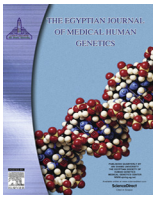
The Egyptian Journal of Medical Human Genetics 19 (2018) 333–344

HOSTED BY

Contents lists available at ScienceDirect

The Egyptian Journal of Medical Human Genetics

journal homepage: www.sciencedirect.com



Original article

Pathogenic predictions of non-synonymous variants and their impacts: A computational assessment of *ARHGEF6* gene

Yashvant M. Khimsuriya, Jenabhai B. Chauhan *

Post Graduate Department of Genetics, Ashok & Rita Patel Institute of Integrated Study & Research in Biotechnology & Allied Sciences (ARIBAS), New Vallabh Vidyanagar, Anand, India

ARTICLE INFO

Article history:
Received 21 March 2018
Accepted 8 May 2018
Available online 23 August 2018

Keywords:
Computational methods
ARHGEF6
Intellectual disability
Missense mutation

ABSTRACT

Introduction: ARHGEF6, a key member and activator of RhoGTPases family that is involved in G-Protein Coupled receptor (GPCR) pathway and stimulate Rho dependent signals in the brain, and mutations in this gene can cause intellectual disability (ID) in Human. Therefore, we aimed to study the consequences of *ARHGEF6* non-synonymous mutations by using advanced computational methods.
Methods: Classification of the genetic mutations in *ARHGEF6* gene was performed according to Ensembl Genome Database and data mining was done using ensemble tools. The functional and disease effect of missense mutations, and pathogenic characteristics of amino acid substitutions of *ARHGEF6* were analyzed using eleven diversified computational tools and servers.
Results: Overall, 47 *ARHGEF6* non-synonymous (NS) variants were predicted to be deleterious by SIFT, Polyphen2 and PROVEAN scores. Above that, SNPs&GO and PhD SNP were further graded 21 customarily pathogenic NS-variants. Protein stability analysis resulted in the significant change in terms of $\Delta\Delta G$ of most identified NS-variants, except K609I. Seven variants were analyzed to be located on most potential domain RhoGEF/DH, whereas the remaining 14 were distributed on CH, SH3, PH and BP domains. Furthermore, pathogenic effects of mutations on protein was presented with different parameters using MutPred2 and PROJECT HOPE. Additionally, STRING network data predicted GIT2 and PARVB as most interacted partners of ARHGEF6.
Conclusion: These findings can be supportive of genotype-phenotype research as well as the development in pharmacogenetics studies. Finally, this study revealed a significance of computational methods to figure out highly pathogenic genomic variants linked with the structural and functional relationship of ARHGEF6 protein.

1. Introduction

The Rho-guanine nucleotide exchange factor 6 (ARHGEF6) protein is known for its involvement in the Rho GTPase cycle, which mediates the organization of cytoskeleton, cell shape, and motility. It is identified as third responsible X-linked intellectual disability (XLID) gene, after Oligophrenin 1 (*OPHN1*) and P21-protein activated kinases 3 (*PAK3*) [1]. It is also known as PAK-interacting exchange factor, alpha (*αPIX*) and *COOL2*. ARHGEF6 is 87.5 kDa protein of 776 amino acids, which belongs to a family of cytoplasmic proteins (RhoGTPases) that activate the Ras-like family of Rho proteins by exchanging bound GDP for GTP. ARHGEF6 in complex with BIN2 and GIT2, forms a complex with G-proteins and stimulate Rho-dependent signals [2]. It also acts as a Ras-related C3 botulinum toxin substrate 1 (RAC1) guanine nucleotide exchange factor. The RhoGTPases are critical regulators of the actin cytoskeleton, where they often mediate signaling from the external environment. In the central nervous system, their function has been linked to axonal growth, development of dendritic arborizations and spine morphogenesis [3,4].

As an activator protein, ARHGEF6 plays a significant role in the cellular mechanisms of Rho-GTPases. The biological mechanisms through which *ARHGEF6* mutations causes the intellectual disability are still not well recognized, although defective plasticity of synaptic networks have been previously proposed. However, several studies reported that this protein is primarily expressed in neuropil regions of the hippocampus and the deregulations can alter neuronal connectivity and impaired synaptic function and cognition [5]. ARHGEF6 located in dendritic spines regulate spine

Peer review under responsibility of Ain Shams University.
* Corresponding author at: Associate Professor & Head, Post Graduate Department of Genetics, Ashok & Rita Patel Institute of Integrated Study & Research in Biotechnology & Allied Sciences (ARIBAS), New Vallabh Vidyanagar, Anand, Gujarat 388121, India. Fax & Tel.: +91 2692 229 189.
E-mail address: jenabhaichauhan@aribas.edu.in (J.B. Chauhan).

https://doi.org/10.1016/j.ejmhg.2018.05.002
1110-8630/

morphogenesis by acting through downstream activation of p21-activated kinase (PAK3). The mutations in both of these genes could induce intellectual disability [6]. A recent study reported that αPIX promotes dendritic Golgi translocation in hippocampal neurons [7].

The human *ARHGEF6* gene, spanning over 22 exons is located on X-chromosome at sub-band q26.3. In humans, rare and common genomic mutations of *ARHGEF6* are constantly being diagnosed with the help of conventional as well as high throughput sequencing technologies [8]. But, the evaluation and correlation of these broad spectrum of clinical phenotypes and their connection with molecular alterations of *ARHGEF6* gene are not yet well examined.

The disease causing mutations may usually affect the size, charge and hydrophobicity value of each encoded amino acid variant, which can successively change the hydrogen bonding and conformational dynamics of the protein. Accordingly, the ability to better interpret the clinical complications of every mutation depends on identifying the real constructive pathogenic mutations from the correlated markers. Despite the fact that molecular validation of such mutations by *in vitro* and *in vivo* studies is more time consuming, and often requires technical expertise and huge expenses. The alternate approach to defeat this challenge is to examine the impact of each genetic variation using a recently developed advanced computational algorithms approaches. Different types of bioinformatics programs and servers have been designed to discover the consequences of genetic mutations on biophysical characteristics, structure and functional properties of proteins [9].

Therefore, we aimed this study to analyze pathogenic variants of *ARHGEF6* gene in exonic positions and to predict the structural and functional implications of ARHGEF6 protein by subjecting the gene sequences along with non-synonymous mutations to the various computational methods.

2. Methods and datasets

The Single Nucleotide Variants (SNVs) and protein sequence of the *ARHGEF6* gene (transcript ID:ENST00000250617.6) were obtained from NCBI dbSNP available at http://www.ncbi.nlm.nih.gov/SNP/ [10], NCBI protein (https://www.ncbi.nlm.nih.gov/protein/) [11] and Ensembl genome browser (https://asia.ensembl.org/index.html) [12]. The classification of all collected SNVs was done as non-coding and coding, depending on the variant nature and position. Only missense variants (non-synonymous) were chosen for further computational analysis because of their potential to disturb the structural conformation of proteins.

2.1. Functional prediction of missense variants

Damaging and deleterious effect of missense variants were predicted using the scores Sorting Intolerant from Tolerant (SIFT) (http://sift.jcvi.org), Polymorphism Phenotyping v2 (PolyPhen-2) (http://genetics.bwh.harvard.edu/pph2/) and Protein Variation Effect Analyzer (PROVEAN) (http://provean.jcvi.org/index.php) tools. SIFT is a sequence homology based tool which predicts the tolerated and deleterious SNVs and identifies the impact of amino acid substitution on protein functions. The results can be deleterious or tolerated substitutions demonstrating threshold ≤ 0.05 score [13]. Polyphen2 is a sequence and structure evolutionary conservation based tool to classify damaging effect of amino acid substitutions and estimates position specific independent count (PSIC) score demonstrating 0.801–1.00 probably damaging index [14]. PROVEAN is a software to obtain pairwise sequence alignment (PSA) score and to identify non-synonymous variants [15]. Furthermore, we used Single Nucleotide Polymorphisms & Gene Ontology (SNPs&GO) (http://snps.biofold.org/snps-and-go/snps-and-go.html) and Predictor of human deleterious single nucleotide polymorphisms (PhD-SNP) (http://snps.biofold.org/phd13 9snp/phd-snp.html) those are Support Vector Machine (SVM)-based tools that used evolutionary information, protein sequence and functions to predict if a given mutation can be classified as disease-related or neutral [16,17].

2.2. Structural conformation and conservation analysis

The Consurf server available at http://consurf.tau.ac.il/, was used for high-throughput characterization and evolutionary conservation of amino acid positions based on the phylogenetic relationship between homologous sequences [18]. The degree of conservation of the amino-acid sites among 50 homologs with similar sequences was estimated. The conservation grades were then projected onto the molecular surface of the human *ARHGEF6* to reveal the stripes with highly conserved residues that are usually essential for biological function.

