

Reckoning the *SIX1* mutation's effects in branchio-oto-renal syndrome — A bioinformatics approach

B. Preethi, V. Shanthi, K. Ramanathan (✉)

Industrial Biotechnology Division, School of Bio Sciences and Technology, VIT University, Vellore – 632014, Tamil Nadu, India

© Higher Education Press and Springer-Verlag Berlin Heidelberg 2015

Abstract Branchio-oto-renal syndrome (BOR) is autosomal dominant disorder which generates hearing impairment and kidney failures in affected individuals. The disease genomic maps were drawn back in recent years, demonstrating, missense mutations responsible in disease were located in *SIX1*, *EYA1* and *EYA2* genes. We try to uncover molecular biology of the syndrome with bioinformatics perspective, taking *SIX1* and *EYA2* protein interaction at center point. The study initiated with 23 natural mutations of *SIX1* gene. They were first analyzed with prediction servers like SIFT, PolyPhen2, I Mutant, SNPs&GO, PHD-SNP and Panther, to identify their impact on their structural stability and function. Subsequently it narrowed down to seven consistent with our quest. They were analyzed on IUPred disorder prediction server. Later *SIX1* and its all mutant proteins were docked with *EYA2* protein using GRAMM-X server. The binding affinity of docked structures was analyzed using DFIRE2 algorithm. The results justify the earlier wet laboratory studies and indicate the reason behind them. Finally we summarize that the proven inactivity of all other mutants is due to the structural disorder created by mutations, hence usual molecular interaction is hindered; strangely protein interaction takes place at DNA binding site of *SIX1* mutants.

Keywords branchio-oto-renal syndrome (BOR), hearing loss, damaging mutations, *SIX1*, *EYA2*, protein-protein interactions

Introduction

Branchio-oto-renal syndrome or BOR (OMIM 113650, 602588, 601653) (Kumar et al., 2000; Kochhar et al., 2007) is a phenotypic disorder characterized by hearing loss, branchial arch and urinary tract malformations. This is known as Melnick-Fraser syndrome, an autosomal dominant developmental disorder which creates hearing impairment and renal failures in the affected individuals (Fraser et al., 1978; Fraser et al., 1980; Fraser et al., 1983; Abdelhak et al., 1997; Kumar et al., 2000; Kawakami et al., 2000; Ruf et al., 2004; Kochhar et al., 2007; Christensen et al., 2008; Patrick et al., 2009). The mutated and faulty genes end up in to faulty protein function during organogenesis, thus causing structural defects. The recent studies emerge that functionally defective missense mutations responsible in disease are located in *SIX1*, *EYA1* and *EYA2* genes (Kumar et al., 2000; Ruf et al., 2004;

Kochhar et al., 2007).

We take a brief review of BOR syndrome and *SIX1-EYA* protein regulatory system. Roughly these genes were reported on broad regions of chromosome 8q¹ in 1994. The relationship of syndrome and *SIX1-EYA* regulatory system was established around 2004 (Ruf et al., 2004; Christensen et al., 2008). Mutations in EYA proteins are the cause of 40 percent of BOR related cases; fact established around 2002. Later studies in 2004–2005 established that the SIX family is another contender. Kochhar et al. (2007) gave a concise review on the pre-2006 stands up. The confirmatory remark on the *SIX1-EYA-DNA* complex disruption and BOR link was pronounced in the literature (Ruf et al., 2004). Recently author Patrick et al. (Patrick et al., 2009) studied all contemporary *SIX1* mutations, their binding with EYA protein and DNA. The study demonstrated that only V17E mutation completely abolished the *SIX1-EYA1* complex formation; whereas other mutations were able to form stable complexes. Even so, the remaining five were unable to keep the DNA complex and the V17E was able to bind the same. Concluding that, all of the *SIX1* mutants were defective in their transcriptional activity (Patrick et al., 2009).

Received June 19, 2015; accepted August 11, 2015

Correspondence: K. Ramanathan

E-mail: kramanathan@vit.ac.in

Let us zoom in to the protein–protein interaction. SIX family proteins are characterized by two major domains first is a DNA binding homeodomain (HD) and second is for protein interaction, SIX domain (SD) (Kumar et al., 2000; Ruf et al., 2004; Kochhar et al., 2007; Christensen et al., 2008; Patrick et al., 2009). EYA protein belongs to the EYA gene family (transcriptional cofactors). They contain N-terminal transactivation domain and highly conserved EYA domain (ED) (Patrick et al., 2009; Kumar et al., 2009). ED shows phosphatase activity and involves in protein–protein interaction with SIX family homeo-domain proteins. *SIX1* protein attaches to EYA to form a transcriptional complex, where *SIX1* confers DNA binding and EYA confers transactivation activity respectively. SIX domain (SD) is important for interaction with EYA and a homeodomain (HD) is responsible for DNA binding (Christensen et al., 2008; Patrick et al., 2009). Thus overall a transcription activity is carried out on a specific DNA site. All of these records come from molecular biology laboratories, and there has been no computational studies performed in near instance.

Methodology

Following methodological algorithms and tools were employed.

Data set

Protein sequences corresponding to SIX1 and EYA2 protein were obtained from the Uniprot website (<http://www.uniprot.org>) (Bairoch and Apweiler, 1996). Protein structures extracted from the Protein Data Bank; Protein IDs (for EYA2 protein) are as follows: 3HB0, 3HB1, 3GEB, 4EGC. All of these proteins except 4EGC were found to be homo-tetramers; 4EGC is a Hetero-dimer; confirmed by pairwise sequence alignment (Larkin et al., 2007). For the analysis of SNPs on the human *SIX1* gene, we have retrieved the sequence data from the NCBI (DBSNP) and SWISS-Prot databases. SWISS-MODEL is used for modeling; it is a fully automated protein structure homology-modeling server. The amino acid sequence of a protein is submitted to construct a 3D model (Schwede et al., 2003). Afterwards NOMADRef program (<http://lorenz.immstr.pasteur.fr>) has been employed to generate the energy minimized structure files of wild and mutant proteins.

