Heliyon 5 (2019) e02709

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.heliyon.com

Research article

The impact of missense mutation in PIGA associated to paroxysmal nocturnal hemoglobinuria and multiple congenital anomalies-hypotonia-seizures syndrome 2: A computational study

Ashish Kumar Agrahari [a,b,**], Enrico Pieroni [c], Gianluca Gatto [d], Amit Kumar [d,*]

[a] Department of Integrative Biology, School of Biosciences and Technology, VIT, Vellore, Tamil Nadu 632014, India
[b] Research Center for Computer-Aided Drug Discovery, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China
[c] CRS4 – Modeling & Simulation Group, Biosciences Department, 09010, Pula, Italy
[d] Department of Electrical and Electronic Engineering, University of Cagliari, via Marengo 2, 09123 Cagliari, Italy

ARTICLE INFO

Keywords:
Biophysics
Condensed matter physics
Computational chemistry
Theoretical chemistry
Bioinformatics
Missense mutations
PNH
Homology modeling
PIGA
Molecular dynamics simulation
MCAHS2

ABSTRACT

Paroxysmal nocturnal hemoglobinuria (PNH) is an acquired clonal blood disorder that manifests with hemolytic anemia, thrombosis, and peripheral blood cytopenias. The disease is caused by the deficiency of two glycosylphosphatidylinositols (GPI)-anchored proteins (CD55 and CD59) in the hemopoietic stem cells. The deficiency of GPI-anchored proteins has been associated with the somatic mutations in phosphatidylinositol glycan class A (*PIGA*). However, the mutations that do not cause PNH is associated with the multiple congenital anomalies-hypotonia-seizures syndrome 2 (MCAHS2). To best of our knowledge, no computational study has been performed to explore at an atomistic level the impact of PIGA missense mutations on the structure and dynamics of the protein. Therefore, we focused our study to provide molecular insights into the changes in protein structural dynamics upon mutation. In the initial step, screening for the most pathogenic mutations from the pool of publicly available mutations was performed. Further, to get a better understanding, pathogenic mutations were mapped to the modeled structure and the resulting protein was subjected to 100 ns molecular dynamics simulation. The residues close to C- and N-terminal regions of the protein were found to exhibit greater flexibility upon mutation. Our study suggests that four mutations are highly effective in altering the structural conformation and stability of the PIGA protein. Among them, mutant G48D was found to alter protein's structural dynamics to the greatest extent, both on a local and a global scale.

1. Introduction

Paroxysmal nocturnal hemoglobinuria (PNH) is an acquired clonal hematopoietic stem cell (HSC) disorder that affects 1–2 people per million, with an average age of 35–40 years. Even though PNH is rare, there is substantial knowledge about its pathophysiology and significant molecular defects (Lee and Abdel-Wahab, 2014). If deprived of the treatment, PNH is a life-threatening disease that manifests with intravascular hemolysis, pancytopenia, bone marrow failure, and venous thrombosis (Brodsky, 2008; Nishimura et al., 2004; Parker, 2002; Socié et al., 1996). Hemolysis of PNH is complement-mediated, which triggers due to the deficiency of complement regulatory proteins on the surface of PNH cells (Brodsky, 2008). The disease commences with the growth of a hematopoietic stem cell that has a severe lack or absence for glycosylphosphatidylinositols (GPI), a glycolipid moiety that anchors above 150 different proteins to the cell surface. The deficiency of GPI anchor in almost all PNH cases is the result of a somatic mutation in phosphatidylinositol glycan class A (PIGA), a X-linked gene that encodes an enzyme crucial for the first step involved in the biosynthesis of GPI anchor proteins (GPI-APs) (Bessler et al., 1994; Miyata et al. 1993, 1994; Takeda et al., 1993).

The deficiency in two of complement inhibitory GPI-APs (CD55 and CD59) in the erythrocytes results in a chronic complement-mediated hemolysis and additionally leads to the activation of platelets, monocytes, and granulocytes (Medof, 1984; Rollins and Sims, 1990). Similarly, complement activation occurs because of the loss of CD55 and CD59 in PNH patients that elucidates the susceptibility to thrombosis in this disorder (Lee and Abdel-Wahab, 2014).

* Corresponding author.
** Corresponding author.
E-mail addresses: ashish@siat.ac.cn, ashish07791@gmail.com (A.K. Agrahari), amit369@gmail.com (A. Kumar).

https://doi.org/10.1016/j.heliyon.2019.e02709
Received 18 November 2018; Received in revised form 12 August 2019; Accepted 17 October 2019

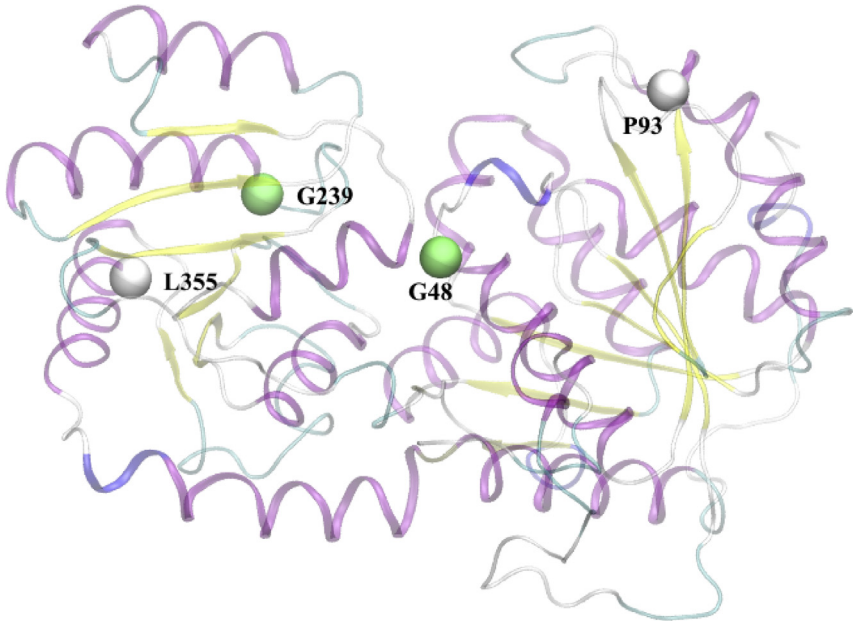

Fig. 1. Structure of modeled protein in cartoon representation. The deleterious residues subjected to mutation are shown.

Table 1
Screening of deleterious missense SNPs associated to *PIGA* gene. The four deleterious mutations found from all the tools are highlighted in bold and italics.

rsID	Variants	SIFT	PolyPhen 2	SNAP	mutationassessor	PROVEAN	I-mutant 3.0	AlignGVGD	iStable	Predict SNP
Paroxysmal nocturnal hemoglobinuria (PNH)										
VAR_015442	R19W	0.04	0	effect	neutral	Neutral	-0.27	Class C65	Decrease	neutral
VAR_015436	D40H	0.18	1	effect	medium	Deleterious	-1.36	Class C65	Decrease	Deleterious
VAR_015437	G48A	0	1	effect	High	Deleterious	-1.08	Class C55	Decrease	Deleterious
VAR_015438	***G48D***	***0***	***1***	***effect***	***High***	***Deleterious***	***-0.44***	***Class C65***	***Decrease***	***Deleterious***
VAR_015439	G48V	0	1	effect	High	Deleterious	0.13	Class C65	Decrease	Deleterious
VAR_015440	H128R	0	1	effect	High	Deleterious	-0.37	Class C25	Increase	Deleterious
VAR_005531	S155F	0.02	1	effect	High	Deleterious	0.12	Class C65	Increase	Deleterious
VAR_015441	***G239R***	***0***	***1***	***effect***	***medium***	***Deleterious***	***-1.31***	***Class C65***	***Increase***	***Deleterious***
VAR_005532	N297D	1	0.001	neutral	neutral	Neutral	-0.41	Class C0	Decrease	neutral
Multiple congenital anomalies-hypotonia-seizures syndrome 2 (MCAHS2)										
VAR_071069	R77L	0.54	1	effect	medium	Deleterious	-0.60	Class C65	Increase	Deleterious
VAR_071070	***P93L***	***0***	***1***	***effect***	***medium***	***Deleterious***	***-0.76***	***Class C65***	***Decrease***	***Deleterious***
VAR_071071	R119W	0.02	1	effect	medium	Deleterious	0.05	Class C65	Decrease	Deleterious
VAR_071072	I206F	0	0.999	effect	medium	Deleterious	-1.75	Class C15	Decrease	Deleterious
VAR_078721	***L355S***	***0***	***1***	***effect***	***High***	***Deleterious***	***-0.50***	***Class C65***	***Decrease***	***Deleterious***

