Front. Biol. 2015, 10(4): 377–385
DOI 10.1007/s11515-015-1346-2

RESEARCH ARTICLE

Functional and structural characterization of missense mutations in *PAX6* gene

S. Udhaya Kumar, N. Priyanka, P. Sneha, C. George Priya Doss (✉)

Medical Biotechnology Division, School of Biosciences and Technology, VIT University, Vellore, Tamil Nadu 632014, India

Abstract **The *PAX6* gene belongs to the Paired box (PAX) family of transcription factors that is tissue specific and required for the differentiation and proliferation of cells in embryonic development. PAX6 regulates the pattern formation in early developmental stages. This function of PAX6 protein enables the successful completion of neurogenesis and oculogenesis in most animals such as mice, *Drosophila* and some other model organisms including humans. A variation in the sequence of *PAX6* gene may alter the function and structure of the protein. Such changes can produce adverse effects on functioning of the PAX6 protein which were clinically observed to occur in a broad range of ocular defects such as aniridia in humans. We employed *in silico* prediction methods such as SIFT, PolyPhen 2; I mutant 3.0, SNAP, SNPs&GO, and PHD-SNP to screen the pathogenic missense mutation in PAX6 and DNA binding sites by BindN and BindN + . Furthermore, we employed KD4V server to examine the structural and functional modifications that occur in the PAX6 protein as a result of mutation. Based on the results obtained from the *in silico* prediction methods, we carried out modeling analysis for V53L, I56T, G64V, and I87R to visualize the impact of mutation in structural context.**

Keywords PAX6, missense mutation, DNA-protein

Introduction

PAX6 is a member of vertebrate paired box family of genes. Paired box genes are tissue-specific transcription factors which play a significant role in early animal development (Maulbecker and Gruss, 1993). This gene contains two DNA binding sites, a paired domain (PD) and a homeodomain (HD) (usually partial or complete) (Puk et al., 2013; Mishra et al., 2002). PAX6 protein is involved in several developmental pathways such as that of the eye, brain and pancreas. The *PAX6* gene consists of 22k base pairs which comprises of 14 exons, an alternatively spliced exon 5a and codes for a protein of 422 amino acids. PAX6 contains two DNA binding domains (a paired domain and a homeodomain) along with a PST (proline/serine/threonine) transactivation domain (Davis et al., 2008). Several mutations have been associated with aniridia (Hill et al., 1991; Matsuo et al., 1993; Osumi et al., 1997; Mishra et al., 2002), which is a rare ocular disorder which causes incomplete formation of the iris. Aniridia has a global prevalence of 1 in 64000 to 1 in 96000. *PAX6* gene has been identified as one of the candidate genes in which if mutations arise can cause this disease (van Heyningen et al., 2002). Loss of function of the gene leads to lack of an eye, and the gain of function leads to an ectopic eye in the model organism *Drosophila* (Halder et al., 1995). Mutations resulted in several developmental anomalies in the model organisms, as well as in humans. To this end, we aimed to provide a functional and structural analysis of the common missense mutations in *PAX6* gene that result in disease conditions.

Single amino acid polymorphisms (SAPs) are found to be the most common and simplest type of genetic variations that have adverse clinical effects. Missense mutations are a type of SAPs that can cause maximum damage to the function of the protein. This is mainly because mutations in the encoded amino acids will have a direct impact on the structure and function of the protein. SAPs may undergo subnormal or abnormal folding and disrupts its function (McCulley et al.,

Received October 20, 2014; accepted December 28, 2014
Correspondence: C. George Priya Doss
E-mail: georgecp77@yahoo.co.in; georgepriyadoss@vit.ac.in

2005; Tzoulaki et al., 2005). Fifty-five percent of missense mutations are found to cause variant phenotypes. The phenotypes may range from mild iris defects to more severe classical aniridia such as optic nerve malformations, micropthalmia, and Peters anomaly (Hanson et al., 1994; Grønskov et al., 1999; van Heyningen et al., 2002; Azuma et al., 2003; McCulley et al., 2005). The identification of the missense mutations responsible for specific phenotype variations requires multiple testing of hundreds or thousands of SAPs in the candidate gene. Compared to experimental approaches, *in silico* methods have priority in characterizing the variants, as they can be employed for the systematic screening of representative variations in samples of the human population. *In silico* methods use alternative techniques for classification, either sequence or structure. These techniques provide quick and precise predictions of the deleterious amino acid substitutions that may have an impact on protein structure and activity on a large scale. In this context, we used the Sorting Intolerant From Tolerant (SIFT) (Ng and Henikoff, 2003), Polymorphism Phenotyping 2.0 (Adzhubei et al., 2010), I mutant 3.0, SNAP (Bromberg and Rost, 2007), SNPs&GO (Calabrese et al., 2009), and PHD-SNP (Capriotti et al., 2006) and KD4V (Wawrocka et al., 2013) server to study the conservation patterns and secondary structure features in the mutated protein. Prediction of DNA binding site is important to understand the binding of transcription factors (Stromo, 2000). We employed BindN (Wang et al., 2006) and BindN + (Wang et al., 2010) to predict the protein-DNA interactions in PAX6. To provide the impacts of mutations on protein structure, mutational analysis was performed using Swiss pdbViewer (Kaplan and Littlejohn, 2001).