SIFT

SIFT (Sorting Intolerant from Tolerant) is a sequence homology based tool, which presumes that important amino acids will be conserved in the protein family. Hence, changes at well-conserved positions tend to be deleterious. SIFT takes a query sequence and uses multiple alignment information to predict tolerated and deleterious substitutions for every

position of the query sequence. Substitutions with normalized probabilities less than cutoff are predicted to be deleterious. The cutoff value in SIFT program is > 0.05 below which mutations are predicted deleterious (Ng and Henikoff, 2003).

PolyPhen2

PolyPhen2 tool uses eight sequence-based and three structure-based features to predict the damaging effects of missense mutations. It analyzes the profiles of homologous sequences and substitution site for a known protein's 3-dimensional structures. In turn they are the parameters taken into account by PolyPhen server to calculate the Position Specific Independent Counts (PSIC) score for each of the two variants, and then compute the PSIC score difference between them. Higher the PSIC score difference, higher is the functional impact a particular amino acid substitution (Adzhubei et al., 2010).

I Mutant 2.0

I Mutant 2.0 is a support vector machine (SVM)-based tool for prediction of protein stability changes upon single point mutations, starting either from the protein structure or sequence. SVM was trained and tested on a data set derived from ProTherm. The predicted free-energy change value ($\Delta\Delta G$) is calculated from the unfolding Gibbs free energy value, of the mutated protein minus the native type (kcal/mol). Positive $\Delta\Delta G$ values mean that the mutated protein possesses high stability and vice versa (Capriotti et al., 2005).

SNPs&GO

SNPs&GO program predicts the disease related mutations from the input of protein sequences with a scoring accuracy and Matthews correlation coefficient of 82% and 0.63 respectively. It is a SVM-based classifier consisting of a single SVM that takes in an input protein sequence, profile and functional information. Of note, we have employed two different SVM based algorithms mainly to eliminate the false positive data in computational prediction. The server predicts if given variation can be classified disease-related or neutral (Calabrese et al., 2009).

PHD-SNP

PHD-SNP is another SVM trained and tested on protein sequence and profile information. It is optimized to predict point mutation to classify as disease-related or as neutral polymorphism (Capriotti et al., 2006, 2013).

PANTHER

PANTHER algorithm is based on a library of Hidden Markov Models (HMMs) obtained from the multiple sequence

alignments of different protein families. PANTHER estimates the likelihood of a particular nsSNP causing a functional impact. The software computes the predicted free energy change value ($\Delta\Delta G$) similar to I Mutant. A positive $\Delta\Delta G$ value is a measure of protein stability (Thomas et al., 2003a; 2003b).

IUPred server

The basis of predicting protein disorder is the difference in sequence characteristics between folded and disordered proteins. Typically, Intrinsically Unstructured Proteins (IUPs) exhibit a strong bias in their amino acid composition. It uses a quadratic expression in the amino acid composition. Applying this calculation for IUP sequences, their estimated energies are clearly shifted toward less favorable energies compared to globular proteins, enabling the prediction of protein disorder on this ground. This tested criterion works out in the algorithm and it provides a probability to amino acids individually (Dosztányi et al., 2005a, 2005b; Dosztányi et al., 2009; Mészáros et al., 2009).

Molecular docking

The docking study was utilized in order to understand the binding affinity. Initially the complex structures were formed by using GRAMM-X docking algorithm (Tovchigrechko and Vakser, 2006). This server is the prolongation of the original GRAMM Fast Fourier Transformation methodology by employing smoothed potentials, refinement stage, and knowledge-based scoring. The server was extensively tested by benchmarking, several months of public employment. Subsequently we have used DFIRE2 server (Tovchigrechko and Vakser, 2005) for the calculation of free energy scores. The PyMOL view of the docked complex is shown in Fig. 1.

PIC server

Protein Interactions Calculator (PIC) is a server, which recognizes several kinds of molecular interactions at protein–protein interaction interface; such as disulphide bonds, hydrophobic interactions, ionic interactions, hydrogen bonds, aromatic- aromatic interactions, aromatic-sulfur interactions and cation- π interactions within a protein or between proteins in a complex. It reads a PDB file of docked protein as input (Tina et al., 2007).

Results

In the original data set obtained from Uniprot and NCBI SNP database we have achieved 27 natural variations out of which 23 were unique; They are as follows V17E, E28D, H44Q, H73P, A95V, R99C, V106G, R110Q, R110W, R112C, W122R, Y129C, T165I, V168I, N193I, P211L, E217K, D227E, D227Y, P249Q, P249L, T252A, L260Q and the wild type. We have performed computational analysis of deleterious missense mutations with 6 major servers; they are SIFT, Polyphen2, I Mutant2, PHDSNP, PANTHER and SNP&GO. We started with putting them on mutation analysis servers; this has reduced the number to one third of the original figure.

