

Research Article



SNP Analysis of NDP Gene that interacts with FZD4 to Causes Familial Exudative Vitreoretinopathy (FEVR)

P. Karthiyayini, Huma Praveen, Jincy Anna Jomy, Rao Sethumadhavan*

Bioinformatics Division, School of Bio Sciences and Technology, VIT University, Vellore, Tamil Nadu, India.

*Corresponding author's E-mail: rsethumadhavan@vit.ac.in

Accepted on: 14-11-2015; Finalized on: 31-12-2015.

ABSTRACT

Here we have done the SNP Analysis of NDP Gene, single protein building blocks (amino acids) changes due to mutations causes an eye disorder Familial exudative vitreoretinopathy (FEVR). For the analysis of the NDP gene, various computational methods and parameters were employed namely, SIFT, Polyphen-2.0, I-Mutant 3.0, SNPs&GO, Root Mean Square Deviation (RMSD), Total Energy, Stabilizing residues and it was inferred that out of 14 variants, 14 were found to be deleterious by SIFT, 13 were found to be damaged by Polyphen2.0, 12 were found to be less stable by I-Mutant3.0 and 5 were found to be diseased by SNPs&GO. The number of commonly affected variants by all the four software was 5. Using Swiss Modelling 3D protein structure of 4MY2_native was generated. Then the obtained model was viewed in Ramachandran Plot using Swiss PDB viewer 4.04 for seeing the residues which are in disallowed region and their PROCHECK was performed using SAVES for analyzing protein structures for validity and assessing how correct they are. Using SwissPdbViewer, the RMSD and Total Energy Values were calculated of the minimized structures of the variants. The minimized structures of all the variants were obtained from Nomad-Ref. Using STRIDE and PIC Webserver. Secondary Structures and Intramolecular Interactions, Protein-Protein Interactions were calculated respectively. Docking was done by Gramm-X. Finally free energy was calculated by using DFIRE2. Further, Native and variants was docked and the view was taken by PyMol.

Keywords: Missense mutation, 4MY2, RMSD, Total energy, docking, FEVR.

INTRODUCTION

Familial exudative vitreoretinopathy (FEVR) is a rare congenital disorder of retinal angiogenesis categorized by deviation of peripheral retinal vascularization that affect the growth and development of blood vessels in the retina leads to retinal detachment, folding, tearing, and/or complete retinal dysplasia¹. Familial exudative vitreoretinopathy (FEVR) is caused due to mutations in the following four genes; NDP (MIM 300658), FZD4 (MIM 604579), LRP5 (MIM 603506), TSPAN12 (MIM 613138)². Wnt signalling pathway is an essential constituent of the cell machinery that regulates cellular proliferation, differentiation, migration, and polarity in the retina³. Generally, low density lipoprotein receptor protein 5 (LRP5) and frizzled 4 (FZD4), acting as coreceptors for Wnt ligands, When the encoding genes of these proteins mutated cause familial exudative vitreoretinopathy (FEVR)⁴. Likewise, NDP, a ligand for these Wnt receptors, when the encoding genes mutated causes FEVR⁵. Specifically, frizzled-4 (FZD4) takes part in the Wnt signalling pathway⁶. At the cell surface, norrin protein (produced from the NDP gene) interacts with frizzled-4 protein (produced from the FZD4 gene) and they both fit together like a key in a lock⁷. Disruption of Wnt signalling inhibits the formation of blood vessels at the edges of the retina⁸. This results in anomalous blood supply to the tissue leads to retinal destruction or visual deficiency and sometimes complete blindness in one or both eyes⁹. Thus, Norrin and Frizzled-4 protein function as a ligand-receptor pair to control vascular development

in the retina and inner ear¹⁰. In frizzled-4 protein, single protein building blocks (amino acids) changes due to mutations¹¹. Sometimes, insertion or deletion in genetic material of the FZD4 gene causes changes to the protein¹². Mostly, the amount of frizzled-4 protein is reduced due to mutations in the FZD4 gene within cells¹³. Sometimes, an unstable protein that cannot bind to norrin protein is supposed to result in due to other mutations¹⁴. Chemical signalling in the developing eye interrupts due to reduction in the amount of frizzled-4 protein which inhibits the formation of blood vessels at the edges of the retina¹⁵. Thus, the resulting asymmetrical blood supply to the tissue leads to retinal devastation and optical insufficiency in several people with familial exudative vitreoretinopathy. Accordingly Norrin production leads to early retinal vascular invasion and delayed Norrin production leads to typical imperfections in intra retinal vascular architecture¹⁶.

Familial exudative vitreoretinopathy (FEVR) can be of two types:

- 1) Autosomal dominant which is caused by mutations in the *FZD4* (11q14-q21) or *LRP5* (11q13.4) genes.
- 2) Autosomal recessive or X-linked recessive which is caused by mutations in the *NDP* gene (Xp11.4-p11.3), the same gene involved in Norrie disease.