Materials and methods

In this work, we have analyzed the functional and structural effects of the missense mutations that are found to occur most commonly in PAX6. Four-fifths of missense mutations were located in the paired domain of the protein and believed to cause changes in the binding affinity of the PAX6 protein to its target gene (Hanson et al., 1994; Azuma et al., 2003; Hanson et al., 1999). Owing to their clinical importance, we obtained the missense mutations that are commonly seen in the *PAX6* gene from recognized public SNP databases such as NCBI- dbSNP (Sherry et al., 1999) and UniProtKB (The UniProt Consortium, 2008). We analyzed a total of 55 missense mutations for their deleterious effect. For pathogenicity testing, we used six computational tools, namely, SIFT, PolyPhen 2.0, SNAP, SNPs&GO, and PHD-SNP. The PAX6 protein stability was tested using I mutant 3.0, and the physicochemical properties such as secondary structure-charatersitics and solvent accessibility features of the mutated protein were studied using KD4V server.

Pathogenicity testing

Sorting Intolerant From Tolerant (SIFT) is a computational tool which predicts whether a mutation is deleterious that may affect the protein function. The results are obtained by the calculation of degrees of conversation of the amino acid residues in comparison to closely related sequences by performing multiple sequence alignments. SIFT predicts the mutations as deleterious or intolerant and non-deleterious or tolerant. A SIFT score of $\leqslant 0.05$ indicates that the amino acid substitution is intolerant and, therefore, deleterious. Whereas a SIFT score of $\geqslant 0.05$ is considered to be tolerant and hence non-deleterious (Ng and Henikoff, 2003). PolyPhen2 is a tool that predicts the impact of an amino acid substitution present by comparing the structure and function of a protein. Multiple sequence alignments fed into the PSIC software (Position-Specific Independent Counts), and a profile matrix was calculated. Profile scores for both allelic variants are calculated by the PolyPhen2 software which then predicts depending on the difference in the scores. An SNP is considered as benign, possibly damaging and probably damaging when the results range between 0.00 and 0.14, 0.15–0.84 and 0.85–1 (Adzhubei et al., 2010). Screening for NonAcceptable Polymorphisms (SNAP) is a sequence, function and structure-based method which uses the neural network algorithm to predict the gain or loss of function of a protein, as a result, of SAPs (Bromberg and Rost, 2007). This tool was developed using SAP annotations from the Protein Mutant Database (PMD) (Nishikawa et al., 1994). SNAP is highly sensitive and proved to predict over 80% of non-acceptable polymorphisms with 77% accuracy (Bromberg et al., 2008) and classifies the mutation as neutral or non-neutral (deleterious). Each SNAP prediction comes with a reliability index that correlates with accuracy. PHD-SNP is another testing tool used for the classification of variants into disease related and non-disease related mutations. It is a support vector machine (SVM) based method that can identify if a new phenotype caused by mutation is susceptible to the disease condition. PHD-SNP categorized the mutations as either neutral or disease-associated polymorphisms (Capriotti et al., 2006). SNPs&GO predicts whether the mutations are deleterious or non-deleterious. SVM based tool in which protein functional annotations are used for prediction. SNPs&GO provides a reliability index along with a neutral or disease prediction (Calabrese et al., 2009).

Protein-DNA binding prediction

BindN tool provides information on protein-nucleic acid interactions and also helps to understand the function of these proteins from primary data sequence. This works on SMV based approach and utilizes three sequence features whcih includes the side chain pK_a value, molecular mass and hydrophobicity index of the given amino acid to predict the

binding efficiency of the DNA/RNA to the protein. The input is given in the form of FASTA format, the output report is recorded labels such as ' + ' for DNA binding site and ' − ' is given for non binding site (Wang et al., 2006). BindN + SMV based method which uses evolutionary information of the three biochemical features that were earlier described in the BindN. Combining these evolutionary features with PSSM (tool used to predict the conserved sequences in the amino acid sequence) showed more accuracy in making the DNA/RNA binding site prediction of the amino acid sequence (Wang et al., 2010). The results from these tools were further compared with the binding sites available in the PDBSUM.

Protein stability testing

I mutant 3.0 was used to measure the extent that the mutation affects the protein stability is important in protein structural studies (Capriotti et al., 2008). I mutant 3.0 detects the change in Gibbs free energy between the native and mutant structures of the protein. The sequence based version used in this study classifies the mutations into either neutral mutation ($-0.5 \leqslant DDG \leqslant 0.5$ kcal/mol) or large decrease mutation ($\leqslant -0.5$ kcal/mol) or large increase mutation (> 0.5 kcal/mol) resepectively. The value denotes the Gibbs free energy of the protein native or mutant structures (Capriotti et al., 2008).

KD4V

Highly conserved regions in proteins, that is, amino acids that are conserved over generations are indicative of structural and functional importance of the protein (Wawrocka et al., 2013). Knowledge of the structural and functional importance of distinct regions in the proteins helps us concentrate on these areas for drug development. In our study, we used KD4V computational tool to characterize the phenotypic changes of missense mutations. KD4V (Comprehensible Knowledge Discovery System for Missense Variant) server helps in identifying conservation patterns and secondary structure properties of a protein (Luu et al., 2012). The KD4V server uses Induction Logic Programming (ILP) and obtains information from both protein structure and sequence. Physicochemical properties such as size, charge, hydrophobicity, polarity and the accessibility of the protein along with a pathogenicity prediction are provided.