SNP data analysis results

SIFT gave 15 out of 23 mutations; Similarly Polyphen2 yielded 13 out of 23 mutations; I Mutant2 predicted 14 probably damaging mutations (Table 1). Only 7 mutations are consistent with our strategy, that they should show apparent loss of function, we took them for further analysis. Then PHD-SNP, PANTHER and SNPs&GO servers calculated similar reliability index of 9 or above for all of the 7 mutations, with the minimum probability score > 0.9

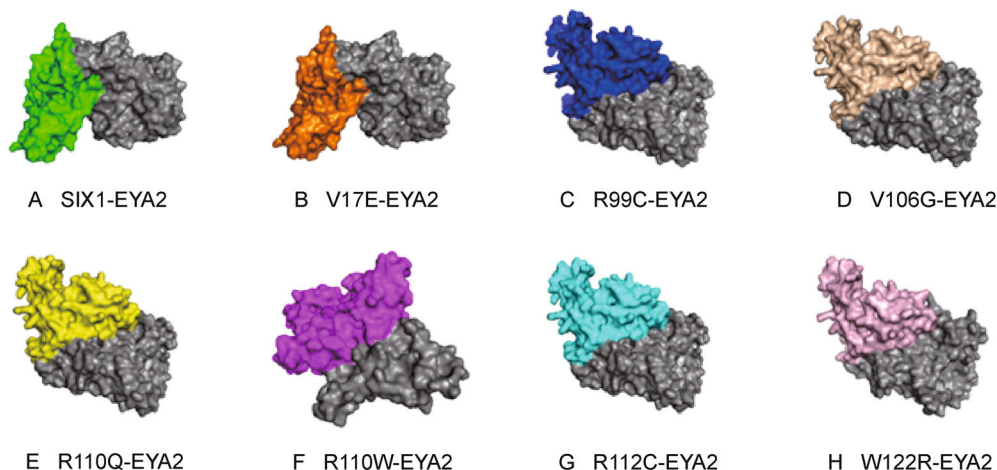


Figure 1 Docked complexes of the native and mutant SIX1 with EYA2 protein.

(Table 2). Conclusively the remaining mutations are: V17E, R99C, V106G, R110Q, R110W, R112C and W122R.

IUPred results

The original wild type protein and all the seven mutants were analyzed on ANCHOR and IUPred servers for their disorder (Dosztányi et al., 2005a, 2005b; Dosztányi et al., 2009; Mészáros et al., 2009). The study reveals that a single amino acid change in a protein structure here has significant impact in protein disorder profile (Figs. 2–4). This elaborates probability of a domain being disordered, based on IUPred and ANCHOR algorithm. The difference is not visible at first sight, however when we overlap these mutant graphs with the wild type we can identify the dissimilarity. Just a single amino

acid change from the original structure leads to the distortion of the disorder probability graph at various places. All these profiles show a commonality, as in all the mutation there is just a single change in amino acid, whereas the probability change is observed in nearer amino acids as well as that are far away from sight of change. Following inference can be taken out:

1) V17E shows no affect of disturbance around 120th amino acid and onwards, this is fairly visible; thus the DNA binding site does not affect much, we predict usual DNA binding.

2) R99C, V106G, R110Q, R110W may have similar in binding pattern as their IUPred disorder profile is distinctly similar. Regions of both binding sites are affected by disturbance.

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

S.No	Uniprot ID and rs ID	Amino acid change	Tolerance index	Predictions	PSIC	Predictions	I Mutant 2.0	Predictions
1	Q15475	V17E	0	Deleterious	0.999	Probably damaging	-0.03	Destabilizing
2	Q15475	H73P	0.01	Deleterious	0.621	Possibly Damaging	0.36	Stabilizing
3	Q15475	V106G	0	Deleterious	1	Probably damaging	-4.2	Destabilizing
4	Q15475	R112C	0	Deleterious	1	Probably damaging	-1.31	Destabilizing
5	Q15475	P249L	0.69	Non-deleterious	0.001	Benign	-0.95	Destabilizing
6	Q15475	R99C	0	Deleterious	1	Probably damaging	-1.07	Destabilizing
7	Q15475	R110Q	0	Deleterious	1	Probably damaging	-0.97	Destabilizing
8	Q15475	R110W	0	Deleterious	1	Probably damaging	-0.37	Destabilizing
9	Q15475	W122R	0	Deleterious	0.97	Probably damaging	-1.85	Destabilizing
10	Q15475	Y129C	0	Deleterious	1	Probably damaging	0.13	Stabilizing
11	rs374638294	P211L	0.28	Non-deleterious	0	Benign	0.72	Stabilizing
12	rs371998997	V168I	0.07	Non-deleterious	0.987	Probably damaging	-0.26	Destabilizing
13	rs370259896	L260Q	0.09	Non-deleterious	0.025	Benign	0.24	Stabilizing
14	rs368974927	P249Q	0.32	Non-deleterious	0.242	Benign	-1.93	Destabilizing
15	rs200835641	A95V	0	Deleterious	0.984	Probably damaging	0.98	Stabilizing
16	rs200205240	T252A	0.95	Non-deleterious	0	Benign	-0.37	Destabilizing
17	rs149265761	D227E	0.61	Non-deleterious	0	Benign	0.64	Stabilizing
18	rs149166341	H44Q	0.87	Non-deleterious	0.006	Benign	-0.12	Destabilizing
19	rs146357380	E28D	0.84	Non-deleterious	0.022	Benign	0.54	Stabilizing
20	rs146129487	E217K	0.91	Non-deleterious	0.704	Possibly damaging	-0.49	Destabilizing
21	rs144481204	D227Y	0.01	Deleterious	0.874	Possibly damaging	0.44	Stabilizing
22	rs143516729	T165I	0.45	Non-deleterious	0.155	Benign	-0.35	Destabilizing
23	rs142301715	N193I	0.24	Non-deleterious	0.138	Benign	0.73	Stabilizing

Table 2 List of variants that were predicted to be functionally significant by SNP&GO, PHD-SNP and PANTHER, mutants in bold letter are significantly deleterious and highly damaging

