Metabolic Brain Disease (2019) 34:1661–1677
https://doi.org/10.1007/s11011-019-00473-6

ORIGINAL ARTICLE

Rare Angiogenin and Ribonuclease 4 variants associated with amyotrophic lateral sclerosis exhibit loss-of-function: a comprehensive in silico study

Aditya K. Padhi[1,2] · Priyam Narain[1] · James Gomes[1]

Received: 9 May 2019 / Accepted: 18 July 2019 / Published online: 31 July 2019

Abstract

Amyotrophic Lateral Sclerosis (ALS), a debilitating neurodegenerative disorder is related to mutations in a number of genes, and certain genes of the Ribonuclease (RNASE) superfamily trigger ALS more frequently. Even though missense mutations in Angiogenin (ANG) and Ribonuclease 4 (RNASE4) have been previously shown to cause ALS through loss-of-function mechanisms, understanding the role of rare variants with a plausible explanation of their functional loss mechanisms is an important mission. The study aims to understand if any of the rare ANG and RNASE4 variants catalogued in Project MinE consortium caused ALS due to loss of ribonucleolytic or nuclear translocation or both these activities. Several in silico analyses in combination with extensive molecular dynamics (MD) simulations were performed on wild-type ANG and RNASE4, along with six rare variants (T11S-ANG, R122H-ANG, D2E-RNASE4, N26K-RNASE4, T79A-RNASE4 and G119S-RNASE4) to study the structural and dynamic changes in the catalytic triad and nuclear localization signal residues responsible for ribonucleolytic and nuclear translocation activities respectively. Our comprehensive analyses comprising 1.2 μs simulations with a focus on physicochemical, structural and dynamic properties reveal that T11S-ANG, N26K-RNASE4 and T79A-RNASE4 variants would result in loss of ribonucleolytic activity due to conformational switching of catalytic His114 and His116 respectively but none of the variants would lose their nuclear translocation activity. Our study not only highlights the importance of rare variants but also demonstrates that elucidating the structure-function relationship of mutant effectors is crucial to gain insights into ALS pathophysiology and in developing effective therapeutics.

Keywords Amyotrophic lateral sclerosis · Angiogenin · Loss-of-functions · Molecular dynamics · Physicochemical properties · Rare variants · Ribonuclease 4

Introduction

ALS is a debilitating terminal neurodegenerative disorder, and yet there is no therapy to arrest its progression (Andersen and Al-Chalabi 2011). Although substantive progress has been achieved in determining the genetic basis of this disease (Renton et al. 2011; Rosen et al. 1993; Sreedharan et al. 2008; Brown and Al-Chalabi 2017; Vance et al. 2009; Padhi and Hazra 2019; Chow et al. 2009; DeJesus-Hernandez et al. 2011), it is not well understood how the genetic variations mediate ALS onset and progression based on epidemiological data. In the past decades, more than 100 genes involved in the manifestation of ALS have been identified, though mutations in *SOD1* and hexanucleotide repeat expansion in the

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11011-019-00473-6) contains supplementary material, which is available to authorized users.

✉ Aditya K. Padhi
adityapadhi.iitd@gmail.com

✉ James Gomes
jgomes@bioschool.iitd.ac.in

[1] Kusuma School of Biological Sciences, Indian Institute of Technology Delhi, Hauz Khas, New Delhi 110016, India

[2] Present address: Laboratory for Structural Bioinformatics, Field for Structural Molecular Biology, Centre for Biosystems Dynamics Research, RIKEN, 1-7-22 Suehiro, Tsurumi, Yokohama, Kanagawa 230-0045, Japan

C9ORF72 are primarily associated with ALS (Abel et al. 2012; Vasaikar et al. 2013; Ferraiuolo et al. 2011; Hardiman et al. 2019; DeJesus-Hernandez et al. 2011; Renton et al. 2011; Narain et al. 2017). Recent studies suggest that although genetic factors may not be directly involved in disease manifestation, they could significantly increase the risk of ALS under the influence of environmental and other related factors. These variations thereby play an important role in ALS pathogenesis by acting as disease modifiers or risks factors (Renton et al. 2014; Krüger et al. 2016). Till date a notable fraction of genetic variations contributing to ALS remain unknown and pose a challenge in defining the genetic paradigm of ALS. Recent studies have led to the identification of a number of rare variants in both familial and sporadic ALS, with an increased burden of these variants in cases than controls (Marangi and Traynor 2015; Narain et al. 2018). Further, these studies suggest that the missing genetic basis of ALS could lie in rare variants; which act as susceptibility factors. Additionally, the unknown heritability of many human traits can be attributed to common variants having low-frequency (MAF 1–5%) or having minor to modest consequence (MAF <1%), or an amalgamation of both.

Although a number of reports have shed light on the increased association of rare variants in ALS, finding out the actual causal role of such variants is still in infancy. Large scale genome wide association studies (GWAS) of ALS patients suggests that rare variants primarily define the genetic architecture of ALS in contrast to other disease like schizophrenia that have association with large number of common variants (Hardiman et al. 2019). Additionally, rare variants that are associated with the increased risk of ALS are likely to be confined to specific individuals, families and ancestral population (Hardiman et al. 2019). Furthermore, not only the discovery of disease causing rare variants but also understanding the functional consequences and disease predisposition due to the variants warrants detailed investigation. To address this critical gap, we focused our study on the variants from Project MinE consortium that aims to sequence whole genome of >15,000 ALS patients and > 7,500 control individuals to provide a coherent genetic architecture of ALS (Van Rheenen et al. 2018).

In this study, we mined the genetic variation data from Project MinE Variant Browser for the two genes of ribonuclease superfamily, namely ANG and RNASE4, and identified a total of six novel and rare uncharacterized variants from different populations (Table 1, Fig. 1). The functional sites of ANG and RNASE4 with a superposition between them showing the high structural similarity is shown in Supplementary Information Fig. 1. As missense mutations in ANG and RNASE4 are frequently associated with ALS through loss of either ribonucleolytic or nuclear translocation activity or both these functions (Greenway et al. 2004; Greenway et al. 2006; Padhi et al. 2014a, b, 2019), we were intrigued if any of the Project MinE derived rare ANG and RNASE4 variants could also exhibit loss-of-function characteristics. In order to understand their role in ALS development, we carried out several in silico analyses and a total of 1.2 μs all-atom MD simulations for wild-type and six rare variants (two from ANG and four from RNASE4). Our comprehensive analyses from MD simulations and related in silico approaches reveal that certain structural and dynamic attributes might be involved in loss of ribonucleolytic activity governed by the catalytic triad residues (His13, Lys40, His114 in ANG and His12, Lys40, His116 in RNASE4), thus predisposing the individuals carrying these mutations to ALS. As our recent reports employing MD simulations and functional assays have effectively shown the molecular mechanisms by which ANG and RNASE4 variants leads to the loss of ribonucleolytic and nuclear translocation activity (Padhi et al. 2014a, b, 2018, 2012, 2013a, b, 2019), our inclusive analysis here showed that certain rare variants cataloged in Project MinE are also loss-of-function in nature, thereby causing ALS.

