

Computational modelling of human Sarcomeric Telethonin protein and predicting the functional effect of missense single nucleotide polymorphisms

P. P. Anand*

Department of Zoology, University of Calicut, Malappuram 673 635, India

Telethonin (T-cap) is a 19 kDa protein and in humans it is encoded by the TCAP gene *chr17q12*. Telethonin is expressed in cardiac and skeletal muscle at Z-discs. Telethonin binds to the titin Z1–Z2 domains and is a substrate of titin kinase, interactions thought to be critical to sarcomere assembly. Variation in Telethonin implicates several diseases, including limb-girdle muscular dystrophy, cardiomyopathy, dilated and idiopathic cardiomyopathy. This protein variation study helps us understand the several cardiac and muscular disorders related to missense single nucleotide polymorphisms (SNPs) mutation and specified drug designing. Here we evaluate computational analysis of missense SNPs present in the TCAP gene. It was performed by considering 13 prediction tools in order to select the deleterious variations and followed by molecular modelling, evolutionary conservation analysis, signal peptide prediction and model validation. It has been found that eight deleterious variants, out of which three (Arg70Trp, Pro90Leu and Arg106Cys) are previously associated with human diseases.

Keywords: Computational analysis, genetic diseases, human Sarcomeric protein, missense single nucleotide polymorphism.

IN cardiac and skeletal muscles, actin and myosin proteins are organized into sarcomeres, which are the fundamental contractile units of striated muscle. The sarcomere is composed of ordered, thick (myosin) and thin (actin, tropomyosin, troponin) filaments that slide past each other during contraction¹. The Z-disc is central and plays a unique role, being the main anchoring point of the molecular machinery that underlies muscle contraction. Major Z-disc proteins are actin, alpha-actinin 2, cap-Z, titin and nebulin and the recently discovered Z-disc proteins are telethonin and myotilin. The emerging structural complexity of the Z-disc is a good example of intricate molecular architecture, while the mechanisms of its assembly and adaptation remain mostly unknown¹.

*e-mail: anandpp633@gmail.com

In cardiac muscle, telethonin co-localizes and interacts with MLP, that may have a role in sensing mechanical stretch. A mutation (W4R) of the human MLP protein is found to be associated with dilated cardiomyopathy as it loses the ability to bind telethonin. As a result, a complex between telethonin and MLP is not formed with consequent misplacement of Telethonin and misalignment of the Z-disc². Telethonin that can bind to the N-terminal domain of titinand is a substrate of titin serine kinase. It has also been found to interact with myostatin, a protein that negatively regulates the growth of skeletal muscle, as well as FATZ (calsarcin, myozenin) and minimal potassium channel subunit (MinK), the beta-subunit of a component of the delayed rectifier potassium current I(Ks) channel. The cytoplasmic domain of minK specifically interacts with the C-terminus of Telethonin and in cardiac muscle both minK and Telethonin are co-localized in the Z-disc. It has been proposed that telethonin may act as a link between myofibrillar components of the sarcomere and beta-subunit of the I(Ks) channel found in the T-tubular membrane surrounding the Z-disc in the inner ventricular myocardium. This suggests the possibility of a stretch-dependent regulation of potassium flux in the cardiac muscle³.

The sarcomeric telethonin has been reported to bind with several other proteins, including FATZ/calsarcin-3 (refs 4 and 5), MLP (ref. 6) the β -subunit of the potassium channel subunit minK covering the phosphorylation site at the C-terminal of telethonin sequence⁷, and the Ankrd2 protein³. Remarkably, telethonin has also been detected to serve via its phosphorylation site as a substrate for the serine/threonine kinase domain of titin in developing myocytes⁸. Moreover, there is evidence that telethonin interacts with GDF-8/myostatin, creating a possible control mechanism for the secretion of this transforming growth factor-beta (TGF-beta) superfamily member⁹.

Mutations in telethonin have been linked to several hereditary myopathies, including limb-girdle muscular dystrophy⁹ and dilated and hypertrophic cardiomyopathy^{10,11}. Although there are considerable variations in

the type of myopathies caused by different telethonin mutations, the missense single nucleotide polymorphisms (SNPs) affect the interaction of telethonin with one or several of its partners. This study helps achieve a better understanding of the molecular role of telethonin as a mediator of Z-disk protein interaction networks.

Methodology

Datasets

The Telethonin missense SNPs and protein sequences in FASTA format were retrieved from NCBI-dbSNP database during January 2018 (NCBI accession: NP_003664.1 and Uniprot id p015273). Only human-validated Telethonin missense SNPs were filtered. The initial SNPs were validated by the following conditions: (i) it has been sequenced in 1000 genome project (<http://www.1000genomes.org>); (ii) it has genotype or frequency data and (iii) it has independent and multiple submissions to the refSNP cluster.

Deleterious missense SNPs selection

Thirteen prediction tools under different categories helped filter all missense SNPs and select those showing convergent deleterious effect in more than nine prediction tools. The different categories are as follows: (i) Sequence homology and evolutionary conservation – SIFT (<http://sift.jcvi.org>)¹², Provean (protein variant effect analyzer; <http://provean.jcvi.org>)¹³, PANTHER (<http://www.pantherdb.org/tools>)¹⁴, Align GV-GD (http://agvgd.hci.utah.edu/agvgd_input.php) and MAPP (<http://www.ngri.org.uk/>). (ii) Supervised learning methods – SNAP 2 (<http://www.Hiv.lanl.Gov/content/sequence/SNAP/SNAP.html>)¹⁵, SNPs & Go (<https://snps-and-go.biocomp.unibo.it/snps-and-go/>), Phd-SNP (<http://SNPs.Biofold.org/phd-snp/phd-snp.html>)¹¹ and SusPect (www.sbg.bio.ic.ac.uk/suspect)¹⁶. (iii) Protein sequence and structure-based methods–PolyPhen-2 (<http://genetics.bwh.harvard.edu/pph2>)¹⁷. (iv) Consensus-based methods–Meta-SNP (<http://snps.biofold.org/meta-snp/>)¹⁸, PONP2 (<http://structure.bmc.lu.se/PON-P2/>)^{19,20} and predict-SNP (<http://loschmidt.chemi.muni.cz/predictsnp>)^{19,20}.

Evolutionary conservation analysis

The ConSurf server²¹ is a tool for estimating the evolutionary conservation of amino acid positions in a protein molecule based on the phylogenetic relations between homologous sequences¹⁸. Using Telethonin protein sequences (NCBI Accession: NP_003664.1) and ConSurf in ConSeq mode, a search was carried out for close homologous sequences using CSI-BLAST (three iterations and 0.0001 *e*-value cut-off) against the UNIREF-90 protein database^{22,23}.

Signal peptide prediction

Phobius and SignalP 4.0 servers were used to analyse the impact of convergent deleterious SNPs in the signal peptide topology prediction¹⁶.

