Available online at www.derpharmachemica.com

Scholars Research Library

Der Pharma Chemica, 2013, 5(6):250-258
(http://derpharmachemica.com/archive.html)

ISSN 0975-413X
CODEN (USA): PCHHAX

In silico mutation analysis of TAR DNA binding protein-43 causing fronto temporal lobar degeneration

Sree Vishnupriya K. and Rajasekaran R.*

Bioinformatics Division, School of Biosciences and Technology, VIT University, Vellore, Tamil Nadu, India

---

ABSTRACT

The most detrimental missense mutations of TAR DNA Binding Protein 43 causing Fronto Temporal Lobar Degeneration were identified computationally and the substrate binding efficiencies of these mutations were analyzed. Out of 24 variants, I-Mutant 2.0, SIFT and PolyPhen programs identified 1 variant (D169G) that was less stable, deleterious and damaging respectively. Modeling of this one variant was performed to understand the change in their conformations with respect to the native TAR DNA Binding Protein 43 by computing their RMSD and Total energy. The native and the variant were docked with RNA to explain the binding efficiencies of those detrimental missense mutations. The loss of binding affinity with their interacting protein namely RNA was investigated by computing the flexibility of binding amino acids of TAR DNA Binding Protein 43 with their interacting proteins and computing the binding free energy (ΔG) between native and mutant complexes. The novelty of our work is to identify and validate the detrimental missense mutations based on structural stability which could be reliable and competent with other computational programs.

Key words: Missense mutation, TAR DNA Binding Protein 43, RNA, Atomic Contact Energy, Flexibility.

INTRODUCTION

Fronto temporal lobar degeneration (FTLD) is a neurodegenerative disease that selectively affects the anterior portions (i.e. frontal lobes, temporal lobes and amygdala) of the brain. FTLD typically has its onset in mid adulthood and is estimated to account for up to 20% of all cases of pre-senile dementia. It has prevalence similar to that of early-onset of Alzheimer's disease which is associated with dramatic changes in emotion [1]. The major pathological change in FTLD is the substantial gliosis which is associated with cell loss. Unlike other degenerative diseases affecting the cortex, FTLD is also characterized by substantial gliosis in the white matter as indicated by the descriptive diagnosis of progressive subcortical gliosis [2]. FTLD collectively constitutes a common cause of dementia, particularly in younger age groups and these can be broadly classified according to the major constituents of the cellular inclusions present [tau, TARDNA-binding protein 43 (TDP-43) or fused-in-sarcoma (FUS)protein, designated FTLD-tau, FTLD-TDP or FTLD-FUS, respectively. A number of causative genes have been identified, of which the genes coding microtubule-associated protein tau (MAPT) and progranulin (GRN) are the most important [3].

Fronto temporal dementia (FTD) is the second most common young-onset dementia and is clinically characterized by progressive behavioral change, executive dysfunction and language difficulties. Three clinical syndromes, behavioral variant FTD, semantic dementia and progressive non-fluent aphasia, form part of a clinicopathological