Familial Exudative Vitreoretinopathy (FEVR) is an eye disorder of retina which gradually at the part of the on and the bosom vessels zigzag nourish the retina. Scientists attack unfold ramble 85% of pretended tight-



fisted tushie be asymptomatic, focus is, showing no symptoms and causes no eye problems. When FEVR arises in its most destructive form many harm can happen. **1.** Retina grows aberrantly, bends over onto itself leads to Retinal fold causes incomplete blindness **2.** Expansion and brakeage of anomalous blood vessels (vascularization) in the outer edges of the retina leads to retinal detachment due to the pulling of the retina and retinal traction due to dragging of the retina **3.** Development of thin fibrous membranes into the vitreous gel in the centre of the eye leads to retinal detachment, folding, tearing, and/or complete retinal dysplasia **4.** Iris is not able to expand due to its Dispersion **5.** Blood vessels inhibition at the edges of the retina¹⁷. Results in anomalous blood supply of the tissues prompts retinal decimation and dream reduction sometimes complete visual deficiency. Frizzled-4 and the LRP5 protein contribute in the Wnt signalling pathway¹⁸. The LRP5 protein is also involved in forming a blood supply to the retina and the inner ear¹². Also, helps normalize bone mineral density. More than 15 mutations in the LRP5 gene have been recognised in people with the eye malady familial exudative vitreoretinopathy¹⁹. In LRP5 protein, single protein building blocks (amino acids) changes due to mutations¹². Sometimes, insertion or deletion in genetic material of the LRP5 gene causes changes to the protein²⁰. Mostly, the amount of LRP5 protein is reduced due to mutations in the LRP5 gene within cells²⁰. Sometimes, chemical signalling in the developing eye interrupts due to reduction in the amount of LRP5 protein which inhibits the formation of blood vessels at the edges of the retina²¹.

MATERIALS AND METHODS

Datasets

Natural Variants and the protein sequence for NDP gene was obtained from UniProt and PDB ID was selected from www.uniprot.com keeping in mind the Method is Xray, Chain is A and all the positions of all the Natural Variants are covered²²⁻²⁴ for SNP analysis.

SIFT; Sequence Homology based method to detect the deleterious coding non synonymous SNPs

SIFT is a sequence homology based tool which assumes that significant amino acids will be conserved in the protein family. Therefore, changes at well-conserved locations tend to be deleterious. We submit the query in the form of SNP IDs or Protein Sequences. The fundamental principle of this program is that, SIFT takes a query sequence and uses multiple alignment information to predict tolerated and deleterious substitutions for every position of the query sequence. SIFT is a multistep procedure that, given a protein sequence, (a) searches for similar sequences, (b) chooses closely related sequences that may share similar function, (c) obtains the multiple alignment of these chosen sequences, and (d) calculates normalized probabilities for all possible substitutions at each position from the alignment. Substitutions at each

position with normalized probabilities less than a chosen cut-off are predicted to be deleterious and greater than or equal to cut-off are predicted to be tolerated. The cut-off value in SIFT program is tolerance index of ≥ 0.05 . Higher the tolerance index, less functional impact a particular amino acid substitution is likely to have²⁵.

PolyPhen 2.0; Structure Homology based method to analyse the possible impact of an amino acid substitution on the structure and function

To analyse the possible impact of an amino acid substitution on the structure and function of Norrin protein we used PolyPhen v2. The protein sequence with mutational position and two amino acid variants were submitted to the server. PolyPhen generates multiple sequence alignment of homologous protein structures, calculates the position-specific independent counts (PSIC) scores for each of the two variants, and then calculates the PSIC score difference between both the allelic variants. The higher the PSIC score difference, the higher the functional impact a particular amino acid substitution is likely to have or the more likely it is to be damaging. The PolyPhen server discriminates nsSNPs into three main categories, benign, possibly damaging, or probably damaging, and provides the corresponding specificity and sensitivity values. The probably damaging nsSNPs are those that are predicted with high confidence and are expected to affect protein structure or function. Therefore, we selected the nsSNPs that were determined to be probably damaging and possessed PSIC scores > 0.997 . Thereafter, we examined nsSNPs predicted to be deleterious or to cause disease both by the SIFT and PolyPhen programs²⁶.

I-MUTANT 3.0; Suite of Support Vector Machine to predict automatically protein stability changes upon single-site mutations

I-Mutant is a suite of Support Vector Machine based predictors integrated in a unique web server. It offers the opportunity to predict automatically protein stability changes upon single-site mutations starting from protein sequence alone or protein structure when available. Moreover it gives the possibility to predict human deleterious Single Nucleotide Polymorphism starting from the protein sequence. The method was trained and tested on a data set derived from ProTherm, which is presently the most comprehensive available database of thermodynamic experimental data of free energy changes of protein stability upon mutation under Different Conditions. I-Mutant3.0 can be used both as a classifier for predicting the sign of the protein stability change upon mutation and as a regression estimator for predicting the related $\Delta\Delta G$ values²⁸.

SNPs & GO, Server for predicting Human Disease-related Mutations in Proteins with Functional Annotations

The genetic basis of human variability is mainly due to Single Nucleotide Polymorphisms (SNPs). The most investigated SNPs are missense mutations resulting in



residue substitutions in the protein²⁹. Here we propose SNPs & GO, an accurate method based on support vector machines, to predict disease related mutations from the protein sequence, scoring with accuracy=82% and Matthews correlation coefficient=0.63. SNPs & GO collects in unique framework information derived from protein sequence, protein sequence profile, and protein function. The server is based on Support Vector Machines (SVM) and for a given protein, its input comprises: the sequence and/or its three-dimensional structure (when available), a set of target variations and its functional Gene Ontology (GO) terms³⁰. The output of the server provides, for each protein variation, the probabilities to be associated to human diseases.