Molecular modelling

A new threading algorithm MUSTER (<http://zhang.bioinformatics.ku.edu/MUSTER>)²³ analyse the previous sequence profile–profile alignment (PPA) method⁵. The best models were evaluated using PROCHECK software²⁴. The characteristic feature of good protein model is 90% of amino acid residue present in the most favoured and additional allowed regions and visualized in PyMOL and model validates by PDBsum^{16,25}.

Project hope

This is a web server that helps analyse the structural effects of point mutation in protein sequences. It collects and combines available information from a series of web services and databases, giving results with figures and animations.

Analysis of native and variant models

NOMAD-Ref Gromacs server: Energy minimization is done for simulating *in vivo* conditions. Energy minimization of the native T-cap protein model and the variant protein models was carried out using NOMAD-Ref Gromacs server (http://lorenz.immstr.pasteur.fr/gromacs/minimization_submission.php)²⁶.

FT-site: This is an accurate method for identifying binding sites. It is based on experimental evidence that ligand-binding sites also bind small organic molecules of various shapes and polarities¹⁷.

VADAR server: This is a compilation of more than 15 different algorithms and programs for analysing and assessing peptide and protein structures from their PDB coordinate data¹⁷.

I-Mutant server: This is used to predict the stability of mutant proteins (<http://folding.biofold.org/cgi-bin/i-mutant2.0.cgi>)²⁷.

Results

Distribution frequency of TCAP SNPs

The human *TCAP* gene contains a total of 162 SNPs. Among these, 45 are missenses, 21 are coding synonymous,

six are nonsense variants, one is present in 3' splice site, 59 in 3'-UTR, two in 5'-UTR, 17 in intron and six SNPs are present in stop gained regions (Figure 1).

The missense SNPs were validated by 1000 genome project frequency or genotype, and refSNP cluster. They were used for initial analysis in predicting the deleterious missense SNPs (Table 1).

Predicted convergent deleterious T-cap SNPs

Thirteen computational tools were used for predicting the effect of missense SNPs on protein function as mentioned earlier. All missense SNPs were filtered based on the natural variants in Uniprot. More than 9 out of 13 prediction tools showed deleterious effects (Table 2).

A total of eight deleterious variants (Arg33Trp, Pro90Leu, Arg70Trp, Glu49Lys, Cys57Trp, Arg106Cys, Arg158Ser and Arg87Trp; highlighted as blue colour bold letter in Table 2), and no SNPs were located in signal peptide regions (Figure 2).

The evolutionary conservation of all missense SNPs within mature region of T-cap was examined using ConSurf server. Three variations (Glu49Lys, Pro90Leu and Arg158Ser) occur in highly conserved regions (Figure 3). Disease-causing missense SNPs tend to occur at evolutionarily conserved positions that have an essential role in the structure and/or function of the encoded protein^{28,29}.

Impact of variations on protein structure

Modelling of telethonin protein: The complete structure of telethonin protein model is not present in protein data bank (PDB). The complete structure of native and deleterious telethonin proteins is mandatory for analysing their

structural and functional aspects. Comparative homology modelling of T-cap protein was done using MUSTER tool. Then a good model template was selected to develop a full protein model. The template used for protein modelling was 1YA5(T) and 2F8V(T). A good template (PDB ID 1YA5(T)) was finally selected for further protein modelling. Native telethonin and mutant protein modelling was visualized using PyMol tool (Figure 4).

Validation of protein model: For a good protein structure, it is expected that there should be more than 90% of the residues in the core or favoured regions of the protein^{26,30}. The generated model was analysed by PDBsum (Table 3). The selected deleterious mutant proteins are highly stable, because the 90% of residue occurred in the core region of Ramachandran plot (Figure 5).

Model validation by PDBsum: The simulated telethonin protein model was further analysed using PDBsum, based on the motif, helices, domain, angles, positions, errors, etc. of protein^{12,25}. PDBsum²⁵ analysis of native protein showed one beta sheet, four beta hairpins, five beta bulges, five strands, nine beta turns and one gamma turn (Table 4).

Mapping missense amino acid variation into the protein: A total of eight mutant proteins were chosen to study the effects, in order to mimic the *in vivo* folding conditions and parameters of the protein. Energy minimization of both native and all mutant proteins was done with the help of Nomad-RefGromacs server using conjugant gradient force fields (Table 5)²⁶.

The total energy of native protein structure was -8955.652 kJ/mol. Among the eight mutants, E49K was found to have the highest energy (-9125.286 kJ/mol), after energy minimization when compared with the naïve structure. The mutants P90L (-9078 kJ/mol) and C57W (-9001.259 kJ/mol) were found to have considerably higher energy values, whereas the mutants R70W (-8954.804 kJ/mol), R87W (-8848.714 kJ/mol), R158S (-8804.438 kJ/mol), R33W (-8775.802 kJ/mol) and R106C (-8716.647 kJ/mol) showed very small variations compared to native protein after energy minimization. The energy variation on comparing mutant protein to native protein, can damage the protein structure and also affect the protein function and stability (Table 5).

RMSD value assessment of the modelled proteins: Among the eight mutants analysed, majority of variants showed RMSD value above 4 Å. Variant R106C showed the highest RMSD value 5.46 Å, followed by C57W (4.99 Å), R158S (4.59 Å), R70W (4.40 Å), R87W (4.29 Å) and R33W (4.08 Å). E49K and P90L showed low RMSD values (1.90 and 2.68 Å respectively). The structural deviation between native and mutant

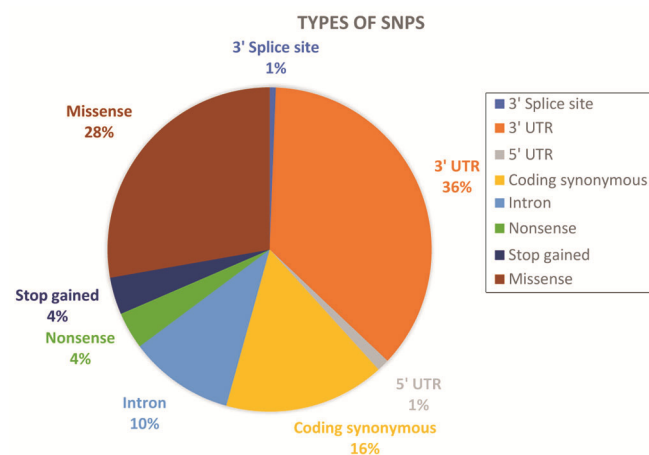


Figure 1. Distribution frequency of Telethonin (T-cap) single nucleotide polymorphisms (SNPs): 36% in 3' UTR, 28% in missense, 16% in coding synonymous, 10% in intron, 4% in stop gained, 4% in nonsense, 1% in 3'-splice site and 1% in 5'-UTR.