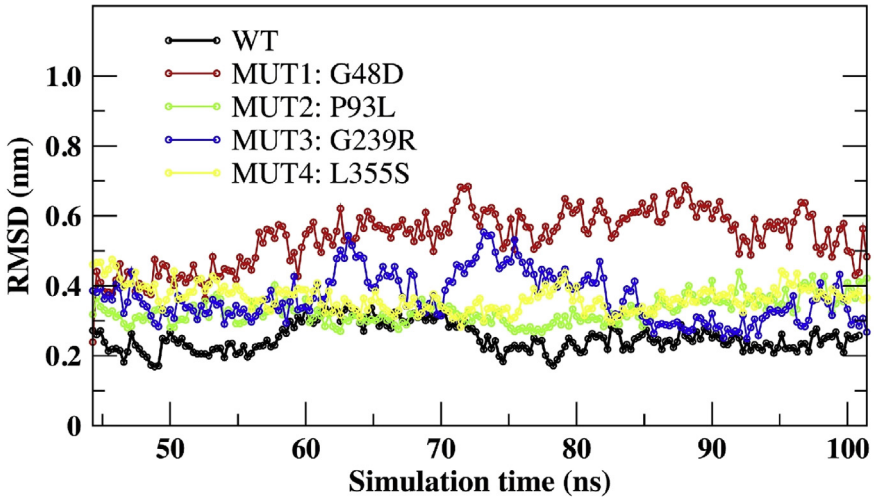

Fig. 2. RMSD plot of native and the four mutants (G49D, P93L, G239R, and L355S) of PIGA protein.

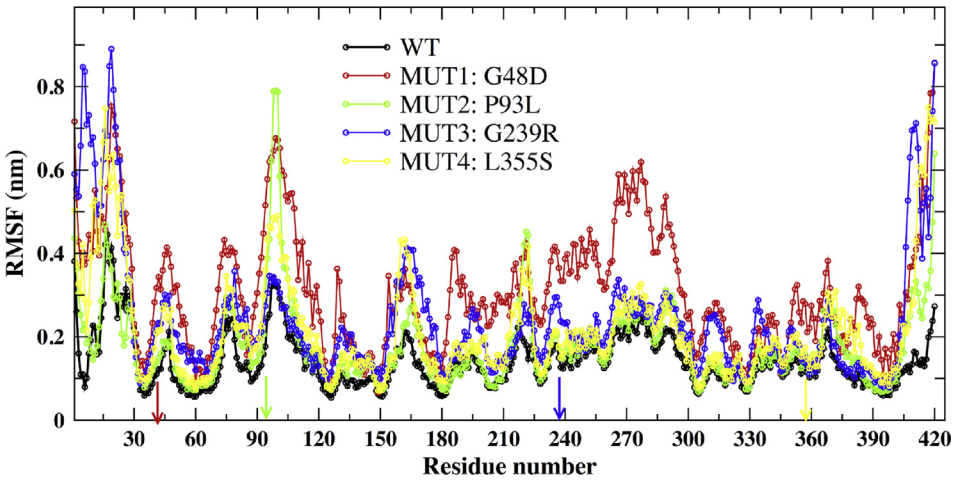

Fig. 3. RMSF plot of native and all mutants (G49D, P93L, G239R, and L355S) of PIGA protein.

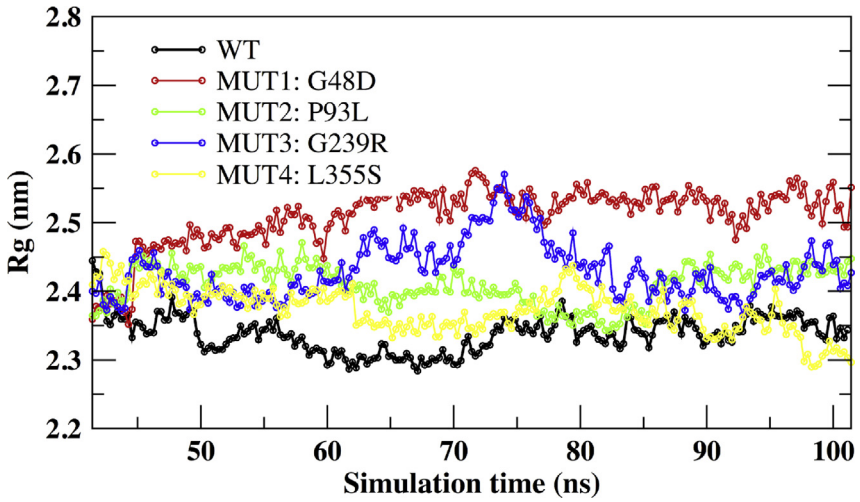

Fig. 4. Rg plot of native and investigated mutants (G49D, P93L, G239R, and L355S) of PIGA protein.

PIGA gene encodes a protein with 484 amino acids, which is expressed in a wide variety of tissues inclusive of brain, liver, heart, and blood cells (Belet et al., 2014). PNH associated Somatic *PIGA* mutations have been documented in OMIM (OMIM: 300818, PNH1) and UNIPROT (UniProtKB – P37287) databases. Contrasting somatic *PIGA* mutations, germline mutations were not yet seen until recently. Based on trials in mice as well as both murine and human embryonic stem cells it has been suggested that germline *PIGA* mutations could be lethal (Tarailo-Graovac et al., 2015).

Based on X-chromosome exome next-generation sequencing

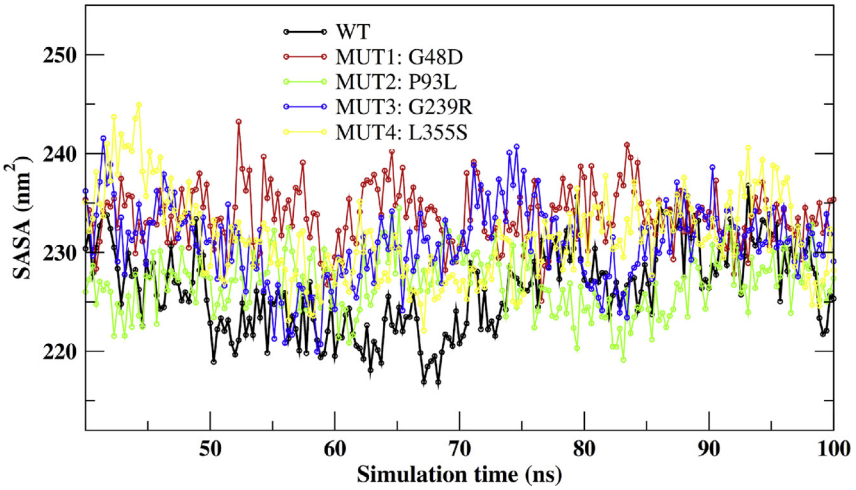

Fig. 5. SASA plot of native and all mutants (G49D, P93L, G239R, and L355S) of PIGA protein.