Table 1 Rare variants retrieved from ProjectMinE and analyzed in this study

ANG/RNASE4 variants with signal peptide sequence	ANG/RNASE4 variants after removal of signal peptide sequence	Chromosomal location	Nucleotide change	Allele frequency (cases)*	Allele frequency (controls)*	Allele frequency (all)*
T35S-ANG	T11S-ANG	chr14:21161827	C > G	0.00	0.000273	0.0000807
R146H-ANG	R122H-ANG	chr14:21162160	G > A	0.000115	0.00	0.0000807
D30E-RNASE4	D2E-RNASE4	chr14:21167620	T > A	0.00	0.000273	0.0000807
N54K-RNASE4	N26K-RNASE4	chr14:21167692	C > G	0.000115	0.00	0.0000807
T107A-RNASE4	T79A-RNASE4	chr14:21167849	A > G	0.00	0.000273	0.0000807
G147S-RNASE4	G119S-RNASE4	chr14:21167969	G > A	0.000115	0.00	0.0000807

*Allele frequency (cases): Allele Frequency among all Project MinE cases, Allele frequency (controls): Allele Frequency among all Project MinE controls, Allele frequency (all): Allele Frequency among all Project MinE cases and controls

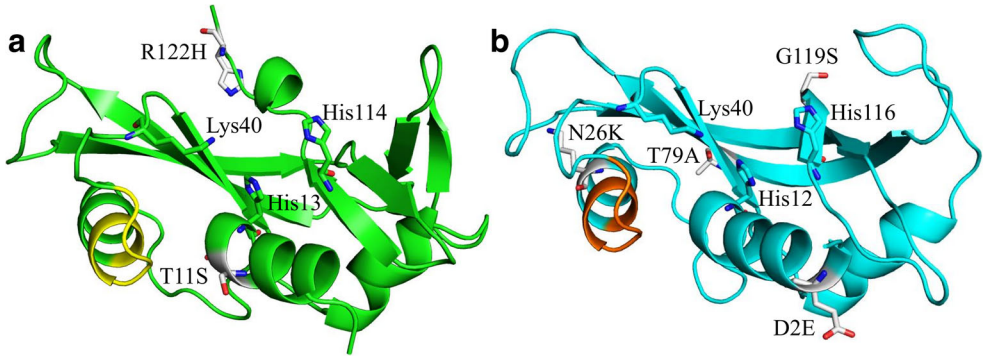

Fig. 1 Rare variants of ANG and RNASE4 identified from the ALS patients. a Cartoon represented structure of Human Angiogenin (PDB entry: 1B1I) showing the two rare variants T11S and R122H. Rare variants are labelled and shown in stick models in grey color. **b** Cartoon represented structure of Human Ribonuclease 4 (PDB entry: 1RNF) showing the four rare variants D2E, N26K, T79A and G119S. Rare variants are labelled and shown in stick models in grey color

Materials and methods

Retrieval of ANG and RNASE4 rare variants

Human ANG and RNASE4 variants were retrieved from Project MinE using the data browser (http://databrowser.projectmine.com/), which includes whole genome sequencing data from 4,366 ALS cases and 1,832 controls. In the first step, we retrieved the variant information of ANG and RNASE4 and filtered with the following criteria available at Project MinE Data browser: missense variants, disruptive and damaging, rare variant burden 0.5%. The variants selected for analysis were those for which the functional mechanism was not known. The final list of variants included ANG (T35S, R146H) and RNASE4 (D30E, N54K, T107A, G147S) (Table 1). Additionally, the allele frequency data retrieved from Project MinE cases and controls was used to classify the variants as rare variants (MAF < 0.05%). Next, human ANG (Accession number: P03950) and human RNASE4 (Accession number: P34096) protein sequences were retrieved from Uniprot to locate the rare variants of mutated residues. After excluding the N-terminal 24 amino acids from ANG and 28 amino acids from RNASE4 corresponding to the signal peptide sequence (not part of the mature protein), the six resulting rare variants (two in ANG and four in RNASE4) were mapped and used subsequently (Table 1). The ANG and RNASE4 rare variants along with their nucleotide change and allele frequency information are presented in Table 1.