Table 1. List of missense single nucleotide polymorphisms (SNPs) used for *in silico* analysis

Variant ID	NCBI accession no.	Aminoacid variation	Position	Native residue	Mutant residue
rs45458802	NP_003664.1	S64L	64	S	L
rs45495192	NP_003664.1	S11W	11	S	W
rs45509691	NP_003664.1	P141A	141	P	A
rs45578741	NP_003664.1	R106C	106	R	C
rs45614536	NP_003664.1	R18Q	18	R	Q
rs121434298	NP_003664.1	R87Q	87	R	Q
rs143233087	NP_003664.1	A118V	118	A	V
rs145524909	NP_003664.1	R33W	33	R	W
rs146906267	NP_003664.1	E105Q	105	E	Q
rs147503632	NP_003664.1	R130H	130	R	H
rs149585781	NP_003664.1	R153H	153	R	H
rs201664428	NP_003664.1	L6M	6	L	M
rs372312912	NP_003664.1	L113F	113	L	F
rs374886575	NP_003664.1	R130C	130	R	C
rs375310569	NP_003664.1	C38F	38	C	F
rs397516863	NP_003664.1	R158S	158	R	S
rs397516864	NP_003664.1	R158H	158	R	H
rs552865793	NP_003664.1	R70Q	70	R	Q
rs572836774	NP_003664.1	R76C	76	R	C
rs727503455	NP_003664.1	A101T	101	A	T
rs727504427	NP_003664.1	P90L	90	P	L
rs753744791	NP_003664.1	G75S	75	G	S
rs755395354	NP_003664.1	M160I	160	M	I
rs755539784	NP_003664.1	R154C	154	R	C
rs762850913	NP_003664.1	G150S	150	G	S
rs773317399	NP_003664.1	T137K	137	T	K
rs775636212	NP_003664.1	R70W	70	R	W
rs777518512	NP_003664.1	R87W	87	R	W
rs779699520	NP_003664.1	E35K	35	E	K
rs45513698	NP_003664.1	E49K	49	E	K
rs76130885	NP_003664.1	I93L	93	I	L
rs143465226	NP_003664.1	M71T	71	M	T
rs369447207	NP_003664.1	C57W	57	C	W
rs558480997	NP_003664.1	Q126R	126	Q	R
rs576098128	NP_003664.1	R106H	106	R	H
rs749685550	NP_003664.1	Q126H	126	Q	H
rs750796201	NP_003664.1	R17H	17	R	H
rs758048577	NP_003664.1	R63C	63	R	C
rs761498487	NP_003664.1	R153S	153	R	S
rs763965899	NP_003664.1	A151V	151	A	V
rs764200782	NP_003664.1	M71V	71	M	V
rs764350236	NP_003664.1	C15S	15	C	S
rs775080606	NP_003664.1	T3N	3	T	N
rs777384494	NP_003664.1	R63H	63	R	H
rs780544264	NP_003664.1	R154H	154	R	H

Human Telethonin (T-cap) protein missense SNPs, validated by 1000 genome, frequency and cluster. Data collected in January 2018 from NCBI.

proteins was determined by the help of RMSD value. The RMSD value indicated the backbone movement of native and mutant protein models of T-cap (Table 5).

Analysis of structural effects of mutant missense SNPs: The structural effects were studied using Project Hope web server service (www.cmbi.ru.nl/hope/). By analysing structural variation of C57W, the mutant residue is larger in size compared to native residue. This residue is located on the surface of the protein, so mutation of this residue can disturb interactions with other molecules or other

parts of the protein. For mutant E49K, the native residue forms a salt bridge with arginine at position 47. In both the PDB file and in the PISA-assembly (PISA database contains protein assemblies that are highly likely to be biologically relevant), this residue was found to be involved in a multimer contact. This indicates that the residue is indeed in contact with other proteins. This is a rare case of mutation occurring in highly conserved regions and it introduces positive charge at this position. This mutation is introducing a bigger residue at this position, which can disturb the multimeric interactions.

Table 2. Predicted convergent deleterious T-cap SNPs

Variant ID	Amino acid variation	SIFT	PROVEAN	PANTHER	Alignn-GVGD	SNAP	SusPect	PhD-SNP	SNPs and GO	Poly Phen-2	Meta-SNP	PON-P 2	PREDICT-SNP	MAPP
rs45458802	S64L	N	D	PSD	D	D	D	N	N	N	N	D	N	D
rs45495192	S11W	D	D	PSD	D	D	N	N	N	N	N	D	D	D
rs45509691	P141A	D	D	PD	N	D	N	N	N	D	N	N	D	D
rs45578741	R106C	D	D	PD	D	D	N	N	N	D	N	D	D	D
rs45614536	R18Q	D	N	PD	N	D	N	D	N	N	N	N	D	D
rs121434298	R87Q	D	D	PSD	N	D	N	N	N	N	N	N	D	D
rs143233087	A118V	N	N	PSD	PSD	N	N	N	N	N	N	N	N	D
rs145524909	R33W	D	D	PSD	D	D	D	N	D	D	N	D	D	D
rs146906267	E105Q	N	N	N	N	D	N	N	N	N	N	N	N	N
rs147503632	R130H	D	D	N	N	D	N	N	N	D	N	N	D	D
rs149585781	R153H	N	N	PSD	N	D	N	N	N	N	N	N	N	N
rs201664428	L6M	D	N	PD	N	D	N	N	N	N	N	N	D	N
rs372312912	L113F	D	N	PD	N	D	D	N	N	N	N	N	D	N
rs374886575	R130C	D	D	N	D	D	N	N	D	D	N	N	D	D
rs375310569	C38F	N	D	PD	D	D	N	D	D	N	N	N	D	D
rs397516863	R158S	D	D	PD	D	D	N	N	N	D	N	D	D	D
rs397516864	R158H	D	D	PD	N	D	N	N	N	D	N	D	D	D
rs552865793	R70Q	N	N	PSD	N	D	N	N	N	D	N	D	N	D
rs572836774	R76C	D	D	N	D	D	N	N	D	N	N	N	D	D
rs727503455	A101T	N	N	N	PSD	N	N	N	N	N	N	N	N	N
rs727504427	P90L	D	D	PD	D	D	N	D	D	D	D	D	D	D
rs753744791	G75S	N	D	PSD	PSD	D	N	N	N	D	N	N	N	D
rs755395354	M160I	D	N	PSD	N	D	N	N	N	N	N	N	D	D
rs755539784	R154C	D	D	PSD	D	D	N	N	N	N	N	N	N	D
rs762850913	G150S	N	N	PSD	PSD	D	N	N	N	N	N	N	N	D
rs773317399	T137K	D	D	N	D	N	N	N	N	N	N	N	N	D
rs775636212	R70W	D	D	PSD	D	D	D	D	D	D	D	D	D	D
rs777518512	R87W	D	D	PSD	D	D	N	N	D	D	N	D	D	D
rs779699520	E35K	D	D	PSD	PSD	D	N	N	N	N	N	N	D	D
rs45513698	E49K	D	D	PSD	PSD	D	N	D	D	D	D	D	D	D
rs76130885	I93L	N	N	PD	N	D	D	D	N	N	D	N	N	D
rs143465226	M71T	D	D	PSD	D	D	D	N	N	N	N	N	D	D
rs369447207	C57W	D	D	PSD	D	D	D	D	D	D	D	N	D	D
rs558480997	Q126R	N	N	PSD	N	D	N	N	N	N	N	N	N	D
rs576098128	R106H	D	D	D	N	D	N	N	N	N	N	N	N	D
rs749685550	Q126H	N	N	PSD	N	D	N	N	N	N	N	N	N	N
rs750796201	R17H	D	D	PSD	N	D	N	N	N	N	N	N	D	D
rs758048577	R63C	D	D	PSD	D	D	N	N	D	N	N	N	D	D
rs761498487	R153S	N	N	PSD	D	D	N	N	N	N	N	N	N	N
rs763965899	A151V	D	N	PSD	PSD	N	N	N	N	N	N	N	N	D
rs764200782	M71V	N	N	PSD	N	D	N	N	N	N	N	N	N	D
rs764350236	C15S	N	N	N	D	D	N	N	N	N	N	N	N	N
rs775080606	T3N	N	N	N	PSD	N	N	N	N	N	N	N	N	D
rs777384494	R63H	N	N	PSD	N	D	N	N	N	N	N	N	N	N
rs780544264	R154H	D	D	PSD	N	D	N	N	N	N	N	N	D	D