S.No	Amino acid change	PHD-SNP probability	PANTHER probability	SNP&GO probability	Prediction
1	V17E	0.98	0.962	0.991	Disease
2	R99C	0.984	0.987	0.991	Disease
3	V106G	0.939	0.972	0.989	Disease
4	R110Q	0.965	0.976	0.99	Disease
5	R110W	0.978	0.995	0.992	Disease
6	R112C	0.986	0.993	0.993	Disease
7	W122R	0.989	0.985	0.993	Disease

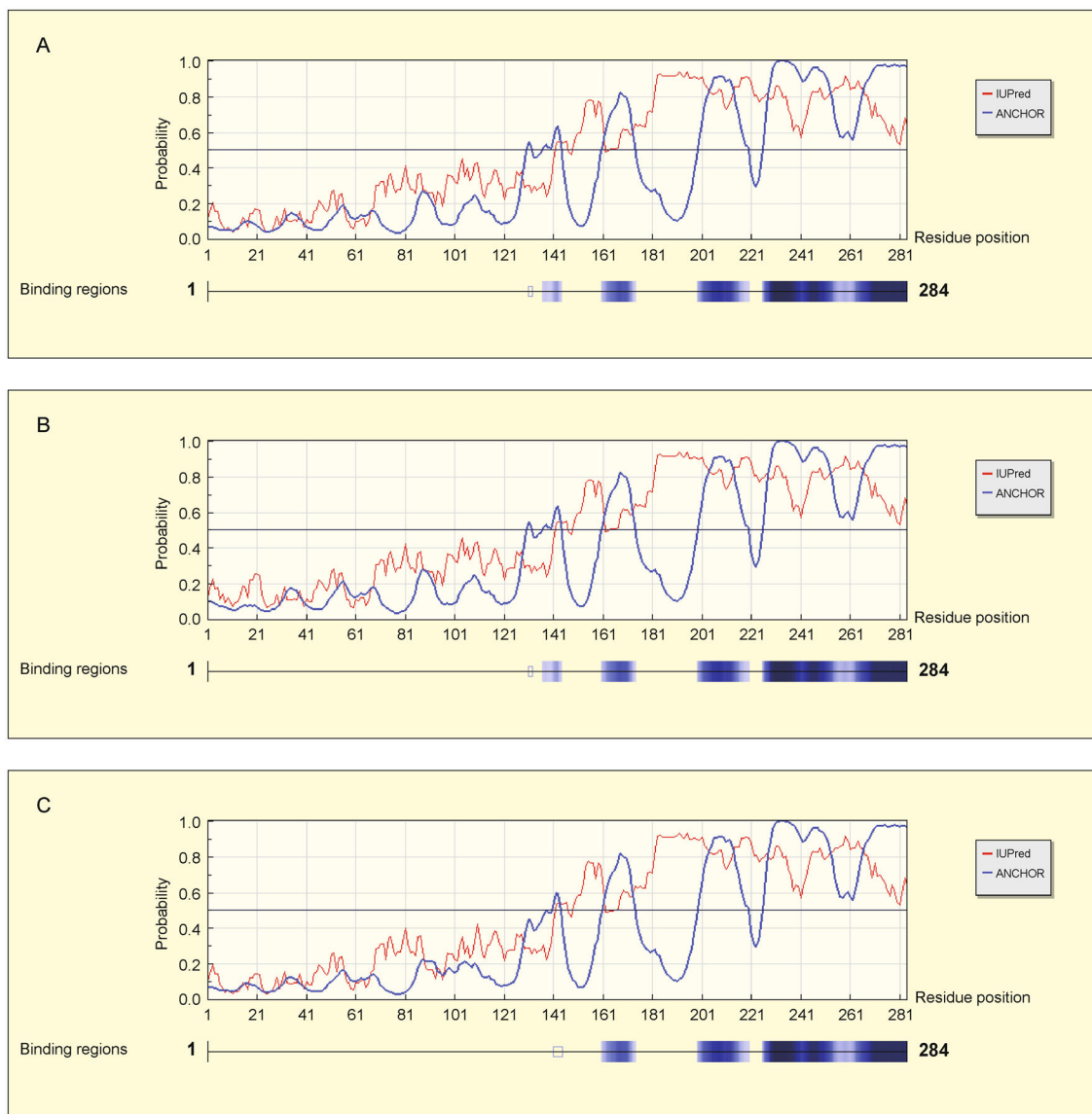


Figure 2 Order and disorder prediction from protein sequence, based on IUPred and ANCHOR algorithm. (A) *SIX1* wild type, (B) V17E and (C) R99C.

3) R112C and W122R were found different among all, the comparative disorder spread was observed to be of long range in both. Regions of both binding sites are affected by disturbance.

4) Except V17E all of them show major fluctuation near around 100 to 140th amino acid, which is the starting of a putative DNA binding region near 124th amino acid.