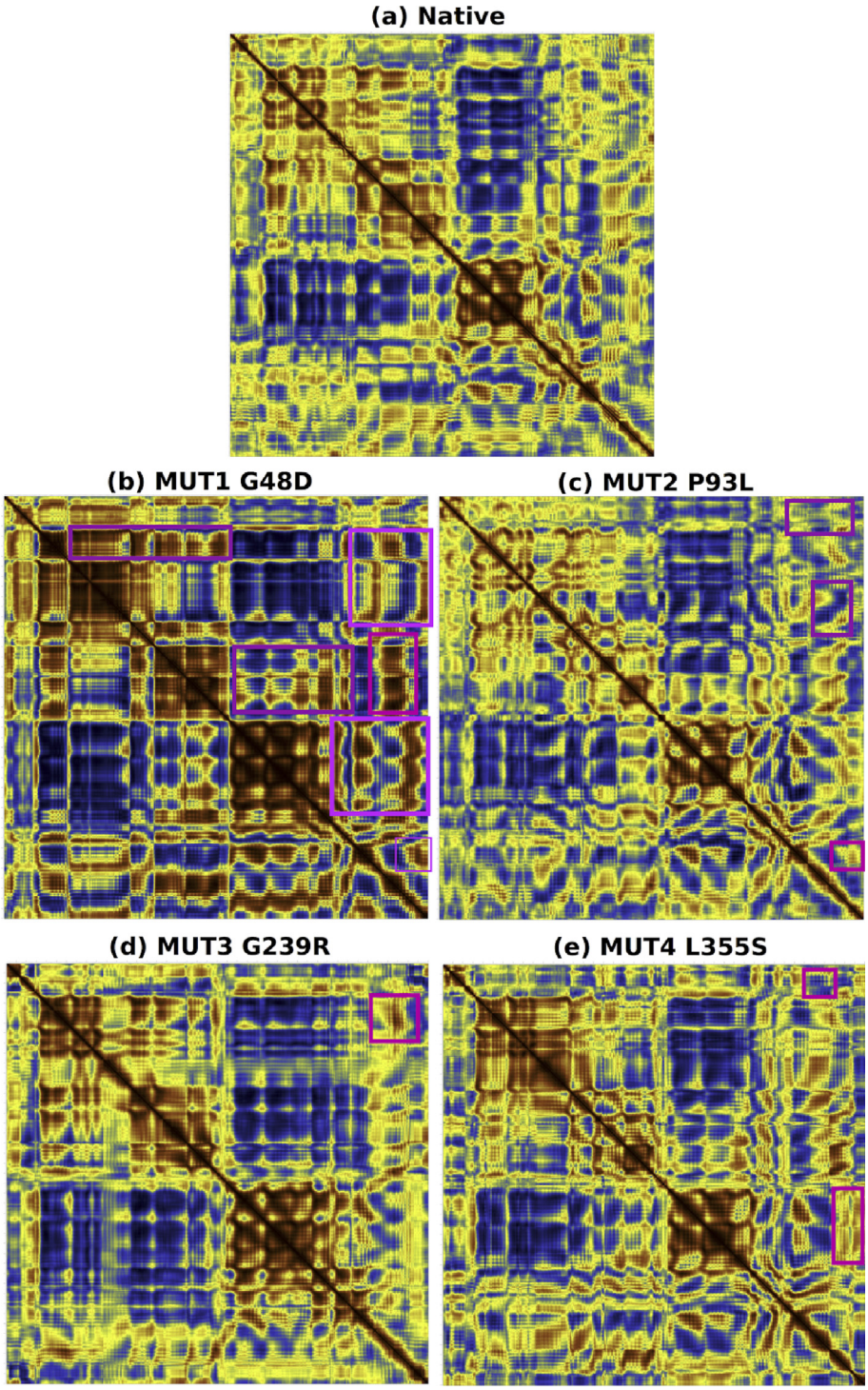

Fig. 6. Cross-correlation plots of C-alpha fluctuations. Maximum of variance-covariance matrix is 1.0 (dark red) and minimum is -0.9 (dark blue), intermediate values are shown in yellow. In (a) for wild type, in (b) for mutant G48D, in (c) for mutant P93L, in (d) for mutant G239R and in (e) for mutant L355S, protein complexes. In plots (b–d), regions that displayed significant differences with respect to wild type protein are highlighted with pink rectangle.

screening, the authors (Johnston Jennifer et al., 2012) recognized a *PIGA* germline nonsense mutation in two siblings to be associated with early epileptic encephalopathy and hypotonia, cleft palate, brain anomalies (myelination abnormalities and thin corpus callosum), cardiac anomalies and early death. In addition, recently, four clinical reports were also reported in patients with germline *PIGA* mutations displaying a wide range of phenotypes and clinical diagnoses (Belet et al., 2014; Kato et al., 2014; Swoboda et al., 2014; van der Crabben et al., 2014), inclusive of Ferro-Cerebro-cutaneous syndrome (FCCS) (Swoboda et al., 2014), multiple congenital anomalies-hypotonia-seizures syndrome 2 (MCAHS2) (Kato et al., 2014; van der Crabben et al., 2014), and West syndrome (Kato et al., 2014). Frameshift and missense mutations non-synonymous SNPs (nsSNPs) of *PIGA* gene were also reported, which abolished the function of the encoded protein (Brodsky, 2014; Nafa et al., 1995).

In this study, our primary interest is to prioritize the nonsynonymous SNPs (nsSNPs) that are most likely to alter the structure and function of the protein, since their appearance in more than half of gene

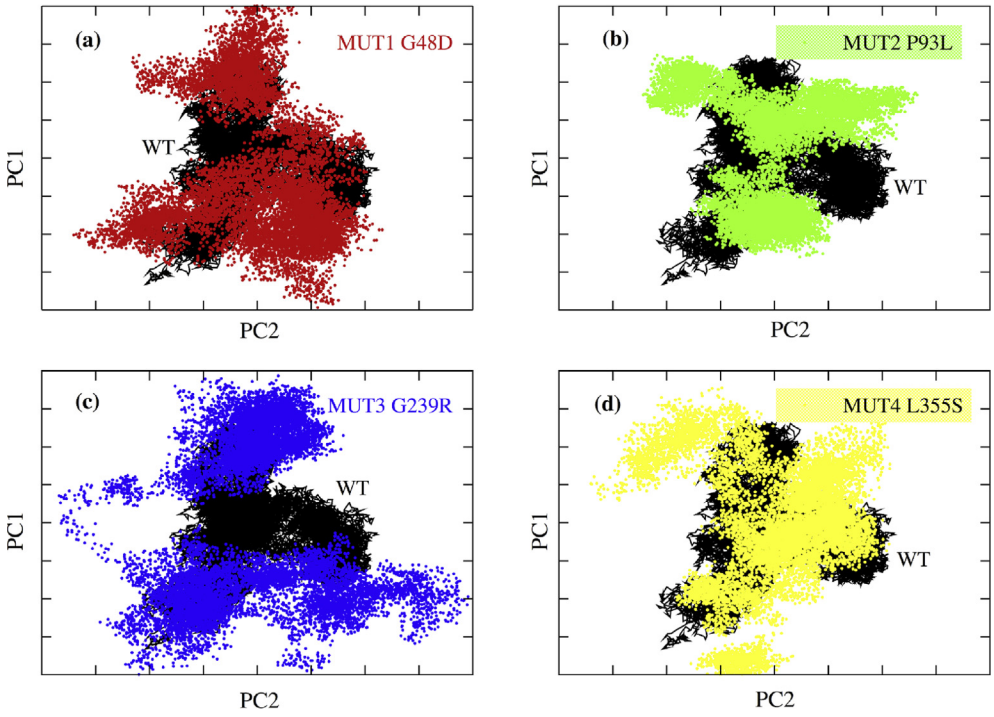

Fig. 7. Protein fluctuation along top two principal components (PC1, PC2) for WT and mutant systems over last 50 ns simulation period. Wild type (WT) protein is used as a reference system for comparison with (a) mutant G48D, in (b) for mutant P93L, in (c) for mutant G239R and in (d) for mutant L355S, respectively.

abnormalities, associated with the development of the disease (Krawczak et al., 2000), have been reported. To the best of our knowledge, this is the first comprehensive *in-silico* investigation in comprehending the impact of the PNH and MCAHS2 associated PIGA mutations on the protein structure and function. To do so, we applied a sequence-based and a structure-based analysis; along with the other tools that allowed integrating the features of both sequence and structure to prioritize the disease-associated nsSNPs of PIGA protein. We used a suite of many software such as, SIFT (Kumar et al., 2009), PolyPhen2 (Adzhubei et al., 2010), SNAP (Bromberg and Rost, 2007), mutationassessor (Reva et al., 2011), PROVEAN (Choi and Chan, 2015), I-Mutant3.0 (Capriotti et al., 2008), Align GVGD and (Tavtigian, 2005) to predict the pathogenicity level of 14 nsSNPs related to PIGA, which have been reported in the UNIPROT databases (UniProtKB – P37287).

Computer simulation, at the molecular level, has drawn considerable interest over the past decade (Elber, 2016), and serves as a complement to traditional experiments, permitting us to study something new, which cannot be determined in different ways. In this context, to obtain molecular insight into the structure-function relationship, we constructed the homology models for the PIGA protein using the Robetta server (Kim et al., 2004). The model structures were validated using RAMPAGE (Lovell et al., 2003) and the best structure was selected. Furthermore, for the most pathogenic mutations, *in-silico* mutation was performed to obtain the best model structure for the mutants. Finally, we performed the MD simulation of native and the four mutants for 50 ns each, using NAMD (Phillips et al., 2005). The proposed theoretical, computational approach allows a better comprehension of the atomic system of the functional SNPs related to the PIGA protein associated with PNH disease, and additionally a framework for the future experimental evaluation.

2. Material and methods

2.1. Data retrieval

We fetched the missense mutations data from UNIPROT (Bairoch, 1996) and HGMD databases (Stenson et al., 2013). The sequence information of PIGA protein (484 amino acids) was retrieved from the UNIPROT database (ID: P37287).