D, Deleterious; N, Neutral; PSD, Possibly damaging; PD, Probably damaging. (The complete prediction details available as supplementary data.)

This possibly disrupts the contact with other molecules. For mutant P90L, wild residue is proline and it is known to be rigid and therefore induce a special backbone conformation which might be required at this position and the mutation can disturb this special conformation. For mutant R33W, the residue has multimer contact. The mutation introduces a large-sized residue at this position which can disturb the multimeric interactions. The wild and mutant proteins have positive and neutral charges respectively. For mutant R70W, the wild-type residue forms a hydrogen bond with glutamine at position 62 and

glutamic acid at position 80. Wild residue forms a salt bridge with aspartic acid at position 27 and glutamic acid at position 80. The mutation introduces a different hydrophobicity that can alter the hydrogen-bond formation and a different charge that can disturb the ionic interactions. This mutation is already reported in diseases, viz. cardiomyopathy and familial hypertrophic 25 (CMH 25). For mutant R87W, wild residue forms a salt bridge with leucine at position 89. This mutation occurs on the surface of the protein and will change the ionic interactions because wild and mutant residues have

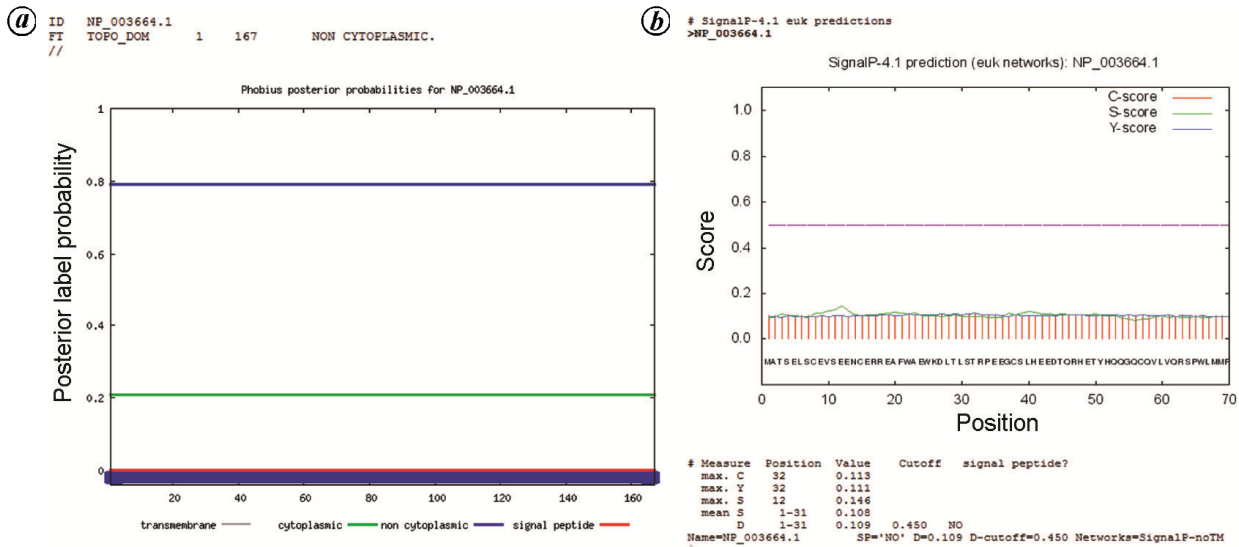


Figure 2. Signal peptide prediction using (a) Phobius and (b) Signal P 4.1.

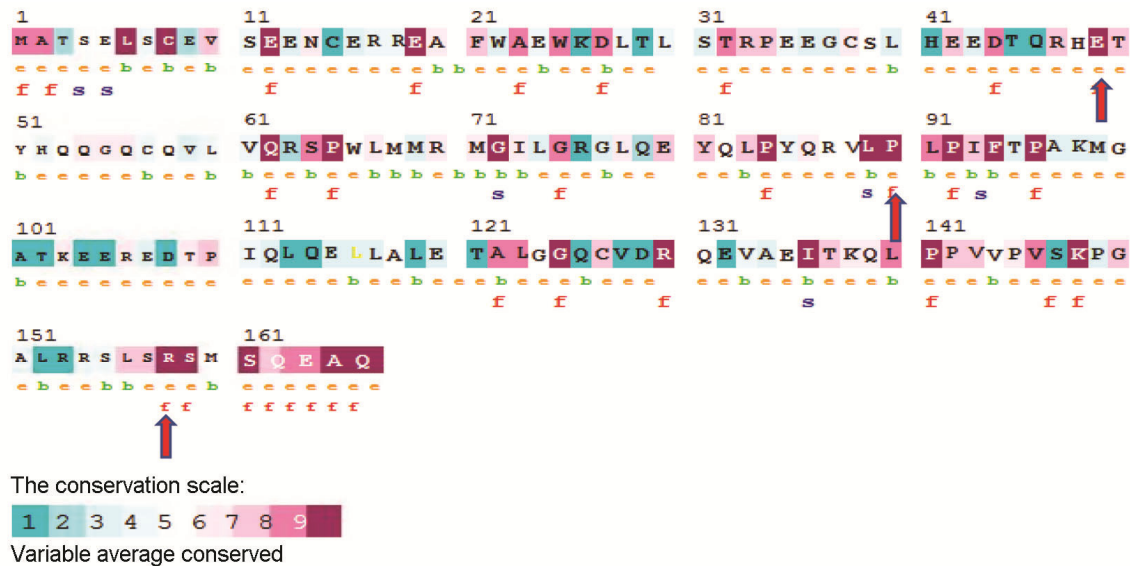


Figure 3. Colour grade indicating the conservation status of amino acids. The grade number (colour) increases with increasing evolutionary conservation (1 is a highly variable and 9 a slowly variable site). Red arrows indicate the three predicted convergent deleterious T-cap missense SNPs.

positive and neutral charge respectively. This mutation is also reported as deleterious and related to dilated cardiomyopathy. For mutant R158S, there are wild (positive charge) and mutant residues (neutral). This mutation introduces more hydrophobic nature and can alter the hydrogen bond or disturb correct protein folding. Comparing the native and mutant residue of R106C, mutant residue is smaller than wild and the mutant residue is more hydrophobic than wild. The wild residue forms hydrogen bond with glutamine at position 112 and salt bridge with glutamic acid at positions 14, 115 and 120. Wild residue is positive and mutant residue is neutral. This mutation is located in a region with known splice

site and also this variant is annotated with severity polymorphism.