Structural analysis

Change in the amino acid sequences can bring about a conformational shift in the protein structure thereby affecting its dynamics in the system. We conducted a study of the conformational changes in the chance of specific mutations by obtaining mutant structures of the proteins. We performed mutations on the native protein PDB ID: 6PAX with 2.5 (Å) by using a computational tool Spdbv version 4.0. Swiss-Pdb viewer also known as DeepView is a computer application that helps analyze several proteins simultaneously (Kaplan and Littlejohn, 2001). Figure 1 displays the structural visualization of the PAX6 protein drawn using the software PyMOL version 1.7.2.1. Total energy of the variants and RMSD values were also calculated using this tool. The mutant structures thus obtained were then superimposed with the native protein and examined the impact of the mutations on the structure and function of the PAX6 protein.

Results

We performed an analysis of the common missense mutations seen in *PAX6* gene that leads to serious clinical manifestations, and a total of 55 missense mutations were collected from NCBI-dbSNP and UniProtKB.

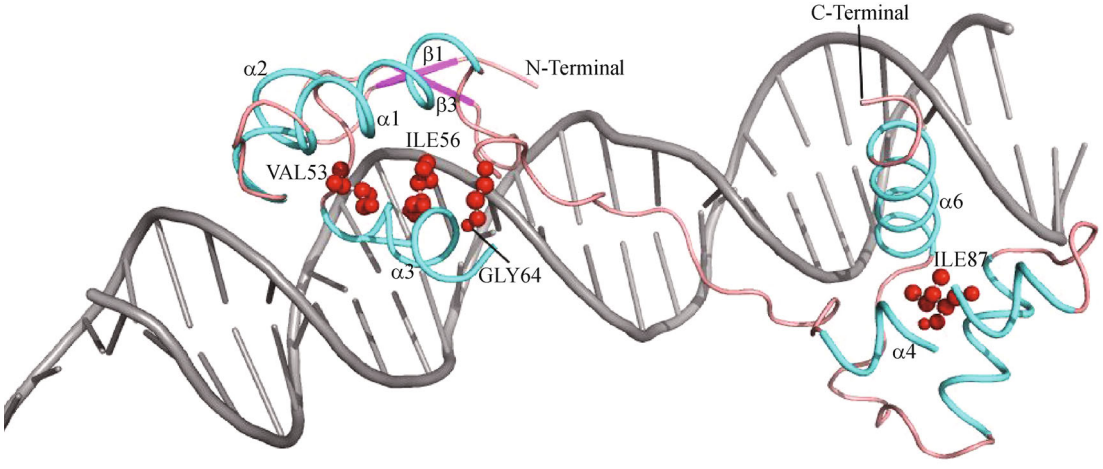

Figure 1 Mutational mapping in PAX6. The position of the four highly deleterious mutations have been mapped on the structure of PAX6 protein (PDB ID: 6PAX) using PyMOL 1.7.2.1. The DNA bound to the protein is visualized in cartoon shape in grey color followed by helices in cyan, β strands in magenta and the loop in salmon color. The native amino acids V53, I56, G64 and I87 are shown in red spheres.

Pathogenicity prediction analysis

Mutational analysis was performed for the 55 missense mutations as collected from databases, by testing for pathogenicity using various *in silico* prediction methods. Out of the 55 mutations, we obtained SIFT scores that characterized 39 mutations as deleterious and 16 as non-deleterious. The SIFT analysis indicated that 79.90% of the mutations are potentially harmful, and 29.09% are harmless. scores characterized 45 of the mutations as deleterious (possibly damaging (10) and probably damaging (36) and the rest nine as non-deleterious. Analysis of the PSIC scores showed us that 81.81% are possibly or probably damaging mutations whereas, 18.18% are benign. SNAP tool scores indicate 40 mutations as non-neutral (deleterious) and 15 mutations as neutral (non-deleterious), which mean that SNAP predicts 72.72% of the mutations as deleterious and 27.27% as non-deleterious. On the other hand, SNPs&GO and PhD-SNP predicted 55 and 30 mutations as disease (deleterious).

Protein stability prediction analysis

Protein stability testing is vital to understand the impact of the mutations on the secondary structure of the protein and the changes caused in free energy of the protein macromolecule. According to I Mutant 3.0, 38 mutations are deleterious, and 17 of the mutations are non-deleterious. The stability of the protein is negatively affected in 69.09% of the mutations while the rest 30.90% do not have any effect on stability of the protein. After a thorough analysis of the scores obtained in each computational tool, the mutations predicted to be deleterious in all the tools simultaneously were picked out (Table 1).

KD4V server

Four missense mutations V53L; I56T, G64V, and I87R were found to be highly deleterious in all the tools used for the study. These four highly deleterious mutations were then queried in the KD4V server to understand their robustness with a detailed analysis of the physicochemical properties of the mutant structures. Table 2 shows the change in the size, charge, hydrophobicity, polarity, difference in the accessibility of the native and deleterious mutants of PAX6 protein. This information will support the impact of mutations on the structure and functionality of the PAX6 protein.