spectrum named fronto temporal lobar degeneration (FTLD). 30-50% of FTD is familial, and mutations in two genes, microtubule associated protein tau and Progranulin (GRN), account for about half of these cases. Rare defects in VCP, CHMP2B, TARDP and FUS genes have been found in a small number of families. Linkage to chromosome 9p13.2e21.3 has been established in familial FTD with motor neuron disease, although the causative gene is yet to be identified [4]. Approximately 40–50% of FTD cases express the microtubule associated protein tau in the inclusion bodies (tau+), and are categorized as Pick's disease, corticobasal degeneration (CBD), progressive supranuclear palsy (PSP) or neurofibrillary tangle dementia. Most of the remaining 50–60% of cases (tau) has inclusions, which are Tau and Synuclein Negative, but intensely labelled by antibodies to Ubiquitin, known as ITSNU, ubiquitin-only inclusions or most commonly motor neuron disease-type inclusions, with the acronym FTD–MND used for the disease. The term fronto temporal lobar degeneration is used in rare tau cases where no inclusion bodies are found [5,6,7]. Fronto temporal lobar degeneration (FTLD) often overlaps clinically with corticobasal syndrome (CBS) and progressive supranuclear palsy (PSP), both of which have prominent eye movement abnormalities. Eye movement abnormalities are sensitive markers of neurological disease and are useful in the differential diagnosis of a variety of clinical neurological syndromes. CBS and PSP are clinically, genetically and pathologically related to fronto temporal lobar degeneration(FTLD), a common cause of dementia in individuals with disease onset at age 65 as well as PSP related oculomotor abnormalities are present in FTLD, comprises three core clinical dementia syndromes, a behavioral and dysexecutive (or frontal) variant called fronto temporal dementia (FTD), and two forms of primary progressive aphasia, a temporal lobe variant, also called semantic dementia (SD), and a progressive non-fluent apahasia [8,9]. Clinical Pick's disease, more recently referred to as fronto temporal lobar degeneration which is often considered 'heterogeneous'. Currently, clinicopathologic studies recognize that ubiquitinated, tau and synuclein negative inclusions, or motor neuron disease type inclusions (MNDI) are also common in the cortex in FTD/Pick complex, In addition, dementia lacking distinctive histology (DLDH) is applied when tau or ubiquitin positive inclusions are lacking. All types include lobar atrophy, neuronal loss, gliosis, superficial spongiosis, and often ballooned neurons, and some glial abnormality, as well as Inclusion body myopathy (IBM) associated with Paget's disease of the bone (PDB) and fronto-temporal dementia (FTD), or IBMPFD (OMIM 167320), is a multisystem degenerative disorder caused by mutations in p97/VCP (valosin containing protein) on chromosome 9p12-13. IBMPFD is autosomal dominantly inherited and primarily affects muscle, brain and bone tissue. The penetrance of FTD in IBMPFD patients is 30% and its onset is at a later age (54 years) than myopathy. IBMPFD CNS pathology is tau negative and ubiquitin-positive consistent with a fronto temporal lobar degeneration with ubiquitinated inclusions (FTLD-U) [10,11]. Up to 40% of the individuals with FTLD have similarly affected first-degree relatives consistent with autosomal dominant inheritance in some patients due to mutations in the tau (MAPT) gene and progranulin gene (PGRN). Reported PGRN mutations include missense mutations generating premature stop codons, insertion or deletion mutations resulting in frame shifts or changes within initiation codons precluding transcription (ex A324T) [12,13]. Mental retardation affects an estimated 2–3% of the population. About 25% of mental retardation is believed to be caused by genetic abnormalities, and up to 10% is estimated to be caused by X chromosome mutations, X-linked mental retardation with progressive supranuclear palsy, corticobasal degeneration and Pick disease, as well as familial forms of fronto-temporal dementia (FTDP-17T) [14].

MATERIALS AND METHODS

Datasets

The protein sequence and variants (single amino acid polymorphisms/missense mutations/point mutations) of TAR DNA binding protein-43 were obtained from the Swissprot database available at http://www.expasy.ch/sprot/. The subsection of each Swissprot entry provides information on polymorphic variants, some of which polymorphic variants may be disease(s) - associated by causing defects in a given protein; most of them were nsSNPs (non-synonymous SNPs) in the gene sequence and SAPs (single amino acid polymorphisms) in the protein sequence [15-17]. The 3D Cartesian coordinates of TAR DNA Binding Protein-43 (TARDBP-43) and its complex were obtained from Protein Data Bank with PDB IDs 2CQG [18] for *in silico* mutation modeling and docking studies based on detrimental point mutants.

Analysis of functional consequence, functional change and stability changes in Sequence level for TAR DNA Binding Protein

To evaluate the data retrieved from Uniprot, we used the program I-Mutant 2.0 for predicting the protein stability changes caused by single point mutations. I-Mutant2.0 is available at:
http://gpcr.biocomp.unibo.it/cgi/predictors/IMutant2.0/ I-Mutant2. 0. cgi [19].

This program was trained and tested on a data set derived from ProTherm [20], which is the most comprehensive available database of thermodynamic experimental data of free energy changes of protein stability caused by mutations under different conditions. The output files showed the predicted free energy change value or sign (ΔΔG). Positive ΔΔG values meant that the mutated protein has higher stability and negative values indicate lower stability. We also used the program SIFT (Sorting Intolerant from Tolerant) [21], which specifically detects deleterious single amino acid polymorphisms, available at http://blocks.fhcrc.org/sift/SIFT.html. SIFT is a sequence homology-based tool, which presumes that important amino acids will be conserved in a protein family; therefore, changes at well-conserved positions tend to be predicted as deleterious [22]. The cutoff value in SIFT program was tolerance index of ≥ 0.05. The higher the tolerance index, the less functional impact a particular amino acid substitution would be likely to have. Finally to analyse the damage caused by point mutation we used the server PolyPhen (Polymorphism Phenotyping) available at http://coot.embl.de/PolyPhen/ [23]. PolyPhen calculates position-specific independent counts (PSIC) scores for each of the two variants and then computes the PSIC score difference between them. The higher the PSIC score difference, the higher the functional impact a particular amino acid substitution would be likely to have.