Swiss Model, Server for automated comparative modelling of three-dimensional (3D) protein structures

It pioneered the field of automated modelling starting in 1993 and is the most widely-used free web-based automated modelling facility today³¹. SWISS-MODEL provides several levels of user interaction through its World Wide Web interface: in the 'first approach mode' only an amino acid sequence of a protein is submitted to build a 3D model. Template selection, alignment and model building are done completely automated by the server. In the 'alignment mode', the modelling process is based on a user-defined target-template alignment. Complex modelling tasks can be handled with the 'project mode' using DeepView (Swiss-PdbViewer), an integrated sequence-to-structure workbench. All models are sent back via email with a detailed modelling report. WhatCheck analyses and ANOLEA evaluations are provided optionally. The reliability of SWISS-MODEL is continuously evaluated in the EVA-CM project. The SWISS-MODEL server is under constant development to improve the successful implementation of expert knowledge into an easy-to-use server³¹.

SAVS and PROCHECK, Server for analyzing protein structure for validity

SAVS (Structural analysis and verification server) is a server for analyzing protein structure for validity and assessing how correct they are. It is run by a research group at UCLA. PROCHECK was performed using SAVES for analyzing protein structures for validity and assessing how correct they are³². From this way, we optimize the protein model to get the stable structure. It utilizes 5 programs for doing this: PROCHECK, WATCHCHECK, ERRAT, VERIFY 3D, PROSa.

Mutant Modeling of Single Amino Acid of Native Protein Structure to compute the RMSD

The mutation was performed by using Swiss-PdbViewer and Energy Minimization for 3D structure was performed by NOMAD-Ref server³³. This server uses GROMACS as default force field for energy minimization based on the methods of steepest descent, conjugate gradient and L-BFGS³⁴. L-BFGS method was used for optimizing the 3D structures. As the server uses Gromacs as the default

force field for energy minimization, based on the methods of steepest descent, conjugate gradient, and limited-memory Broyden-Fletcher-Goldfarb-Shanno (LBFGS). We used conjugate gradient method to minimize the energy of the 3D structure of 4MY2. Structure analysis was performed to evaluate the structural deviation between native proteins and mutant proteins by means of root mean square deviation (RMSD). Deviation of the mutant structure from the native structure could be caused by substitutions, deletions and insertions and the deviation between the two structures could alter the functional activity with respect to binding efficiency of the inhibitors, which was evaluated by their RMSD values³⁵.

Computation of Total Energy and Stabilizing Residues

SRIDE is used to find out the stabilizing residues (SRs) in proteins mainly on the interactions of a given residue with its spatial. A residue is selected as a stabilizing residue if it has high surrounding hydrophobicity, high long-range order, and high conservation score and if it belongs to a stabilization center. The definition of all these parameters and the thresholds used to identify the SRs. The algorithm for identifying SRs was originally developed for TIM-barrel proteins and is now generalized for all proteins of known 3D structure. SRs could be applied in protein engineering and homology modeling and could also help to explain certain folds with significant stability. Total energy indicates the stability between native and mutant modelled structures, and could be computed by the GROMOS96 force field that is embedded in the SWISSPDB viewer³⁶.

Computation of Intramolecular Interactions and Protein-Protein Interactions using PIC Server

Computation of the intra-molecular and protein-protein interactions for native and mutants structures is done by Protein Interactions Calculator³⁷. Protein Interactions Calculator (PIC) is a server which recognizes various kinds of interactions; such as disulphide bonds, hydrophobic interactions, ionic interactions, hydrogen bonds, aromatics-aromatics interactions, aromatics-sulphur interactions and cation- π interactions within a protein or between proteins in a complex. It also determines the accessible surface area as well as the distance of a residue from the surface of the protein. The input should be in the protein data bank (.pdb) format. Interactions are calculated based on empirical or semi-empirical set of rules. Heteroatoms are not included for the analysis. For NMR structure only the first model is considered³⁷.

Secondary structure prediction using STRIDE Webserver

The information about secondary structures is obtained by STRIDE is a software tool for secondary structure assignment from atomic resolution protein structures³⁸. It implements a knowledge-based algorithm that makes combined use of hydrogen bond energy and statistically derived backbone torsional angle information and is optimized to return resulting assignments in maximal



agreement with crystallographers' designations. The STRIDE web server provides access to this tool and allows visualization of the secondary structure, as well as contact and Ramachandran maps for any file uploaded by the user with atomic coordinates in the Protein Data Bank (PDB) format. A searchable database of STRIDE assignments for the latest PDB release is also provided³⁸.

Protein docking using Gramm-x

The program is used for docking of energy minimized Native (ligand) and FZD4 (Receptor). Protein docking software GRAMM-X and its web interface extend the original GRAMM Fast Fourier Transformation methodology by employing smoothed potentials, refinement stage, and knowledge-based scoring³⁹. The web server frees users from complex installation of database-dependent parallel software and maintaining large hardware resources needed for protein docking simulations. Docking problems submitted to GRAMM-X server are processed by a 320 processor Linux cluster. The server was extensively tested by benchmarking, several months of public use, and participation in the CAPRI server track. This server will ignore any small ligands or other non-protein molecules in the input files. It is designed exclusively for docking pairs of protein molecules³⁹.

Analysing the Binding Affinity between Energy Minimized Native/Natural Variants and FZD4

The docked protein complex is given to the DFIRE server as an input for calculating the binding free energy (ΔG) score. *DFIRE2* is used to calculate the Free Energy. Protein Conformation *Free Energy Score* is used to find the difference between two closely related all-atom statistical energy functions⁴⁰. Ab initio folding of terminal segments with secondary structures reveals the fine difference between two closely associated proteins.