2.3. Prediction of disease related amino acid substitution by MutPred2

The MutPred2 (http://mutpred.mutdb.org/), a unique web based tool was developed to predict any amino acid substitution, whether pathogenic or benign [19]. Based on>50 different protein properties we can classify the inference of molecular mechanisms of pathogenicity. It uses SIFT, PSI-BLAST [20], and Pfam profiles [21] along with some structural disorder prediction algorithms, TMHMM [22], MARCOIL [23], and DisProt [24]. Random Forest (RF) classifier was used and obtained g-score for prediction of the probability and the p score for identification of structural and functional properties. As a result, by combining the scores of all programs, the accuracy of prediction ascend to a greater extent.

2.4. Structural analysis of ARHGEF6 protein and mutants

2.4.1. Protein structure prediction and modeling

To succeed in dealing with the absence of crystal protein structure in databases, ARHGEF6 protein structure was built after subjecting the referenced amino acids sequence (NP_004831) to the I-TASSER (Iterative Threading ASSEmbly Refinement), a web based server available at https://zhanglab.ccmb.med.umich.edu/I-TASSER/. It uses basic templates from the PDB by multiple threading approaches and constructs full-length atomic models by iterative template fragment assembly simulations [25]. Then, it has predicted five top models, among which one best model was identified on the basis of confidence score, estimated TM-score and estimated root mean square deviation (RMSD) value. Further, the similar standards was also analyzed for structural deviation prediction for the c-alpha atoms of each amino acid residue in the mutant models. The selected model was eventually subjected for Gromacs energy minimization by Normal Mode Analysis Deformation and Refinement (NOMAD-Ref) Server available at http://lorentz.immstr.pasteur.fr/nomad-Ref.php, to remove disarrangement in the space of a collection of atoms [26]. This energy minimized model was used as a standard template to construct mutant models of ARHGEF6 (manually inserted mutated residues in the referenced protein sequence of *ARHGEF6*) by Modeller v9.19. This software applies comparative protein structure modeling by satisfaction of spatial restraints in the protein of interest. Likewise, RAMPAGE server (http://mordred.bioc.cam.ac.uk/~rapper/rampage.php) was used to check stereo-chemical properties of ARHGEF6 wild type models [27]. PyMol and Chimera programs were used to generate mutated models and visualize interactions of molecules [28].

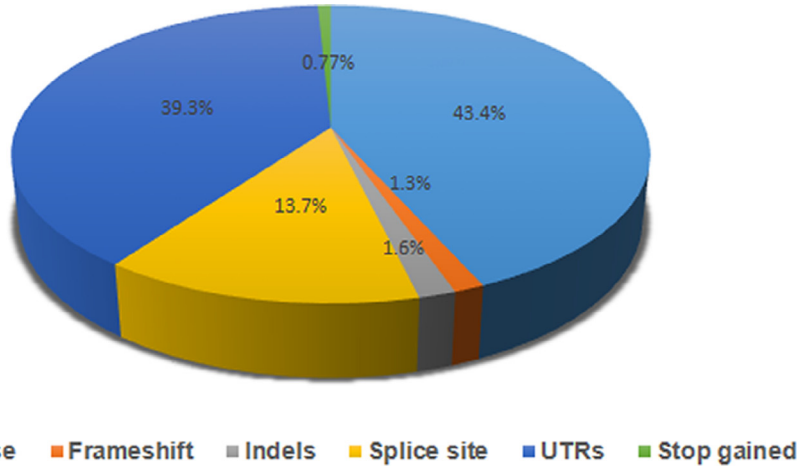


Fig. 1. Distribution of nucleotide variants of *ARHGEF6* gene; Types of variants identified with different color.

Table 1
SIFT, PolyPhen2, PROVEAN, SNP&GO and PhD-SNP prediction for ARHGEF6 mutations.

Sr. No.	Variant ID	Mutation	Functional analysis			Disease relevance	
			SIFT	Polyphen2	Provean	SNP&GO	PhD-SNP
1	rs35747426	L11P	0	1	−2.885	Disease	Disease
2	rs373683437	P27S	0	0.99	−3.109	Neutral	Neutral
3	rs773538676	I164R	0	0.51	−3.872	Disease	Neutral
4	rs748608017	N170K	0.02	0.75	−2.87	Disease	Disease
5	rs767986900	L179V	0	0.99	−2.825	Disease	Disease
6	rs755769516	D185N	0	0.99	−4.408	Disease	Disease
7	rs757130619	E194K	0	0.95	−3.418	Neutral	Disease
8	rs182216332	E194G	0	0.95	−5.488	Neutral	Disease
9	rs764121106	R205S	0	0.59	−3.119	Neutral	Neutral
10	rs139422484	E216G	0	1	−6.49	Disease	Disease
11	rs769209490	S225C	0	0.98	−3.903	Disease	Neutral
12	rs745458792	P226A	0.01	0.46	−3.584	Disease	Neutral
13	rs780994752	P226L	0	0.96	−4.736	Disease	Neutral
14	rs750580186	Y241C	0	0.99	−7.006	Disease	Disease
15	rs868099426	S278G	0.02	0.61	−3.264	Neutral	Neutral
16	rs776248155	E290K	0.01	0.95	−2.815	Neutral	Disease
17	rs187104213	E305G	0.02	0.95	−5.492	Neutral	Neutral
18	rs767066725	Y334N	0	1	−8.853	Disease	Disease
19	rs747044748	F355L	0.04	0.97	−5.18	Disease	Disease
20	rs777956624	G360V	0	1	−8.768	Disease	Disease
21	rs779277722	A361T	0	0.98	−3.853	Neutral	Neutral
22	rs755410840	S363R	0	0.69	−3.534	Neutral	Neutral
23	rs375550477	T371K	0	0.81	−5.228	Neutral	Disease
24	rs752815354	R379Q	0	0.99	−3.885	Disease	Disease
25	rs756705199	R392W	0	1	−7.872	Disease	Disease
26	rs778051195	H393R	0	0.99	−7.305	Neutral	Disease
27	rs375664814	C418Y	0	1	−10.282	Disease	Disease
28	rs776639562	W439L	0	0.9	−12.205	Neutral	Disease
29	rs761952185	I444N	0	0.93	−5.752	Disease	Disease
30	rs371126822	L447F	0	0.78	−3.835	Neutral	Neutral
31	rs147323188	R469W	0	1	−7.012	Disease	Disease
32	rs201896882	R469Q	0.02	0.99	−3.556	Disease	Disease
33	rs755356810	Y470S	0	0.99	−8.585	Neutral	Neutral
34	rs760285231	R486Q	0	0.99	−3.656	Neutral	Disease
35	rs771303937	Y492C	0	0.99	−8.263	Disease	Disease
36	rs757308541	K495E	0	0.79	−3.41	Disease	Disease
37	rs757220742	E524G	0	0.65	−4.76	Neutral	Neutral
38	rs757220742	E524V	0	0.64	−5.636	Disease	Disease
39	rs199878133	C530F	0	0.99	−8.99	Neutral	Disease
40	rs770679754	P588L	0	0.80	−4.018	Disease	Neutral
41	rs771876538	P596T	0	1	−4.958	Disease	Neutral
42	rs748119122	P598S	0.02	0.99	−4.205	Neutral	Neutral
43	rs756204695	K609I	0	0.99	−3.976	Disease	Disease
44	rs765195778	S684F	0.02	0.75	−2.757	Disease	Neutral
45	rs376499517	P686T	0.04	0.57	−2.749	Disease	Neutral
46	rs746586184	L714R	0	0.98	−3.257	Disease	Disease
47	rs144205542	D716A	0	0.96	−4.345	Disease	Disease

(SIFT Prediction: Deleterious score ≤0.05; Tolerated Score >0.05, PolyPhen2 Prediction: Benign = 0.000–0.004; Possibly Damaging = 0.401–0.800; Probably Damaging = 0.801–1.000, PROVEAN Prediction: Deleterious or Neutral score cutoff = −2.5; SNP&GO and PhD-SNP directly predicts disease effect. Commonly found disease related variants highlighted with red color).

2.4.2. Protein stability prediction

I-Mutant 2.0 (http://folding.biofold.org/i-mutant/i163mutant-t2.0.html), a neural network based tool, predicts the change in the stability of the protein upon mutation [29]. This program consequently predicts protein strength changes upon single site transformations. Prediction can be performed by utilizing either protein structure or sequence. The amino acids sequence (FASTA file) of ARHGEF6 (NP_004831) retrieved from NCBI, was used as an input to predict the mutational effect on protein stability. The output is obtained in the form of protein stability change upon mutation and Gibbs-free energy change (ΔΔG/DDG).

DUET server (http://biosig.unimelb.edu.au/duet/), an integrated computational method was used to predict the change in the stability of ARHGEF6 protein at 3D structure levels. The PDB file of ARHGEF6 protein structure with a chain identifier and missense mutation information such as wild-type and mutant residues codes in the one-letter format was submitted as an input to this server. It calculates the combined/consensus predictions of mCSM (mutation Cutoff Scanning Matrix) and SDM (Site Directed Mutator) methods in a nonlinear regression fashion using Support Vector Machines (SVMs). The output is in the form of change in Gibbs free energy (DDG), and negative values denote destabilizing mutations [30].