Computation of Total Energy and RMSD by Modelling the Single Amino Acid Polymorphisms location on Protein 3D structure

The single amino acid polymorphism database (SAAP) [24] was used to recognize the protein encoded by TAR DNA binding protein 43 (PDB ID: 2CQG) available from PDB (Protein Data Bank) and identified single point mutation residue positions. The mutations were manually implemented using SWISSPDB viewer and the energy minimization for 3D structures was performed by NOMAD-Ref server [25]. This server use Gromacs as default force field for energy minimization based on the methods of steepest descent, conjugate gradient and L-BFGS methods [26].To optimize the 3D structure of TAR DNA Binding Protein 43, we used the ifold server [27] for simulated annealing, which is based on discrete molecular dynamics and is one of the fastest strategies for simulating protein dynamics. RMSD (Root Mean Squared Deviation), a parameter was used to analyse the structural level deviation between the native and the mutant modelled structures. Divergence of the mutant structure from the native structure could be caused by substitutions, deletions and insertions [28] and the deviation between the two structures could alter the functional activity [29,30] with respect to the binding efficiency of the inhibitors, which was evaluated by their RMSD values.

Identification of Binding Sites and Computation of Atomic Contact Energy (ACE) between TAR DNA binding protein 43 and its substrate

To compute the ACE between TAR DNA binding protein 43 and its substrate (RNA), we used the program PatchDock for docking the native and mutant TAR DNA binding protein 43 with RNA to compute the ACE by using additional option of binding residue parameter. The underlying principle of this server is based on molecular shape representation, surface patch matching plus filtering and scoring [31]. It finds docking transformations that yield good molecular shape complementarities. Such transformations, when applied, induce both wide interface areas and small amounts of steric clashes. A wide interface ensured that include several matched local features of the docked molecules that have complementary characteristics were included. The PatchDock algorithm divides the Connolly dot surface representation [32] of the molecules into concave, convex and flat patches. Then, complementary patches are matched to generate candidate transformations. Each candidate transformation is further evaluated by a scoring function that considers both geometric fit and atomic desolvation energy [33, 34]. Finally, an RMSD clustering was applied to the candidate solutions to discard redundant solutions. The main reason behind Patch Dock's high efficiency is its fast transformational search, which is driven by local feature matching rather than by brute force searching of the six dimensional transformation spaces. It further speeds up the computational processing time using advanced data structures and spatial pattern detection techniques, such as geometric hashing and poses clustering.

Calculating the Total Number of Intra Molecular Interactions Using PIC server

We used PIC server for computing intra-molecular interactions for both native and mutant structures respectively. PIC (Protein Interactions Calculator) server accepts atomic coordinate set of a protein structure in the standard Protein Data Bank (PDB) format. Interactions within a protein structure and interactions between proteins in an assembly are essential considerations in understanding the molecular basis of stability and functions of proteins and their complexes. There are several weak and strong interactions that render stability to a protein structure or an assembly. It computes various interactions such as interaction between apolar residues, disulphide bridges, hydrogen bond between main chain atoms, and hydrogen bond between main chain and side chain atoms, hydrogen bond

between two side chain atoms, interaction between oppositely charged amino acids (ionic interactions), aromatic-aromatic interactions, aromatic-sulphur interactions and cation-π interactions. The PIC server [35] is available at: http://crick.mbu.iisc.ernet.in/PIC. We further analyzed the intramolecular interactions between the native and the mutant by secondary structure analysis

Analysis of secondary structure elements of native and mutant

We used STRIDE web server for analysis of secondary structure of native and mutant. STRIDE is an automatic algorithm for protein secondary structure assignment from atomic coordinates. 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 [36]. STRIDE considers both hydrogen bonding patterns and backbone geometry. The hydrogen bond energy is calculated using an empirical energy function which takes into account the distance between the donor and the acceptor and the deviations from linearity of the bond angles. A weighted product of hydrogen bond energy and torsion angle probabilities for a-helix and b-sheet is used to determine the start and stop positions of secondary structure elements based on empirically optimized recognition thresholds.