Molecular docking

In the beginning we modeled *SIX1* protein and *EYA2* protein structure. All the modeling has been done using SWISS-MODEL service on the ExPASy server in automatic mode. 4EGC protein was employed as template to model *SIX1* protein. Protein 3HB1 chain A was taken as template to model

EYA2 protein structure. *SIX1* mutations at V17E, R99C, V106G, R110Q, R110W, R112C and W122R positions were then modeled. 3D model were energy minimized using NOMADRef server. Consequently we had 9 structural files ready at hands; original wild type *SIX1*, *EYA2* protein and 7 *SIX1* mutant proteins. Further these files were uploaded to GRAMM-X docking server to proceed with docking. We had 8 combinations of docking in between *EYA2* and *SIX1* wild type and its 7 analog mutants. As a result a docked file has been generated in each experiment. The first result out of 10 output structures was taken in the interpretation. To gain understanding of the types of molecular interactions occurring at the interface we have used PIC server. The Table 3 briefly shows the number of ionic interactions taking place at the interface and the DEFIRE free energy score. To compute

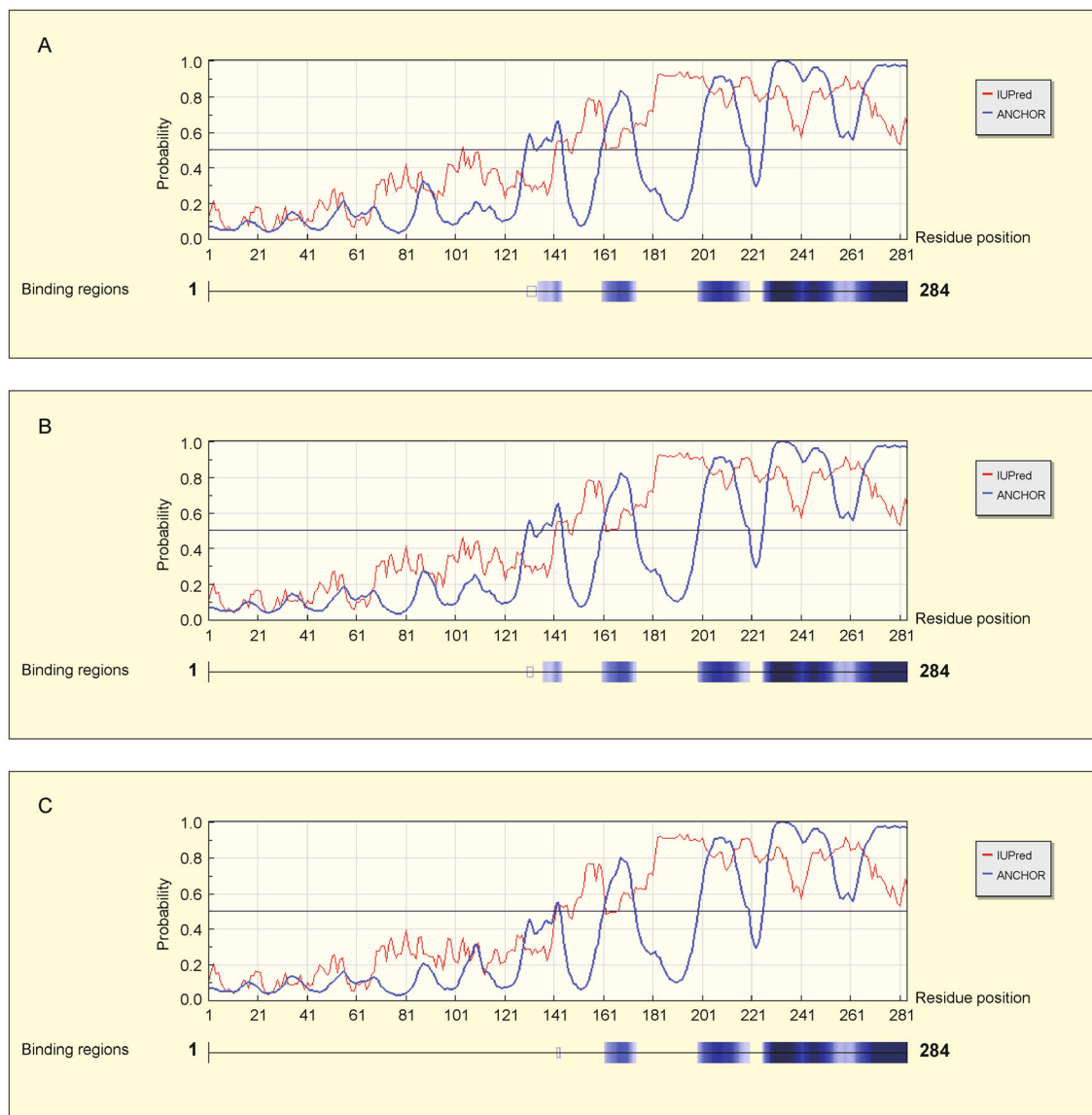


Figure 3 Order and disorder prediction from protein sequence, based on IUPred and ANCHOR algorithm. (A) V106G, (B) R110Q and (C) R110W.

free energy scores DFIRE server has been employed. The results are shown in Table 4.

Discussion

Figure 5 depicts the relation between the numbers of ionic interactions and the normalized free energy. For the sake of scaling free energy score divided (normalized) by a factor of 100. The Figure clearly indicates there is one to one relationship between the two variables; therefore we deduce out reasoning that ionic interactions play a major role in protein–protein interaction. Looking at the Fig. 5 and Table 5, originally only one ionic interaction was present in native protein while docking (41 ASP of *SIX1* and 210 LYS of

EYA2). This alone interaction is visible in V17E and R110W mutant proteins. While in all the other mutant (docked complexes), we can witness five or six ionic interactions. All other mutations are interacting out of their usual places as compared with wild type interactions. Original *SIX1*-*EYA2* complex has 4 hydrophobic interactions in the former region at 35-39 position that was seen absent from all mutant proteins except V17E (Fig. 6). Besides these, there is a significant increase in the number of hydrogen bonds in all *SIX1* mutants except V17E and number of hydrogen bonds in the interface has almost doubled. V17E has been an exception, where hydrogen bonds and free energy has decreased. We bother mentioning one more interaction has been established, that was originally not present, its cation- π interaction. Two cation- π interactions have been found in