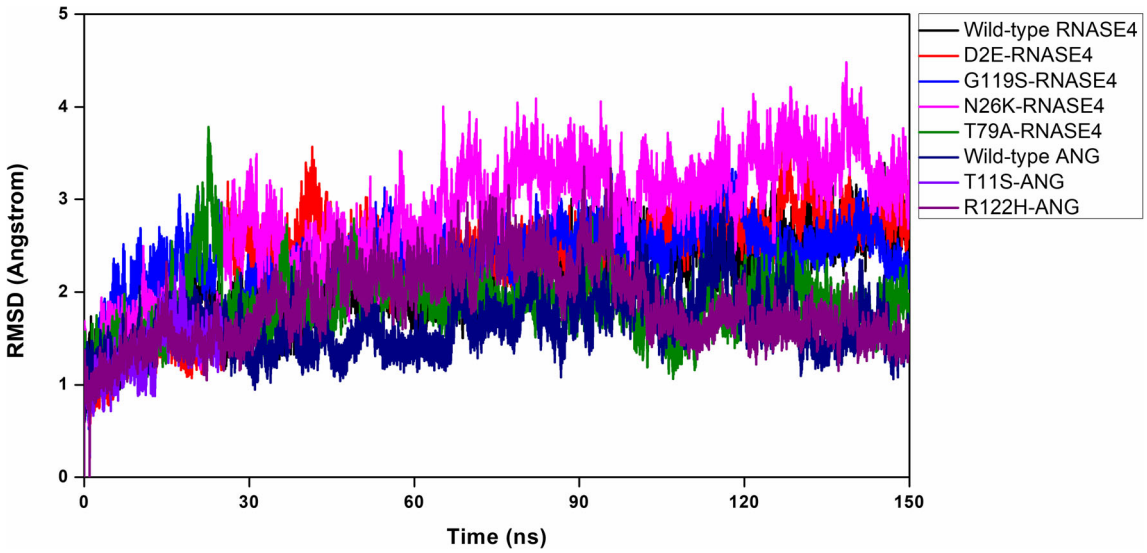

Fig. 2 Root mean square deviation (RMSD) profiles of ANG and RNASE4 rare variants. Plots showing the backbone RMSD profiles of wild-type ANG, wild-type RNASE4 and rare variants during the course of MD simulations. RMSD values of the backbone atoms obtained from the respective equilibrated conformations of the variants are computed and shown. RMSD profiles of wild-type ANG, T11S-ANG, R122H-ANG, wild-type RNASE4, D2E-RNASE4, N26K-RNASE4, T79A-RNASE4 and G119S-RNASE4 are shown with their respective color legends

Prediction of pathogenicity and physicochemical properties from in silico tools

In recent years, several methods have been developed to determine the effect of mutations on proteins and their function. These approaches are mainly based on physical and chemical properties of amino acids and their side chains (Mueller et al. 2014). In this study, we aimed at assessing the pathogenicity of rare variants through a combination of different prediction tools available at Variant Effect Predictor (VEP) webserver (McLaren et al. 2010). We predicted the effect of the mutations using the several widely used such as SIFT, PolyPhen 2.0, Mutation Assessor, Mutation Taster, Condel, LRT and PROVEAN (Table 2) (Ng and Henikoff 2003; Adzhubei et al. 2010; Reva et al. 2011; Schwarz et al. 2014; González-Pérez and López-Bigas 2011; Chun and Fay 2009; Choi and Chan 2015). Briefly, SIFT (sorting intolerant from tolerant) uses sequence homology based on the conceivability value (value<0.05; deleterious and value>0.05; tolerated) to predict the effect of amino acid substitution on protein function whereas PolyPhen 2.0 uses a combination of structural and sequence features using naïve Bayesian classifier with position specific independent count (PSIC) to differentiate a mutation as likely damaging (PSIC 0.85–1), possibly damaging (PSIC 0.02–0.85) and benign (PSIC 0–0.02). The LRT_pred (Likelihood ratio test) identifies a subset of deleterious mutations that disrupt highly conserved amino acids within protein-coding sequences, where lower scores represent more deleterious mutations. Mutation Assessor predicts the functional impact of amino acid substitutions in proteins using evolutionary conservation of the affected amino acid in protein homologs and gives a prediction as either 'neutral', 'low', 'medium' and 'high', where variants with higher scores are more likely to be deleterious. Mutation Taster uses Bayes Classifier to classify a mutation as deleterious or neutral, where value close to 1 indicates a high 'security' of the prediction. PROVEAN (Protein Variation Effect Analyzer) predicts whether an amino acid substitution or indel has an impact on the biological function of a protein, where a score ≤ -2.5 classifies a mutation to be "deleterious" and score > -2.5 classifies it to be "neutral". Condel (Consensus deleteriousness score of missense mutations) uses a weighted average of the normalized scores of the individual methods (WAS), where the scores of different methods are weighted using the complementary cumulative distributions of approximately 20.000 missense single nucleotide polymorphisms (SNPs), both deleterious and neutral. VEP also provides the data for variants from dbNSFP and therefore, we used the dbNSFP functional prediction tool on VEP for generating individual scores for each of the filtered variant for the tools: Mutation Assesor, Mutation Taster, LRT and PROVEAN. Finally, we generated a consensus score for the listed variants using Condel tool at VEP web interface. For understanding the physicochemical and structural properties due to rare variants, the HOPE server, an automatic mutant analysis server that accumulates and merges information from a series of web-servers and databases was used to provide further insights (Venselaar et al. 2010).

Modeling of mutant proteins and all-atom MD simulations

The X-ray crystal structures of human ANG and human RNASE4 with Protein Data Bank (PDB) entries 1B1I and 1RNF respectively were obtained and used as the starting structures for MD simulations (Leonidas et al. 1999;

Table 2 Prediction of pathogenicity of ANG and RNASE4 rare variants using in silico tools

ANG/RNASE4 rare variant	SIFT	PolyPhen 2.0	LRT_pred	Mutation Assessor	Mutation Taster	PROVEAN	Condel
T11S-ANG	T (0.76)	B (0)	U (0.28)	M (2.6)	N (1)	N (−0.92)	N (0)
R122H-ANG*	**D (0.02)**	B (0)	U (0.04)	N (0)	N (1)	N (−0.21)	N (0.38)
D2E-RNASE4	T (1)	B (0)	N (0.59)	N (−0.22)	N (1)	N (1.95)	N (0)
N26K-RNASE4*	**D (0.02)**	**D (0.9)**	**D (0)**	**M (2.98)**	N (0.71)	**D (−5.42)**	**D (0.74)**
T79A-RNASE4*	**D (0.01)**	**D (0.63)**	**D (0)**	**H (3.96)**	**D (0.99)**	**D (−4.7)**	**D (0.62)**
G119S-RNASE4	T (0.07)	B (0.04)	N (0.91)	N (0.48)	N (1)	N (0.49)	N (0.32)