Visualization of docked complex using Pymol

The docked complex can be visualized using Pymol. PyMOL is one of the best visualization software. Visualization is the most direct route to understanding in structural biology. The Graphical User Interface (GUI) is divided into two- the main window (PyMOL Viewer window) and the menu window (External GUI). This is a single OpenGL window where all 3D graphics are displayed and where all direct user interaction with 3D models takes place⁴¹.

RESULTS AND DISCUSSION

Single Amino Acid Polymorphism dataset from Uniprot

The 4MY2 Native and a total of 14 Natural Variants namely I18K, R38C, R41K, H42R, K54N, K58N, L61I, L103V, R115L, R121G, R121W, R121Q, Y120C and L124F¹⁹ were retrieved from Uniprot database (Table 1).

Deleterious SNPs by SIFT

Protein sequence was submitted to SIFT program to check the tolerance index. Higher the tolerance index, less functional impact a particular amino acid substitution is likely to have and vice versa²⁵. Among 14 natural variants, All 14 were found to be deleterious having the tolerance index score of ≤ 0.05 . The results are shown in table 1. We observed that, out of 14 deleterious nsSNPs, All 14 nsSNPs showed a highly deleterious tolerance index of 0.00.

Damaged nsSNPs by Polyphen 2.0

How damaging a particular SNP is. Values of Damaging SNPs can be derived from the Polyphen program²⁶. 14 natural variants were submitted to the Polyphen 2.0 server and the results are shown in table 1 score > 0.5 is considered to be damaging. It can be seen that out of 14 SNPs, 13 nsSNPs were found to be damaging.

Identification of Stability by I-MUTANT 3.0

More negative is the DDG value, less stable the given point mutation is likely to be as predicted by I-Mutant 3.0²⁷. From this analysis, we obtained that out of 14 natural variants, 12 were found to be less stable as shown in table 1.

Prediction of Human Disease-related Mutations in Proteins with Functional Annotations by SNPs & GO

To predict disease related mutations from the protein sequence, scoring with accuracy=82% and Matthews correlation coefficient=0.63. SNPs&GO collects in unique framework information derived from protein sequence, protein sequence profile, and protein function [29]. Among 14 natural variants, 5 were found to be diseased having the missense mutations.

Rational Consideration of Detrimental Point Mutations

We got the 5 most potential detrimental Point mutations (R115L, R121G, R121W, R121Q and Y120C) for further course of studies because they were commonly found to be less stable, deleterious, damaging and diseased by the I-Mutant3.0, SIFT and Poly Phen-2 and SNPs&GO respectively. As we study the statistical accuracy of these four programs, I-Mutant gives the possibility to predict human deleterious Single Nucleotide Polymorphism starting from the protein sequence. The method was trained and tested on a data set derived from ProTherm, which is presently the most comprehensive available database of thermodynamic experimental data of free energy changes of protein stability upon mutation under different conditions. The percentage of Natural Variants affected with I-Mutant is 35.7%. We observed that, out of 14 deleterious nsSNPs, All 14 nsSNPs showed a highly deleterious tolerance index of 0. Thus, SIFT correctly predicted 100% of the substitutions associated with the disease that affect protein function. PolyPhen-2 allows the submission of protein sequence with mutational position and two amino acid variants and generates



multiple sequence alignment of homologous protein structures, calculates the position-specific independent counts (PSIC) scores for each of the two variants, and then calculates the PSIC score difference between both the allelic variants. The higher the PSIC score difference, the higher the functional impact a particular amino acid substitution is likely to have or the more likely it is to be damaging. PolyPhen-2 was informed to attain a rate of true positive predictions of 92.8%. SNPs&GO, to predict disease related mutations from the protein sequence, scoring with accuracy=82% and Matthews correlation coefficient=0.63 was provided with 85.7% of affected Natural Variants. To attain precise and accurate measures of the detrimental effect of Natural variants, comprehensive parameters of all these four programs could be more significant than individual tool parameters. Hence, we further studied these detrimental missense mutations by structural analysis. Figure 1 shows the list of functionally significant mutations with the commonly affected ones.

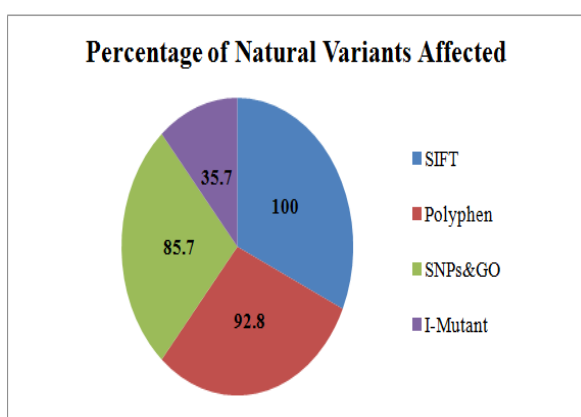


Figure 1: Pie Chart For percentage of natural variants affected

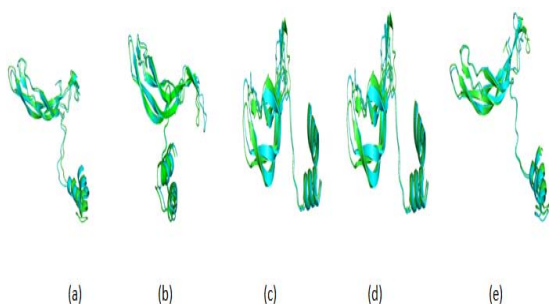


Figure 2: Superimposed Structure of the Native Protein (Green) with Mutant

(A) Superimposed structure of native 4MY2 (green) with mutant R115L (cyan) structure showing RMSD of 2.01Å
 (B) Superimposed structure of native 4MY2 (green) with mutant R121G (cyan) structure showing RMSD of 2.35Å
 (C) Superimposed structure of native 4MY2 (green) with mutant R121Q (cyan) structure showing RMSD of 1.85Å
 (D) Superimposed structure of native 4MY2 (green) with mutant R121W (cyan) structure showing RMSD of 2.26Å.
 (E) Superimposed structure of native 4MY2 (green) with mutant Y120C (cyan) structure showing RMSD of 3.15Å.