2.2. Screening of missense mutations

The analysis of the captured mutations was executed using pathogenic and stability based *in silico* prediction methods. SIFT (Kumar et al., 2009), PolyPhen-2 (Adzhubei et al., 2010), SNAP (Bromberg and Rost, 2007), mutationassesor (Reva et al., 2011), PROVEAN (Choi and Chan, 2015), I-Mutant 3.0 (Capriotti et al., 2008) and Align GV-GD (Tavtigian, 2005) tools were utilized for predicting the deleterious missense mutations, which might have a phenotypic effect. SIFT tool uses the sequence homology to predict if the mutation alters the function by designating amino-acid residues as functionally deleterious or neutral. SIFT program is based on the well funded hypothesis that essential amino acid residues will be conserved in protein family and alterations at well-conserved positions have a tendency to be predicted as deleterious. The SIFT tools provide a normalized probability score for each substitution. If the probability score is lesser than 0.05, then the observed mutation is categorized as deleterious, and for the score greater than 0.05 as neutral (Kumar et al., 2009). PolyPhen 2.0 utilizes sequence as well as structure-based attributes and uses a naive Bayesian classifier to predict the pathogenic effect of an amino acid substitution on stability and function of the protein. PolyPhen 2.0 functionally annotates the SNPs, extracts annotations and structural attributes of the protein sequence, maps coding SNPs to gene transcripts, and builds conservation profiles. Based on a combination of all these properties the probability of the missense mutation is computed. Based on the position specific independent count (PSIC) score difference, Polyphen 2.0 tool categorizes the mutants with scores between 0 to 1, such as benign [0, 0.02], possibly damaging [0.02, 0.85] and probably damaging [0.85, 1] (Adzhubei et al., 2010). SNAP uses the neural network methodology (a machine learning technique) to predict the effect of nsSNPs. It only requires the sequence information, however improves the prediction from structural as well as functional annotations, if accessible. SNAP additionally has incorporated the different class of solvent accessibility (buried, intermediate, and

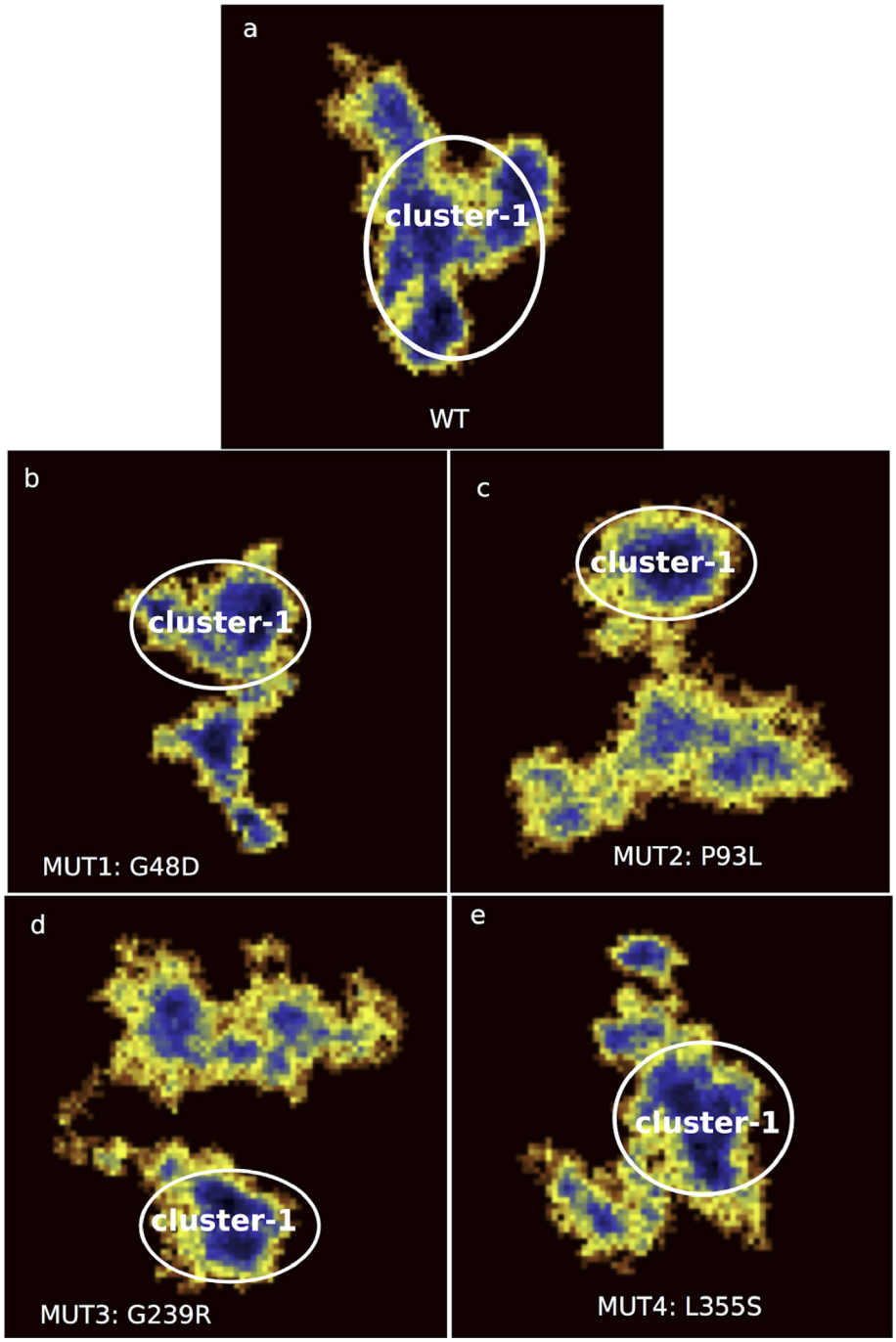

Fig. 8. Free energy landscape representation for the systems investigated. The most populated cluster (cluster-1) has been highlighted in each of the cases. In (a) for wild type, in (b) for mutant G48D, in (c) for mutant P93L, in (d) for mutant G239R and in (e) for mutant L355S, protein complexes.

exposed) for nsSNPs analysis. For each amino acid SNAP generates the three outputs, which are reliability index (RI, ranges from 0 to 9), binary prediction (neutral or non-neutral), and expected accuracy (Bromberg and Rost, 2007). Moreover, we used mutation assessor, which classified the mutations that show high, medium, and neutral impact on protein based on conserved evolutionary patterns (Reva et al., 2011). Additionally, we also used the PROVEAN tool to assess the pathogenic effect of the missense mutations, which classified the mutations into two categories, either deleterious or neutral. Next, we used the I-Mutant3.0 (Capriotti et al., 2008) to predict the stability of the protein after mutation. I-Mutant3.0 uses a support vector machine (SVM) algorithm to categorize the mutations based on their stability and computes the free energy change (DDG) of the mutant protein. A DDG score less than -0.5, represents that the mutation largely destabilizes the protein, greater than 0.5, represents that the mutation primarily stabilizes the protein and a DDG score $\geq$ -0.5 and $\leq$0.05; shows a weak effect on protein stability (Capriotti et al., 2008). Further, we predicted the pathogenic missense mutations based on a physicochemical property of amino acids by using the Align-GVGD server. The server calculates the degree of likeliness of a mutation to be deleterious or neutral based on the output spectrum ranges from C0, C15, C25, C35, C45, C55, and C65. In the spectrum, if the obtained score is C0, then the mutation is least likely to be deleterious, and if the score is C65, then the mutation is most likely to alter the function (Tavtigian, 2005). Furthermore, PredictSNP (Bendl et al., 2014), and istable (Chen et al., 2013) tools were also used to validate our findings.