Exploring the Flexibility of Binding Pocket by Normal Mode Analysis

A quantitative measure of the atomic motions in proteins could be obtained from the mean square fluctuations of the atoms relative to their average positions. These could be related to the B-factor [37, 38]. Analysis of B-factors, therefore, could provide fresh insights into protein dynamics, the flexibility of amino acids, and protein stability [39]. Protein flexibility is important for protein function and for rational drug design [40]. In addition the flexibility of certain amino acids in a protein is useful for various types of interactions. Moreover, the flexibility of amino acids in the binding pocket is considered a significant parameter for understanding the binding efficiency. In fact, loss of flexibility impairs the binding effect [41] and vice versa. Hence, this can be analyzed by the B-factor, which is computed from the mean-square displacement $<R^2>$ of the lowest-frequency normal mode using the ElNémo server [42].

RESULTS AND DISCUSSION

The SAP Data Set from Swissprot

The TAR DNA binding protein-43 and 24 variants, namely, A90V, D169G, N267S, G287S, G290A, G294A, G294V, G295R, G295S, G298S, A315T, Q331K, S332N, G335D, M337V, Q343R, G348C, G357R, R361T, S379P, A382T, N390S and S393L investigated in this work were retrieved from the Swissprot database.

Identification of Functional Variants by I-mutant 2.0

Of the 24 variants, 7 variants were found to be less stable using the I-Mutant 2.0 server (Table 1). Among these 7 variants, two variants showed a ΔΔG value < -1.0 and five variants showed a ΔΔG value > -1.0 as depicted in (Table 1). Of the seven variants that showed a negative ΔΔG, four variants (A90V, N269S, S379P and N390S) retained their amino acid properties. Two variants (A315T and A382T) changed from non-polar to polar uncharged and one variant (D169G) changed from negatively charged to non polar. Indeed, by considering only amino acid substitution based on physico-chemical properties, we could not be able to identify the detrimental effect. Rather, by considering the sequence conservation along with the above said properties could have more advantages and reliable to find out the detrimental effect of missense mutations.

Deleterious Single Point Mutants Identified by the SIFT Program

The degree of conservation of a particular position in a protein was determined using sequence homology based tool SIFT. The protein sequences of the 24 variants were submitted to SIFT to determine their tolerance indices. As the tolerance level increases, the functional influence of the amino acid substitution decreases and vice versa. Among the 24 variants, 10 variants were found to be deleterious, having tolerance index scores of ≤0.05 (Table 1). Among these 24 variants, 4 variants showed a very high deleterious tolerance index score of 0.00. One variant had a tolerance index score of 0.02, three variants had tolerance index scores of 0.03, and one had a tolerance index score of 0.04 (Table 1). Interestingly, two deleterious variants identified by SIFT also were seen to be less stable by the I-Mutant 2.0 server.

Damaging Single Point Mutations identified by the PolyPhen Server

Structural level alterations were determined by PolyPhen program. Protein sequence with mutational position and amino acid variants associated with the 24 single point mutants were submitted to the PolyPhen server. A PSIC score difference of 0.5 and above was considered to be damaging. It could be seen from Table 1 that, out of 24 variants, 11 were considered to be damaging by PolyPhen. These variants also exhibited a PSIC score difference from 0.01 to 0.999.

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

VARIANTS	ΔΔG	Tolerance index	PSIC SD
A90V	**-1.49**	0.38	0.149
D169G	**-0.87**	**0.03**	**0.635**
N267S	**-0.51**	0.88	0.007
G287S	0.81	0.45	0.404
G290A	0.47	0.98	0.02
G294A	0.67	0.65	0.028
G294V	0.08	**0.05**	0.139
G295R	0.62	0.17	**0.733**
G295S	0.95	0.59	0.01
G298S	0.85	0.59	0.063
A315T	**-1.1**	0.31	0.063
Q331K	0.38	**0.03**	0.462
S332N	1.84	**0.03**	0.027
G335D	0.15	**0**	**0.999**
M337V	0.67	0.15	**0.635**
Q343R	0.63	**0**	**0.967**
G348C	1.59	**0.02**	**0.992**
G357R	1.33	0.35	**0.917**
R361T	0.31	**0**	**0.604**
S379C	0.16	**0**	**0.951**
S379P	**-0.53**	**0.02**	0.001
A382T	**-0.48**	0.2	0.267
N390S	**-0.07**	1	**0.956**
S393L	0.17	**0.04**	**0.804**