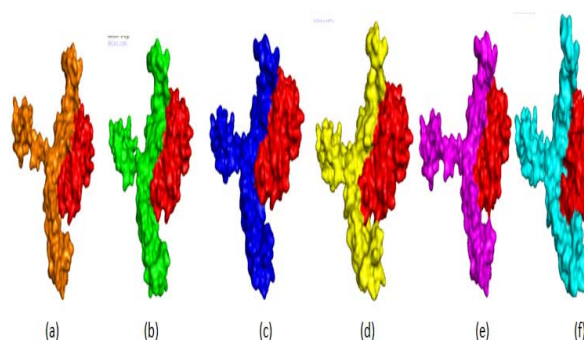


Figure 3: Docked Complexes of Native and Mutant 4MY2 with FZD4

(A) Docked complex of Native 4MY2 (orange) and FZD4 (red) having the Free energy of -6977.974, (B) Docked complex of R115L (green) and FZD4 (red) having the Free energy of -6249.923, (C) Docked complex of R121G (blue) and FZD4 (red) having the Free energy of -6672.942, (D) Docked complex of R121Q (yellow) and FZD4 (red) having the Free energy of -6877.22, (E) Docked complex of R121W (magenta) and FZD4 (red) having the Free energy of -6789.153, (F) Docked complex of Y120C (cyan) and FZD4 (red) having the Free energy of -6919.693.

Computing the RMSD and TOTAL ENERGY by Modeling of Mutant Structures

Mutation at specified was performed by SwissPDBViewer independently to get modeled structures. Then, energy minimization is performed by NOMAD-Ref server³³ of both native structure and mutant modeled structures.

In order to find out the deviation between the native and the mutant, the native structure was superimposed with all the energy refined mutant structures to get RMSD. The higher the RMSD Value, the more is the deviation between the native and the mutant structure.

In order to find out the structural stability of native and mutants, the total energy this includes bonds, angles, torsions, non-bonded and electrostatic constraints from GROMOS96 force field implemented in SwissPdbViewer to check their stability.

Table 2 shows the RMSD values for native structure with each mutant modelled structure. Table 2 shows that, one variant, Y120C exhibited a high RMSD >3.00 Å and the other three variants exhibited an RMSD >2.00 Å.

The total energy was calculated for both native and mutant structures. Table 2 shows that total energy of native structure was -6977.974KJ/MOL. whereas the 5 mutant structures all had slightly higher total energies than the native structure. Note that the higher the total energy, the lesser the stability and vice versa.

Computation of Stabilizing Residues using SRIDE

SRIDE is used to find out the stabilizing residues (SRs) in proteins mainly on the interactions of a given residue with its spatial. A residue is selected as a stabilizing residue if it has high surrounding hydrophobicity, high

long-range order and high conservation score and if it belongs to a stabilization centre. The definition of all these parameters and the thresholds used to identify the SRs. The native structure and the other 4 mutant structures have only one stabilizing residue (ARG64). Whereas, Y120C has 3 stabilizing residue (ARG64, TYR122, ILE123). This indicates that all the 5 mutants were less stable than the native structure. We further estimated the outcome of these detrimental missense mutations by executing binding analysis between 4MY2 and FZD4 using docking studies³⁶.

Computation of Intramolecular Interactions and Protein-Protein Interactions using PIC Server

We further calculated the various kinds of interactions; such as disulphide bonds, hydrophobic interactions, ionic interactions, hydrogen bonds, aromatics-aromatics interactions, aromatics-sulphur interactions and cation- π interactions within a protein or between proteins in a complex to study the stability of protein structure for both native and mutant structures by using the PIC server (Table 3).

Interactions within a protein structure and the interactions between proteins in an assembly were essential considerations in understanding molecular basis of stability and functions of proteins and their complexes. There were several weak and strong intra-molecular interactions that render stability to a protein structure. Therefore these intra-molecular interactions were computed by PIC server in order to further substantiate the stability of protein structure. Based on this analysis, we found that a total number of 336 intra-molecular interactions were obtained in the native structure of 4MY2. On the other hand, 5 mutant structures of 4MY2

established the intra-molecular interactions between the ranges of 328 to 350 as shown in Table 3. We further evaluated the effect of these 5 detrimental missense mutations by performing binding analysis between 4MY2 and FZD4 through protein-protein docking studies in order to understand the functional activity of 4MY2³⁷.

Analysing the Binding Efficiency for Native and Mutant

To find out the binding efficiency of native and mutant FZD4, we applied molecular dynamics approach for justifying the functional activity of these 5 mutants. In this examination, we executed 5 missense mutations (R115L, R121G, R121Q, R121W and Y120C) in the chain A of the PDB IDs 4MY2 by swisspdb viewer individually and energy minimization was achieved for the entire complex (both native and mutant complex) by GROMACS (Nomad-ref) followed by simulated annealing to get the enhanced structures using a discrete molecular dynamics approach (ifold). We used Grammx to dock 4MY2 native and mutant structures with FZD4 and moreover; we used DFire, for finding the protein conformation free energy source for the docked complex retrieved from Grammx⁴⁰. We used this server for the missense mutation study with respect to finding the free energy source of both native and mutants. In this study, we found that the binding free energy for FZD4 with native 4MY2 protein was found to be -336.16kJ/mol, has a lower binding affinity compared to the mutants. This study reveals that native 4MY2 exhibited lower binding affinity with FZD4. Hence, the lesser binding free energies may probably be due to loss of intermolecular non-covalent interactions. This study clearly revealed that native complex had low intermolecular non covalent interactions than mutant complexes.

