Urso E, Sala P, Fornasarig M, Grammatico P, Piepoli A, Host C, Turchetti D, Viel A, Memo L, Giunti L, Stigliano V, Varesco L, Bertario L, Genuardi M, Lucci Cordisco E, Tibiletti MG,Di Gregorio C, Andriulli A, Ponz de Leon M; AIFEG. Cancer risk associated with STK11/LKB1 germline mutations in Peutz-Jeghers syndrome patients: results of an Italian multicenter study. Dig Liver Dis. 2013; 45:606-11.
11. Amos CI, Keitheri-Cheteri MB, Sabripour M, Wei C, McGarrity TJ, Seldin MF,Nations L, Lynch PM, Fidder HH, Friedman E, Frazier ML. Genotype-phenotype correlations in Peutz-Jeghers syndrome. J Med Genet. 2004 ;41:327-33.
12. Schumacher V, Vogel T, Leube B, Driemel C, Goecke T, Möslein G, Royer-PokoraB. STK11 genotyping and cancer risk in Peutz-Jeghers syndrome. J Med Genet. 2005; 42:428-35.
13. Boudeau J, Sapkota G, Alessi DR. LKB1, a protein kinase regulating cell proliferation and polarity. FEBS Lett. 2003; 546:159-65.
14. Zeqiraj E, Filippi BM, Deak M, Alessi DR, van Aalten DM. Structure of the LKB1-STRAD-MO25 complex reveals an allosteric mechanism of kinase activation. Science. 2009; 326:1707-11.
15. Baas AF, Boudeau J, Sapkota GP, Smit L, Medema R, Morrice NA, Alessi DR, Clevers HC. Activation of the tumour suppressor kinase LKB1 by the STE20-like pseudokinase STRAD. EMBO J. 2003; 22:3062-72.
16. Boudeau J, Scott JW, Resta N, Deak M, Kieloch A, Komander D, Hardie DG, Prescott AR, van Aalten DM, Alessi DR. Analysis of the LKB1-STRAD-MO25 complex. J Cell Sci. 2004; 117:6365-75.
17. Hawley SA, Boudeau J, Reid JL, Mustard KJ, Udd L, Mäkelä TP, Alessi DR, Hardie DG. Complexes between the LKB1 tumor suppressor, STRAD alpha/beta and MO25 alpha/beta are upstream kinases in the AMP-activated protein kinase cascade. J Biol. 2003;2(4):28.
18. Verma D, Jacobs DJ, Livesay DR. Changes in Lysozyme Flexibility upon Mutation Are Frequent, Large and Long-Ranged. PLoS Comput Biol. 2012;8(3):e1002409.
19. Ramensky V, Bork P, Sunyaev S. Human non-synonymous SNPs: server and survey.Nucleic Acids Res. 2002; 30:3894-900.

20. Yip YL, Scheib H, Diemand AV, Gattiker A, Famiglietti LM, Gasteiger E, Bairoch A. The Swiss-Prot variant page and the ModSNP database: a resource for sequence and structure information on human protein variants. Hum Mutat. 2004; 23:464-70.
21. Berman HM, Battistuz T, Bhat TN, Bluhm WF, Bourne PE, Burkhardt K, Feng Z,Gilliland GL, Iype L, Jain S, Fagan P, Marvin J, Padilla D, Ravichandran V, Schneider B, Thanki N, Weissig H, Westbrook JD, Zardecki C. The Protein Data Bank. Acta Crystallogr D Biol Crystallogr. 2002; 58: 899-907.
22. Capriotti E, Fariselli P, Casadio R. I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. Nucleic Acids Res. 2005; 33: W306-10.
23. Bava KA, Gromiha MM, Uedaira H, Kitajima K, Sarai A. ProTherm, version 4.0:thermodynamic database for proteins and mutants. Nucleic Acids Res. 2004; 32:D120-1.
24. Ng PC, Henikoff S. Predicting deleterious amino acid substitutions. Genome Res.; 11:863-74.
25. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res. 2003; 31:3812-4.
26. Ramensky V, Bork P, Sunyaev S. Human non-synonymous SNPs: server and survey. Nucleic Acids Res. 2002; 30:3894-900.
27. Lindahl E, Azuara C, Koehl P, Delarue M. NOMAD-Ref: visualization, deformation and refinement of macromolecular structures based on all-atom normal mode analysis. Nucleic Acids Res. 2006 ;34:W52-6.
28. Delarue M, Dumas P. On the use of low-frequency normal modes to enforce collective movements in refining macromolecular structural models. Proc Natl Acad Sci U S A. 2004; 101:6957-62.
29. Han JH, Kerrison N, Chothia C, Teichmann SA. Divergence of interdomain geometry in two-domain proteins. Structure. 2006; 14:935-45.
30. Varfolomeev SD, Uporov IV, Fedorov EV. Bioinformatics and molecular modeling in chemical enzymology. Active sites of hydrolases. Biochemistry (Mosc). 2002 ;67:1099-108.
31. Duhovny D, Nussinov R, Wolfson H J .Efficient unbound docking of rigid molecules. In: Proceedings of the 2nd Workshop on Algorithms in Bioinformatics (WABI) Lecture Notes in Computer Science, 2002 Rome, Italy, 2452: 185–200
32. Connolly ML. Solvent-accessible surfaces of proteins and nucleic acids. Science. 1983; 221:709-13.
33. Zhang C, Vasmatzis G, Cornette JL, DeLisi C. Determination of atomic desolvation energies from the structures of crystallized proteins. J Mol Biol.1997; 267:707-26.
34. Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ. PatchDock and SymmDock: servers for rigid and symmetric docking. Nucleic Acids Res. 2005 ;33: W363-7.
35. Porollo A, Meller J. Prediction-based fingerprints of protein-protein interactions. Proteins. 2007 ;66:630-45.
36. Magyar C, Gromiha MM, Pujadas G, Tusnády GE, Simon I. SRide: a server for identifying stabilizing residues in proteins. Nucleic Acids Res. 2005; 33:W303-5.
37. Carlson HA, McCammon JA. Accommodating protein flexibility in computational drug design. Mol Pharmacol. 2000 ;57:213-8.
38. Suhre K, Sanejouand YH. ElNemo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement. Nucleic Acids Res. 2004 ;32:W610-4.
39. Boeckmann B, Bairoch A, Apweiler R, Blatter MC, Estreicher A, Gasteiger E,Martin MJ, Michoud K, O'Donovan C, Phan I, Pilbout S, Schneider M. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. Nucleic Acids Res. 2003;31:365-70.
40. Yip YL, Famiglietti M, Gos A, Duek PD, David FP, Gateau A, Bairoch A. Annotating single amino acid polymorphisms in the UniProt/Swiss-Prot knowledgebase. Hum Mutat. 2008 ;29:361-6.
41. Hinkle A, Tobacman LS. Folding and function of the troponin tail domain. Effects of cardiomyopathic troponin T mutations. J Biol Chem. 2003;278:506-13.