*Letters in **bold** indicate mutants predicted to be less stable, deleterious and damaging by I-Mutant 2.0, SIFT and PolyPhen respectively*

Rational Consideration of Detrimental Point Mutations

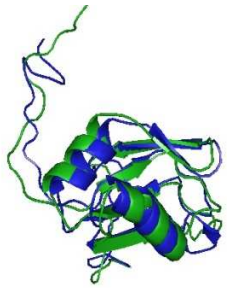
We rationally considered one most potential detrimental point mutation (D169G) for further course of investigations because it was commonly found to be less stable, deleterious, and damaging by the I-Mutant2.0, SIFT and PolyPhen servers respectively. We considered the statistical accuracy of these three programs, I-Mutant improves the quality of the prediction of the free energy change caused by single point protein mutations by adopting a hypothesis of thermodynamic reversibility of the existing experimental data. The accuracy of prediction for sequence and structure based values were 78% and 84% with correlation coefficient of 0.56 and 0.69, respectively [44]. SIFT correctly predicted 69% of the substitutions associated with the disease that affect protein function. PolyPhen-2 evaluates rare alleles at loci potentially involved in complex phenotypes, densely mapped regions identified by genome-wide association studies, and analyses natural selection from sequence data, where even mildly deleterious alleles must be treated as damaging. PolyPhen-2 was reported to achieve a rate of true positive predictions of 92% [43, 44, 45]. To obtain precise and accurate measures of the detrimental effect of our variants, comprehensive parameters of all these three programs could be more significant than individual tool parameters. Hence, we further investigated this detrimental missense mutation by structural analysis.

Computing Total Energy and RMSD by Modelling of Mutant Structures

Mapping the one variant namely, D169G into TAR DNA binding protein-43 structure information was obtained from SAAPdb. The available structure for TAR DNA Binding protein has a PDB ID 2CQG. The mutational position and amino acid variants were mapped in the native structure. Mutation at specified position was performed by SWISSPDB viewer independently to get modeled structures. Then, energy minimization was performed by the NOMAD - Ref server for both the native structure (PDB 2CQG) and mutant modeled structures. In order to find out the structural stability of TAR DNA binding protein-43 of native and mutant, we computed the total energy, which included bonds, angles, and torsions, non-bonded and electro-static constraints from GROMOS 96 force field implemented in DeepView to check their stability. It could be seen from Table 2 that the total energy of the native

protein had -10091.175KJ/Mol whereas all the 3 mutants had the total energy higher than native protein. The higher the total energy, lesser was the stability of the protein structure. In order to find out the deviation between the two structures, we superimposed the native structure (PDB 2CQG) with all the mutant modeled structures to get the RMSD. The higher was the RMSD value, more was the deviation between the native and mutant structure, which in turn change the binding efficiency with its interacting partners due to deviation in the 3D space of the binding residues of TAR DNA binding protein-43. Table 2 showed the RMSD for native structure with all the mutant modeled structures. Figure 2 showed the superimposed structure of native TAR DNA Binding Protein 43 (green) with mutant D169G (Blue)

Figure 2. Superimposed structure of the native protein (green) with mutant



Superimposed structure of native TAR DNA Binding Protein 43 (green) with mutant D169G (Blue) structure showing RMSD of 2.75 Å.

Computing the Intra-molecular interactions in TAR DNA Binding protein

We further evaluate the stability of protein structure by using the PIC server to identify the number of intra-molecular interactions for both native and mutant structures. Interactions within a protein structure and the interactions between proteins in an assembly were essential considerations in understanding the molecular basis of stability and functions of proteins and their complexes. There are several weak and strong intra-molecular interactions that render stability of a protein structure. Therefore these intra-molecular interactions were computed by the PIC server in order to further substantiate the stability of protein structure. Based on this analysis, we found that a total number of 167 intra-molecular interactions were obtained in the native structure of TAR DNA binding protein-43. On the other hand, the mutant structure (D169G) of TAR DNA binding protein-43 established the intra-molecular interactions of 268 as shown in Table 2. We further evaluated the effect of this detrimental missense mutation by studying the secondary structure elements of both native and mutant protein. Since the secondary structure elements are different in native and mutant this could be the reason for alteration of conformation of mutant structure.