Table 1: List of Functionally Significant Mutants predicted to be deleterious, damaging, less stable and diseased by SIFT, I-Mutant 3.0, PolyPhen-2 and SNPs & GO respectively.

S. No.	Natural Variants	SIFT	Polyphen 2.0	I-Mutant 3.0	SNPs & GO
1	I 18 K	0	0.008	-0.95	0.190 (Neutral)
2	R 38 C	0	0.999	-2.18	0.388(Neutral)
3	R 41 K	0	0.979	-0.54	0.318(Neutral)
4	H 42 R	0	0.991	-0.5	0.140(Neutral)
5	K 54 N	0	0.999	-0.81	0.202(Neutral)
6	K 58 N	0	0.999	-1.36	0.195(Neutral)
7	L 61 I	0	0.997	-0.43	0.282(Neutral)
8	L 103 V	0	0.997	-0.51	0.142(Neutral)
9	R 115 L	0	0.997	-0.12	0.550(Disease)
10	Y 120 C	0	1.000	-0.7	0.507(Disease)
11	R 121 G	0	0.997	-0.93	0.612(Disease)
12	R 121 Q	0	0.998	-0.27	0.588(Disease)
13	R 121W	0	1.000	0.05	0.621(Disease)
14	L 124 F	0	0.999	0.45	0.267(Neutral)

NOTE: Letters in bold indicate mutants predicted to be less stable, deleterious and damaging by I-Mutant 3.0, SIFT and PolyPhen-2, SNPs & GO respectively.



Table 2: Total Energy, RMSD, SRIDE, No. of SR

S. No.	Natural Variants	Total Energy KJ/MOL	RMSD	SRIDE	No. Of S.R
1	Native_4my2	-6977.974		ARG64	1
2	R115L	-6249.923	2.01Å	ARG64	1
3	R121G	-6672.942	2.35Å	ARG64	1
4	R121Q	-6877.22	1.85Å	ARG64	1
5	R121W	-6789.153	2.26Å	ARG64	1
6	Y120C	-6919.693	3.15Å	ARG64, TYR122, ILE123	3

Note: RMSD- Root Mean Square Deviation; SR- Stabilizing residues; the common stabilizing residues are shown in bold

Table 3: Intra Molecular Interactions Studies Using PIC Webserver

Natural Variants	HI	DB	MM	MS	SS	II	AA	AS	CI	Total
Native_4MY2	47	4	95	49	9	8	0	0	4	216
Native_4MY2 EM	48	4	125	87	57	10	1	0	4	336
R115L_EM	49	4	129	84	61	10	1	0	4	342
R121G_EM	49	4	130	79	64	9	1	0	4	340
R121Q_EM	51	4	133	89	60	9	0	0	4	350
R121W_EM	50	4	132	79	63	9	1	0	4	342
Y120C_EM	44	4	130	80	57	9	0	0	4	328

Note: Total no of intramolecular interactions. HI- Hydrogen Interactions, MM- Main chain-Main chain interaction, MS- Main chain Side chain interaction, SS- Side chain side chain interactions, II- Ionic-Ionic interaction, AA- Aromatic-Aromatic interactions, AS- Aromatic-Sulphur interactions, CI Cation- π interactions

Table 4: Number of Protein-Protein Interactions of the Native Protein and Mutants

Natural Variants	HI	DB	MM	MS	SS	II	AA	AS	CI	Total
Native_4MY2 EM	3	0	0	7	16	2	0	0	1	29
R115L_EM	3	0	0	8	16	2	0	0	1	30
R121G_EM	3	0	0	8	25	2	0	0	1	39
R121Q_EM	5	0	1	10	14	2	0	0	1	33
R121W_EM	4	0	1	4	13	2	0	0	1	25
Y120C_EM	6	0	1	12	6	5	1	0	0	31

Table 5: Secondary Structure Distribution

S.No.	Natural Variants	Total	No. Of α - Helix	No. Of Coils	No. Of Turns	No. Of Strands	No. Of Bridges
1	Native_EM	133	21	16	37	57	2
2	R115L_EM	133	21	18	37	55	2
3	R121G_EM	133	21	20	33	57	2
4	R121Q_EM	133	21	12	37	61	2
5	R121W_EM	133	21	20	33	57	2
6	Y120C_EM	133	21	18	32	60	2

Table 6: Free Energy Calculation

S. No.	Natural Variants	Free Energy
1	Native_EM	-336.16
2	R115L_EM	-336.161
3	R121G_EM	-335.222
4	R121Q_EM	-339.15
5	R121W_EM	-341.065
6	Y120C_EM	-338.599