2.5. Structural and functional analysis of ARHGEF6 mutant models

2.5.1. Domain identification

By using Sanger Pfam web server (https://pfam.xfam.org/), the functional domains of ARHGEF6 protein were searched. The native protein sequence and default settings were used for Sanger Pfam domain prediction. The default threshold (E-value) was 1.0 [21].

2.5.2. Solvent stability

The relative surface and solvent accessibility of amino acids substitutions of *ARHGEF6* were calculated using NetSurfP server available at http://www.cbs.dtu.dk/services/NetSurfP/. The input

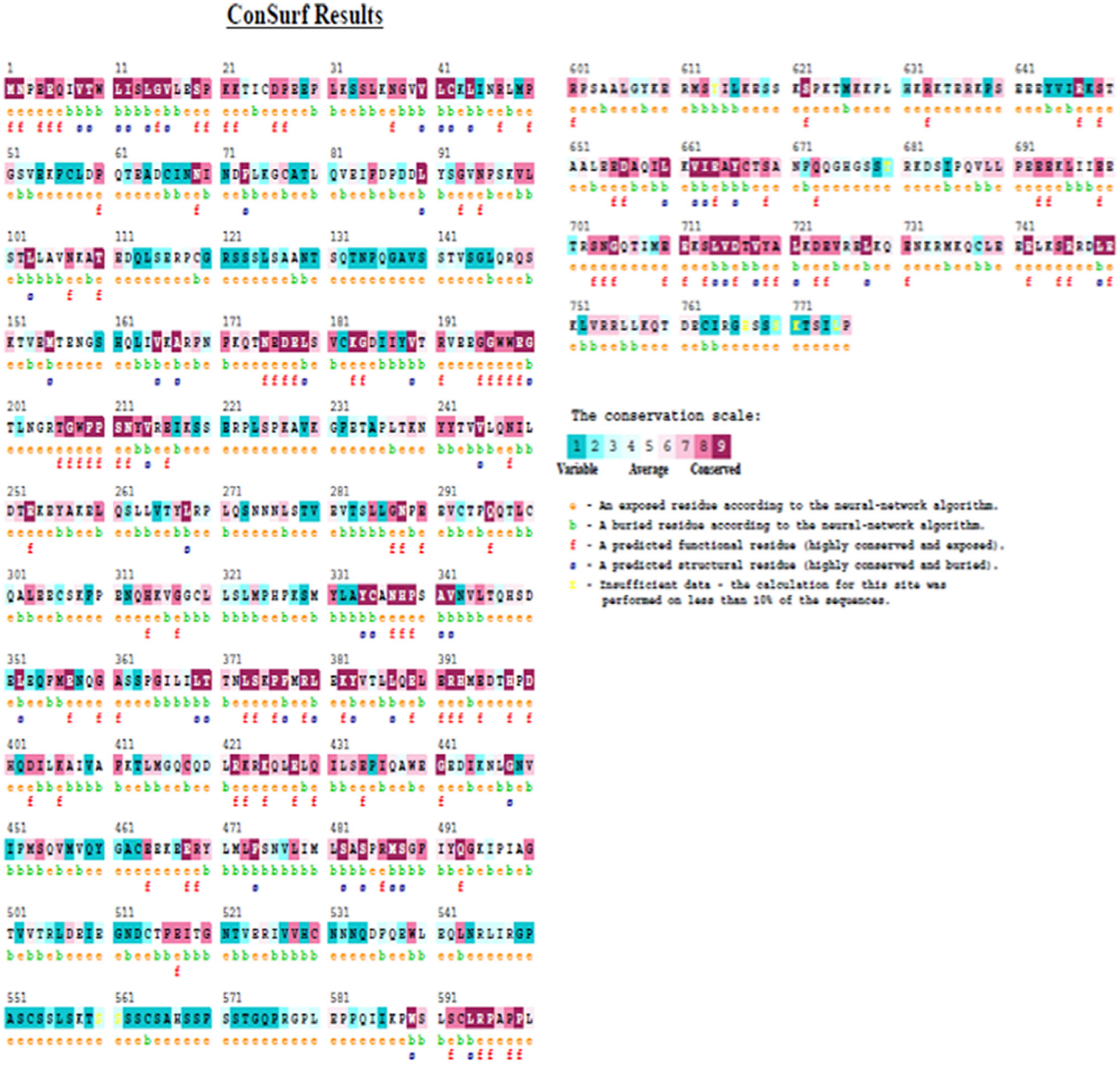


Fig. 2. The Evolutionary Conservation prediction analysis of amino acid residues of ARHGEF6 by Consurf server. Conservation scale represented with color coding bar.

was FASTA format of both native and mutant protein sequences. An artificial neural network method that is trained to predict the relative surface accessibility of a mutation and the reliability of each, in the form of a Z-score [31].

2.5.3. Project HOPE

Project HOPE (Project Have yOur Protein Explained) (http://www.cmbi.ru.nl/hope/home) is a stand-alone automatic mutant analysis service that can provide insight into the structural implications of a mutation regarding molecular features of native and mutant protein [32]. The input was protein sequence in the single letter code along with native and mutant residues, and the results were demonstrated by figures and animations that showed structural impacts of native to mutant type amino acid residue.

2.6. Protein-Protein interactions analysis

The Search Tool for the Retrieval of Interacting Genes (STRING) database search available at https://string-db.org/, is used to show physical and functional interacting partners of ARHGEF6 protein concerning of confidence score (>0.99). The input options consist of ARHGEF6 in Human (protein name + Species name). The output consists of the network of predicted interacting proteins of ARHGEF6 based on stem from computational prediction, from knowledge transfer between organisms, and from interactions aggregated from other databases. Based on confidence (C) score, the highly interacting proteins of ARHGEF6 was differentiated from all interacting proteins network [33].

3. Results

3.1. Mutation spectrum of ARHGEF6 gene

The total of 389 (out of 853 pre-curated) mutations of various classes were found in *ARHGEF6* gene. Among the coding region mutations, missense are seen to be more frequently (n = 169; 43.4%) compared to frameshift (n = 5; 1.3%), indels (n = 6; 1.6%), splice site (n = 53; 13.7%), UTRs (n = 153; 39.3%) and stop gained (n = 3; 0.77%) (Fig. 1).

3.2. Functional identification of deleterious missense mutations

Functional missense mutations score of SIFT uncovered that 66 (39%) mutations are extremely-intolerant (Score 0.00), 34 (20%) are intolerant (Score 0.01–0.05) and 69 (41%) are tolerated (>0.05), implicating that the lower the SIFT score the higher the deleterious impact of that particular non-synonymous mutation on structure and function of *ARHGEF6*. With PolyPhen-2 score, 55 (32.5%) missense mutations were predicted to be probably damaging (Score 0.801–1.000), 35 (20.7%) were possibly damaging (Score 0.401–0.800), and 79 (46.8%) were benign (Score 0.000–0.400). With PROVEAN score, 58 (34.3%) mutations were predicted to be deleterious, and 111 (65.7%) were tolerated (Table 1).

The PhD-SNP 2.0 and SNPs&GO tools classified the mutation as a disease-related or neutral polymorphism. Among the ns-SNPs in the *ARHGEF6* gene analyzed, 26 were predicted to be disease related by PhD-SNP 2.0, and 29 ns-SNPs using SNPs&GO, out of 47 commonly predicted damaging and deleterious by SIFT, PolyPhen, and PROVEAN (Table 1).

3.3. Conservation of amino acid residue

The results generated by the ConSurf tool consist of a structural representation of the protein (Fig. 2) with a colorimetric conservation score. ConSurf identifies functional areas in proteins, taking into account the evolutionary relationships among theirs sequence homologs. As expected, the ConSurf analysis has revealed, that the functional regions of the protein are highly conserved. We observed that variants having different conservational scales include 9 (L11P, L179V, Y334N, R379Q, R392W, L714R, D716A), 8 (E216G, Y241C, G360V, C418Y, I444N, R469W, R469Q), 7 (D185N, Y492C), 6 (F355L, K609I) and 5 (N170K, K495E, E524G).

3.4. Prediction of disease related amino acid substitution by MutPred2

The probable deleterious mutation score (≥0.500) were detected for L11P, N170K, L179V, D185N, E216G, Y241C, Y334N, F355L, G360V, R379Q, R392W, C418Y, I444N, R469W, R469Q, Y492C, K495E, E524G, K609I, L714R, D716A, was 0.950, 0.759, 0.758, 0.722, 0.790, 0.729, 0.917, 0.616, 0.571, 0.623, 0.716, 0.919, 0.746, 0.860, 0.667, 0.838, 0.662, 0.387, 0.514, 0.858, and 0.742, respectively. Except E524G amino acid variant, the

Table 2
MutPred2 server prediction for the effect of ARHGEF6 amino acid substitutions.