*Results suggesting pathogenicity are marked in bold

SIFT: D = Deleterious, T = Tolerated

PolyPhen 2.0: D = Possibly damaging, B=Benign

LRT: D = Deleterious, N=Neutral, U=Unknown

Mutation Assessor: H=High, M = Medium, N=Neutral

Mutation Taster: D = Disease causing, N = polymorphism

PROVEAN: D = Deleterious, N=Neutral

Condel: D = Deleterious, N=Neutral

Terzyan et al. 1999). Disulphide bonds in ANG formed between residues Cys26-Cys81, Cys39-Cys91 and Cys57-Cys107 and in RNASE4 formed between residues Cys57-Cys107, Cys64-Cys71, Cys25-Cys81 and Cys39-Cys92 were processed according to AMBER protocols. The N-terminal signal peptide sequence that is not part of the mature protein was removed from ANG and RNASE4 structures and in silico mutagenesis was performed. The T11S and R122H-ANG, and D2E, N26K, T79A and G119S-RNASE4 variant structures were modelled by mutating the source residues with desired amino acids while keeping the secondary structures intact. The wild-type structures of ANG and RNAE4 along with variant structures were subjected to addition of Hydrogen atoms and subsequently solvated in an octahedral box of TIP4P-Ew water model with ~10 Å distance between the protein surface and box boundary (Horn et al. 2004). Following solvation, each system was electrostatically neutralized by adding Cl- counter ions. Simulations were carried out using the SANDER module of AMBER 14 package with ff14SB forcefield (Case et al. 2014; Maier et al. 2015). A standard step comprising 2500 steps of steepest decent followed by 1000 steps of conjugate gradient energy minimization to eliminate van der Waals contacts, followed by a gradual heating from 0 to 300 K in 200 ps and constant temperature equilibration at 300 K for 1000 ps was employed for all MD simulations. Subsequently, 150 ns production run each for wild-type ANG, wild-type RNASE4, T11S-ANG, R122H-ANG, and D2E-RNASE4, N26K-RNASE4, T79A-RNASE4 and G119S-RNASE4 variants were carried out for a total of 1.2 μs with periodic boundary conditions at 300 K with Berendsen temperature coupling and a constant pressure of 1 atm with isotropic molecule-based scaling (Berendsen et al. 1984). All covalent bonds containing hydrogen atoms were fixed using SHAKE algorithm and long range electrostatic forces were calculated using particle-mesh Ewald (PME) (Ryckaert et al. 1977; Essmann et al. 1995). The protonation states of the pH-sensitive residues were kept default as described before and the stabilization of energy and the root mean square deviations (RMSD) of the wild-type ANG, wild-type RNASE4 with rare variants were analysed by sampling the coordinates of the trajectory at every 1 ps. Structural and dynamic analyses of the MD simulated trajectories was carried out using CPPTRAJ module of AMBER 14 (Roe and Cheatham 2013).

Molecular docking simulations

Molecular docking simulations were performed using AutoDock 4.2 software suite (Morris et al. 1998; Huey et al. 2007). Snapshots from T11S-ANG, N26K-RNASE4 and T79A-RNASE4 rare variants with native and altered His114 and His116 conformations respectively were extracted from the MD trajectories and subsequently used for docking. The R122H-ANG, D2E-RNASE4 and G119S-RNASE4 variants which didn't exhibit any conformational change of catalytic His114 or His116 during MD simulations were treated as WT-ANG and WT-RNASE4 and snapshots with only native His114 and His116 were extracted and used for docking. Molecular docking of the extracted snapshots were carried out with ligand molecule NCI-65828, retrieved from NCBI-PubChem Compound database [PubChem: 5351857], which is a known inhibitor of ribonucleolytic activity. A standard docking protocol as described before was followed where Hydrogen atoms were added to the extracted proteins for correct ionization and then tautomeric states of the amino acids and non-polar hydrogens were merged. Charge assignment and solvation parameters to the proteins and ligand were prepared by Kollman united atom charges and Gasteiger charges respectively and converted to PDBQT format afterwards. For the docking simulations, Lamarckian Genetic Algorithm with a population size of 150 dockings was used and 2.5 million energy evaluations were carried out and remaining parameters including crossover rate and mutation rate were kept default. The grid size of 80 × 80 × 80 centered on the protein with a default grid point spacing of 0.375 Å was used for search space and next, pre-calculated grid maps that store the grids of interaction energy were obtained using AutoGrid. The results were then clustered into bins of similar conformations based on root mean square deviation and orientation, and docking logs were analysed using the AutoDockTools (ADT) 1.5.4 (Sanner 1999).

Solvent accessible surface area

The nuclear translocation activity of wild-type ANG, wild-type RNASE4, T11S-ANG, R122H-ANG, and D2E-RNASE4, N26K-RNASE4, T79A-RNASE4 and G119S-RNASE4 variants was evaluated by computing the solvent accessible surface area of ^{31}RRR^{33} residues for ANG and ^{30}QRR^{32} residues for RNASE4. The VolArea: a Visual Molecular Dynamics (VMD) Plug-in was used to compute the accessible surface areas from the MD trajectories of wild-type and rare variants (Ribeiro et al. 2013; Humphrey et al. 1996). A default probe radius of 1.4 Å was used for the calculation of accessible surface areas.

Graphics and rendering

All relevant figures were rendered and generated using PyMOL (http://www.pymol.org) and VMD. The HBonds plugin of VMD was used for the calculation of Hydrogen bonds (H-bond) and percentage H-bond occupancy. Cytoscape was used to visualize the H-bond interactions and networks (Shannon et al. 2003).

Results

In this study, we analysed the rare variants of ANG and RNASE4 identified from ALS patients and catalogued in Project MinE consortium, whose goal is to systematically sequence and map full DNA profiles of at least 15,000 ALS patients and 7,500 control subjects to perform comparative analyses. As most of the variants in ANG and RNASE4 cause ALS through loss of ribonucleolytic or nuclear translocation or both these activities, our goal here was to examine if any of the Project MinE derived ANG and RNASE4 rare variants would result in loss-of-functions and held responsible being the causal mechanism for ALS. To achieve this, we first used a series of in silico tools to predict the pathogenicity of the rare variants and subsequently carried out 1.2 μs MD simulations to examine the structural and dynamic changes of the catalytic triad and nuclear localization signal residues accountable for ribonucleolytic and nuclear translocation activity respectively.

Assessment of pathogenicity of rare variants by in silico tools

Initial assessment of the pathogenicity of the rare variants was carried out using a number of popular web-tools and servers. Out of the six rare variants in total, while SIFT predicted T79A-RNASE4, N26K-RNASE4 and R122H-ANG variants as deleterious, PolyPhen 2.0 that uses sequence homology as well as structural information, predicted only the N26K-RNASE4 and T79A-RNASE4 variants as "possibly damaging" and "probably damaging" respectively (Table 2). On the other hand LRT_pred and PROVEAN both predicted the N26K-RNASE4 and T79A-RNASE4 as the only rare variants to be pathogenic. Interestingly, Condel which uses results from both MutationAssessor and FATHMM predicted N26K-RNASE4 and T79A-RNASE4 variants as deleterious in agreement with most other methods. Overall, our collective analyses of the rare variants by several in silico methods suggested that N26K-RNASE4, T79A-RNASE4 and R122H-ANG variants were possibly damaging by unfavourably affecting the structure and function of ANG and RNASE4.