ASA analysis using VADAR server: The total accessible surface area (ASA) of R70W, R87W and R158S is less than native T-cap protein, whereas for other variants total ASA is higher than native proteins. The percentage of side ASA hydrophobic of the variant model is higher than the native protein. This indicates the structural deviation of protein and loss of interaction (Table 6).

Ligand-binding validation: FT-site is an accurate method for identifying binding sites. It is based on

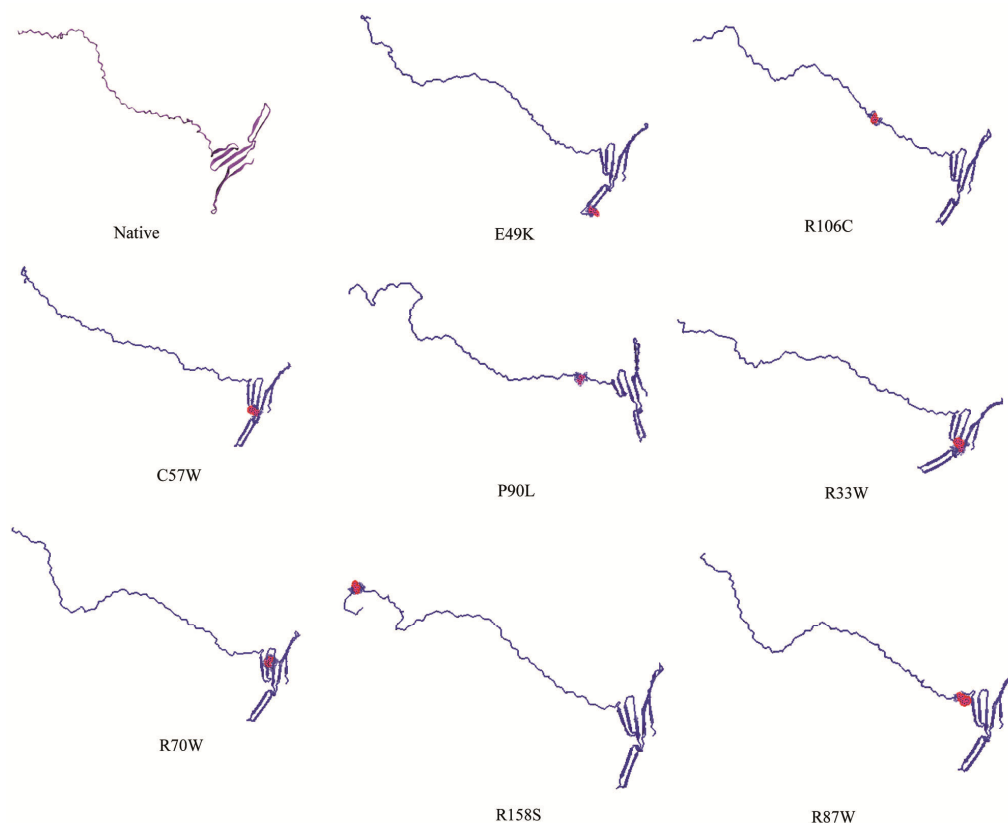


Figure 4. Native and mutant protein validation. Red colour mesh is mutant site. 3D structures are visualized using PyMol tool.

Table 3. Results of Ramachandran plot analysis and TM-ALIGN score ($0.0 < \text{TM score} < 0.30$ indicate random structural similarity and $0.5 < \text{TM score} < 1.00$ in about the same fold)

Model	Ramachandran plot		TM ALIGN
	Most favoured regions	Additional allowed regions	TM score
NATIVE	95.2	4.1	–
C57W	95.2	4.8	0.45964
E49K	93.1	6.2	0.89954
P90L	97.3	2.7	0.42126
R33W	93.1	6.2	0.44366
R70W	94.5	4.8	0.56271
R87W	93.1	5.5	0.55415
R106C	95.9	4.1	0.44723
R158S	94.5	4.8	0.55868

Table 4. Data generated by PDBsum and PROCHECK: structural variation assessment of native T-cap and mutant proteins. Native, P90L and R33W proteins show the same structural similarities, as is also the case of R70W and R87W

Molecule	Beta sheet	Beta hairpin	Beta bulges	Strands	Beta turns	Gamma turns	Helix
NATIVE	1	4	5	5	9	1	0
Mutant C57W	2	4	4	7	9	0	1
Mutant E49K	1	4	5	5	9	2	0
Mutant P90L	1	4	5	5	9	1	0
Mutant R33W	1	4	5	5	9	1	0
Mutant R70W	2	4	4	7	8	0	1
Mutant R87W	2	4	4	7	8	0	1
Mutant R106C	1	4	5	5	8	0	0
Mutant R158S	2	4	4	7	8	0	1