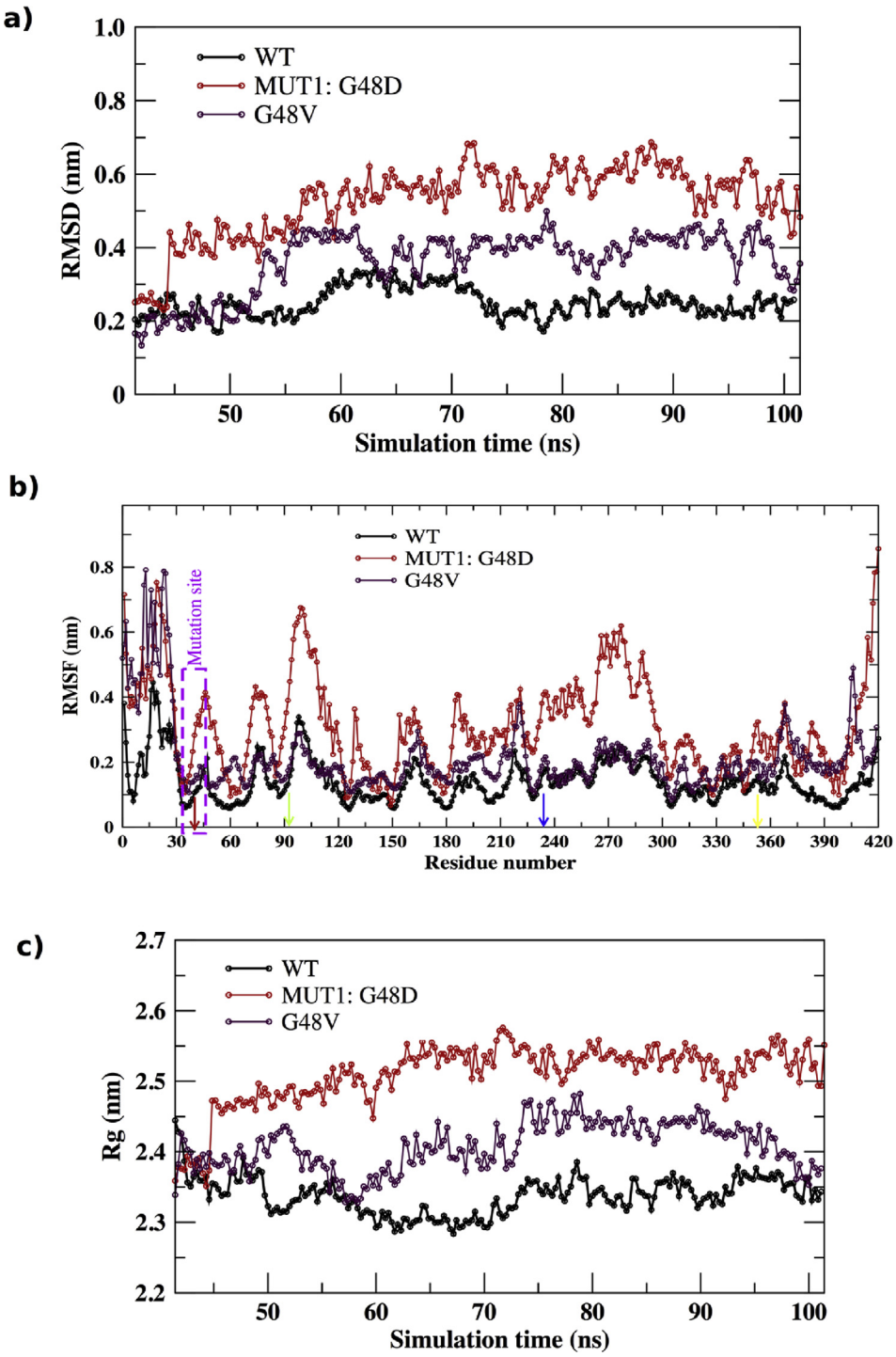

Fig. 9. Comparative analysis for the native (WT) protein and the mutants (G48V, G48D). In a) RMSD plot, b) RMSF plot, and the arrows indicate the position of other missense mutants and the region close to residue 48 has been highlighted with a pink box, and in c) radius of gyration plot.

2.3. Homology modeling

Because of the absence of three-dimensional crystal structure PIGA protein, we performed the homology modeling using Robetta server (Kim et al., 2004). The sequence length of the PIGA protein is 484 amino acids, whereas the model structure has 420 amino acids. The Robetta server generates the total five models for PIGA protein; we selected the best model based on Ramachandran plot validation obtained from RAMPAGE server (Lovell et al., 2003). The best model showed 96.9% residues in the favoured region, 2.9% residues in allowed region and 0.2% residues in outlier region, which suggest that the modeled structure is good enough to proceed for further analysis. Further, the mutations predicted to be most pathogenic were mapped to the best model structure using Swiss PDB Viewer (Schwede, 2003). Further, each structure was subjected to energy minimization to make the structure relaxed by removing any bad contacts and steric hindrance.

2.4. Molecular dynamics simulation

The missing hydrogen atoms in the modeled native PNH and four mutants: G48D, P93L, G239R, L355S (Fig. 1), were built using psfgen package of VMD software (Humphrey et al., 1996). Each system was then

immersed in a box of ~9250 water molecules, and counterions were added to obtain a neutral system. The initial dimension of the simulation box was 102 × 85 × 86 $Å^3$, for a total of ~70000 atoms. We used TIP3P parameters (Jorgensen et al., 1983) for water molecules and Charmm22 force-field parameters for protein atoms. The correct protonation state of the residues was assigned using Propka software (Rostkowski et al., 2011). Each individual complex was energy minimized and heated to 300 K in steps of 30 K with initial positional constraints of 50 kcal/(mol $Å^2$) on carbon alpha atoms. The positional constraints were then gradually released in steps of 10 kcal/(mol $Å^2$). After completely releasing the constraints an equilibration run of 3ns was performed. The simulations were performed for 100 ns in NPT ensemble with T = 300 K and 1 atm pressure. The analysis was performed on the last 50 ns. Further details of the parameters used in molecular dynamics (MD) simulations have been described in our previous works (Kumar et al. 2013, 2015, 2018). MD simulations were performed employing NAMD (Phillips et al., 2005) software package on 64 processors cluster (Kumar et al., 2014). Analysis was performed using Carma software package (Glykos, 2006).

3. Results and discussion

The aim of our study is to examine the potential mechanisms by which disease-related missense mutations may influence PIGA proteins, employing relevant computational tools, such as structural bioinformatics, molecular modeling, and MD simulations. Missense mutations have been related to several genetic diseases, though; some of the missense mutations do not alter the protein function, and hence does not show the disease consequences (Minde et al., 2011). Prevalent sequence-based predictors (e.g., SIFT (Ng and Henikoff, 2003) PolyPhen-2 (Adzhubei et al., 2010), SNAP (Bromberg and Rost, 2007), mutationassesor (Reva et al., 2011), PROVEAN (Choi and Chan, 2015), and Align GV-GD (Tavtigian, 2005) suggest whether or not a missense mutation can be deleterious to the function of the encoded protein. Whereas, structure-based predictors (e.g., I-Mutant 3.0 (Capriotti et al., 2008)) evaluate the protein structure for particular mechanistic alterations. The time-dependent and three-dimensional dynamical and structural information revealed by MD simulation adds value to sequence-based as well as structure-based computational predictions, and further allows a more detailed inference to the structures at the molecular level. In this study, by using a series of sequence as well as structure based *in-silico* tools, we screened the pathogenic mutations from the pool of missense mutations available in the UNIPROT (Bairoch, 1996), and HGMD (Stenson et al., 2013) databases.

3.1. In silico screening of mutants

Out of 14 missense mutations, SIFT predicted eight mutations as deleterious with a score of zero and three mutations with a score range of 0.1 to 0.05 (Table 1). PolyPhen-2 tool predicted 12 mutations as probably damaging with a score ranging between 0.85 and 1, and two mutations were predicted as benign with a score ranging between 0 and 0.2 (Table 1). SNAP predicted 13 mutations as non-neutral and one mutation as neutral. Moreover, mutation assessor predicted six mutations with highly destabilizing effect, six mutations with medium effect and two mutations with neural effect (Table 1). PROVEAN predicted 12 mutations as deleterious and two mutations as neutral (Table 1).

Further, I-Mutant 3.0 predicted seven mutations that could destabilize the protein structure with a delta G score less than -0.5. Rest of the seven mutations showed less impact on protein structure with delta G score ranging between -0.5 and 0.5. Hence, all the captured mutations predicted by I-Mutant3.0, could decrease the stability of the protein (Table 1). Align-GVGD predicted ten mutations as most deleterious and classified as class C65 (most likely to alter the function, see Methods) and the remaining four mutations as less likely to be interfering with the function of the protein (Table 1). Finally, consensus mutations that were predicted to be deleterious using different computational tools were found to be most pathogenic and utilized for further analysis. In detail, the four missense mutations G48D, P93L, G239R, and L355S, were predicted to be deleterious and destabilizing by *in silico* prediction methods, which could be pathogenic and disease-causing by altering the stability of the PIGA protein (Table1). Among these four mutations, G48D and G239R have been reported to be associated with paroxysmal nocturnal hemoglobinuria 1 (PNH) (Nafa et al., 1998). While, P93L and L355S mutants have been reported to be associated with multiple congenital anomalies-hypotonia-seizures syndrome 2 (MCAHS2) (Trump et al., 2016; van der Crabben et al., 2014). Furthermore, the four missense mutations were also confirmed using PredictSNP (Bendl et al., 2014), and istable (Chen et al., 2013) tools. Moreover, in previous studies relation of missense mutations to other diseases has been also reported (Ali et al., 2017; P et al., 2017).