Impact of rare variants on physicochemical properties of wild-type ANG and RNASE4

Next, the impact of missense variants on the structure and physicochemical properties of ANG and RNASE4 has been investigated using HOPE server. The following properties were analysed after substitutions of the source residues with the mutated ones to obtain a meaningful interpretation.

i. *Contact made by the mutated residues*: Analysis of the residue-residue contact revealed that the D2E-RNASE4, G119S-RNASE4 and R122H-ANG variants did not make contact with the catalytic residues and residues important for disulphide bond formation. However, the N26K-RNASE4 variant was found to be located very close to Cys25 residue that forms disulphide bond with Cys81, and hence the N26K variant could affect bond formation. Further Asp26 forms H-bond with Tyr94 and thus a substitution with Lys26 in N26K may obstruct the H-bond with Tyr94. The other important mutant is T79A-RNASE4, where Thr79 forms H-bonds with Gly20, His47 and His48, and its substitution with hydrophobic Ala79 is expected to affect H-bond with critical residues in its vicinity. In the case of T11S-ANG, Ser11 is unlikely to favour H-bond with Thr31 as Thr11 does. Further, being a proximal residue to catalytic His13, substitution of Thr11 in T11S may not facilitate the preferred catalytic conformation of ANG, and thus affecting its ribonucleolytic activity. Analysis of residue-residue contact analyses suggest that N26K-RNASE4, T79A-RNASE4 and T11S-ANG variants will disrupt the critical H-bonds, disulphide bonds and catalytic residue conformations in their respective mutant structures and could negatively affect their ribonucleolytic activities.
ii. *Conservation*: Analysis of the conservation profile revealed that in D2E and N26K and G119S-RNASE4 variants, the mutant residues Asp2, Lys26 and Ser119 are observed more often at these positions in other homologous sequences, suggesting that they are unlikely to alter the structure and function of the proteins. However, in the case of T79A-RNASE4, while Thr79 is highly conserved, neither the mutated residue Ala79 nor another hydrophobic residue has been observed at this position in other homologous sequences, suggesting this mutation will probably have an adverse effect on its structure and function. In T11S-ANG, the wild-type residue is not conserved at this position and another residue type not similar to Ser11 was observed more often at this position in other homologous sequences, suggesting this mutation to have a detrimental role in its structure and function. However, for R122H-ANG, it was found that along with Arginine, certain other residues were found in this position different from Histidine, suggesting this mutation may be tolerated.

Overall, analysis of physicochemical properties for the rare ANG and RNASE4 variants suggested that T11S-ANG, N26K-RNASE4 and T79A-RNASE4 variants could affect the physicochemical properties of the mutant proteins ultimately affecting their structure and function.

MD simulation based structural, dynamic and stability analysis

The structural and dynamic attributes from MD simulations of all the rare variants were investigated to examine if they

exhibit loss of ribonucleolytic and nuclear translocation activities. Different computational control experiments were performed to validate the time scale of the MD simulations and the conformational switching of catalytic residue His114 and His116 in ANG and RNASE4. As described in our recent reports (Padhi et al. 2012, 2019), the 150 ns time period simulation for each variant was validated by rotating the catalytic residue His114 in wild-type ANG and His116 in wild-type RNASE4 and the altered structures when subjected to MD simulations, it was found that His114 and His116 acquire their native conformations and continued to be in the native state thereafter indicating their most favourable energetic conformation (Supplementary Information Fig. 2). Next, the conformational switching of His114 and His116 observed in T11S-ANG, N26K-RNASE4 and T79A-RNASE4 variants was validated, where snapshots of T11S, N26K and T79A variants with altered His114 and His116 conformations respectively were extracted from MD trajectories, the variant sequence was then mutated back to wild-type and a fresh set of MD simulations were carried out. Interestingly, we found that catalytic His114 and His116 returned to their native conformations and maintained in those states afterwards. Our previous reports (Padhi et al. 2012, 2019), along with the present control experiments demonstrating the reliability and time frames of the MD simulations strongly indicate that our simulation models are accurate enough to carry out subsequent prediction and analyses.

Next, the effect of missense variations on stability was examined during the course of MD simulations. Backbone RMSDs of the ANG and RNASE4 variants were determined using CPPTRAJ and it was found that T79A-RNASE4 and N26K-RNASE4 variants experienced higher deviations as compared to other variants. The R122H-ANG variant exhibited a dip in RMSD at about 95 ns but stabilized afterwards. Overall, after 110 ns of MDs, RMSDs of all the variants converged and exhibited nearly similar profile to wild-types (Fig. 2).

Ribonucleolytic activity of ALS associated ANG and RNASE4 rare variants

One of the critical functions of ANG and RNASE4 is the ribonucleolytic activity that is governed by the catalytic triad residues His13, Lys40 and His114 in ANG and by residues His12, Lys40 and His116 in RNASE4 (Supplementary Information Fig. 1). As the conformational fluctuations of these catalytic residues are important for their ribonucleolytic activity, we first visualized the structural and dynamic changes of the catalytic residues for wild-type and all the rare variants from each of the 150 ns MD simulated trajectories. Upon examination, it was found that only the T11S-ANG and N26K-RNASE4 and T79A-RNASE4 variants undergo a characteristic conformational switching of His114 in ANG and His116 in RNASE4 (Fig. 3). The other variants such as D2E-RNASE4, G119S-RNASE4 and R122H-ANG did not display any conformational alterations and behaved similar to wild-types. Conformational alteration of the variants at the start of the simulation and one during the course of MD simulations are shown in Fig. 3. This conformational alteration of His114 in ANG and His116 in RNASE4 was quantified by computing the HA-CA-CB-CG dihedral angle during the MD simulations. It was observed that while the T11S-ANG, N26K-RNASE4 and T79A-RNASE4 variants exhibited a change in dihedral angle from the mean − 80° position to −179° during the simulations, the remaining variants D2E-RNASE4, G119S-RNASE4 and R122H-ANG behaved similar to wild-types, possibly retaining their ribonucleolytic activities (Fig. 4). Interestingly, a nearly similar shift in dihedral angle of His114 and His116 was observed for T11S-ANG and N26K-RNASE4, T79A-RNASE4 variants indicating that they may exhibit a partial loss of ribonucleolytic activity.