DNA binding prediction: *PAX6* gene functions as a transcription factor (Mishra et al. 2002) and hence understanding the DNA binding sites becomes a prime importance. Computational methods BindN and BindN + tools were used to predict the DNA binding sites available for *PAX6* gene. BindN tool predicted about 159 residues to have DNA binding site, whereas a more accurate BindN + tool predicted that 62 residues have DNA binding site. The results obtained were further compared with already predicted binding sites in the PDBSUM. By comparing PDBSUM, BindN, and BindN + a total of 18 residues in PAX6 were found to be in DNA binding site as tabulated in Table 3.

In silico mutational analysis

We used PDB ID: 6PAX (4 – 136 amino acids) as the initial structure for modeling analysis. Mutational analysis was carried out for four missense mutations namely V53L, I56T, G64V, I87R that showed highly deleterious mutations using above *in silico* prediction tools. We obtained mutant protein using SPDB Viewer and superimposed the native and mutant structures using PyMOL. The total energy of native and

Table 1 Summary of nsSNP along with their prediction scores of by computational methods in *PAX6* gene

SNPs/Variants	Position	SIFT	PolyPhen2	I Mutant 3.0	SNAP	SNPs&GO	PhD-SNP
VAR_003808	N17S	0.04	0.535	Large decrease	Non-neutral	Disease	Disease
VAR_003809	G18W	0	1	Large increase	Non-neutral	Disease	Disease
VAR_047860	R19P	0	1	Large decrease	Non-neutral	Disease	Neutral
VAR_003810	R26G	0	1	Decrease	Non-neutral	Disease	Disease
VAR_008694	I29S	0	1	Large decrease	Non-neutral	Disease	Neutral
VAR_003811	I29V	0	0.996	Large increase	Neutral	Disease	Neutral
VAR_008695	A33P	0	1	Large decrease	Non-neutral	Disease	Neutral
VAR_008697	I42S	0	0.999	Large decrease	Non-neutral	Disease	Neutral
VAR_008698	S43P	0	0.999	Large increase	Non-neutral	Disease	Disease
VAR_003812	R44Q	0	0.999	Large decrease	Non-neutral	Disease	Neutral
VAR_047861	L46R	0	1	Large increase	Non-neutral	Disease	Disease
VAR_047862	C52R	0	1	Large increase	Non-neutral	Disease	Disease
VAR_008700	V53D	0	1	Large Increase	Non-neutral	Disease	Disease
VAR_008699	**V53L**	**0**	**0.999**	**Large decrease**	**Non-neutral**	**Disease**	**Disease**
VAR_047863	**I56T**	**0**	**1**	**Large decrease**	**Non-neutral**	**Disease**	**Disease**
VAR_008701	T63P	0	1	Large decrease	Non-neutral	Disease	Neutral
VAR_008702	**G64V**	**0**	**1**	**Large decrease**	**Non-neutral**	**Disease**	**Disease**
VAR_017540	P68S	0	1	Decrease	Non-neutral	Disease	Disease

(*Continued*)

SNPs/Variants	Position	SIFT	PolyPhen2	I Mutant 3.0	SNAP	SNPs&GO	PhD-SNP
VAR_047864	G73D	0	1	Large Increase	Non-neutral	Disease	Disease
VAR_008703	A79E	0	1	Large Decrease	Non-neutral	Disease	Disease
rs372222637	V83I	0	0.998	Large decrease	Neutral	Disease	Neutral
VAR_047865	I87K	0	0.989	Large decrease	Non-neutral	Disease	Neutral
VAR_003813	**I87R**	**0**	**1**	**Large decrease**	**Non-neutral**	**Disease**	**Disease**
VAR_015065	P118R	0	1	Large increase	Non-neutral	Disease	Disease
VAR_008704	S119R	0	1	Large decrease	Non-neutral	Disease	Disease
VAR_017541	R125C	0	0.214	Large decrease	Non-neutral	Disease	Disease
VAR_008705	V126D	0	0.999	Large decrease	Non-neutral	Disease	Disease
VAR_003814	R128C	0	0.999	Large decrease	Non-neutral	Disease	Disease
rs371018133	M137I	0.36	0.011	Large decrease	Neutral	Disease	Neutral
rs201983312	Y143C	0	0.926	Large decrease	Neutral	Disease	Disease
rs372143889	R159C	0.06	0.007	Large increase	Non-neutral	Disease	Disease
rs141021880	T166S	0.36	0	Large decrease	Neutral	Disease	Neutral
rs151086737	S167L	0.43	0.012	Large decrease	Non-neutral	Disease	Neutral
VAR_003815	Q178H	0.33	0.77	Large decrease	Neutral	Disease	Neutral
rs138803897	E181K	0.09	0.001	Large decrease	Neutral	Disease	Neutral
rs374396492	G194R	0.01	0.791	Large increase	Neutral	Disease	Disease
VAR_008706	R208Q	0	1	Large increase	Non-neutral	Disease	Disease
VAR_003816	R208W	0	1	Large increase	Non-neutral	Disease	Disease
VAR_047866	R242T	0	0.992	Large increase	Non-neutral	Disease	Disease
VAR_017542	F258S	0	0.986	Large decrease	Non-neutral	Disease	Disease
rs201846044	R267G	0.01	1	Large increase	Non-neutral	Disease	Disease
rs199610944	R267I	0	1	Large decrease	Non-neutral	Disease	Neutral
rs201439078	K270E	0.01	0.751	**NA**	Non-neutral	Disease	Disease
VAR_017543	S292I	0.39	0.907	Large decrease	Non-neutral	Disease	Disease
VAR_047867	A321T	0.56	0.736	Large decrease	Non-neutral	Disease	Neutral
VAR_008707	S353A	0.27	0.214	Large increase	Neutral	Disease	Neutral
VAR_017544	S363P	0.08	0	Large increase	Neutral	Disease	Neutral
rs201690439	T373S	0.73	0.232	Large decrease	Neutral	Disease	Neutral
VAR_015066	P375Q	0.1	0.498	Large decrease	Neutral	Disease	Disease
VAR_017545	Q378R	0.73	0.22	Large decrease	Neutral	Disease	Neutral
VAR_017546	M381V	0.51	0	Large decrease	Neutral	Disease	Neutral
VAR_047868	G387D	0.71	0.058	Large decrease	Non-neutral	Disease	Neutral
VAR_017547	T391A	0.62	0.001	Large decrease	Neutral	Disease	Neutral
VAR_067698	G395R	0	0.891	Large increase	Non-neutral	Disease	Neutral
VAR_008708	Q422R	0	0.993	Large decrease	Non-neutral	Disease	Neutral