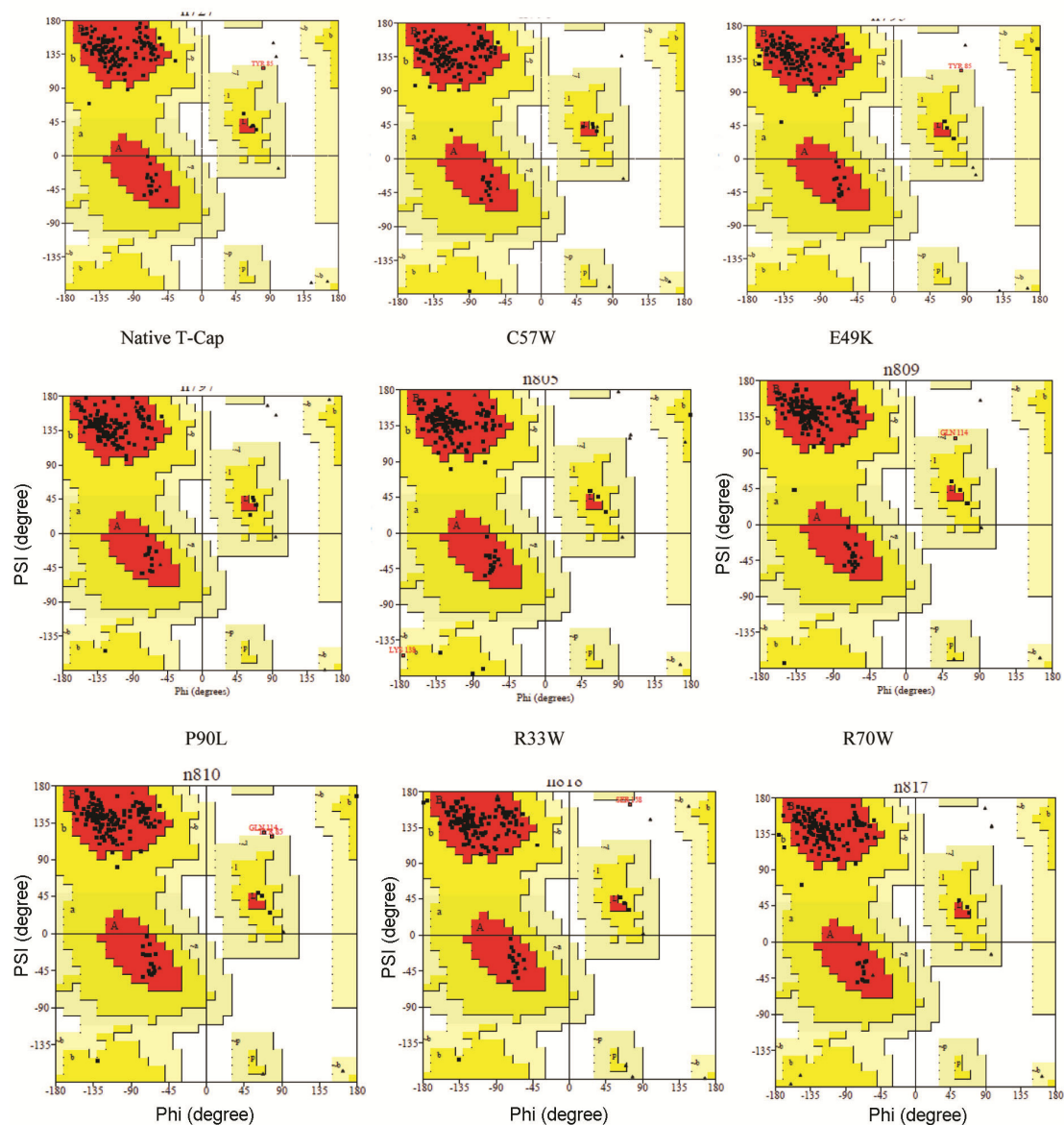


Figure 5. Ramachandran plot for native T-cap and mutant proteins. All the mutant and native proteins and most of residues occur in core regions.

Table 5. RMSD, energy minimization, protein stability assessment of native and mutant T-cap proteins

Protein	Variant ID	Nucleotide variation	Amino acid variation	Total energy after EM (kJ/mol)	RMSD (Å)	I-Mutant results		
						I-Mutant score	RI	Protein stability free energy change value (kcal/mol)
Telethonin-native	****	****	****	-8955.652	0.00	****	****	****
Telethonin-mutant	rs45578741	C-T	R106C	-8716.647	5.46	Increase	1	0.44
Telethonin-mutant	rs145524909	C-T	R33W	-8775.802	4.08	Decrease	5	-0.33
Telethonin-mutant	rs397516863	C-A	R158S	-8804.438	4.59	Decrease	7	-1.31
Telethonin-mutant	rs727504427	C-T	P90L	-9078.910	2.68	Decrease	6	-0.85
Telethonin-mutant	rs775636212	C-T	R70W	-8954.804	4.40	Decrease	4	0.20
Telethonin-mutant	rs777518512	C-T	R87W	-8848.714	4.29	Decrease	6	-0.82
Telethonin-mutant	rs45513698	G-A	E49K	-9125.286	1.90	Decrease	3	-0.07
Telethonin-mutant	rs369447207	C-G	C57W	-9001.259	4.99	Increase	1	0.84

EM, Energy minimization; RI, Reliability index.

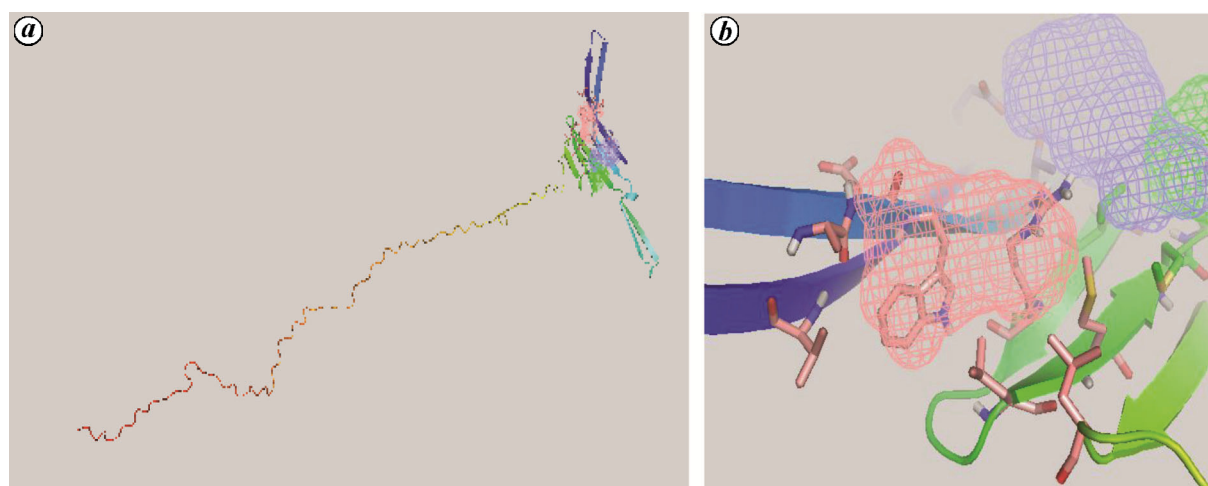


Figure 6. *a*, Ligand-binding site generated by FT-site sever depicted in blue, green and pink mesh-like structures. *b*, Enlarged view of ligand-binding site.