Analysis of secondary structure elements of native and mutant using STRIDE web server

We further evaluate the distribution of secondary structure elements in native and mutant protein (D169G). Using the STRIDE web server we calculated the number of secondary structure elements for native and mutant structures. From table 2 we can see that the distribution of secondary structure elements in native is distributed as 14 coils, 31 turns, 30 strands, and 28 alpha helixes where as for variant (D169G) the secondary structure elements are distributed as 13 coils, 32 turns, 31 strands and 28 alpha helixes. Since the secondary structure elements are different in native and mutant this could be the reason for alteration of conformation of mutant structure. We further analyzed the effect of this detrimental missense by performing binding analysis between TAR DNA binding protein-43 and RNA through docking studies in order to understand the functional activity of TAR DNA binding protein-43.

Investigating the Rationale of Binding Efficiency for Native and Mutant TAR DNA binding protein-43 with RNA.

In order to find out the binding efficiency of native and mutant TAR DNA binding protein-43 with its interacting partner RNA, we implemented molecular dynamics approach for rationalizing the functional activity of this one mutant D169G. In this analysis, we performed 1 missense mutation (D169G) in the chain A of the PDB ID 2CQG

by swisspdb viewer independently and energy minimization was performed for the entire complex (both native and mutant complex) by GROMACS (Nomad-ref) followed by simulated annealing to get the optimized structures using a discrete molecular dynamics approach (ifold).

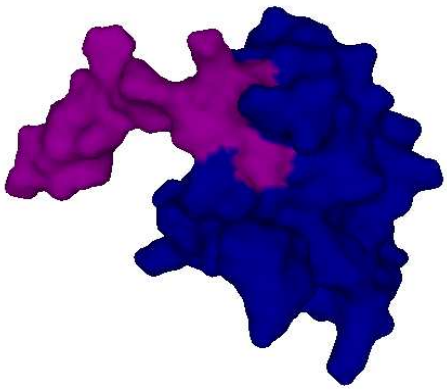
Docking was performed using the PatchDock server between TAR DNA Binding protein 43 and RNA with both native and mutant modeled structures of TAR DNA Binding protein 43 to find out the binding efficiency in the form of Atomic Contact Energy (ACE). By this analysis, we found that, the ACE between RNA and native TAR DNA Binding protein 43 was found to be -24.70kcal/mol, whereas with mutant, the ACE was found to be -16.77kcal/mol (Table 2). Figure 3 shows the docked complex of Mutant D169G TAR DNA Binding Protein (blue) and RNA (magenta). This data clearly portrays that the maximum binding effect of RNA with the mutant (D169G) might be due to the 3D conformation of RNA which exclusively made a comfortable fit with less ACE into the 3D space of the binding residues of these mutants as compared to the native.

Table 2. RMSD, Total Energy, Number of Intramolecular Interactions, ACE and secondary structure distribution of native and mutant

Tool	Native	D169G
RMSD (Å)	0 Å	2.75 Å
Total Energy (KJ/mol)	-4574.801	-4471.873
No. of Intramolecular interactions	167	268
ACE	-24.70	-16.77
Secondary Structure Distribution		
Coil	14	12
Turn	31	32
Strand	30	31
AlphaHelix	28	28

*RMSD root mean square deviation, **ACE**-Atomic Contact Energy*

Figure 3. Docked complexes of Native and Mutant TAR DNA Binding protein 43 with RNA



Docked complex of Mutant D169G TAR DNA Binding Protein (blue) and RNA (magenta) having the ACE score of -16.17,

The majority of amino acids in active site showed loss of flexibility

To understand the cause of the lower substrate binding efficiency of the detrimental missense mutation, we used the program ElNemo to compare the flexibility of amino acids that were involved in binding with RNA of both the native protein and the mutants. Table 3 and depicted the flexibility of the amino acids in the substrate binding pocket (active site) of both the native and mutant proteins by means of the normalized mean square displacement $<R^2>$. These data were further sorted into three different categories of flexibility. One was where the $<R^2>$ of the amino acids in the substrate binding pocket of the mutant was the same as that of the native protein (termed identical flexibility). The second category was where the $<R^2>$ of the amino acids in the substrate binding pocket of the mutant was higher than that of the native protein (termed increased flexibility). The last category was where the <R2> of the amino acids in the substrate binding pocket of a mutant was lower than that of the native protein (termed decreased flexibility). From this analysis, we found that the substrate binding amino acids of these 3 mutants have increased flexibility (Table 4). Thus the majority of the amino acids participating in substrate binding

of these mutants lost their flexibility, leading to a loss of binding efficiency with the substrate. Thus this study showed that, increased flexibility of the protein was the cause for the loss in substrate binding affinity.