NA-Not available

mutant proteins V53L, I56T, G64V, I87R were found to be −7356.003, −7580.859, −7566.213, −7498.595, and −7502.002 respectively. A comparative study of the structure of native and the mutant was carried out by calculating Root Mean Square Deviation (RMSD) using Swiss pdbViewer. The RMSD values between native and V53L, I56T, G64V, I87R were calculated as 1.36Å, 1.3Å, 1.2Å and 1.2 Å respectively with a range of 1–2 Å This superimposition was carried out for all the four highly deleterious mutants and polar contact interactions of V53L, I56T, G64V, and I87R is displayed in Fig. 2.

Discussion

Missense mutations commonly cause adverse effects on the structure and functioning of the PAX6 was obtained from db-SNP database of NCBI and UniProt KB. Fifty-five mutations were subjected to extensive pathogenicity and protein stability testing using various computational prediction methods. For pathogenicity testing we used, SIFT, PolyPhen 2.0, SNAP, SNPs&GO, and PHD-SNP all of which grouped the mutations into disease-causing and non- disease-causing mutations. For protein stability testing, we used I mutant 3.0

Table 2 Prediction of physico-chemical properties by KD4V server

Mutation	Size	Charge	Hydrophobicity	Polarity	Accessibility	Prediction
V53L	Increase	Unchanged	Unchanged	Unchanged	1.51(Buried) 1.34(Buried)	Deleterious
I56T	Decrease	Unchanged	Decrease	Increase	11.26 (Intermediate) 12.92 (Intermediate)	Deleterious
G64V	Increase	Unchanged	Increase	Unchanged	32.07 (Accessible) 52.04 (Accessible)	Deleterious
I87R	Increase	Increase	Decrease	Increase	0.07 (Buried) 0.00 (Buried)	Deleterious

Table 3 DNA Binding site determination using PDBSUM, BindN and BindN + tool

Mutation	DNA binding site (PDBSUM)	BindN	Confidence score	BindN +	Confidence score
3S	Present	Present	0.5244	Present	0.6925
22R	Present	Present	0.647	Present	0.6885
29R	Present	Present	0.871	Present	0.5622
41R	Present	Present	0.7567	Present	0.47
52S	Present	Present	0.5889	Present	0.4876
57S	Present	Present	0.779	Present	0.5972
58K	Present	Present	0.8645	Present	0.7604
62R	Present	Present	0.6959	Present	0.7972
72R	Present	Present	0.5116	Present	0.5668
77S	Present	Present	0.8737	Present	0.9198
78K	Present	Present	0.22	Present	0.9309
80R	Present	Present	0.3559	Present	0.9521
111S	Present	Present	0.5484	Present	0.93
124S	Present	Present	0.5594	Present	0.5916
125S	Present	Present	0.4949	Present	0.2874
127W	Present	Present	0.4313	Present	0.703
128R	Present	Present	0.811	Present	0.5084
137R	Present	Present	0.7014	Present	0.4059

to predict the change in Gibbs free energy between the native and mutant proteins thereby detecting change in the stability of the proteins. These two tools, also grouped the mutations into deleterious and non-deleterious. With these predictions, we were able to narrow down the mutations that were predicted to be susceptible to cause a disease condition. The mutations that were found to be deleterious in all the computational methods simultaneously were termed as, "Highly Deleterious" mutations which were then subjected to further analysis by the KD4V server. The KD4V server can identify conservation patterns in missense mutations as well as characterize the physicochemical properties and accessibility of the mutant proteins. Using KD4V server, we studied the physicochemical properties such as size, charge, hydrophobicity and polarity of the highly deleterious missense mutations V53L, I56T, G64V, and I87R. From this result, we recognized the change in the accessibility of the mutant proteins as opposed to the accessibility of the wild type PAX6 protein. V53L-The valine at position 53 is a hydrophobic amino acid which when compared to leucine has the same structure but a shorter side chain. The valine residue in the native protein interacts with L57, K55, C52 (α-helix 1), S49 (loop) and N50 (α-helix 2), thereby stabilizing the protein structure. This interaction is however not affected when the valine gets substituted by a leucine. In this case, the mutation is pathogenic even though the wild-type amino acid is replaced by another amino acid of similar properties. I56T-At position 56, the wild-type amino acid isoleucine is being substituted with threonine. The hydrophobic amino acid isoleucine in the native protein is buried in the structure and interacts with Y60, R59 and C52 (α-helix 3). When substituted for the hydrophilic amino acid threonine, another interaction is introduced with the hydrophilic amino acid L57 (α-helix 3). The introduction of a hydrophilic amino acid in the place of a hydrophobic amino acid can cause a structural change as the hydrophilic residue will try to attain a more exposed position in the structure. Further, a hydrophobic–hydrophilic mismatch can also cause the helix distortion in the protein. G64V-The glycine at position 64 in the native protein is part of α-helix 3 of the protein's secondary structure. This glycine that is ambivalent interacts with T63, Y60 (α-helix 3) and S65 (loop) playing a role in the protein structure stabilization. When substituted for a hydrophobic amino acid valine, this stabilization could be lost. Further, a