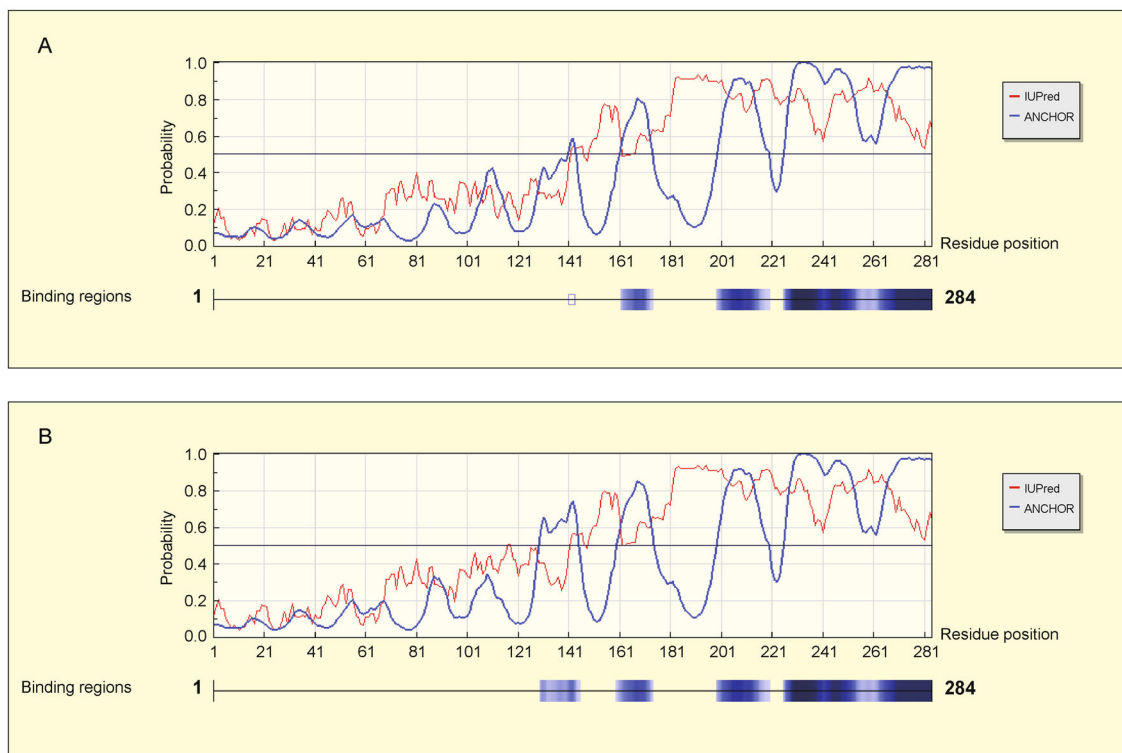


Figure 4 Order and disorder prediction from protein sequence, based on IUPred and ANCHOR algorithm. (A) R112C, and (B) W122R.

Table 3 List of intermolecular hydrogen bonds essential for the stability of SIX1-EYA2 docked complex

S.No	Docked protein complexes	Main chain-main chain interactions	Main chain-side chain interactions	Side chain-side chain interactions	Total number of interactions
1	SIX1-EYA2	2	5	2	9
2	V17E-EYA2	2	4	0	6
3	R99C-EYA2	0	10	7	17
4	V106G-EYA2	0	10	7	17
5	R110Q-EYA2	0	10	7	17
6	R110W-EYA2	1	2	7	10
7	R112C-EYA2	1	10	7	18
8	W122R-EYA2	0	10	11	21

Table 4 Details of number of ionic interactions, total interactions and binding energy of docked complex

S.No	Docked protein complexes	Number of ionic interactions	Total number of interactions	Free energy of binding (kcal/mol)
1	SIX1-EYA2	1	9	-843.913
2	V17E-EYA2	1	6	-841.117
3	R99C-EYA2	5	17	-854.855
4	V106G-EYA2	5	17	-854.141
5	R110Q-EYA2	5	17	-854.932
6	R110W-EYA2	1	10	-640.758
7	R112C-EYA2	6	18	-855.918
8	W122R-EYA2	6	21	-852.954

every mutant except V17E. We can also conclude that the binding in all the mutant cases is stronger than its wild type relative.

Conclusion

The present study aims to report the *SIX1* mutations with their

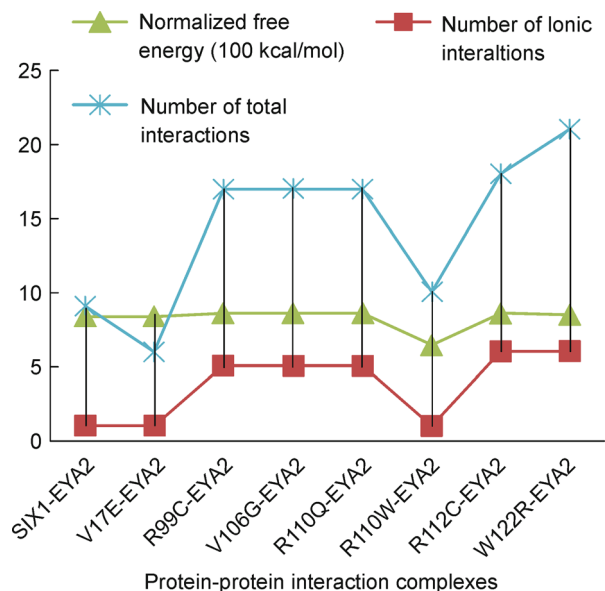


Figure 5 Showing the similarity pattern between normalized free energy, number of total interactions and number of ionic interactions in corresponding docked files.

disease expressions at structural level. Generally this has been observed that whenever we find a critical mutation in a

protein, we shall find it in vicinity of the active sites and active site structure distort such that the binding force loosens up hence its binding energy decreases. In the present investigation of SIX1 protein particularly, we report following three revelations.