Table 3 Comparison of normalized mean square displacement of binding amino acids for native and mutant TAR DNA Binding Protein 43 protein

Binding residues	Normalized mean square displacement $<R^2>$		
	Native	**Native EM**	**D169G**
Pro112	0.0121	0.0078	0.0066
Trp113	0.0104	0.0026	0.0023
Lys114	0.0131	0.0078	0.0085
Thr115	0.0114	0.0068	0.0071
Lys137	0.0007	0.0093	0.0108
His 143	0.0055	0.0062	0.0084
Ser 144	0.0048	0.0040	0.0051

*Bold indicates increased flexibility of mutants than native. *indicates decreased flexibility of mutants than native.*

Table 4 Substrate binding amino acids of mutants with different ranges of flexibility based on$<R^2>$

Mutants	A	B	C
	$<R^2>$		
D169V	2	5	0

Where the letter 'A' denotes amino acids with decreased flexibility of mutants than native; 'B' denotes amino acids with increased flexibility of both native and mutants; 'C' denotes amino acids with identical flexibility of both native and mutants.

CONCLUSION

Of the 24 variants that were retrieved from Swissprot, 7 variants found less stable by IMutant2.0, 11 variants were found to be deleterious by SIFT and 11 variants were considered damaging by PolyPhen. One variant was selected as potentially detrimental point mutation because it was commonly found to be less stable, deleterious and damaging by the I-Mutant 2.0, SIFT and PolyPhen servers, respectively. The structure of this variant was modeled and the RMSD between the mutants and native structure was found to be 2.75 Å. Docking analysis between RNA and the native and mutant modeled structures generated Atomic Contact Energy scores between -24.70 and -16.77. Finally, we concluded that the lower binding affinity of mutant (D169G) with RNA compared with TAR DNA Binding protein 43 in terms of their ACE and RMSD scores identified them as deleterious mutations. Normalized mean square displacement $<R^2>$ by normal mode analysis allowed us to conclude that the majority of amino acids in the mutants bound to RNA (i.e. are in the active site) had increased flexibility which could be the cause for their loss in substrate binding affinity. Thus the results indicate that our approach successfully allowed us to (1) consider computationally a suitable protocol for the 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 TAR DNA Binding protein mutants in depth.

Acknowledgment
The authors thank the management of Vellore Institute of Technology for providing the facilities to carry out this work.

REFERENCES

[1] V. E. Sturm, H. J. Rosen, S. Allison and L. Bruce, *Brain* **2006**, 129, 2508–2516
[2] E. Schofield, C. Kersaitis, C. E. Shepherd and J. Jillian, *Brain* **2003**, 126, 827-840
[3] J. D. Rohrer, T. Lashley, J. M. Schott and J. E. Warren, *Brain* **2011**, 134; 2565–2581
[4] H. Seelaar, J. D. Rohrer, Y. A. L. Pijnenburg and N. C. Fox, *J Neurol Neurosurg Psychiatry* **2011**, 82:476-486
[5] D. G. Munoz, D. S. Teodoro and J. Woulfe, *Brain* **2005**, 128, 1962–1963
[6] M. Neumann, R. Rademakers, S. Roeber and M. Baker, *Brain* **2009**, 132; 2922–2931
[7] R. Zahn, J. Moll, V. Iyengar and E. D. Huey, *Brain* **2009**, 132; 604–616
[8] S. Garbutt, A. Matlin, J. Hellmuth and A. K. Schenk, *Brain* **2008**, 131, 1268-1281
[9] J. C. L. Looi, O. Lindberg, B. B. Zandbelt and P. Ostberg, *AJNR Am J Neuroradiol*, **2009,** 29:1537– 43