Molecular docking simulations to ascertain loss of ribonucleolytic activity

Molecular docking experiments were carried out to ascertain if the conformational switching of the catalytic residue His114 in T11S-ANG and His116 in N26K-RNASE4 and T79A-RNASE4 are indeed responsible for loss of ribonucleolytic activity. Our previous findings using computational and experimental characterization have effectively demonstrated that the conformational switching of catalytic residue His114 or His116 is responsible for the loss of ribonucleolytic activity in ANG or RNASE4 variants respectively. In this study, we were intrigued if the similar conformational changes of His114 or His116 observed in T11S-ANG and N26K-RNASE4, T79A-RNASE4 would play a role in loss of ribonucleolytic activity. Our docking results from the extracted snapshots of the variants with native and altered His114 and His116 conformations docked against NCI-65828, a potent inhibitor of ribonucleolytic activity, revealed that in T11S-ANG and N26K-RNASE4 and T79A-RNASE4, the HD1 atom of His114 and His116 failed to form salt bridge and hydrogen bond interactions with azo-group of NCI-65828 as a result of conformational alteration of catalytic Histidine (Fig. 5). This failed interaction resulted in significantly lower binding energies compared to wild-type ANG and RNASE4, and D2E-RNASE4, G119S-RNASE4 and R122H-ANG variants (Table 3). Our docking results suggested that the conformational switching of His114 and His116 is a molecular level determinant for loss of ribonucleolytic activity in ANG and RNASE4.

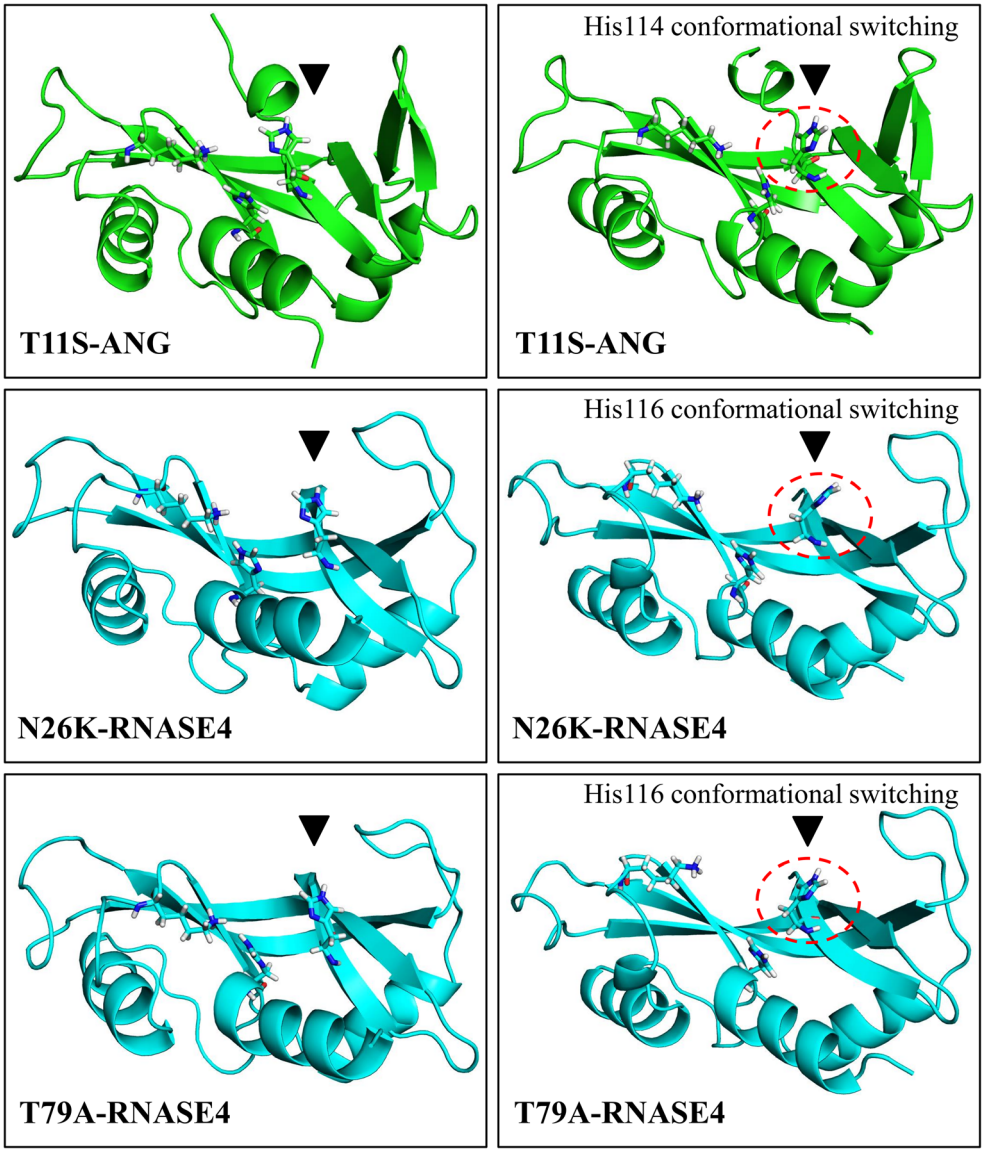

Fig. 3 Conformational switching of catalytic His114 and His116 in ANG and RNASE4 rare variants. Two snapshots of T11S-ANG, N26K-RNASE4 and T79A-RNASE4 variants showing the native, stable His114 and His116 conformations at the start of production phase (left) and exhibiting a characteristic conformational switching of catalytic residue His114 and His116 during the simulations. Conformational switching of catalytic His114 and His116 is shown by black arrow and highlighted in red dotted circle

Nuclear translocation activity of ALS associated ANG and RNASE4 rare variants

Ribonuclease family of proteins, specifically ANG and RNASE4 undergo nuclear translocation using their nuclear localization signal comprising residues ^{29}IMRRRGL^{35} (in ANG) and ^{28}MMQRRKM^{34} (in RNASE4) respectively. Among the nuclear localization signal residues, it has been established that the ^{31}RRR^{33} residues are extremely important for the nuclear translocation in ANG. Through our earlier studies, we have succinctly shown that the conformational dynamics of ^{31}RRR^{33} residues in ANG and ^{30}QRR^{32} residues in RNASE4 determine the fate of nuclear translocation activity. Therefore, we analysed the structural and conformational dynamics of ^{31}RRR^{33} and ^{30}QRR^{32} residues from our MD trajectories of the variants. We observed that none of the variants displayed a local folding and close packing of ^{31}RRR^{33} and ^{30}QRR^{32} residues unlike some of the well-established ANG and RNASE4 variants exhibiting loss of nuclear translocation activity (Fig. 6). We quantified the accessible surface area for all the variants and found that these exhibited higher accessible surface area profiles similar to wild-type proteins. The open conformation and loose packing of ^{31}RRR^{33} and ^{30}QRR^{32} residues, suggest that none of the rare variants would lose their nuclear translocation activity (Fig. 7).