CONCLUSION

Of the 14 variants that were retrieved from Uniprot, 12 variants were found less stable by I-Mutant2.0, 14 variants were found to be deleterious by SIFT, 13 variants were considered damaging by PolyPhen and 5 variants were found to be diseased by SNPs&GO. Five variants were selected as potentially detrimental point mutations because they were commonly found to be less stable, deleterious and damaging by the I-Mutant 3.0, SIFT, Poly-Phen-2.0 and SNPs&GO servers, respectively. The structures of these 5 variants were modelled and the RMSD between the mutants and native structures ranged from 1.85Å to 3.15Å. Docking analysis between FZD4 and the native and mutant modelled structures generated Free Energy scores between -335.222 and -341.065. Finally, we concluded that the lower binding affinity of 5 mutants (R115L, R121G, R121Q, R121W and Y120C) with FZD4 compared with 4MY2 in terms of their Free energy and RMSD scores identified them as deleterious mutations. Thus the results indicate that our approach successfully allowed us to (1) consider computationally a suitable protocol for missense mutation (point mutation/single amino acid polymorphism) analysis before wet lab experimentation and (2) provided an optimal path for further clinical and experimental studies to characterize 4MY2 mutants in depth.

REFERENCES

- Rishi PS, Jonathan ES, Rumneek B, Andrew PS, Justis PE, and Peter K Kaiser. Oral eplerenone for the management of chronic central serous chorioretinopathy. *Int J Ophthalmol.* 8, 2015, 310–314.
- Gilmour DF. Familial exudative vitreoretinopathy and related retinopathies. *Eye.* 29, 2014, 1–14.
- Albert BM, Signaling in Cell Differentiation and Morphogenesis, 4, 2012, a008151.
- Nikopoulos K, Venselaar H, Collin RW, Riveiro-Alvarez R, Boonstra FN, Hooymans JM, Mukhopadhyay A, Shears D, van Bers M, de Wijs IJ, van Essen AJ, Sijmons RH, Tilanus MA, van Nouhuys CE, Ayuso C, Hoefsloot LH, Cremers FP, Overview of the mutation spectrum in familial exudative vitreoretinopathy and Norrie disease with identification of 21 novel variants in FZD4, *Hum Mutat.*, LRP5, and NDP, 31, 2010, 656-666.
- Miki Hiraoka, Hiroshi Takahashi, Hideo Orimo, Miina Hiraoka, Tsutomu Ogata, Noriyuki Azuma, Genetic screening of Wnt signaling factors in advanced retinopathy of prematurity, *Mol Vis.*, 16, 2010, 2572-2577.
- Ricken A, Lochhead P, Kontogiannea M, Farookhi R, Wnt signaling in the ovary: identification and compartmentalized expression of wnt-2, wnt-2b, and frizzled-4 mRNAs Wnt signaling in the ovary: identification and compartmentalized expression of wnt-2, wnt-2b, and frizzled-4 mRNAs, *Endocrinology*, 143, 2002, 2741-2749.
- James AP, Manir A, David FG, Aine R, Hiroyuki K, Kenshi H, David AM, Lisa S K, Jonathan BR, Jamie EC, Eric AP, Louise MD, Moin DM, Alexander FM, Chris FI, Carmel T, Mutations in TSPAN12 Cause Autosomal-Dominant Familial Exudative Vitreoretinopathy, *Am J Hum Genet.*, 86, 2010, 248-253.
- Jing C, Andreas S, Nathan MK, Molly RS, Roberta JD, Przemyslaw S, Jing H, Colman JH, Aimee MJ, Christopher MA, Keirnan LW, Karen IG, Akiko M, Matthew C, Lois EH, Wnt signaling mediates pathological vascular growth in proliferative retinopathy, 124, 2011, 1871-1881.
- Penn JS, Madan A, Caldwell RB, Bartoli M, Caldwell RW, Hartnett ME, Vascular Endothelial Growth Factor in Eye Disease, *Prog Retin Eye Res*, 27, 2008, 331-371.
- Xu Q, Wang Y, Dabdoub A, Smallwood PM, Williams J, Woods C, Kelley MW, Jiang L, Tasman W, Zhang K, Nathans J, Vascular development in the retina and inner ear: control by Norrin and Frizzled-4, a high-affinity ligand-receptor pair, *Cell*, 116, 2004, 883-895.
- Toomes C, Bottomley HM, Jackson RM, Towns KV, Scott S, Mackey DA, Craig JE, Jiang L, Yang Z, Trembath R, Woodruff G, Gregory-Evans CY, Gregory-Evans K, Parker MJ, Black GC, Downey LM, Zhang K, Inglehearn CF, Mutations in LRP5 or FZD4 underlie the common familial exudative vitreoretinopathy locus on chromosome 11q, *Am J Hum Genet*, 74, 2004, 721-730.
- Hiroyuki K, Shunji K, Aki Y, Eiichi U, Akihiko T, Tomoko T, Genetic variants of FZD4 and LRP5 genes in patients with advanced retinopathy of prematurity, *Mol Vis*, 19, 2013, 476-485.
- Kondo H, H Hayashi H, Oshima K, Tahira T, Hayashi K, Frizzled 4 gene (FZD4) mutations in patients with familial exudative vitreoretinopathy with variable expressivity, *Br J Ophthalmol*, 87, 2003, 1291-1295.
- Tao-Hsin C, Fu-Lien H, Matthias Z, Karl H, Jonathan E, Yvonne JE, Structure and functional properties of Norrin mimic Wnt for signalling with Frizzled4, Lrp5/6, and proteoglycan, *eLife*, 4, 2015, e06554.
- Nancy LP, Jan K, Wnt Signaling in Angiogenesis, 9, 2008, 558-565.
- Yanshu W, Amir R, Yulian Z, John W, Philip MS, Jeremy N, Norrin/Frizzled4 signaling in retinal vascular development and blood brain barrier plasticity, *Cell*, 151, 2012, 1332-1344.
- Steel DHW, Lotery AJ, Idiopathic vitreomacular traction and macular hole: a comprehensive review of pathophysiology, diagnosis, and treatment, *Eye (Lond)*, 27, 2013, S1-S21.
- MacDonald BT, He X, Frizzled and LRP5/6 receptors for Wnt/ β -catenin signaling, *Cold Spring Harb Perspect Biol*, 4, 2012, a007880.
- Xiaodong J, Valerio V, Michael TT, Barkur SS, Fielding HJ, Autosomal Recessive Familial Exudative Vitreoretinopathy Is Associated with Mutations in LRP5, *Am J Hum Genet*, 75, 2004, 878-884.
- Alessandra P, Eveline B, Elke P, Giuseppe D, Nadia LI, Angela V, Paolo V, Wim VH, Anna V, Cristina S, Identification of the first deletion in the LRP5 gene in a patient with Autosomal Dominant Osteopetrosis type I, *Bone*, 49, 2011, 568-571.
- John TS, Gene Transfer for the treatment of Neovascular Ocular Disease (An American Ophthalmological Society Thesis), *Trans Am Ophthalmol Soc*, 104, 2006, 530–560.