[10] A. Kertesz, P. McMonagle, M. Blair and W. Davidson, *Brain* **2005**, 128, 1996–2005
[11] J. S. Ju and C. C. Weihl, *Human Molecular Genetics*, **2010**, 19, R38-R 45
[12] M. Stuart, P. Brown, S. Rollinson and D. Du Plessis, *Brain* **2008**, 131, 721-731
[13] I. R. A. Mackenzie, M. Baker, S. P. Brown and G. Y. R. Hsiung, *Brain* **2006** 129, 3081–3090
[14] J Y Garbern, M. Neumann, J. Q. Trojanowski and M. Y. Virginia, *Brain*, **2008** 131, 1268-1281
[15] Y. L. Yip, H. Scheib and A. V. Diemand, *Hum Mutat*. **2004**, 23,464–470
[16] Y. L. Yip, M. Famiglietti and A. Gos, *Hum Mutat*. **2008**, 29,361–366
[17] B. Boeckmann, A. Bairoch and R. Apweiler, *Nucleic Acids Res* **2003**, 31,365–370
[18] H. M. Berman, J. Westbrook and Z. Feng, *Nucleic Acids Res.* **2000**, 28,235–242
[19] E. Capriotti, P. Fariselli and R. Casadio, *Nucleic Acids Res*, **2005**, 33,306–310
[20] K. A. Bava, M. M. Gromiha and H. Uedaira, Nucleic Acids Res 2004, 32:120–121
[21] P. C. Ng and S. Henikoff, *Nucleic Acids Res.* **2003**, 31,3812–3814
[22] P. C. Ng and S. Henikoff, *Genome Res.* **2001**, 11,863–874
[23] V. Ramensky, P. Bork and S. Sunyaev, *Nucleic Acids Res.* **2002**, 30, 3894–3900
[24] A. Cavallo and A. C. Martin, *Bioinformatics*, **2005**, 21,1443–1450
[25] E. Lindahl, C. Azuara and P. Koehl, *Nucleic acids Res.*, **2006**, 34,52–56
[26] M. Delarue and P. Dumas, *Proc Natl Acad Sci*, **2004**, 101, 6957–6962
[27] S. Sharma, F. Ding and H. Nie, *Bioinformatics*, **2006**, 22, 2693-2694
[28] J. H. Han, N. Kerrison and C. Chothia, *Structure*, **2006**, 14, 935–945
[29] S. D. Varfolomeev, I. V. Uporov and E. V. Fedorov, *Biochemistry (Mosc)*, **2002**, 67, 1099–1108
[30] K. C. Chou and L. Carlacci, *Protein Eng*, **1991**, 4, 661-7
[31] D. Duhovny, R. Nussinov and H. J. Wolfson, *Lecture Notes in Computer Science, Springer, Rome, Italy*, **2002**, 2452, 185–200
[32] M. L. Connolly, *Science*, **1983**, 221, 709–713
[33] D. Schneidman-Duhovny, Y. Inbar and R. Nussinov, *Nucleic Acids Res*, **2005**, 33, 363–367
[34] C. Zhang, G. Vasmatzis and J. L. Cornette, *J Mol Biol.* **1997**, 267, 707–726
[35] K. G. Tina, R. Bhadra and N. Srinivasan, *Nucleic Acids Res*. **2007,** 35, 473-476
[36] M. Heinig and D. Frishman, *Nucl.Acids Res*, **2004**, 32, W500-502
[37] Z. Yuan, T. L. Bailey and R. D. Teasdale, *Proteins*, **2005**, 58, 905–912
[38] D. Ringe and G. A. Petsko, *Methods Enzymol*. **1986**, 131, 389–433
[39] S. Parthasarathy and M. R. Murthy, *Protein Eng*. **2000**, 13, 9–13
[40] H. A. Carlson and J. A. McCammon, *Mol Pharmacol*, **2000**, 57, 213–218
[41] A. Hinkle and L. S. Tobacman, *J Biol Chem*, **2003**, 278, 506–513.
[42] K. Suhre and Y. H. Sanejouand, *Nucleic Acids Res.* **2004**, 32, 610–614
[43] P. Kumar, S. Henikoff and P. C. Ng, *Nat Protoc*. **2009**, 4(7), 1073-81
[44] E. Capriotti, P. Fariselli and I. Rossi, *BMC Bioinformatics*. **2008**, 26, S6-59
[45] I. A. Adzhubei, S. Schmidt and L. Peshkin, *Nat Methods*. **2010**, 7(4), 248–249