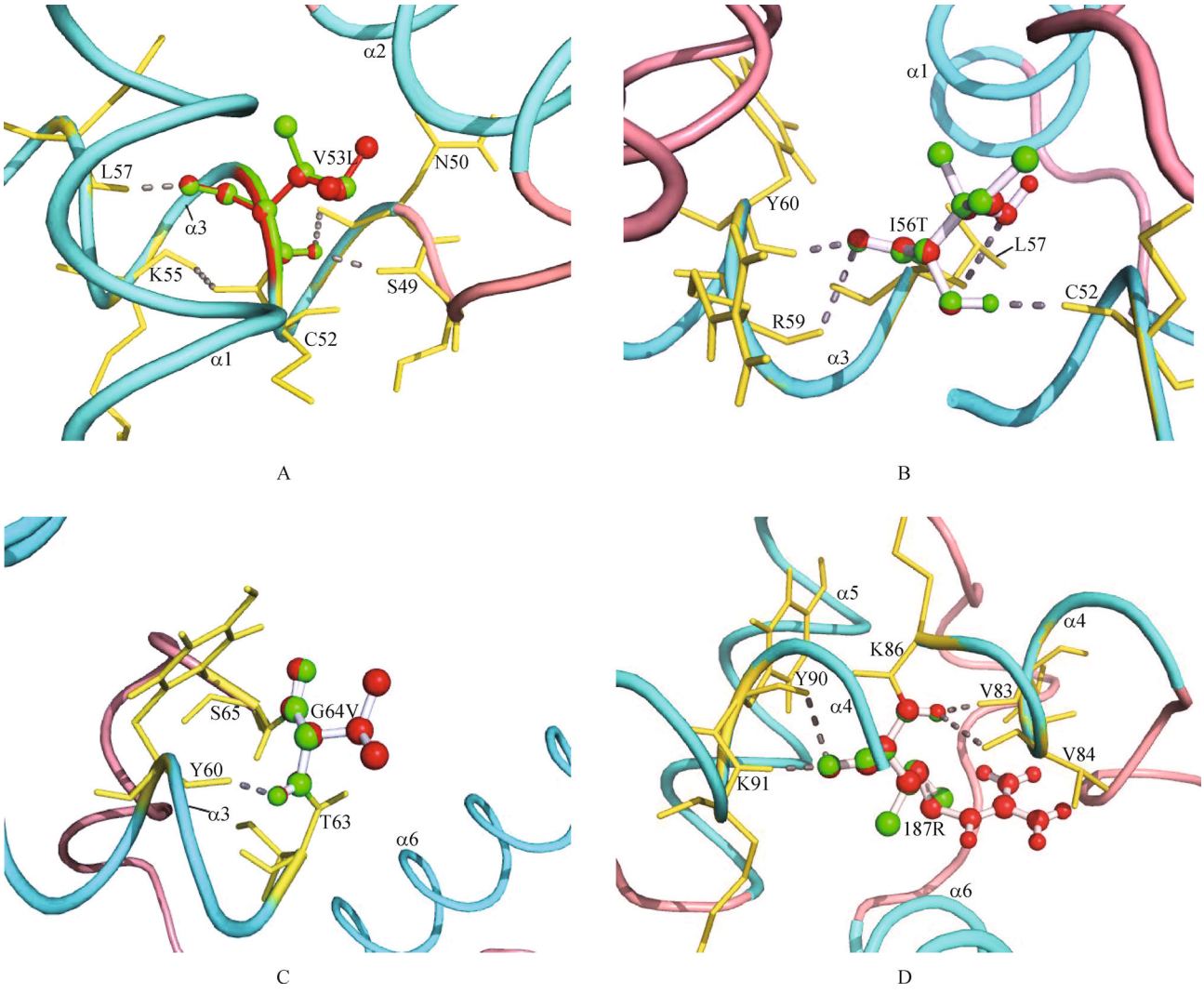

Figure 2 Closer view of the deleterious mutations and their interactions. The structure of the PAX6 protein is shown using cartoon feature in PyMOL 1.7.2.1. The helices were shown in cyan colour, the β stands in magenta and the loop in salmon color. Further, the helices and β strands were labeled to better understand the location of the mutations in the protein structure. (A) At position 53, the native amino acid valine is substituted with leucine. The native and mutant residues are undergoing polar contact with the surrounding amino acids L57, K55, C52, S49 and N50. (B) At position 56, the native amino acid isoleucine is substituted with threonine. The native residue undergoes polar contact with Y60, R59 and C52, whereas the mutant residue undergoes polar contact with L57. (C) At position 64, the native amino acid glycine is substituted with valine. The native and mutant residues undergo polar contact with the surrounding amino acid Y60. Further the amino acids S65 and T63 are close to the affected residue and will be affected by variations occurring in that position. (D) At position 87, the native amino acid isoleucine is substituted with arginine. At this position, the native residue undergoes polar contact with the surrounding amino acids Y90, K91, V83 and V84.