Mutation	MutPred2 score	Predicted molecular mechanism (P-value)
L11P	0.950	–
N170K	0.759	–
L179V	0.758	Altered ordered interface (P = 0.04) Altered metal binding(P = 0.03) Loss of relative solvent accessibility(P = 0.03) Gain of catalytic site at E178 (P = 0.03) Altered transmembrane protein (P = 0.03)
D185N	0.722	Altered metal binding (P = 0.03) Altered ordered interface (P = 0.04) Altered transmembrane protein (P = 0.03)
E216G	0.790	Altered stability (P = 0.04) Altered transmembrane protein (P = 0.03)
Y241C	0.729	Altered ordered interface (P = 0.04)
Y334N	0.917	Altered ordered interface (P = 0.03) Altered metal binding (P = 0.02)
F355L	0.616	–
G360V	0.571	–
R379Q	0.623	Altered coiled coil (P = 0.03) Loss of helix (P = 0.04) Loss of acetylation at K382 (P = 0.05)
R392W	0.716	Altered coiled coil (P = 0.04) Loss of intrinsic disorder (P = 0.05) Gain of loop (P = 0.03) Altered metal binding (0.03)
C418Y	0.919	–
I444N	0.746	Gain of B factor (P = 0.03) Altered coiled coil (P = 0.02)
R469W	0.860	Loss of SUMOylation at K466 (P = 0.02) Gain of sulfation at Y470 (P = 0.03)
R469Q	0.667	Loss of SUMOylation at K466 (P = 0.02) Altered stability (P = 0.05) Gain of sulfation at Y470 (P = 0.03)
Y492C	0.838	Loss of strand (P = 0.05)
K495E	0.662	Gain of strand (P = 0.05)
E524V	0.387	–
K609I	0.514	Altered disordered interface (P = 0.02) Loss of phosphorylation at Y614 (P = 0.02) Loss of helix(P = 0.02) Gain of strand(P = 0.02)
L714R	0.858	Altered coiled coil (P = 0.03) Gain of intrinsic disorder (P = 0.01) Gain of phosphorylation at Y719 (P = 0.02) Altered disordered interface (P = 0.04) Gain of SUMOylation at K712 (P = 0.04) Loss of ubiquitylation at K712 (P = 0.04) Loss of proteolytic cleavage at D716 (P = 0.02)
D716A	0.742	Loss of phosphorylation at Y719 (P = 0.02) Altered disordered interface (P = 0.05) Altered coiled coil (P = 0.02) Loss of proteolytic cleavage at D716 (P = 0.01) Loss of ubiquitylation at K712 (P = 0.04)

(Probability of deleterious mutation score ≥ 0.500).

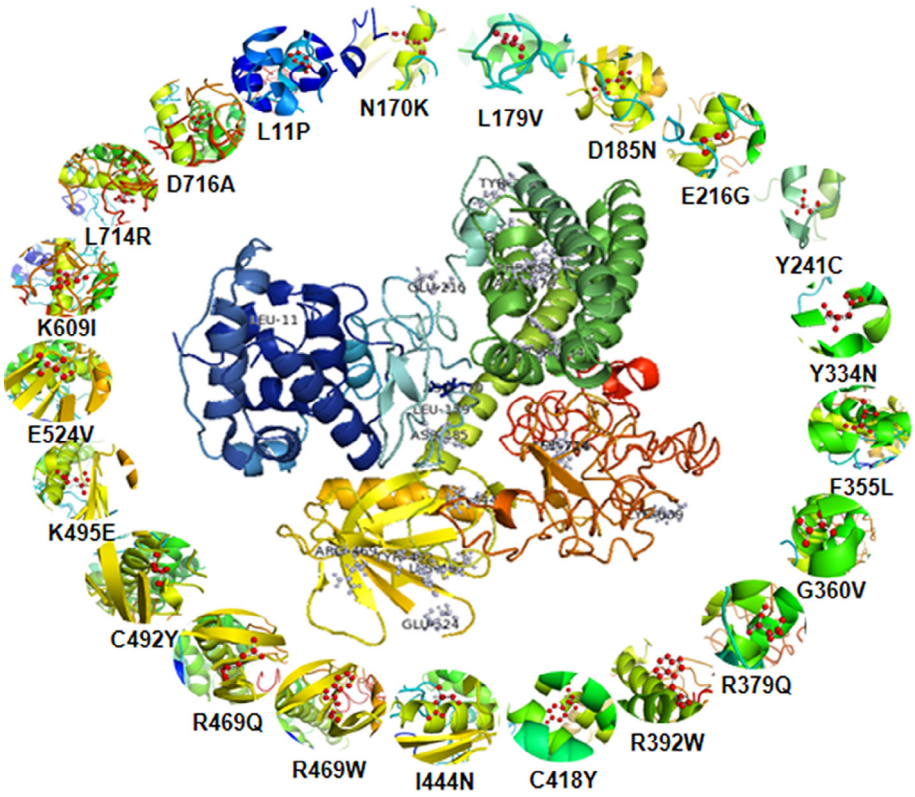


Fig. 3. The central image represent ARHGEF6 wild type model generated by I-TASEER; Peripheral images showed all (21) missense mutations of ARHGEF6 protein identified with ball-stick structure in red color.

remaining were predicted as various abnormal molecular mechanisms. The prediction of molecular mechanisms along with P-value also predicted, and the result was summarized in Table 2.

3.5. ARHGEF6 protein modeling

I-Tasser server was used to build the 3-D structure for ARHGEF6 protein (Fig. 3). The C-score of polypeptide chains was -2.84, estimated TM-score and RMSD score was 0.39 and 15.5 ± 3.3 respectively. Further, the energy minimization of the ARHGEF6 built model was checked after applying Gromacs96 force field in Nomad-Ref server.

Ramachandran Plot assessment for geometrical validation of predicted protein structure has indicated that relatively small percentage of amino acid residues possessed phi/psi angles in the disallowed regions. The percentage of amino acid residues in favored, allowed and outlier regions of the native ARHGEF6 protein are found to be 93.3%, 4.7%, and 2.1% respectively (Fig. 4). Thus, structural assessment of ARHGEF6 protein model described that predicted structure is significantly similar to other homologous structure of ARHGEF6 protein.

In the progression, we assigned twenty-one out of forty-seven accordant deleterious mutations (based on SIFT, PolyPhen2, PROVEAN, SNP&GO and PhD-SNP results) located in different domains and exonic region of *ARHGEF6* for structural analysis. The mutant protein models were built by manual insertion of altered amino acid in the primary sequence and template structure of a native-type ARHGEF6 protein. Energy minimization and stereo chemical property checking can be defined the traditional values of native-type protein structures with anticipations of Ramachandran plot. It revealed that >95% of the residues in the built model is within the favored and allowed regions.

3.6. Stability effect of amino acid residues

The location and type of a mutated residue affect the stability and structure of the protein. Testing the protein stability of 21 mutations concerning free energy values as per I-Mutant 2.0 and DUET web server revealed the results of stability analysis that all 21 mutant models has shown the free energy change (DDG/ΔΔG) values ranging between −0.307 and −3.626 kcal/Mol (Table 3) and predicted as decreased stability to form a protein structure. The negative DDG value suggests that the given amino acid substitution is deleterious to the stability of the protein.

3.7. Structural deviation predictions

In order to measure structural deviation, RMSD values of c-alpha atom and TM-Score standards were analyzed by incorporating native and mutant models using TM-Score online sever [34]. The RMSD values at both, protein structure (5.60–21.59 Å) and amino acid residue level (0.82–1.44 Å) demonstrated significant structural coast based on super-positioning prediction of 21 assorted variants models on native ARHGEF6. At the same side, TM-Score of all mutant models revealed deviations in the structure that confirms morbific impact of genomic changes on ARHGEF6 protein (Table 4).

3.8. ARHGEF6 mutant protein characteristics

3.8.1. Domain identification

The Pfam Server revealed five known functional domain spanning between 1 and 112, 167 to 215, 245 to 419, 442 to 548 and 681 to 769 amino acids (Fig. 4). The N-terminal domain between 1 and 112 amino acids known as Calponin Homology (CH)