Discussion

Disease causing power of rare genetic variations cannot be downplayed because of their increased burden on susceptibility of complex diseases such as ALS. One of the largest

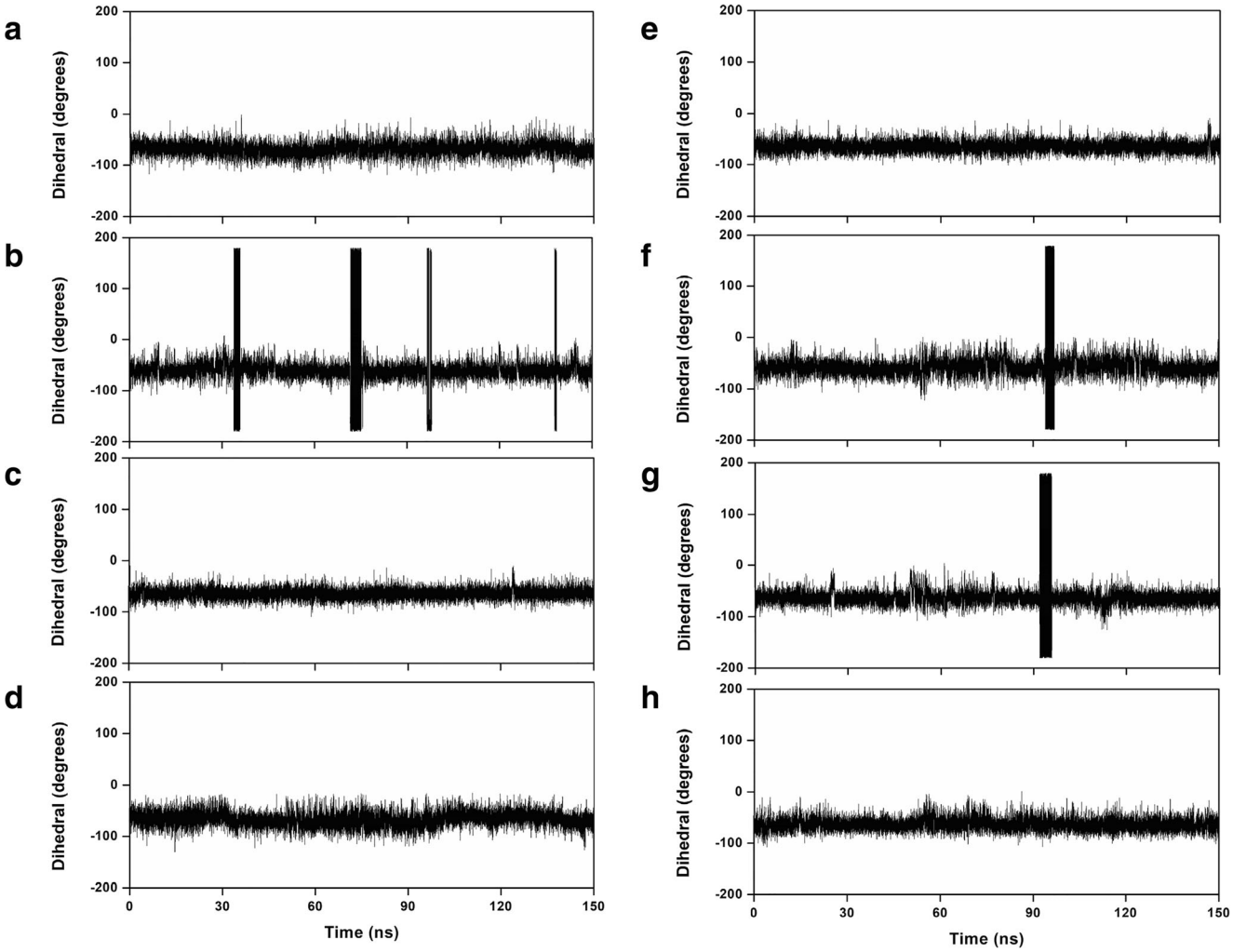

Fig. 4 Computed HA-CA-CB-CG dihedral angle of His114 and His116 in ANG and RNASE4 rare variants. Computed HA-CA-CB-CG dihedral angles of catalytic residue His114 in ANG variants and His116 in RNASE4 variants as a function of time are shown. Catalytic residue His114 in (B) T11S-ANG and His116 in (F) N26K-RNASE4 and (G) T79A-RNASE4 changed their conformation from −80° during the simulations, resulting in dihedral angle shift, while the (C) R122H-ANG, (E) D2E-RNASE4 and (H) G119S-RNASE4 variants retained their native, stable His114, His116 conformation and behaved similar to (A) wild-type ANG and (D) wild-type RNASE4 as seen from the simulated trajectories

GWAS study performed by Du et al. on 12,577 ALS cases and 23,475 controls revealed an important role of low-frequency variants in increasing the overall risk of ALS (Du et al. 2018). Till date the GWAS studies have primarily focused on common variations (MAF > 1%), thereby leaving the identified rare and low frequency variants untested. To establish the genetic basis of diseases like ALS, there is an exigent need to analyze the rare variants in cases with and without a family history. These rare variants could be tolerating variants but when combined with other variants they would be exerting synergistic deleterious effect (Cady et al. 2015; Pang et al. 2017).