smaller amino acid when replaced with a larger amino acid can also cause change in the positioning of nearby amino acids disrupting the overall protein structure. I87R-The uncharged isoleucine at position 87 is found to interact with K91, Y90, V83 and V84 (α-helix 4) in the native protein. When substituted with a larger and positively charged amino acid, these interactions are maintained. However, the sudden introduction of a charged amino acid in the place of an uncharged amino acid can cause the helix distortion within the protein. With the cumulative results of all these tools, it is evident that these four mutations can alter the structure and functionality of PAX6 protein. To understand the change at the structural level, we performed *in silico* mutational analysis using Swiss pdb Viewer and generate the four mutant modeled structures. With the help of PyMOL, we superimposed the native and mutant amino acid to understand the changes taking place within the protein macromolecule. As it is evident in Fig. 2, the superimposed structures clearly points out the change in the local environment of the amino acid sequence of the protein. There were also observable changes seen in the total energy of the variants from the native protein, where the total energy of native protein was

found to be −7356.003 and variants V53L, I56T, G64V, I87R with total energy as −7580.859, −7566.213, −7498.595, −7502.002 respectively. The RMSD values between native and V53L, I56T, G64V, I87R were calculated to be 1.36Å, 1.3Å, 1.2Å and 1.2 Å respectively which has a range of 1–2 Å and may further affect the functionality or stability of the protein. These changes will in turn influence the overall functionality of the protein. DNA binding site in the amino acids sequence is further analyzed using various *in silico* methods. BindN tool showed 159 residues that had protein-nucleic interaction, whereas a more accurate tool BindN + which utilizes PSSM in combination with the evolutionary biochemical features showed 52 residues to be DNA binding regions. These results were compared with the PDBSUM, which showed around 22 residues to be in DNA binding residues (between the amino acids 4 and 133). Conformational change that occurred in the protein sequence either disrupts the binding ability of the protein to its gene targets or the function of the protein as a regulator of transcription. The latter is seen in the case of *PAX6* gene as it is a master control gene. The protein product of this gene due to mutational changes does not perform its function as a transcription factor and also the ability of the DNA binding is also found to be affected due to these mutations. This causes either a complete or partial disruption of the developmental pathways, especially to that of the eye resulting in conditions such as aniridia. Hence, our study of mutations and their structural impact on PAX6 will provide a basis for further development of drugs and treatments specific to diseases such as aniridia.

Acknowledgements

The authors take this opportunity to thank the management of VIT University for providing the facilities and encouragement to carry out this work. The authors have no conflict of interest to declare.

Compliance with ethics guidelines

This article does not contain any studies with human or animal subjects performed by any of the authors.

References

Acharya V, Nagarajaram H A (2012). Hansa: an automated method for discriminating disease and neutral human nsSNPs. Hum Mutat, 33 (2): 332–337

Adzhubei I A, Schmidt S, Peshkin L, Ramensky V E, Gerasimova A, Bork P, Kondrashov A S, Sunyaev S R (2010). A method and server for predicting damaging missense mutations. Nat Methods, 7(4): 248–249

Azuma N, Yamaguchi Y, Handa H, Tadokoro K, Asaka A, Kawase E, Yamada M (2003). Mutations of the *PAX6* gene detected in patients with a variety of optic-nerve malformations. Am J Hum Genet, 72(6): 1565–1570

Bromberg Y, Rost B (2007). SNAP: predict effect of non-synonymous polymorphisms on function. Nucleic Acids Res, 35(11): 3823–3835

Bromberg Y, Yachdav G, Rost B (2008). SNAP predicts effect of mutations on protein function. Bioinformatics, 24(20): 2397–2398

George Priya Doss C, Rajith B, Chakraborty C (2013). Predicting the impact of deleterious mutations in the protein kinase domain of FGFR2 in the context of function, structure, and pathogenesis—a bioinformatics approach. Appl Biochem Biotechnol, 170(8): 1853–1870

Calabrese R, Capriotti E, Fariselli P, Martelli P L, Casadio R (2009). Functional annotations improve the predictive score of human disease-related mutations in proteins. Hum Mutat, 30(8): 1237–1244

Capriotti E, Calabrese R, Casadio R (2006). Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. Bioinformatics, 22(22): 2729–2734

Capriotti E, Fariselli P, Rossi I, Casadio R (2008). A three-state prediction of single point mutations on protein stability changes. BMC Bioinformatics, 9(2 Suppl 2): S6

Davis L K, Meyer K J, Rudd D S, Librant A L, Epping E A, Sheffield V C, Wassink T H (2008). Pax6 3′ deletion results in aniridia, autism and mental retardation. Hum Genet, 123(4): 371–378

George Priya Doss C, Rajith B, Garwasis N, Mathew P R, Raju A S, Apoorva K, William D, Sadhana N R, Himani T, Dike I P (2012). Screening of mutations affecting protein stability and dynamics of FGFR1—A simulation analysis. App Transl Genomics, 1: 37–43