3.2. Molecular dynamics simulation analysis

These four most pathogenic mutations obtained from *in-silico* tools were further mapped to the structure. Molecular dynamics (MD) simulation technique has been progressively utilized to reveal the impact of mutations over the stability of protein structure at atomistic level and the role of each residue in the native as well as in the mutant structures (Thirumal Kumar et al. 2019a, 2019b). Hence, to have the more detailed insight on the effect of mutations over the structure and dynamics of the protein, MD simulation of native wild type and the four pathogenic mutants (screened by series of *in-silico* tools) were performed. The resultant trajectory files generated from last 50ns run of MD simulation were subjected to several analyses, such as root mean Square Deviation (RMSD), root mean square fluctuations (RMSF), radius of gyration (Rg), and solvent accessible surface area (SASA) calculations as a preliminary step to analyse the convergence, protein stability, compactness, hydrophobic and hydrophilic nature of the protein systems. Furthermore, the overall variations in fluctuation were measured using the covariance matrix and principal component analysis (PCA).

3.3. Root mean square deviation

RMSD of carbon alpha (C-alpha) atoms of native and mutant proteins were analysed to examine the convergence, i.e. stable conformation of the trajectory files. We observed the lowest value of RMSD for the native case. Among the four mutants G48D displayed highest RMSD value, while the other three mutants displayed quite similar RMSD values (Fig. 2). A higher RMSD value indicates a decrease in the stability of the protein (Agrahari et al., 2017; Yun and Guy, 2011) and a lower RMSD value illustrates relatively stable protein structure ((Agrahari et al., 2018a; Agrahari et al. 2018b; Agrahari et al. 2017). Thus, with this preliminary and very simple tool the four-missense single point mutations investigated were found to have a potential impact on the protein stability.

3.4. Root mean square fluctuation

Fluctuations of the residues in the protein arise as a crucial element in determining the biological function, instructing that functional positions of the protein are often uniquely coupled with structural fluctuations (Eaton et al., 1991; Ikeguchi et al., 2005; Levy and Onuchic, 2006). At different time interval, the subtle and substantial flexibility differences can be correlated to the functional dynamics of the protein (Agrahari et al., 2018b). We calculated the flexibility of each residue of the native and all the mutants to inspect the local and overall dynamic changes.

The c- and n-terminal residues in all the cases display the highest fluctuation (Fig. 3). Overall, protein in the native form displayed lower flexibility than all the four mutants (G48D, P93L, G239R and L355S). Mutation G48D was found to influence protein flexibility, both on a local and global scale. Interestingly, this same mutant displayed highest RMSD value. Mutation P93L resulted mainly in local modification of protein

flexibility. For mutant G239R protein we note increase in fluctuation in the range of residues 150–180 and 300–330. Lastly, for mutation L355S we observed modification in protein flexibility for residues in ranges 90–100, 210–220 and 270–280. The RMSF analysis suggested that all the investigated mutants are able to alter the local and overall flexibility of the protein and in consequence change the stability, which further potentially change the function or the interactions with other proteins.

3.5. Radius of gyration

The radius of gyration (Rg) is an important parameter that allows measuring structural compactness, overall folding and shape of the protein. In Fig. 4, variation of Rg for PIGA native and mutant proteins at a different time interval has been shown. We examined the Rg to inspect the conformational changes and dynamic stability of native and mutant PIGA proteins. A lower Rg score designates better compactness of the protein structure (Agrahari et al., 2017; Lobanov et al., 2008; Sneha and George Priya Doss, 2016).

The Rg (Fig. 4) of native protein displayed the lowest value than the four mutants (G48D, P93L, G239R, and L355S). Mutant G48D protein displayed the highest value of Rg, consistent with highest RMSF and RMSD values. During the last 10 ns, mutant G239R and P39L proteins displayed similar values. While, mutant L355S displayed similar values of native for the last 10 ns of MD simulations. The Rg analysis suggested that the observed differences between the native and mutant cases, could affect the overall structure conformation and folding pattern of the protein.

3.6. Solvent accessibility surface area

Furthermore, we calculated the SASA to examine the behavior of hydrophobic and hydrophilic residues of PIGA protein. SASA predicts the residues present at the surface (hydrophilic) and residues present in the core of the protein (hydrophobic).

It has been shown previously that changes in the SASA pattern can explain the alteration in the structure of the protein (Agrahari and George Priya Doss, 2015; Agrahari et al., 2018b; Agrahari et al., 2017). We noted a similar trend for SASA (Fig. 5) as Rg. Native and mutant proteins (G48D, P93L, G239R, and L355S) were also found to exhibit quite similar pattern, with variation in their SASA values between 210-245 nm^2. The SASA analysis indicated all mutant protein systems to display subtle change of their solvent accessibility compared to the native. Thus, these subtle structural rearrangements could affect the stability and function of the PIGA protein.

3.7. Covariance matrix analyses

The covariance matrix calculation provides details of collective motions of the atoms in the protein instead of its local fluctuations. The knowledge of correlated motions is crucial to understand the critical biological functions, such as protein-ligand and protein-protein interactions as well as dynamics stability of protein (Ichiye and Karplus, 1991). Collective motion of the atoms plays a major role in the more substantial fluctuations of the protein atoms. Atoms with similar long-time behavior belong in a same collective group and move in a correlated manner (Ichiye and Karplus, 1991). Many covalent and non-covalent interactions interconnect atoms within the protein. These huge complex interconnected networks give rise to correlated dynamics of atoms, where interference or motion of one structural component displays the covariance with the positional displacements of other components. Hence, over a given time-scale, protein atoms exist in sub-states of a correlated ensemble that covers a huge configurational space. The amino acid mutations alter protein structural dynamics in the range of picosecond time in MD simulation studies. In this way, the knowledge of correlated motions of such a system is essential for understanding the structure-function relationships (Theobald and Wuttke, 2008).

The dynamic cross-correlation fluctuation calculation involves generation of covariance matrix whose elements (C_{ij}) can be represented as a cross-correlation map (Fig. 6). The cross-correlation coefficients were computed on C-alpha atoms of the last 50 MD trajectories of the native and mutant systems.

As expected, we observe a strong correlation along the diagonal for the native (Fig. 6a) as well as the mutant systems (Fig. 6b-e). However, we also observed strong correlated fluctuations for off-diagonal regions. The regions highlighted in pink for the mutant cases are the ones that exhibit significant differences with respect to the native simulations. Consistent with the above observations, for the mutant G48D (Fig. 6b) case we found many regions to display divergent correlated fluctuations with respect to the native one. However, we also observe some regions in the other three mutant cases to display different correlation with respect to the native protein. Thus, from the analysis of the correlation matrix, we can conclude that no mutant system showed a similar pattern of correlations as observed in the native protein. Thus, from the analysis of the correlation matrix, we can conclude that no mutant system showed a similar pattern of correlations as observed in the native protein. In a recent study, the authors reported that altered protein structural conformations were due to a change in the correlated movements and dynamics pattern of the atom pairs (Ndagi et al., 2017). In this context, the observed altered correlation confirms the altered structural conformation due the atomic rearrangements induced by the G48D, P93L, G239R, and L355S mutations.

The outcome of covariance analysis suggested that each mutation has the potential to affect the function by changing the structural conformation of the native PIGA protein.

3.8. Principle component analysis

The PCA was calculated using the outcome from the covariance matrix on 25000 frames of final 50ns MD simulation trajectory. The estimation of functionally significant global aggregate movement of the protein is a very demanding task. PCA minimizes the difficulty of identifying global aggregate motions of protein, as it filters collective motions (often slow) from the local fast motions (Amadei et al., 1993). These crucial atomic fluctuations of a protein can be directly associated with the dynamic stability and function of the protein.

The PCA plot (Fig. 7) allows representing the global dynamics of the protein in the essential subspace of the full system phase space (Kumar and Delogu, 2017). Each point represents fluctuations of protein during MD simulations. In general, we note the global fluctuations in the four mutant systems to cover larger subspaces. With respect to wild-type simulations, protein dynamics analysed in mutant proteins G48D, G239R and L355S covered much larger subspace along both PC1 and PC2 components, while only along PC2 for mutant P93L protein was observed. In this manner, PCA provides a clear picture of atomic fluctuations of native and all the mutants (G49D, P93L, G239R, and L355S) of PIGA protein. Moreover, the PCA analysis also proved that G49D, P93L, G239R, and L355S mutations, might have an impact on altering the structure and function of the PIGA protein. More substantial changes in the collective motions are liable to reduce the dynamic stability of a protein (Agrahari et al., 2017), which were detected in the case of mutants. Subsequently, free energy landscapes on the principal component planes defined by eigenvectors pair 1–2 were obtained for the native and mutant cases were obtained (Fig. 8). The color spectrum of energy landscape plot ranges from blue to red, where blue color dictates the global minima conformation associated to the highest stable state of the protein and the red color dictates the lower stable state of the protein. Cluster analysis was performed using the top three principal components to identify distinctive grouping of the protein conformations. Trajectory frames that constitute the core of the most populated cluster was denoted as cluster 1, which is considered to be the most relevant state and corresponds to a specific chemical configuration of the protein, as sampled during MD simulation. The RMSD from the average structure in cluster 1

was highest for G48D mutant case (~1.5 Å). Therefore, the altered conformation of the mutant proteins structure in comparison to the native protein suggested a changed functional behaviour of the protein (Agrahari et al., 2017; Kamaraj and Purohit, 2013; Nagasundaram et al., 2015; Tavtigian, 2005). FEL analysis outcomes were in concordance with the results of RMSF (Fig. 3) and Rg analysis (Fig. 4), SASA (Fig. 5), and covariance fluctuation (Fig. 6).