42. Capriotti E, Fariselli P, Rossi I, Casadio R. A three-state prediction of single point mutations on protein stability changes. BMC Bioinformatics. 2008 ;9 Suppl 2:S6.
43. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc. 2009;4:1073-81.
44. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. A method and server for predicting damaging missense mutations. Nat Methods. 2010 ;7:248-9.
45. Kann MG. Protein interactions and disease: computational approaches to uncover the etiology of diseases. Brief Bioinform. 2007;8:333-46.
46. Ye Y, Li Z, Godzik A. Modeling and analyzing three-dimensional structures of human disease proteins. Pac Symp Biocomput. 2006:439-50.
47. Kuragaki C, Enomoto T, Ueno Y, Sun H, Fujita M, Nakashima R, Ueda Y, Wada H,Murata Y, Toki T, Konishi I, Fujii S. Mutations in the STK11 gene characterize minimal deviation adenocarcinoma of the uterine cervix. Lab Invest. 2003;83:35-45.
48. Lim W, Hearle N, Shah B, Murday V, Hodgson SV, Lucassen A, Eccles D, Talbot I,Neale K, Lim AG, O'Donohue J, Donaldson A, Macdonald RC, Young ID, Robinson MH, Lee PW, Stoodley BJ, Tomlinson I, Alderson D, Holbrook AG, Vyas S, Swarbrick ET, Lewis AA, Phillips RK, Houlston RS. Further observations on LKB1/STK11 status and cancer risk in Peutz-Jeghers syndrome. Br J Cancer. 2003;89:308-13.
49. Wingo SN, Gallardo TD, Akbay EA, Liang MC, Contreras CM, Boren T, Shimamura T, Miller DS, Sharpless NE, Bardeesy N, Kwiatkowski DJ, Schorge JO, Wong KK,Castrillon DH. Somatic LKB1 mutations promote cervical cancer progression. PLoS One. 2009;4:e5137.
50. Dong SM, Kim KM, Kim SY, Shin MS, Na EY, Lee SH, Park WS, Yoo NJ, Jang JJ,Yoon CY, Kim JW, Kim SY, Yang YM, Kim SH, Kim CS, Lee JY. Frequent somatic mutations in serine/threonine kinase 11/Peutz-Jeghers syndrome gene in left-sided colon cancer. Cancer Res. 1998 ;58:3787-90.
51. Imielinski M, Berger AH, Hammerman PS, Hernandez B, Pugh TJ, Hodis E, Cho J,Suh J, Capelletti M, Sivachenko A, Sougnez C, Auclair D, Lawrence MS, Stojanov P,Cibulskis K, Choi K, de Waal L, Sharifnia T, Brooks A, Greulich H, Banerji S,Zander T, Seidel D, Leenders F, Ansén S, Ludwig C, Engel-Riedel W, Stoelben E,Wolf J, Goparju C, Thompson K, Winckler W, Kwiatkowski D, Johnson BE, Jänne PA,Miller VA, Pao W, Travis WD, Pass HI, Gabriel SB, Lander ES, Thomas RK, Garraway LA, Getz G, Meyerson M. Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. Cell. 2012;150:1107-20.

*****