Grønskov K, Rosenberg T, Sand A, Brøndum-Nielsen K (1999). Mutational analysis of *PAX6*: 16 novel mutations including 5 missense mutations with a mild aniridia phenotype. Eur J Hum Genet, 7(3): 274–286

Halder G, Callaerts P, Gehring W J (1995). Induction of ectopic eyes by targeted expression of the eyeless gene in Drosophila. Science, 267 (5205): 1788–1792

Hanson I, Churchill A, Love J, Axton R, Moore T, Clarke M, Meire F, van Heyningen V (1999). Missense mutations in the most ancient residues of the PAX6 paired domain underlie a spectrum of human congenital eye malformations. Hum Mol Genet, 8(2): 165–172

Hanson I M, Fletcher J M, Jordan T, Brown A, Taylor D, Adams R J, Punnett H H, van Heyningen V (1994). Mutations at the PAX6 locus are found in heterogeneous anterior segment malformations including Peters' anomaly. Nat Genet, 6(2): 168–173

Hill R E, Favor J, Hogan B L, Ton C C, Saunders G F, Hanson I M, Prosser J, Jordan T, Hastie N D, van Heyningen V (1991). Mouse small eye results from mutations in a paired-like homeobox-containing gene. Nature, 354(6354): 522–525

Kaplan W, Littlejohn T G (2001). Swiss-PDB Viewer (Deep View). Brief Bioinform, 2(2): 195–197

Lill M A, Danielson M L (2011). Computer-aided drug design platform using PyMOL. J Comput Aided Mol Des, 25(1): 13–19

Luu TD, Rusu A, Walter V, Linard B, Poidevin L, Ripp R, Moulinier L, Muller J, Raffelsberger W, Wicker N, Lecompte O, Thompson JD, Poch O, Nguyen H (2012). KD4V: comprehensible knowledge discovery system for missense variant. Nucl Acids Res, W71–W75

Matsuo O Y N. Noji S, Ohuchi H, Koyama E, Myokai F, Matsuo N, Taniguchi S, Doi H, Iseki S, Ninomiya Y, Fujiwara M,Watanabe T , Eto K (1993). A mutation in the *pax6* gene in rat small eye is

associated with impaired migration of mid-brain crest cells. Nat Genet, 3: 299–304

Maulbecker C C, Gruss P (1993). The oncogenic potential of *Pax* genes. EMBO J, 12(6): 2361–2367

McCulley T J, Mayer K, Dahr S S, Simpson J, Holland E J (2005). Aniridia and optic nerve hypoplasia. Eye (Lond), 19(7): 762–764

Mi H, Guo N, Kejariwal A, Thomas P D (2007). PANTHER version 6: protein sequence and function evolution data with expanded representation of biological pathways. Nucleic Acids Res, 35 (Database issue): D247–D252

Mishra R, Gorlov I P, Chao L Y, Singh S, Saunders G F (2002). PAX6, paired domain influences sequence recognition by the homeodomain. J Biol Chem, 277(51): 49488–49494

Ng P C, Henikoff S (2003). SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res, 31(13): 3812–3814

Nishikawa K, Ishino S, Takenaka H, Norioka N, Hirai T, Yao T, Seto Y (1994). Constructing a protein mutant database. Protein Eng, 7(5): 733

Osumi N, Hirota A, Ohuchi H, Nakafuku M, Iimura T, Kuratani S, Fujiwara M, Noji S, Eto K (1997). Pax-6 is involved in the specification of hindbrain motor neuron subtype. Development, 124 (15): 2961–2972

Puk O, Yan X, Sabrautzki S, Fuchs H, Gailus-Durner V, Hrabě de Angelis M, Graw J (2013). Novel small-eye allele in paired box gene 6 (Pax6) is caused by a point mutation in intron 7 and creates a new exon. Mol Vis, 19: 877–884

Sherry S T, Ward M, Sirotkin K (1999). dbSNP-database for single nucleotide polymorphisms and other classes of minor genetic variation. Genome Res, 9(8): 677–679

Stromo G D (2000). DNA binding sites: Representation and discovery. Bioinformatics, 16(1): 16–23

The UniProt Consortium (2008). The Universal Protein Resource (UniProt). Nucleic Acids Res, 36: D190–D195

Tzoulaki I, White I M, Hanson I M (2005). PAX6 mutations: genotype-phenotype correlations. BMC Genet, 6(1): 27

van Heyningen V, Williamson K A (2002). PAX6 in sensory development. Hum Mol Genet, 11(10): 1161–1167

Wang L, Brown S J(2006). Bind N: A web based tool for efficient prediction of DNA and RNA binding site in amino acid sequences. Nucleic Acids Res, 34: W243–248

Wang L, Huang C, Yang MQ, Yang JY (2010). BindN + for accurate prediction of DNA and RNA-binding residue from protein sequence features. BMC Syst Biol, 4: S3

Wawersik S, Maas R L (2000). Vertebrate eye development as modeled in *Drosophila*. Hum Mol Genet, 9(6): 917–925

Wawrocka A, Sikora A, Kuszel L, Krawczynski M R (2013). 11p13 deletions can be more frequent than the PAX6 gene point mutations in Polish patients with aniridia. J Appl Genet, 54(3): 345–351