3.9. MD simulation analysis for mutant G48V

In order to validate our analysis, we performed simulation of a null mutant G48V complex and compared the results with missense mutant G48D. In detail, we provide a comparison between the WT, G48D (MUT1) and G48V systems, by analyzing root mean square deviation (RMSD, Fig. 9a), root mean square fluctuation (RMSF, Fig. 9b), radius of gyration (Rg, Fig. 9c).

It is evident from Fig. 9a that the impact of mutation G48V is less pronounced than G48D reflected from lower RMSD value. RMSF plot in Fig. 9b clearly indicates a much lower fluctuation for the G48V mutant system than G48D. In particular, for the missense mutant G48D mutant, we note higher fluctuations than G48V in the region close to mutation site. Furthermore, we note that G48V displays lower Rg than mutant G48D, which is consistent with lower RMSD and RMSF values. Therefore, our analysis indicated the impact of G48V mutation on the overall structural dynamics of the protein to be lesser than for G48D mutant system. This behavior was expected, as mutation from glycine (G) to aspartic acid (D) causes change in both size and hydrophobic nature, while G to valine (V) predominantly changes only in the size.

4. Conclusion

In the current study, we have performed molecular dynamics (MD) simulations to examine the impact of four most pathogenic mutations (G49D, P93L, G239R, and L355S) screened from a suite of *in-silico* tools. We explored several molecular properties of native and mutant proteins, such as atomic flexibility, compactness, and correlated motions. From these analyses, we can conclude that all the mutants result in altering the structural conformation and dynamical stability of the PIGA protein. The impact of null mutation G48V on the structural dynamics of the protein was significantly lower with respect to the G48D mutant. The impact of mutation G48D on the structural dynamics of the protein was found to be the highest.

In this way, our outcomes confirmed the pathogenic nature of (G49D and G239R) as well as (P93L, and L355S) mutations and their association to Paroxysmal nocturnal hemoglobinuria 1 (for G49D, P93L) and Multiple congenital anomalies-hypotonia-seizures syndrome 2 (for MCAHS2), respectively. The advancement of more modern computational prediction tools is required in the future for the assessment of variations, which will reveal insight into the potential genotype-phenotype relationship, and will aid in drug designing and novel personalized drug discovery for genetic diseases.

Declarations

Author contribution statement

A. Agrahari: Conceived and designed the experiments; Analyzed and interpreted the data; Wrote the paper.

E. Pieroni and G. Gatto: Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data.

A. Kumar: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Wrote the paper.

Funding statement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Competing interest statement

The authors declare no conflict of interest.

Additional information

No additional information is available for this paper.

Acknowledgements

A.K. thanks the high performance computing facility of CRS4 (Pula), Italy, for providing remote access to the computational resources. A special thanks to HPC staffs Carlo Podda, Marco Moro and Michele Muggiri at CRS4 for their continuous support and assistance.

References

Adzhubei, I.A., et al., 2010. A method and server for predicting damaging missense mutations. Nat. Methods 7, 248–249.
Agrahari, A., George Priya Doss, C., 2015. Impact of I30T and I30M substitution in MPZ gene associated with Dejerine–Sottas syndrome type B (DSSB): a molecular modeling and dynamics. J. Theor. Biol. 382, 23–33.
Agrahari, A.K., Kumar, A., R, S., Zayed, H., C, G.P.D., 2018a. Substitution impact of highly conserved arginine residue at position 75 in GJB1 gene in association with X-linked Charcot–Marie-tooth disease: a computational study. J. Theor. Biol. 437, 305–317.
Agrahari, A.K., Muskan, M., George Priya Doss, C., Siva, R., Zayed, H., 2018b. Computational insights of K1444N substitution in GAP-related domain of NF1 gene associated with neurofibromatosis type 1 disease: a molecular modeling and dynamics approach. Metab. Brain Dis. 33, 1443–1457.
Agrahari, A.K., Sneha, P., George Priya Doss, C., Siva, R., Zayed, H., 2017. A profound computational study to prioritize the disease-causing mutations in PRPS1 gene Metab. Brain. Dis. 33, 589–600.
Ali, S.K., Sneha, P., Priyadharshini Christy, J., Zayed, H., George Priya Doss, C., 2017. Molecular dynamics-based analyses of the structural instability and secondary structure of the fibrinogen gamma chain protein with the D356V mutation. J. Biomol. Struct. Dyn. 35, 2714–2724.
Amadei, A., Linssen, A.B.M., Berendsen, H.J.C., 1993. Essential dynamics of proteins Proteins. Struct., Funct., Genet. 17, 412–425.
Bairoch, A., 1996. The SWISS-PROT protein sequence data bank and its new supplement TREMBL. Nucleic Acids Res. 24, 21–25.
Belet, S., et al., 2014. Early frameshift mutation inPIGAIdentified in a large XLID family without neonatal lethality. Hum. Mutat. 35, 350–355.
Bendl, J., et al., 2014. PredictSNP: robust and accurate consensus classifier for prediction of disease-related mutations. PLoS Comput. Biol. 10, e1003440.
Bessler, M., et al., 1994. Paroxysmal nocturnal haemoglobinuria (PNH) is caused by somatic mutations in the PIG-A gene the. EMBO J. 13, 110–117.
Brodsky, R.A., 2008. Narrative review: paroxysmal nocturnal hemoglobinuria: the physiology of complement-related hemolytic anemia. Ann. Intern. Med. 148.
Brodsky, R.A., 2014. Paroxysmal nocturnal hemoglobinuria. Blood 124, 2804–2811.
Bromberg, Y., Rost, B., 2007. SNAP: predict effect of non-synonymous polymorphisms on function. Nucleic Acids Res. 35, 3823–3835.
Capriotti, E., Fariselli, P., Rossi, I., Casadio, R., 2008. A three-state prediction of single point mutations on protein stability changes. BMC Bioinf. 9.
Chen, C.W., Lin, J., Chu, Y.W., 2013. iStable: off-the-shelf predictor integration for predicting protein stability changes. BMC Bioinf. 14 (Suppl 2), S5.
Choi, Y., Chan, A.P., 2015. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. Bioinformatics 31, 2745–2747.
Eaton, W.A., Henry, E.R., Hofrichter, J., 1991. Application of linear free energy relations to protein conformational changes: the quaternary structural change of hemoglobin. Proc. Natl. Acad. Sci. 88, 4472–4475.
Elber, R., 2016. Perspective: computer simulations of long time dynamics. J. Chem. Phys. 144.
Glykos, N.M., 2006. Software news and updates. Carma: a molecular dynamics analysis program. J. Comput. Chem. 27, 1765–1768.
Humphrey, W., Dalke, A., Schulten, K., 1996. VMD: visual molecular dynamics. J. Mol. Graph. 14, 33–38.
Ichiye, T., Karplus, M., 1991. Collective motions in proteins: a covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations Proteins. Struct., Funct., Genet. 11, 205–217.
Ikeguchi, M., Ueno, J., Sato, M., Kidera, A., 2005. Protein structural change upon ligand binding: linear response theory. Phys. Rev. Lett. 94.
Johnston Jennifer, J., et al., 2012. The phenotype of a germline mutation in PIGA: the gene somatically mutated in paroxysmal nocturnal hemoglobinuria. Am. J. Hum. Genet. 90, 295–300.
Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W., Klein, M.L., 1983. Comparison of simple potential functions for simulating liquid water. J. Chem. Phys. 79, 926–935.