First, we state with evidence that the mutations far away from active site or binding regions can also affect the binding region. IUpred results clarify that impact of single point mutation can be of long range.

Secondly, we report that the free energy or the binding energy of docked complex which is supposed to decrease in a general trend is increasing in many of the mutations (Table 4); thus indicating stronger binding. This also endorses the results from wet laboratory observations empirically.

Thirdly, the mutant proteins bind *EYA2* strangely away from its original active site but occupying the part of protein sites responsible for binding the DNA in native state.

The key finding of the work is that, the binding ability in the *SIX1* mutant has been heightened; however, it may get the DNA binding side functionally affected. Our results are matching and vindicating wet laboratory work; consequently providing the proof at molecular level. Therefore we can predict that all these mutations may wind up as nonfunctional to activate a transcriptional regulator mechanism.

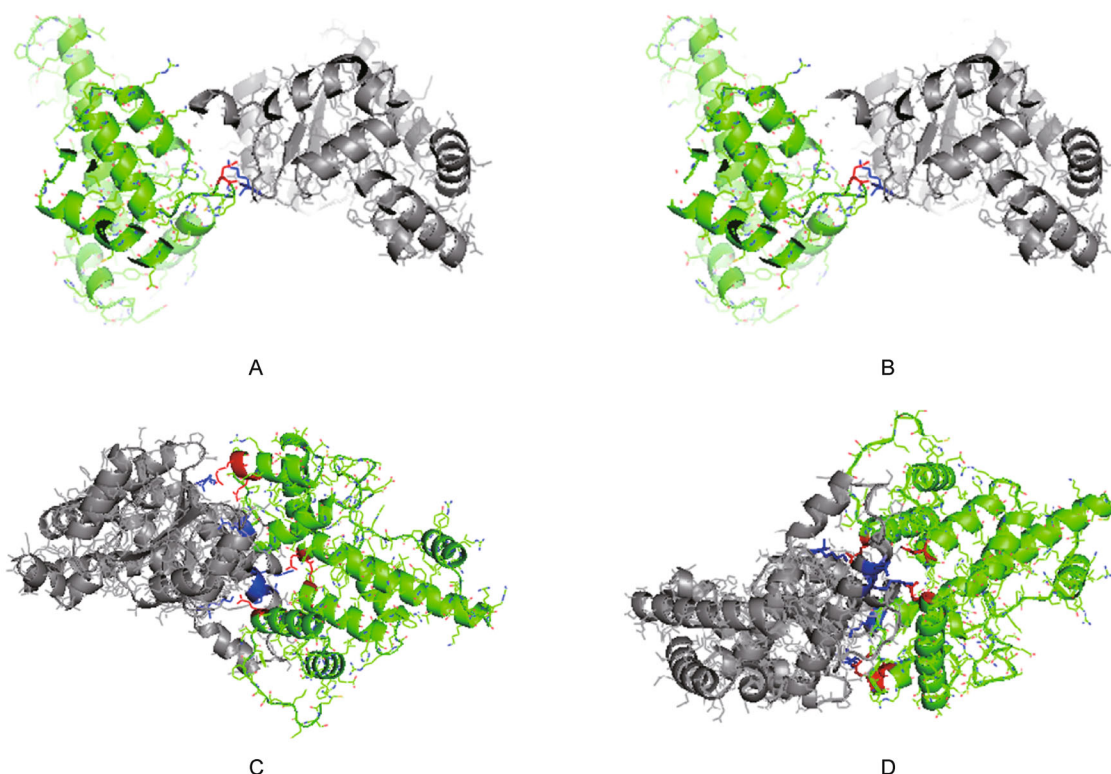


Figure 6 Visualization of ionic interactions between native and mutant SIX1 protein and EYA2.

Table 5 Elaborated Ionic interactions in the docked protein complexes

S.No	Docked protein complexes	Number of ionic interactions	POS (SIX1)	RES	POS (EYA2)	RES
1	SIX1-EYA2	1	41	ASP	210	LYS
2	V17E-EYA2	1	41	ASP	210	LYS
3	R99C-EYA2	5	12	GLU	210	LYS
			153	GLU	223	ARG
			183	GLU	111	ARG
			42	HIS	108	ASP
			152	ARG	1	GLU
4	V106G-EYA2	5	12	GLU	210	LYS
			153	GLU	223	ARG
			183	GLU	111	ARG
			42	HIS	108	ASP
			152	ARG	1	GLU
5	R110Q-EYA2	5	12	GLU	210	LYS
			153	GLU	223	ARG
			183	GLU	111	ARG
			42	HIS	108	ASP
			152	ARG	1	GLU
6	R110W-EYA2	1	185	LYS	125	GLU
7	R112C-EYA2	6	12	GLU	210	LYS
			47	GLU	111	ARG
			153	GLU	223	ARG
			183	GLU	111	ARG
			42	HIS	108	ASP
			152	ARG	1	GLU
8	W122R-EYA2	6	12	GLU	210	LYS
			47	GLU	111	ARG
			153	GLU	223	ARG
			183	GLU	111	ARG
			42	HIS	108	ASP
			152	ARG	1	GLU

Hint: POS: position; RES: residue.

Acknowledgements

The authors express a deep sense of gratitude to the management of Vellore Institute of Technology for all the support, assistance and constant encouragement to carry out this work.