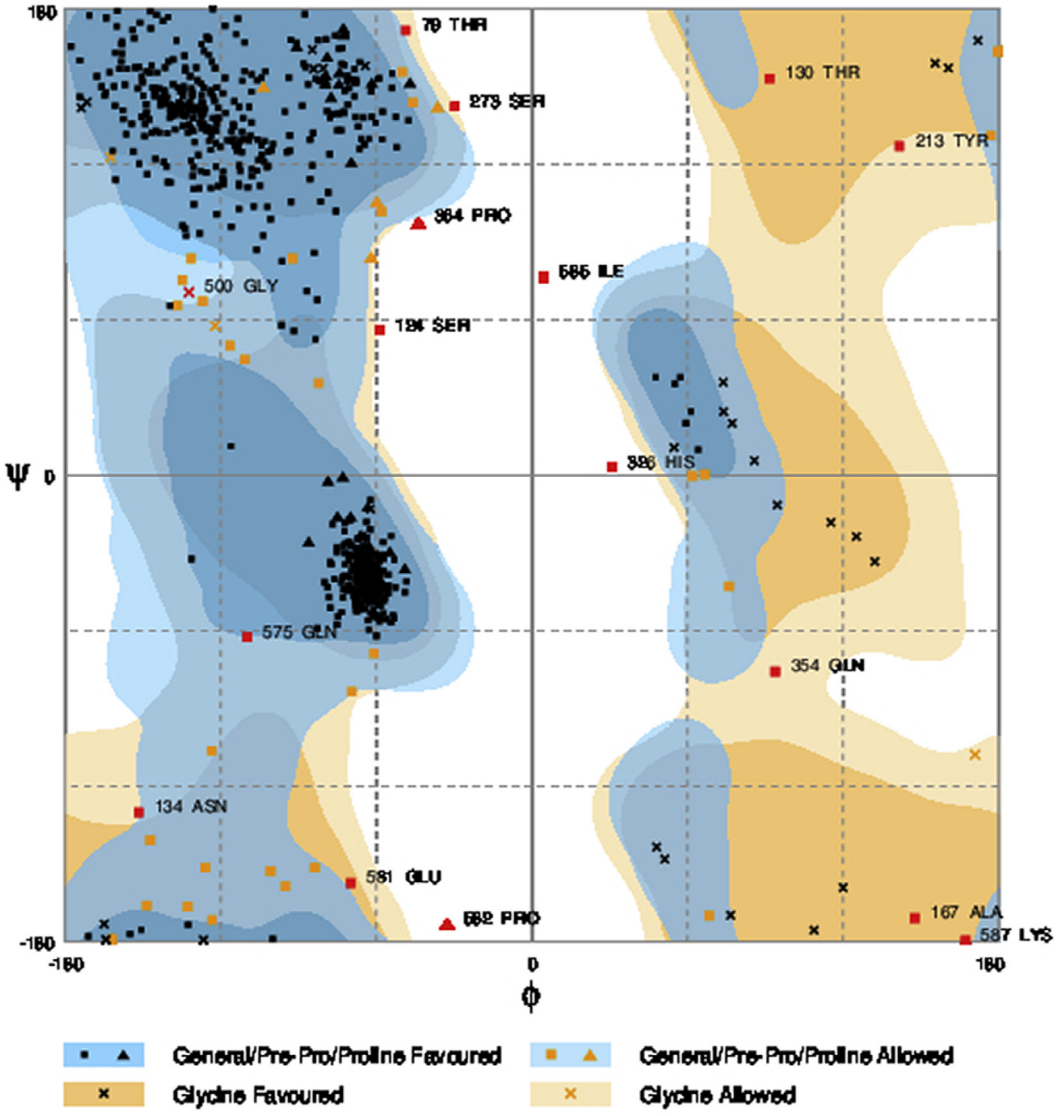


Fig. 4. Ramachandran plot assessment of the predicted model of ARHGEF6 protein.

domain and functioning in ARHGEF6 and PARVB binding lead to activation of the GTPases Rac1 and Cdc42 [35]. The next domain spans from 167 to 215 amino acids known as Src Homology 3 (SH3) domain and function as an interacting junction between ARHGEF6 and, BIN2 and GIT2 protein complex [36]. The third is a RhoGEF domain/Dbl Homology (DH) domain that spans 245 to 419 amino acids. It is responsible for modulating the Rho-GEF activity (release of bound GDP and subsequent binding of GTP) for many Rho GTPases interactions and activations [37–39]. However, protein–protein interaction predictions (PIPs) database evidenced the interaction of ARHGEF6 and RHEB in cytoplasm and plasma membrane [40] with score 2.51. The fourth domain spanned from 442 to 548 amino acids is known as Plekin Homology (PH) domain. The primary function of this domain is to interact with membrane but due to their juxtaposition with a DH domain and role in regulating the GEF activity, they are more likely to be involved in protein–protein interactions [41]. The fifth domain called as BetaPIX (BP) domain is spanning in between 681 and 769 amino acids (Fig. 5). The BP domain interacts with GIT protein and forms PIX-GIT complex, to integrate signaling among Arf, Cdc42, and Rac proteins in response to cues emanating from integrins, heterotrimeric G proteins, receptortyrosine kinases, and cell–cell interactions [2].

3.8.2. Solvent accessibility

The solvent accessibility of the residue decreases because of mutations, and it can affect the stability of the protein. The NetSurfP server was used to test the solvent accessibility of mutated ARHGEF6 residues. In the case of I444N mutation, a significant change from buried to expose orientation property was found that is considered with a potential mutational effect on ARHGEF6 protein stricture. Of note, all mutations showed the significant difference in z-score (Table 5) that can cause structural drifts in residue orientations of ARHGEF6 protein.

3.8.3. Mutational features of ARHGEF6 by project hope

Twenty-one non-synonymous mutations in ARHGEF6 protein were subjected to PROJECT HOPE server and revealed structural and functional identification of protein features. These include amino acid property, 3D structure effects, conservation of residue and effects on domains. The characteristics of twenty-one muta-

Table 3
Stability Prediction of ARHGEF6 protein among amino acids substitutions.

Sr No.	Nucleotide variant	Amino acid variant	I-mutant2.0	DUET Server
1	c.32 T > C	L11P	−0.25	−1.506
2	c.510C > A	N170K	−1.46	−2.094
3	c.535C > G	L179V	−1.81	−2.293
4	c.553G > A	D185N	−0.8	−1.846
5	c.647A > G	E216G	−1.97	−1.601
6	c.722A > G	Y241C	0.13	−1.53
7	c.1000 T > A	Y334N	−1.03	−3.235
8	c.1065C > A	F355L	−1.08	−1.498
9	c.1079G > T	G360V	−8.768	−0.82
10	c.1136G > A	R379Q	−0.74	−1.163
11	c.1174C > T	R392W	−0.7	−0.307
12	c.1253G > A	C418Y	1.38	−1.786
13	c.1331 T > A	I444N	−1.13	−2.669
14	c.1405C > A	R469W	−0.73	−0.788
15	c.1406G > A	R469Q	−0.92	−1.113
16	c.1475A > G	Y492C	1.31	−0.964
17	c.1483A > G	K495E	−0.59	−1.031
18	c.1571A > T	E524V	−1.57	−0.658
19	c.1826A > T	K609I	0.58	0.549
20	c.1994 T > G	L714R	−2.26	−1.298
21	c.2147A > C	D716A	0.06	−3.626

(I-Mutant2.0 prediction: ΔΔG < 0.00 = Decrease stability; ΔΔG > 0.00 = Increase Stability).

Table 4
Protein structural deviation predictions using RMSD value and TM-Score.

Sr. No.	Amino acids variants	RMSD (Å)		TM-score
		Protein structure level	Amino acid residue level	
1	L11P	11.45	1.02	0.7394
2	N170K	18.01	0.95	0.7210
3	L179V	14.98	0.97	0.7039
4	D185N	16.61	0.87	0.6715
5	E216G	14.69	0.98	0.7574
6	Y241C	09.25	1.16	0.8223
7	Y334N	19.11	1.03	0.6849
8	F355L	16.94	1.05	0.7020
9	G360V	05.60	1.44	0.8735
10	R379Q	17.78	1.16	0.7587
11	R392W	17.28	0.89	0.6989
12	C418Y	15.85	1.20	0.7024
13	I444N	21.59	1.13	0.6952
14	R469W	15.49	1.05	0.6858
15	R469Q	19.27	1.01	0.6866
16	Y492C	17.72	1.01	0.6823
17	K495E	10.23	0.99	0.7112
18	E524V	19.45	0.82	0.6572
19	K609I	19.40	0.97	0.7031
20	L714R	09.84	1.09	0.8120
21	D716A	11.33	1.12	0.8429

(RMSD = Root mean square deviation; TM-Score = Template modeling score).

Table 5
Surface accessibility prediction of ARHGEF6 mutations by NetSurfP server.