Table 6. ASA of native and mutant proteins. The total ASA of R70W, R87W and R158S is less than that of native T-cap protein, while for other mutants it is higher than that of native proteins. The percentage of side ASA hydrophobic is higher than native protein

Model	Total ASA (Å ²)	Exposed nonpolar ASA (Å ²)	Exposed polar ASA (Å ²)	Exposed charged ASA (Å ²)	% Side ASA hydrophobic
Native	18130.7	11029.6	3933.4	3167.7	37.83
C57W	18295.4	11199.5	3933.7	3162.2	38.47
E49K	18319.5	11154.8	3963.5	3201.2	38.11
P90L	18303.3	11047.9	4225.0	3030.4	38.72
R33W	18313.7	10992.3	4195.8	3125.6	38.57
R70W	18094.3	11039.5	4021.8	3033.1	38.26
R87W	18091.9	11072.3	4029.7	2990.0	39.99
R106C	18285.5	11086.3	4110.7	3088.6	38.61
R158S	18060.2	10994.0	4000.8	3065.4	38.05

Table 7. FT-site, ligand-binding site score of Telethonin. The predicted variants were not present in the ligand-binding sites. So the variants do not affect the ligand-binding affinity of the protein

Binding site 1	Binding site 2	Binding site 3
VAL × 10	VAL × 59	THR × 3
ALA × 23	LEU × 60	GLU × 5
GLU × 24	VAL × 61	THR × 29
TRP × 25	MET × 71	LEU × 30
ARG × 63	GLY × 72	SER × 31
LEU × 67	ILE × 73	VAL × 59
MET × 69		LEU × 60
LEU × 83		VAL × 61
		ARG × 63
		MET × 69
		MET × 71

experimental evidence that ligand-binding sites also bind small organic molecules of various shapes and polarities. The FT-site server depicted three ligand sites in T-cap and all the predicted variants were not present in the binding sites. So the variants do not affect the ligand-binding affinity of the protein (Table 7 and Figure 6 *a* and *b*).

Discussion

Single nucleic acid variations caused by SNPs are a major cause of variations in populations³¹. The missense SNPs are associated with single amino acid alteration in the coding regions of a gene that may have a drastic effect on the structural and functional properties of the corresponding protein¹. The fundamental objectives of medical genetics research is the identification of SNPs in the coding region of a gene. Several studies have attempted to predict the functional effects of SNPs, namely whether they are deleterious or neutral. The present study mainly focuses on the effect of missense SNPs on T-cap protein. The computational analysis of 45 missense SNPs, showed that eight of them had damaging and deleterious effects on T-cap protein. According to the Human Gene Mutation Database, missense mutations account for almost half of all DNA mutations that are known to cause genetic diseases.

According to ConSurf results, three missense SNPs showed deleterious mutations in the conserved region of the genome²⁹. Evolutionarily, a sequence in the genome

remains conserved only when any mutation in the region is eliminated by natural selection. Normally conserved region contains sequences that code for amino acids which are integral for the functioning of the protein²¹. So, occurrence of the three missense SNPs in the conserved region of the genome would indicate that they are deleterious¹.

The structural effects of mutation analysis clearly show that mutant and wild residues differ in size and charge¹². This affects the interaction between other sarcomeric muscle proteins, muscle assembly and their coordination⁸. E49K mutation alters the salt bridge with arginine at position 47. E49K and R33W mutants disturb the multimeric interactions of the protein. P90L mutation can disturb the special backbone conformation of protein folding. R70W and R106C disturb the hydrogen bond formation. The overall analysis of structural effects of protein shows that the mutant residue affects the protein properties like hydrophobicity, ionic interactions, hydrogen bond and salt bridge formation, multimeric interactions, protein folding and protein function.

Models for the mutated proteins depict varying energies compared to native protein structure. The native protein energy after energy minimization is -8955.652 kJ/mol. Total energy after energy minimization of mutant protein compared with native protein. Some mutant proteins like E49k (-9125.286 kJ/mol), P90L (-9078 kJ/mol) and C57W (-9001.259 kJ/mol) show higher values than other mutants R70W (-8954.804 kJ/mol), R87W (-8848.714 kJ/mol), R158S (-8804.438 kJ/mol), R33W (-8775.802 kJ/mol) and R106C (-8716.647 kJ/mol). The varying energy indicates damage to protein structure leading to abnormal protein function. VADAR server was used to predict ASA, which indicates solvent accessibility of amino acid residues. The total ASA of R70W, R87W and R158S is less than native protein and percentage side ASA hydrophobic of variant model is higher than native protein. This variation indicates the structural deviation of protein and loss of interaction of protein with other sarcomeric proteins^{32,33}.

The differences in topological similarities and structural differences of native and mutant proteins validated by RMSD and TM-score clearly show that the mutant proteins have structural deviation compared to native proteins. This will alter the normal protein function and may cause deleterious effects; some of the mutants are already reported as deleterious¹. If mutation occurs, the stability of protein in the case R106C and C57W increases while in other mutant proteins it decreases. In topological analysis, if mutation occurs on P90L and R33W, it shows correct structural similarity to native protein. In this case there is one beta sheet, four beta hairpins, five beta bulges, five strands, nine beta turns and one gamma turn. In case of P90L, which is already reported as associated with human diseases, mutation occurs in the highly conserved region^{1,32,33}. The

deleterious mutant R33W may show the same effect of P90L because both have the same structural arrangement. R70W and R87W also show the same structural arrangement. In FT-site server analysis, the ligand-binding affinity of protein shows no deleterious mutant located on the ligand-binding site of the protein. It indicates that, the mutation does not affect the ligand binding capacity of protein. But it will alter the multimeric interactions.

Conclusion

In silico analysis of various characteristic features of the protein variant models confirmed their damaging effect on protein structure and function. These mutations can affect the normal functioning of telethonin. The deleterious effects of the five predicted missense SNPs (Arg33Trp, Glu49Lys, Cys57Trp, Arg158Ser and Arg87Trp) are not described in association with genetic diseases in the literature and three deleterious missense SNPs (Arg70Trp, Pro90Leu and Arg106Cys) have already been reported as deleterious. The data available from this study may help in further *in vitro* and population genetics studies to determine the impact of these variations on Telethonin structure and function. A computational study not only reduces the need for time-consuming experimental analysis, but also predicts the effects of mutation on a protein at molecular level. Knowledge of structural and functional impacts of point mutations is significant for the design of drugs against diseases. Structural modelling will lead to advances in rational drug design, if it is of sufficient accuracy, to reflect structural differences among drug-selective receptor subtypes. To achieve this level of accuracy, experimental studies as well as computational modelling are essential.

Declaration: In this study all missense SNPs of human Telethonin protein were retrieved from NCBI GenBank only.

1. Miller, M. P. and Kumar, S., Understanding human disease mutation through the use of interspecific genetic variation. *Hum. Mol. Genet.*, 2001, **10**, 2319, 2328.
2. Fulkner, G., Lanfranchi, G. and Valle, G., Telethonin and other new proteins of the Z disc of skeletal muscle. *IUBMB Life*, 2001, **51**, 275–282.
3. Kojic, S. *et al.*, The Ankrd2 protein, a link between the sarcomere and the nucleus in skeletal muscle. *J. Mol. Biol.*, 2004, **339**, 313–325.
4. Faulkner, G. *et al.*, FATZ. A Filamin, actinin and telethonin binding protein of the Z-disc of skeletal muscle. *J. Biol. Chem.*, 2000, **275**, 41234–41242.
5. Frey, N. and Olson, E. N., Calsarcin-3, a novel skeletal muscle specific member of the calsarcin family, interacts with multiple Z-disc proteins. *J. Biol. Chem.*, 2002, **277**, 13998–14004.
6. Kleywegt, G. J. and Jones, T. A., Phi/Psi-chology: Ramachandran revisited. *Structure*, 1996, **4**, 1395–1400.