Kamaraj, B., Purohit, R., 2013. In SilicoScreening and molecular dynamics simulation of disease-associated nsSNP in TYRP1 gene and its structural consequences. OCA3 BioMed. Res. Int. 2013, 1–13.
Kato, M., et al., 2014. PIGA mutations cause early-onset epileptic encephalopathies and distinctive features. Neurology 82, 1587–1596.
Kim, D.E., Chivian, D., Baker, D., 2004. Protein structure prediction and analysis using the Robetta server. Nucleic Acids Res. 32, W526–W531.
Krawczak, M., Ball, E.V., Fenton, I., Stenson, P.D., Abeysinghe, S., Thomas, N., Cooper, D.N., 2000. Human Gene Mutation Database?A biomedical information and research resource. Hum. Mutat. 15, 45–51.
Kumar, A., et al., 2013. Identification of calcium binding sites on calsequestrin 1 and their implications for polymerization. Mol. Biosyst. 9.
Kumar, A., Delogu, F., 2017. Dynamical footprint of cross-reactivity in a human autoimmune T-cell receptor. Sci. Rep. 7.
Kumar, A., Melis, P., Genna, V., Cocco, E., Marrosu, M.G., Pieroni, E., 2014. Antigenic peptide molecular recognition by the DRB1–DQB1 haplotype modulates multiple sclerosis susceptibility. Mol. Biosyst. 10, 2043–2054.
Kumar, A., et al., 2018. Novel 2-pheynlbenzofuran derivatives as selective butyrylcholinesterase inhibitors for Alzheimer's disease. Sci. Rep. 8, 4424.
Kumar, A., Sechi, L.A., Caboni, P., Marrosu, M.G., Atzori, L., Pieroni, E., 2015. Dynamical insights into the differential characteristics of Mycobacterium avium subsp. paratuberculosis peptide binding to HLA-DRB1 proteins associated with multiple sclerosis. New J. Chem. 39, 1355–1366.
Kumar, P., Henikoff, S., Ng, P.C., 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat. Protoc. 4, 1073–1081.
Lee, S.C.-W., Abdel-Wahab, O., 2014. The mutational landscape of paroxysmal nocturnal hemoglobinuria revealed: new insights into clonal dominance. J. Clin. Investig. 124, 4227–4230.
Levy, Y., Onuchic, J.N., 2006. Mechanisms of protein assembly: lessons from minimalist models. Acc. Chem. Res. 39, 135–142.
Lobanov, M.Y., Bogatyreva, N.S., Galzitskaya, O.V., 2008. Radius of gyration as an indicator of protein structure compactness. Mol. Biol. 42, 623–628.
Lovell, S.C., et al., 2003. Structure validation by Cα geometry: ϕ,ψ and Cβ deviation Proteins. Struct., Funct., Bioinformat. 50, 437–450.
Medof, M.E., 1984. Inhibition of complement activation on the surface of cells after incorporation of decay-accelerating factor (DAF) into their membranes. J. Exp. Med. 160, 1558–1578.
Minde, D.P., Anvarian, Z., Rüdiger, S.G.D., Maurice, M.M., 2011. Messing up disorder: how do missense mutations in the tumor suppressor protein APC lead to cancer? Mol. Cancer 10.
Miyata, T., et al., 1993. The cloning of PIG-A, a component in the early step of GPI-anchor biosynthesis. Science 259, 1318–1320.
Miyata, T., Yamada, N., Iida, Y., Nishimura, J., Takeda, J., Kitani, T., Kinoshita, T., 1994. Abnormalities of PIG-A transcripts in granulocytes from patients with paroxysmal nocturnal hemoglobinuria. N. Engl. J. Med. 330, 249–255.
Nafa, K., Bessler, M., Castro-Malaspina, H., Jhanwar, S., Luzzatto, L., 1998. The spectrum of somatic mutations in thePIG-AGene in paroxysmal nocturnal hemoglobinuria includes large deletions and small duplications blood cells. Mol. Dis. 24, 370–384.
Nafa, K., Mason, P.J., Hillmen, P., Luzzatto, L., Bessler, M., 1995. Mutations in the PIG-A gene causing paroxysmal nocturnal hemoglobinuria are mainly of the frameshift type. Blood 86, 4650–4655.
Nagasundaram, N., Zhu, H., Liu, J., V, K., C, G.P., Chakraborty, C., Chen, L., 2015. Analysing the effect of mutation on protein function and discovering potential inhibitors of CDK4: molecular modelling and dynamics studies. PLoS One 10, e0133969.
Ndagi, U., Mhlongo, N.N., Soliman, M.E., 2017. The impact of Thr91 mutation on c-Src resistance to UM-164: molecular dynamics study revealed a new opportunity for drug design. Mol. Biosyst. 13, 1157–1171.
Ng, P.C., Henikoff, S., 2003. SIFT: predicting amino acid changes that affect protein function. Nucleic Acids Res. 31, 3812–3814.
Nishimura, J.-I., et al., 2004. Clinical course and flow cytometric analysis of paroxysmal nocturnal hemoglobinuria in the United States and Japan. Medicine 83, 193–207.
OMIM: 300818, PNH1. https://omim.org/entry/300818, 2019.
P, S., D, K.T., Tanwar, H., R, S., C, G.P.D., Zayed, H., 2017. Structural analysis of G1691S variant in the human filamin B gene responsible for larsen syndrome: a comparative computational approach. J. Cell. Biochem. 118, 1900–1910.
Parker, C.J., 2002. Historical aspects of paroxysmal nocturnal haemoglobinuria: 'defining the disease. Br. J. Haematol. 117, 3–22.
Phillips, J.C., et al., 2005. Scalable molecular dynamics with. NAMD J. Comput. Chem. 26, 1781–1802.
Reva, B., Antipin, Y., Sander, C., 2011. Predicting the functional impact of protein mutations: application to cancer genomics. Nucleic Acids Res. 39 e118-e118.
Rollins, S.A., Sims, P.J., 1990. The complement-inhibitory activity of CD59 resides in its capacity to block incorporation of C9 into membrane C5b-9. J. Immunol. 144, 3478–3483.
Rostkowski, M., Olsson, M.H.M., Søndergaard, C.R., Jensen, J.H., 2011. Graphical analysis of pH-dependent properties of proteins predicted using PROPKA. BMC Struct. Biol. 11.
Schwede, T., 2003. SWISS-MODEL: an automated protein homology-modeling server. Nucleic Acids Res. 31, 3381–3385.
Sneha, P., George Priya Doss, C., 2016. Molecular dynamics. In: Personalized Medicine. Advances in Protein Chemistry and Structural Biology, pp. 181–224.
Socié, G., et al., 1996. Paroxysmal nocturnal haemoglobinuria: long-term follow-up and prognostic factors the. Lancet 348, 573–577.
Stenson, P.D., Mort, M., Ball, E.V., Shaw, K., Phillips, A.D., Cooper, D.N., 2013. The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. Hum. Genet. 133, 1–9.
Swoboda, K.J., et al., 2014. A novel germline PIGA mutation in Ferro-Cerebro-Cutaneous syndrome: a neurodegenerative X-linked epileptic encephalopathy with systemic iron-overload. Am. J. Med. Genet. 164, 17–28.
Takeda, J., et al., 1993. Deficiency of the GPI anchor caused by a somatic mutation of the PIG-A gene in paroxysmal nocturnal hemoglobinuria. Cell 73, 703–711.
Tarailo-Graovac, M., et al., 2015. The genotypic and phenotypic spectrum of PIGA deficiency Orphanet. J. Rare Dis. 10.
Tavtigian, S.V., 2005. Comprehensive statistical study of 452 BRCA1 missense substitutions with classification of eight recurrent substitutions as neutral. J. Med. Genet. 43, 295–305.
Theobald, D.L., Wuttke, D.S., 2008. Accurate structural correlations from maximum likelihood superpositions. PLoS Comput. Biol. 4.
Thirumal Kumar, D., et al., 2019a. Computational and modeling approaches to understand the impact of the Fabry's disease causing mutation (D92Y) on the interaction with pharmacological chaperone 1-deoxygalactonojirimycin (DGJ). Adv. Protein Chem. Struct. Biol. 114, 341–407.
Thirumal Kumar, D., et al., 2019b. A computational method to characterize the missense mutations in the catalytic domain of GAA protein causing Pompe disease. J. Cell. Biochem. 120, 3491–3505.
Trump, N., et al., 2016. Improving diagnosis and broadening the phenotypes in early-onset seizure and severe developmental delay disorders through gene panel analysis. J. Med. Genet. 53, 310–317.
UniProtKB – P37287. https://www.uniprot.org/uniprot/P37287, 2019.
van der Crabben, S.N., et al., 2014. Expanding the spectrum of phenotypes associated with germlinePIGAmutations: a child with developmental delay, accelerated linear growth, facial dysmorphisms, elevated alkaline phosphatase, and progressive CNS abnormalities. Am. J. Med. Genet. 164, 29–35.
Yun, S., Guy, H.R., 2011. Stability tests on known and misfolded structures with discrete and all atom molecular dynamics simulations. J. Mol. Graph. Model. 29, 663–675.