Amino acid	Position	RSA	ASA	Z-score	Class assigned
Leu	11	0.046	8.459	−0.047	B
Pro	11	0.047	6.655	0.118	B
Asn	170	0.489	71.516	−0.139	E
Lys	170	0.552	113.588	−0.247	E
Leu	179	0.176	23.207	−0.971	B
Val	179	0.122	18.813	−0.896	B
Asp	185	0.406	58.461	−0.899	E
Asn	185	0.455	66.627	−0.952	E
Glu	216	0.438	76.519	−1.231	E
Gly	216	0.440	34.659	−1.092	E
Tyr	241	0.355	75.842	0.701	E
Cys	241	0.512	71.857	−0.753	E
Tyr	334	0.067	14.275	0.602	B
Asn	334	0.062	9.135	0.629	B
Phe	355	0.069	13.868	0.134	B
Leu	355	0.060	11.078	0.496	B
Gly	360	0.561	44.166	−0.528	E
Val	360	0.582	89.500	−0.596	E
Arg	379	0.158	36.182	0.246	B
Gln	379	0.159	28.415	0.214	B
Arg	392	0.605	138.637	1.304	E
Trp	392	0.597	143.482	1.441	E
Cys	418	0.078	10.979	−1.175	B
Tyr	418	0.070	14.874	−0.871	B
Ile	444	0.125	23.181	−2.214	B
Asn	444	0.352	51.606	−1.317	E
Arg	469	0.077	17.702	1.050	B
Trp	469	0.063	15.200	0.895	B
Gln	469	0.065	11.645	0.858	B
Tyr	492	0.134	28.614	0.324	B
Cys	492	0.140	19.600	0.103	B
Lys	495	0.325	66.894	0.303	E
Glu	495	0.322	56.323	0.394	E
Glu	524	0.355	62.001	0.910	E
Val	524	0.343	52.719	0.743	E
Lys	609	0.611	125.662	−0.930	E
Ile	609	0.409	75.591	−0.732	E
Leu	714	0.072	13.183	−0.028	B
Arg	714	0.172	39.480	−0.975	B
Asp	716	0.246	35.477	−0.222	B
Ala	716	0.235	25.853	−0.268	B

(RSA = Relative surface area value < 0.2 Buried residue and > 0.2 Exposed residue; ASA = Absolute surface area value below 25% of ASA_{max} = Buried, and above 25% of ASA_{max}=Exposed; Class assigned: B = Buried and E = Exposed).

tions collectively predicted by PROJECT HOPE server is summarized in Table 6.

3.9. Protein-protein network analysis

The STRING database search has resulted in the direct interaction of ARHGEF6 regarding confidence score (>0.95) with 31

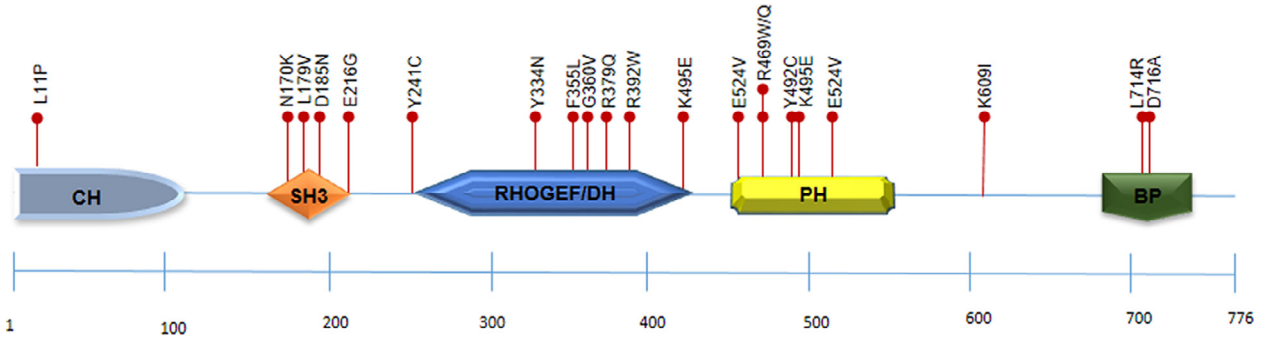


Fig. 5. The schematic representation of identified missense mutations distributed across the functional domains of ARHGEF6 protein.

Table 6
ARHGEF6 protein phenotype features prediction by PROJECT HOPE analysis.

Sr. No.	Mutation code	Characteristics
1	L11P	Proline (mutant) residue is smaller than the Leucine (native) residue, which can cause empty space in the core of the protein The location of this residue is in α-helix and Proline disrupt α-helix. Leucine is highly conserved at this position and located in the CH domain that is important for binding of other molecules
2	N170K	Lysine (mutant) residue is bigger than the Asparagine (native) residue. It has changed in charge from neutral into positive that can create repulsion between the mutant residue and neighboring residues. The residue positioned within SH3 domain that can abolish protein function The native residue is not conserved at this position but it is in contact with residues in another domain and possible that the mutation disturbs these contacts
3	L179V	Valine (mutant) residue is smaller than the Leucine (native) residue. The mutation will cause an empty space in the core of the protein. The native residue is very conserved The mutation is located within SH3 binding domain and introduces Valine with different properties, which can disturb this domain
4	D185N	Aspartic Acid (native) residue charge was negative and Asparagine (mutant) residue charge is neutral. The difference in charge will disturb the ionic interaction The native residue forms a salt bridge with Arginine at position 205 The mutation is located on SH3 domain. The native residue is very conserved
5	E216G	Glycine (mutant) residue is smaller than the Glutamic Acid (native) residue. The native residue charge was negative and mutant charge is neutral. The native residue is more hydrophobic than the mutant residue Glutamic Acid forms a hydrogen bond with Serine at position 211 and a salt bridge with Arginine at position 222. The mutation is located on SH3 domain
6	Y241C	Cysteine (mutant) residue is smaller than the Tyrosine (native) residue. Tyrosine is more hydrophobic than Cysteine, and that will cause a possible loss of external interactions. The mutation is located within the DH domain. The native residue is very conserved.
7	Y334N	Asparagine (mutant) residue is smaller than the Tyrosine (native) residue. Here, Tyrosine is more hydrophobic than Asparagine. The mutated residue is located very close to a residue that makes a cysteine bond. The native-type residue forms a hydrogen bond with Leucine at position 373. The residue is located on DH domain and the interaction with other domains could be disturbed by the mutation This Mutation is of a 100% conserved residue and usually damaging for the protein
8	F355L	Leucine (mutant) residue is smaller than the Phenylalanine (native) residue The mutation is located within the DH domain and is located near a highly conserved position
9	G360V	Valine (mutant) is bigger than the Glycine (native) residue. Here, Glycine residue is more hydrophobic than Valine The mutant residue is located on the DH domain. The torsion angles for this residue are unusual and only Glycine is flexible enough to make these torsion angles The native-type residue is very conserved
10	R379Q	Glutamine (mutant) is smaller than the Arginine (native) residue. The native residue charge was positive the mutant residue charge is neutral. The difference in charge will disturb the ionic interaction The Arginine (native) forms a hydrogen bond with Threonine at position 252, and a salt bridge with Glutamic Acid at position 255 and 259. The mutation is located within the DH domain. The native residue is very conserved
11	R392W	Tryptophan (mutant) is bigger than Arginine (native). Arginine residue charge was positive and Tryptophan residue charge is neutral. The native residue is more hydrophobic than the mutant residue Arginine residue forms a hydrogen bond with Histidine at position 401, and the salt bridge with Glutamic Acid at position 389 and 395, and also with Aspartic Acid at position 396. The difference in charge will disturb the ionic interaction The mutation is located on DH domain and native residue is highly conserved in nature
12	C418Y	The mutant residue (Tyrosine) is bigger than the native-type residue (Cysteine). Cysteine is more hydrophobic than Tyrosine Together with loss of the cysteine bond (formation of cysteine bridge), the differences between the old and new residue can cause destabilization of the structure The mutation is located within DH domain, and based on highly conservation scores this mutation is probably damaging to the protein
13	I444N	Asparagine (mutant) residue is bigger than Isoleucine (native) residue. Isoleucine residue is more hydrophobic than asparagine and might cause loss of hydrophobic interactions with other molecules on the surface of the protein The mutation is located within PH domain. The native residue is very conserved at this position
14	R469W	Tryptophan (mutant) is bigger than Arginine (wild) residue. Arginine residue charge was positive and tryptophan residue charge is neutral. The native residue is more hydrophobic than the mutant residue. Arginine forms a salt bridge with Glutamic Acid at position 467. The mutation is located within a domain, and the native residue is very conserved
15	R469Q	Glutamine (mutant) is smaller than Arginine (native). The native residue charge was positive, the mutant residue charge is neutral. Arginine forms a salt bridge with glutamic acid at position 467. The mutation is located within PH domain. The native residue is very conserved and buried in the core of a domain
16	Y492C	Cysteine (mutant) is smaller than Tyrosine (native). The native residue is more hydrophobic than the mutant The mutation is located on the surface of pH domain. Here, Tyrosine residue is very conserved
17	K495E	Glutamic acid (mutant) is smaller than Lysine (native). The native residue charge was positive, the mutant residue charge is negative The mutation is located on the surface of pH domain. The native residue is very conserved
18	E524V	Valine (mutant) is smaller than Glutamic acid (native). The native residue charge was negative, the mutant residue charge is neutral. Glutamic acid residue is more hydrophobic than the Valine. Glutamic acid forms a hydrogen bond with Asparagine at position 521. The mutation is located within surface of pH domain. The native-type residue is very conserved
19	K609I	Isoleucine (mutant) is smaller than Lysine (native). Lysine residue charge was positive and the mutant residue charge is neutral. Lysine is more hydrophobic than Isoleucine The charge of the native residue will be lost, this can cause loss of interactions with other molecules or residues. The smaller size of mutant leads to loss of interactions
20	L714R	Arginine (mutant) is bigger than Leucine (native). The native residue charge was neutral, the mutant residue charge is positive. Leucine is more hydrophobic than Arginine residue
21	D716A	Aspartic Acid (mutant) is smaller than Alanine (native). Alanine charge was negative Aspartic Acid charge is neutral. The native residue is more hydrophobic than the mutant residue

proteins based on experiments, datasets, and text mining (Fig. 6). All these proteins predicted strong functional associations with ARHGEF6 and manifested to be involved in activation of RhoGTPase pathway. From the interaction network analysis, it is clear that G protein-coupled receptor kinase interacting ArfGAP 2 (GIT2) (score 0.99) and Parvinbeta (PARVB) (score 0.99) are the