7. Furukawa, T. *et al.*, Specific interaction of the potassium channel beta-subunit with the sarcomeric protein T-cap suggests a T-tubule myofibril linking system. *J. Mol. Biol.*, 2001, **313**, 775–784.
8. Mayans, O. *et al.*, Structural basis for activation of the titin kinase domain during myofibrillogenesis. *Nature*, 1998, **395**, 863–869.
9. Moreira, E. S. *et al.*, Limb girdle muscular dystrophy type 2G is caused by mutations in the gene encoding the sarcomeric protein telethonin. *Nature Genet.*, 2000, **24**, 163–166.
10. Hayashi, T., Arimura, T. and Itohsatoh, M., T-cap gene mutations in hypertrophic cardiomyopathy and dilated cardiomyopathy. *J. Am. Coll. Cardiol.*, 2004, **44**, 2192–2201.
11. Knoll, R. *et al.*, The cardiac mechanical stretch sensor machinery involves A Z disc complex that is defective in a subset of human dilated cardiomyopathy. *Cell*, 2002, **111**, 943–955.
12. Kumar, P., Henikoff, S. and Ng, P. C., Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature Protoc.*, 2009, **4**, 1073–1081.
13. Choi, Y. *et al.*, Predicting the functional effect of amino acid substitutions and indels. *PLOS ONE*, 2012, **7**, e46688; doi:10.1371/journal.pone.0046688.
14. Mi, H. *et al.*, The PANTHER database of protein families, subfamilies, functions and pathways. *Nucleic Acids*, 2005, **33**.
15. Bromberg, Y., Yachdav, G. and Rost, B., SNAP predicts effect of mutations on protein function. *Bioinformatics*, 2008, **24**, 2397–2398.
16. Pires, A. S. *et al.*, *In Silico* analyses of deleterious missense SNPs of human apolipoprotein E3. *Nature Sci. Rep.*, 2017, **7**, 2509.
17. Adzhubei, I. A. *et al.*, A method and server for predicting damaging missense mutations. *Nature Methods*, 2010, **7**, 248–249.
18. Capriotti, E., Calabrese, R. and Casadio, R., Predicting the insurgenence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. *Bioinformatics*, 2006, **22**, 2729–2734.
19. Bendl, J. *et al.*, Predict SNP: robust and accurate consensus classifier for prediction of disease-related mutations. *PLOS Comput. Biol.*, 2014, **10**, e1003440.
20. Niroula, A., Urolagin, S. and Vihinen, M., PON-P2: prediction method for fast and reliable identification of harmful variants. *PLOS ONE*, 2015, **10**(2), e0117380.
21. Askenazy, H. *et al.*, ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res.*, 2010, **38**, W529–W533; doi:10.1093/nar/gkq399.
22. Haworth, R. S. *et al.*, Protein kinase D is a novel mediator of cardiac troponin 1 phosphorylation and regulates myofilament function. *Circ. Res.*, 2004, **95**, 1091–1099.
23. Sitao, W. and Zhang, Y., MUSTER: improving protein sequence profile–profile alignments by using multiple sources of structure information. *Proteins*, 2008, **72**(2), 547–556.
24. Laskowski, R. A. *et al.*, PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.*, 1993, **26**, 283–291.
25. Laskowski, R. A., Chistyakov, V. V. and Thornton, J. M., PDBsum more: new summaries and analyses of the known 3D structures of proteins and nucleic acids. *Nucleic Acids Res.*, 2005, **33**, 266–268.
26. Lindahl, E. *et al.*, NOMAD-Ref: visualization, deformation and refinement of macromolecular structures based on all atom normal mode analysis. *Nucleic Acids Res.*, 2006, **34**, 52–60.
27. Gopalakrishnan, K. *et al.*, Ramachandran plot on the web (2.0). *Protein Pept. Lett.*, 2007, **14**, 669–671.
28. Angermuller, C. *et al.*, Discriminating modelling of context-specific amino acid substitution probabilities. *Bioinformatics*, 2012, **28**, 3240–3247.
29. Celniker, G. *et al.*, ConSurf: using evolutionary data to raise testable hypothesis about protein function. *Isr. J. Chem.*, 2013, **53**, 199–206.
30. Sherry, S. T. *et al.*, dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.*, 2001, **29**, 308–311.
31. Porto, W. F., Franco, O. L. and Alencar, S. A., Computational analysis and prediction of guanlylin deleterious SNPs. *Peptides*, 2015, **69**, 92–102.
32. Mazzone, A. *et al.*, A mutation in Telethonin alters Na_v1.5 function. *J. Biol. Chem.*, 2008, **283**, 16537–16544.
33. Francis, A. *et al.*, Novel TCAP mutation C.32C >A causing limb girdle muscular dystrophy 2G. *PLOS ONE*, 2014, **9**(7), e102764.
34. Mues, A. *et al.*, Two immunoglobulin like domains of the Z-disc portion of titin interact in a conformation dependent way with telethonin. *FEBS Lett.*, 1998, **428**, 111–114.
35. Arnold, K. *et al.*, The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics*, 2006, **22**, 195–201.
36. Gopalakrishnan, C. *et al.*, Computational modelling of complete HOXB13 protein for predicting the functional effect of SNPs and the associated role in hereditary prostate cancer. *Nature Sci. Rep.*, 2017, **7**, 43830.
37. Valle, G. *et al.*, Telethonin, a novel sarcomeric protein of heart and skeletal muscle. *FEBS Lett.*, 1997, **415**, 163–168.
38. Li, W. and Godzik, A., Cd hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, 2006, **22**, 1658–1659.
39. Nicholas, G. *et al.*, Titin-cap associates with and regulates secretion of myostatin. *J. Cell Physiol.*, 2002, **193**, 120–131.
40. Nikos, P. *et al.*, Evidence for a dimeric assembly of titin/telethonin complexes induced by the telethonin C-terminus. *J. Struct. Biol.*, 2006, **155**, 239–250.
41. Suzek, B. E. *et al.*, Uniref: comprehensive and non-redundant UniProt reference clusters. *Bioinformatics*, 2007, **23**, 1282–1288.
42. Dakal, T. C., Kala, D. and Dhiman, G., Predicting the functional consequences of non-synonymous single nucleotide polymorphism in IL8 gene. *Nature. Sci. Rep.*, 2017, **7**(1), 6525.
43. Van Oeveren, J. and Janssen, A., A single nucleotide polymorphisms. *Single Nucleotide Polymorph.*, 2009, **578**, 73–91.
44. Yue, P. and Moul, J., Identification and analysis of deleterious human SNPs. *J. Mol. Biol.*, 2006, **356**, 1263–1274.

Received 10 December 2018; revised accepted 6 April 2019

doi: 10.18520/cs/v117/i4/638-648