Compliance with ethics guidelines

B. Preethi, V. Shanthi and K. Ramanathan declare that they do not have any conflicts of interest. This article does not contain any studies with human or animal subjects performed by any of the authors.

References

Abdelhak S, Kalatzis V, Heilig R, Compain S, Samson D, Vincent C, Weil D, Cruaud C, Sahly I, Leibovici M, Bitner-Glindzicz M, Francis M, Lacombe D, Vigneron J, Charachon R, Boven K, Bedbeter P, Van

Regemorter N, Weissenbach J, Petit C (1997). A human homologue of the *Drosophila* eyes absent gene underlies branchio-oto-renal (BOR) syndrome and identifies a novel gene family. *Nat Genet*, 15 (2): 157–164

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

Bairoch A, Apweiler R (1996). The SWISS-PROT protein sequence data bank and its new supplement TREMBL. *Nucleic Acids Res*, 24(1): 21–25

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, Altman R B, Bromberg Y (2013). Collective judgment predicts disease-associated single nucleotide variants. *BMC Genomics*, 14(Suppl 3): 3

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, Casadio R (2005). I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Res*, 33(Web Server): W306–W310
- Christensen K L, Patrick A N, McCoy E L, Ford H L (2008). The six family of homeobox genes in development and cancer. *Adv Cancer Res*, 101: 93–126
- Dosztányi Z, Csizmok V, Tompa P, Simon I (2005a). IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics*, 21(16): 3433–3434
- Dosztányi Z, Csizmók V, Tompa P, Simon I (2005b). The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins. *J Mol Biol*, 347(4): 827–839
- Dosztányi Z, Mészáros B, Simon I (2009). ANCHOR: web server for predicting protein binding regions in disordered proteins. *Bioinformatics*, 25(20): 2745–2746
- Fraser F C, Aymé S, Halal F, Sproule J, Opitz J M (1983). Autosomal dominant duplication of the renal collecting system, hearing loss, and external ear anomalies: a new syndrome? *Am J Med Genet*, 14(3): 473–478
- Fraser F C, Ling D, Clogg D, Nogrady B, Gorlin R J (1978). Genetic aspects of the BOR syndrome – branchial fistulas, ear pits, hearing loss, and renal anomalies. *Am J Med Genet*, 2(3): 241–252
- Fraser F C, Sproule J R, Halal F, Opitz J M (1980). Frequency of the branchio-oto-renal (BOR) syndrome in children with profound hearing loss. *Am J Med Genet*, 7(3): 341–349
- Kawakami K, Sato S, Ozaki H, Ikeda K (2000). Six family genes – structure and function as transcription factors and their roles in development. *BioEssays*, 22(7): 616–626
- Kochhar A, Fischer S M, Kimberling W J, Smith R J H (2007). Branchio-oto-renal syndrome. *Am J Med Genet A*, 143(14): 1671–1678
- 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(8): 1073–1081
- Kumar S, Deffenbacher K, Marres H A, Cremers C W, Kimberling W J (2000). Genomewide search and genetic localization of a second gene associated with autosomal dominant branchio-oto-renal syndrome: clinical and genetic implications. *Am J Hum Genet*, 66(5): 1715–1720
- Larkin M A, Blackshields G, Brown N P, Chenna R, McGettigan P A, McWilliam H, Valentin F, Wallace I M, Wilm A, Lopez R, Thompson J D, Gibson T J, Higgins D G (2007). Clustal W and Clustal X version 2.0. *Bioinformatics*, 23(21): 2947–2948
- Mészáros B, Simon I, Dosztányi Z (2009). Prediction of protein binding regions in disordered proteins. *PLOS Comput Biol*, 5(5): e1000376
- Ng P C, Henikoff S (2003). SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res*, 31(13): 3812–3814
- Patrick A N, Schiemann B J, Yang K, Zhao R, Ford H L (2009). Biochemical and functional characterization of six SIX1 Branchio-oto-renal syndrome mutations. *J Biol Chem*, 284(31): 20781–20790
- Ruf R G, Xu P X, Silvius D, Otto E A, Beekmann F, Muerb U, Kumar S, Neuhaus T J, Kemper M J, Raymond R M, Brophy P D, Berkman J, Gattas M, Hyland V, Ruf E M, Schwartz C, Chang E H, Smith R J, Stratakis C A, Weil D, Petit C, Hildebrandt F (2004). Six1 mutations cause branchio-oto-renal syndrome by disruption of Eya1- Six1-DNA complexes. *Proc Natl Acad Sci USA*, 101(21): 8090–8095
- Schwede T, Kopp J, Guex N, Peitsch M C (2003). SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Res*, 31(13): 3381–3385
- Thomas P D, Campbell M J, Kejariwal A, Mi H, Karlak B, Daverman R, Diemer K, Muruganujan A, Narechania A (2003a). PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res*, 13(9): 2129–2141
- Thomas P D, Kejariwal A, Campbell M J, Mi H, Diemer K, Guo N, Ladunga I, Ulitsky-Lazareva B, Muruganujan A, Rabkin S, Vandergriff J A, Doremieux O (2003b). PANTHER: a browsable database of gene products organized by biological function, using curated protein family and subfamily classification. *Nucleic Acids Res*, 31(1): 334–341
- Tina K G, Bhadra R, Srinivasan N (2007). PIC: Protein Interactions Calculator. *Nucleic Acids Res*, 35(Web Server): W473–W476
- Tovchigrechko A, Vakser I A (2005). Development and testing of an automated approach to protein docking. *Proteins*, 60(2): 296–301
- Tovchigrechko A, Vakser I A (2006). GRAMM-X public web server for protein-protein docking. *Nucleic Acids Res*, 34(Web Server): W310–W314