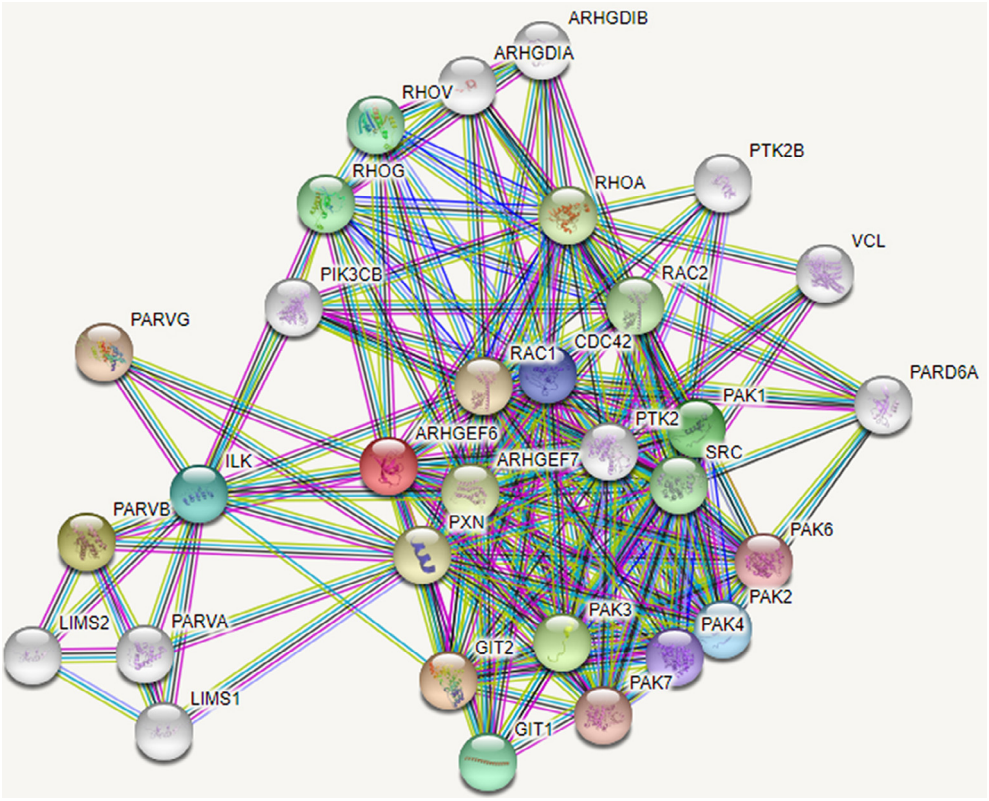


Fig. 6. SRTING interaction network of ARHGEF6 protein showing 30 different interacting protein partners. Representation of nodes and edges described on SRTING web server.

more potential interaction partners, which are involved in integrin-mediated signaling leading to activation of the GTPases Rac1 and Cdc42 [35,42].

4. Discussion

The Rho-guanine nucleotide exchange factor 6 (*ARHGEF6*) is the third X-linked intellectual disability gene, after *OPHN1* and *PAK3*, whose protein was found to be involved in Rho-GTPase pathway, which mediates organization of cytoskeleton, cell shape and cell motility. Studies reported that deregulation in ARHGEF6 protein altered neuronal connectivity and leads to impaired synaptic function and cognition [1,5]. Till date different genetic mutations in *ARHGEF6* gene are found with the help of conventional high throughput sequencing [8]. However, the biological consequences of these alleles are not clearly understood. The association study of genomic variants responsible for specific clinical conditions with their molecular approaches are more expensive as well as time consuming. Whereas, computational analysis is a useful advance approach that can help in the selection of pathogenic variants to be screened in genetic association studies and for the functional and structural protein phenotype prediction.

In the present study, advanced computational strategies has refined missense mutations and predicted structural and functional impacts on ARHGEF6 protein. SIFT, POLYPHEN2.0 and PROVEAN tools were used to screen and identify the most pathogenic missense variants of *ARHGEF6* gene. SNPs&GO and PhD-SNP severs helped to narrow down the missense mutations depending on their disease relevance property. The functional prediction revealed total 47 (24%) missense variants as deleterious and damaging effect to the ARHGEF6 protein. In regard with commonly found disease relevance effect for the mutations, SNPs&GO and PhD-SNP assorted 21 mutations as highly pathogenic.

Evolutionary information is of fundamental importance for detecting mutations that affect human health. The ConSurf analysis for evolutionary conservation of amino acid residues revealed that all twenty-one amino acid variants are highly conserved and alteration in these positions can cause abnormal functional effect of the ARHGEF6 protein. Similarly, predicted molecular mechanisms of amino acid variants revealed altered metal binding ($P = 0.03$) in L179V, D185N, Y334N and R392W, and Loss of helix in R379Q and K609I. Protein stability effect due to amino acid change studied by I-mutant and DUET server revealed that all twenty-one amino acid variants showing DDG values ranging between -0.307 and -3.626 kcal/Mol demonstrated decreased stability of ARHGEF6 protein due to these amino acid variants.

The protein properties considered at a 3D level and mutations mapped on CH, SH3, RHOGEF/DH, PH and BP domains induce biochemically and biophysically significant amino acid changes that are damaging to ARHGEF6 protein phenotypes. These mutations also have induced changes in protein stability and surface accessibility of ARHGEF6 which can directly or indirectly influence intermolecular and intramolecular interactions of amino acids and can be translated into disease risks in the body. The analyzed interaction of ARHGEF6 protein with other proteins has identified GIT2 and PARBV as most interacting partners in cellular protein network.

The comparison of structural deviations in terms of RMSD and Tm-score between mutant and wild type models demonstrated significant deviations in all twenty-one mutated models of ARHGEF6. The deleterious mutations mapped to greatly conserved region that is RHOGEF domain (Y241C, Y334N, F355L, G360V, R379Q, R392W, and C418Y) has shown higher structural divergence compared to mutations on other four domains. The RHOGEF domain spans between 249 and 419 amino acids in human. This domain can bind to a subset of Rho GTPases, but a comprehensive

biochemical characterization is not clear. The significant deviations of amino acid residues may disturbs physical characteristics such as hydrogen bonds, active sites of residue and electrostatic charge.

Numerous studies have demonstrated the utility of advanced computational predictions for emphasizing the variants in many genes associated with intellectual disability and complex genetic conditions [43–45].

5. Conclusion

The results from the present study can be supportive for genotyping, novel drug targets for the relevant protein structure as well as for the pharmacogenetic studies. However, to point out the definitive impact of these mutations in cellular physiology, disease inheritance and pathogenic validation can be better understood if the new study would be carried out with other functional biological assays. Nevertheless, our data may be considered for designing the functional biological assay for *ARHGEF6* dysfunction associated with genetically inherited diseases. Finally, the study demonstrates a significance of various computational methods to figure out highly pathogenic genomic variants linked with the structural and functional relationship of ARHGEF6 protein.

Acknowledgments

We thank Directors' of the ARIBAS and Sophisticated Instrument Centre for Applied Research & Testing (SICART) for providing computational facilities. Financial support provided by Charutar Vidya Mandal (CVM), Vallabh Vidyanagar (partial grant) and University Grants Commission, India (RGN-Fellowship) thankfully acknowledged.

Conflict of interest

The authors declare no conflict of interest.

References

[1] Kutsche K, Yntema H, Brandt A, Jantke I, Nothwang HG, Orth U, et al. Mutations in ARHGEF6, encoding a guanine nucleotide exchange factor for Rho GTPases, in patients with X-linked mental retardation. Nat Genet 2000;26:247–50. https://doi.org/10.1038/80002.

[2] Frank SR, Hansen SH. The PIX – GIT complex : a G protein signaling cassette in control of cell shape. Semin Cell Dev Biol 2008;19:234–44. https://doi.org/10.1016/j.semcdb.2008.01.002.

[3] Ba W, van der Raadt J, Kasri NN. Rho GTPase signaling at the synapse : implications for intellectual disability. Exp Cell Res 2013;319:2368–74. https://doi.org/10.1016/j.yexcr.2013.05.033.

[4] Govek E, Newey SE, Van Aelst L. The role of the Rho GTPases in neuronal development. Genes Dev 2005;19:1–49. https://doi.org/10.1101/gad.1256405.and.

[5] Meyer M. Highly expressed genes within hippocampal sector CA1: implications for the physiology of memory. Neurol Int 2014;6:5388. https://doi.org/10.4081/ni.2014.5388.