22. Yip YL, Scheib H, Diemand AV, Gattiker A, Famiglietti LM, Gasteiger E and Bairoch A. The Swiss-Prot Variant Page And The Modsnr Database: A Resource For Sequence And Structure Information On Human Protein Variants, *Hum Mutat*, 23, 2004, 464–470.
23. Yip YL, Famiglietti M, Gos A, Duek PD, David FP, Gateau A, Bairoch A, Annotating Single Amino Acid Polymorphisms In The Uniprot/Swiss-Prot Knowledgebase, *Hum Mutat*, 29, 2008, 361–366.
24. Brigitte B, Amos B, Rolf A, Marie-Claude B, Anne E, Elisabeth G, Maria JM, Karine M, Claire O'D, Isabelle P, Sandrine P, Michel S, The SWISS-PROT Protein Knowledgebase and its supplement Trembl In 2003. *Nucleic Acids Res*, 31, 2003, 365–370.
25. Ng PC, Henikoff S, SIFT: Predicting Amino Acid Changes That Affect Protein Function, *Nucleic acids res*, 31, 2003, 3812–3814.
26. Ng PC, Henikoff S, Predicting Deleterious Amino Acid Substitutions, *Genome Res*, 11, 2001, 863–874.
27. Ramensky V, Bork P, Sunyaev S, Human non-synonymous SNPs: server and survey, *Nucleic Acids Res*, 30, 2002, 3894–3900.
28. Cavallo A, Martin AC, Mapping Snps to Protein Sequence and Structure Data, *Bioinformatics*, 21, 2005, 1443–1450.
29. Calabrese R, Capriotti E, Fariselli P, Martelli PL, Casadio R, Functional annotations improve the predictive score of human disease-related mutations in proteins, *Hum Mutat*, 30, 2009, 1237-1244.
30. Capriotti E, Calabrese R, Fariselli P, Martelli PL, Altman RB, Casadio R, WS-SNPs&GO: a web server for predicting the deleterious effect of human protein variants using functional annotation, 14, 2013, S3-S6.
31. Schwede T, Kopp J, Guex N, Peitsch MC, SWISS-MODEL: An automated protein homology-modeling server, *Nucleic Acids Res*, 31, 2003, 3381-3385.
32. Laskowski RA, MacArthur MW, Moss DS, Thornton JM, PROCHECK: a program to check the stereochemical quality of protein structures, *J Appl Cryst*, 26, 1993, 283-291.
33. Lindahl E, Azuara C, Koehl P, Delarue M, NOMAD-Ref: Visualization, Deformation and refinement of macromolecular structures based on all-atom Normal Mode Analysis, *Nucleic Acids Res*, 34, 2006, 52–56.
34. Delarue M, Dumas P, On the use of Low-Frequency Normal Modes to Enforce Collective Movements In Refining Macromolecular Structural Models, *Proc Natl Acad Sci USA*, 101, 2004, 6957–6962.
35. Sharma S, Ding F, Nie H, Watson D, Unnithan A, Lopp J, Pozefsky D, Dokholyan NV. iFold: a platform for interactive folding simulation of proteins, *Bioinformatics*, 22, 2006, 2693–2694.
36. Magyar C, Gromiha MM, Pujadas G, Tusnády GE, Simon I, SRide: a server for identifying stabilizing residues in proteins, *Nucleic Acids Res*, 33, 2005, 303-305.
37. Tina KG, Bhadra R, Srinivasan N, PIC: Protein Interactions Calculator, *Nucleic Acids Res*, 35, 2007, 473-476.
38. Heinig M, Frishman D, STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins, *Nucleic Acids Res*, 32, 2004, 500-502.
39. Andrey T, Ilya AV, GRAMM-X public web server for protein–protein docking, 34, 2006, 310-314.
40. Yuedong Y, Yaoqi Z, Ab initio folding of terminal segments with secondary structures reveals the fine difference between two closely related all-atom statistical energy functions, *Protein Science*, 17, 2008, 1212-1219.
41. Jun-Jie L, Matthew AB, Xueqi L, Chu-Ya N, Ailong K, Hong-Wei W, Visualization of distinct substrate-recruitment pathways in the yeast exosome by EM, *nat. struct. mol. Biol*, 21, 2014, 5-102.

Source of Support: Nil, Conflict of Interest: None.