[6] Nodé-Langlois R, Muller D, Boda B. Sequential implication of the mental retardation proteins ARHGEF6 and PAK3 in spine morphogenesis. J Cell Sci 2006;119:4986–93.

[7] Meseke M, Rosenberger G, Forster E. Reelin and the Cdc42/Rac1 guanine nucleotide exchange factor a PIX/Arhgef6 promote dendritic Golgi translocation in hippocampal neurons. Eur J Neurosci 2013;37:1404–12. https://doi.org/10.1111/ejn.12153.

[8] Selvan LDN, Manoj J, Muthusamy B, Nguyen TT, Stawiski EW, Jaiswal BS, et al. Next-generation sequencing reveals novel mutations in X-linked intellectual disability. Omics 2017;21:295–303. https://doi.org/10.1089/omi.2017.0009.

[9] Aamer Mehmood M. Use of bioinformatics tools in different spheres of life sciences. J Data Min Genom Proteom 2014;5:158. https://doi.org/10.4172/2153-0602.1000158.

[10] Sherry S, Ward M, Kholodov M, Baker J, Phan L, Smigielski E, et al. DbSNP: the NCBI database of genetic variation. Nucl Acids Res 2001;29:308–11.

[11] Leary NAO, Wright MW, Brister JR, Ciufo S, Haddad D, Mcveigh R, et al. Reference sequence (RefSeq) database at NCBI : current status, taxonomic expansion, and functional annotation. Nucl Acids Res 2016;44:733–45. https://doi.org/10.1093/nar/gkv1189.

[12] Aken BL, Ayling S, Barrell D, Clarke L, Curwen V, Fairley S, et al. The Ensembl gene annotation system. Database 2016;2016:1–19. https://doi.org/10.1093/database/baw093.

[13] Sim N, Kumar P, Hu J, Henikoff S, Schneider G, Ng PC. SIFT web server : predicting effects of amino acid substitutions on proteins. Nucl Acids Res 2012;40:452–7. https://doi.org/10.1093/nar/gks539.

[14] Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using polyphen-2. Curr Protoc Hum Genet 2013:1–41. https://doi.org/10.1002/0471142905.hg0720s76.

[15] Choi Y, Chan AP, Craig TJ. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. Bioinformatics 2015;31:2745–7. https://doi.org/10.1093/bioinformatics/btv195.

[16] Capriotti E, Calabrese R, Fariselli P, Martelli PL, Altman RB, Casadio R. WS-SNPs&GO: a web server for predicting the deleterious effect of human protein variants using functional annotation. BMC Genom 2013;14:S6. https://doi.org/10.1186/1471-2164-14-S3-S6.

[17] Capriotti E, Calabrese R, Casadio R. Sequence analysis Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. Bioinformatics 2006;22:2729–34. https://doi.org/10.1093/bioinformatics/btl423.

[18] Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, Pupko T, et al. An improved methodology to estimate and visualize evolutionary conservation in macromolecules. Nucl Acids Res 2016;2016:1–7. https://doi.org/10.1093/nar/gkw408.

[19] Pejaver V, Urresti J, Lugo-Martinez J, Pagel KA, Lin GN, Nam H-J, et al. eMutPred2: inferring the molecular and phenotypic impact of amino acid variants. bioRxiv 2017.

[20] Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST : a new generation of protein database search programs. Nucl Acids Res 1997;25:3389–402.

[21] Finn RD, Bateman A, Clements J, Coggill P, Eberhardt Y, Eddy SR, et al. Pfam : the protein families database. Nucl Acids Res 2014;42:222–30. https://doi.org/10.1093/nar/gkt1223.

[22] Krogh A, Larsson È, Von Heijne G, Sonnhammer ELL. Predicting transmembrane protein topology with a hidden markov model : application to complete genomes. J Mol Biol 2001;305:567–80. https://doi.org/10.1006/jmbi.2000.4315.

[23] Delorenzi M, Speed T. An HMM model for coiled-coil domains and a comparison with PSSM-based predictions. Bioinformatics 2002;18:617–25.

[24] Sickmeier M, Hamilton JA, Legall T, Vacic V, Cortese MS, Tantos A, et al. DisProt: the database of disordered proteins. Nucl Acids Res 2007;35:786–93. https://doi.org/10.1093/nar/gkl893.

[25] Yang J, Zhang Y. I-TASSER server: new development for protein structure and function predictions. Nucl Acids Res 2015;43:174–81. https://doi.org/10.1093/nar/gkv342.

[26] Lindahl E, Azuara C, Koehl P, Delarue M. NOMAD-Ref : visualization, deformation and refinement of macromolecular structures based on all-atom normal mode analysis. Nucl Acids Res 2006;34:52–6. https://doi.org/10.1093/nar/gkl082.

[27] Lovell SC, Davis IW, Arendall III WB, de Bakker PIW, Word JM, Prisant MG, et al. Structure validation by Calpha geometry: phi, psi and Cbeta deviation. Proteins Struct Funct Genet 2002;50:437–50.

[28] Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF chimera — a visualization system for exploratory research and analysis. J Comput Chem 2004;25:1605–12. https://doi.org/10.1002/jcc.20084.

[29] Capriotti E, Fariselli P, Casadio R. I-Mutant2. 0: predicting stability changes upon mutation from the protein sequence or structure. Nucl Acids Res 2005;33:306–10. https://doi.org/10.1093/nar/gki375.

[30] Pires DE V, Ascher DB, Blundell TL. DUET : a server for predicting effects of mutations on protein stability using an integrated computational approach. Nucl Acids Res 2014;42:314–9. https://doi.org/10.1093/nar/gku411.

[31] Petersen B, Petersen TN, Andersen P, Nielsen M, Lundegaard C. A generic method for assignment of reliability scores applied to solvent accessibility predictions. BMC Struct Biol 2009;9:51. https://doi.org/10.1186/1472-6807-9-51.

[32] Venselaar H, Ah T, Kuipers RKP, Hekkelman ML, Vriend G. Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. BMC Bioinf 2010;11:548. https://doi.org/10.1186/1471-2105-11-548.

[33] Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-cepas J, et al. STRING v10: protein – protein interaction networks, integrated over the tree of life. Nucl Acids Res 2015;43:447–52. https://doi.org/10.1093/nar/gku1003.

[34] Zhang Y, Skolnick J. Scoring function for automated assessment of protein structure template quality. Protein Struct Funct Bioinform 2004;57:702–10. https://doi.org/10.1002/prot.20264.

[35] Rosenberger G, Jantke I, Gal A, Kutsche K. Interaction of a PIX (ARHGEF6) with b -parvin (PARVB) suggests an involvement of a PIX in integrin-mediated signaling. Hum Mol Genet 2003;12:155–67. https://doi.org/10.1093/hmg/ddg019.

[36] Sanchez-Barrena M, Vallis Y, Clatworthy M, Doherty G, Veprintsev D, Evans P, et al. Bin2 is a membrane sculpting N-BAR protein that influences leucocyte podosomes, motility and phagocytosis. PLoS ONE 2012;7:E52401.

[37] Zheng Y. Dbl family guanine nucleotide. Trends Biochem Sci 2001;26:724–32.

[38] Feng Q, Baird D, Cerione RA. Novel regulatory mechanisms for the Dbl family guanine nucleotide exchange factor Cool-2/a -Pix. Embo J 2004;23:3492–504. https://doi.org/10.1038/sj.emboj.7600331.
[39] Rossman KL, Der CJ, Sondek J, Hill C. GEF means GO: turning on RHO GTPases with guanine nucleotide exchange factors. Mol Cell Biol 2005;6:167–80. https://doi.org/10.1038/nrm1587.
[40] Mcdowall MD, Scott MS, Barton GJ. PIPs : human protein – protein interaction prediction database. Nucleic Acids Res 2009;37:651–6. https://doi.org/10.1093/nar/gkn870.
[41] Lenoir M, Kufareva I, Abagyan R, Overduin M. Membrane and protein interactions of the pleckstrin homology domain superfamily. Membranes (Basel) 2015;5:646–63. https://doi.org/10.3390/membranes5040646.
[42] Zhou W, Li X, Premont RT. Expanding functions of GIT Arf GTPase-activating proteins, PIX Rho guanine nucleotide exchange factors and GIT – PIX complexes. J Cell Sci 2016;129:1963–74. https://doi.org/10.1242/jcs.179465.
[43] Banaganapalli B, Mohammed K, Khan IA, Al-aama JY, Elango R, Shaik NA. A computational protein phenotype prediction approach to analyze the deleterious mutations of human MED12 gene. J Cell Biochem 2016;13:1–13. https://doi.org/10.1002/jcb.25499.
[44] Desai M, Chauhan JB. Computational analysis for the determination of deleterious nsSNPs in human MTHFD1 gene. Comput Biol Chem 2017;70:7–14.
[45] Hussain MRM, Shaik NA, Yousuf Al-Aama J, Asfour HZ, Khan FS, Masoodi TA, et al. In silico analysis of single nucleotide polymorphisms (SNPs) in human BRAF gene. Gene 2012;508:188–96.