Over the last decades, substantial efforts have been made on the identification of novel genetic variants and gain mechanistic insights into their role in motor neuron degeneration. Thanks to these discoveries, several reports have demonstrated that RNA processing pathway plays a vital role in ALS pathophysiology (van Blitterswijk and Landers 2010). In the loss-of-function category, ANG appears as one of the frequently mutated genes associated with ALS (Greenway et al. 2004, 2006; Wu et al. 2007; Thiyagarajan et al. 2012). Missense mutations in ANG cause loss of either ribonucleolytic activity or nuclear translocation activity or both of these functions leading to the development of ALS (Greenway et al. 2004, 2006; Padhi et al. 2014a, b, 2018, 2012, 2013a, b, 2019; Wu et al. 2007; Thiyagarajan et al. 2012). Our research and those carried out by Wu et al. have shown that loss-of-function is a characteristic feature of ANG and RNASE4 variants causal for ALS (Padhi et al. 2014a, b, 2018, 2012, 2013a, b, 2019; Wu et al. 2007; Thiyagarajan

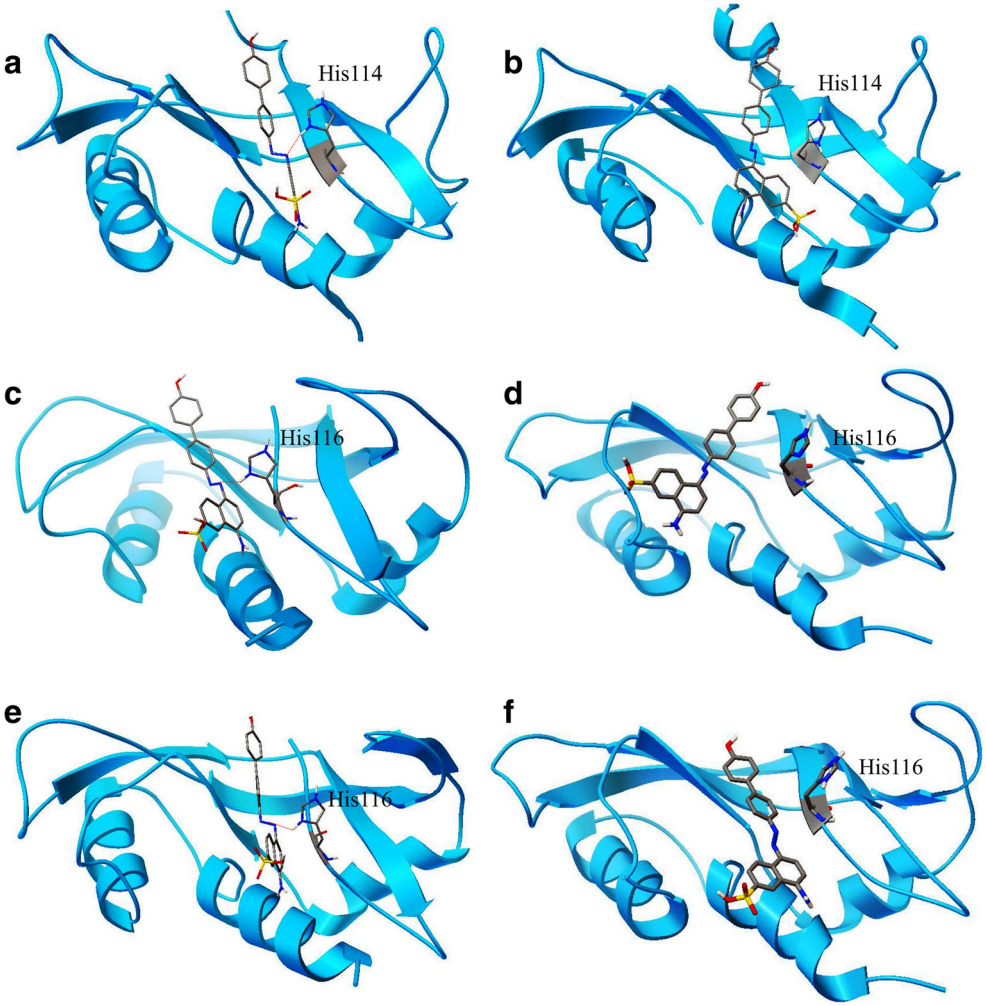

Fig. 5 Lowest-energy molecular docking poses of ANG and RNASE4 rare variants with NCI-65828. Stereo views lowest-energy AutoDock poses of T11S-ANG, N26K-RNASE4 and T79A-RNASE4 variants with NCI-65828. The backbone of ANG and RNASE4 is shown in blue colour along with their catalytic His114 and His116 residues shown as stick model. The hydrogen bond between the azo-group of the inhibitor and His114/His116 is shown in red dashed line. In T11S-ANG, (**a**) the hydrogen bond can be observed in the native His114 conformation while in (**b**) the altered conformation of His114 resulted in termination of hydrogen bond interaction. In N26K-RNASE4 and T79A-RNASE4, (**c** and **e**) the hydrogen bond formed can be observed in the native conformation of His116 while in (**d** and **f**) the hydrogen bond interaction was found to be absent in the altered conformation of His116

et al. 2012; Li et al. 2013). Although the loss-of-function mechanisms of ALS causing variants of ANG and RNASE4 have been well studied (Padhi et al. 2014a, b, 2018), efforts to understand the impact of rare variants of ANG and RNASE4 are not considered markedly. This encouraged us to investigate the loss-of-function mechanisms of yet to be revealed rare variants of ANG and RNASE4 catalogued in Project MinE in order to have a clear picture of functional loss mechanisms and ALS association. Our comprehensive study employing physicochemical features and MD simulations for the six uncharacterized ANG and RNASE4 rare variants showed that three of the variants (T11S-ANG, N26K-RNASE4 and T79A-RNASE4) would result in loss of ribonucleolytic activity due to conformational fluctuation of catalytic Histidine and being associated with ALS. Specifically, the N26K-RNASE4 variant is present only in patients and is absent in controls, thereby indicating a pathogenic role of this variant in ALS. Furthermore, the other two variants T11S-ANG and T79A-RNASE4 are present at extremely low frequency in controls and are predicted to be pathogenic by our analyses, suggesting their role as risk factors in ALS manifestation.

Table 3 Computed binding energies between wild-type ANG, wild-type RNASE4 and rare ANG and RNASE4 variants docked against NCI-65828

ANG/RNASE4 proteins	AutoDock binding energy scores	
	Native His114/ His116 conformation	Altered His114/ His116 conformation
Wild-type ANG	−8.78 kcal/mol	
T11S-ANG*	−7.73 kcal/mol	−4.09 kcal/mol
R122H-ANG	−7.78 kcal/mol	
Wild-type RNASE4	−8.97 kcal/mol	
D2E-RNASE4	−8.68 kcal/mol	
N26K-RNASE4*	−8.45 kcal/mol	−5.33 kcal/mol
T79A-RNASE4*	−8.63 kcal/mol	−5.16 kcal/mol
G119S-RNASE4	−8.66 kcal/mol	

*ANG and RNASE4 variants that exhibit conformational switching of catalytic His114 and His116 respectively

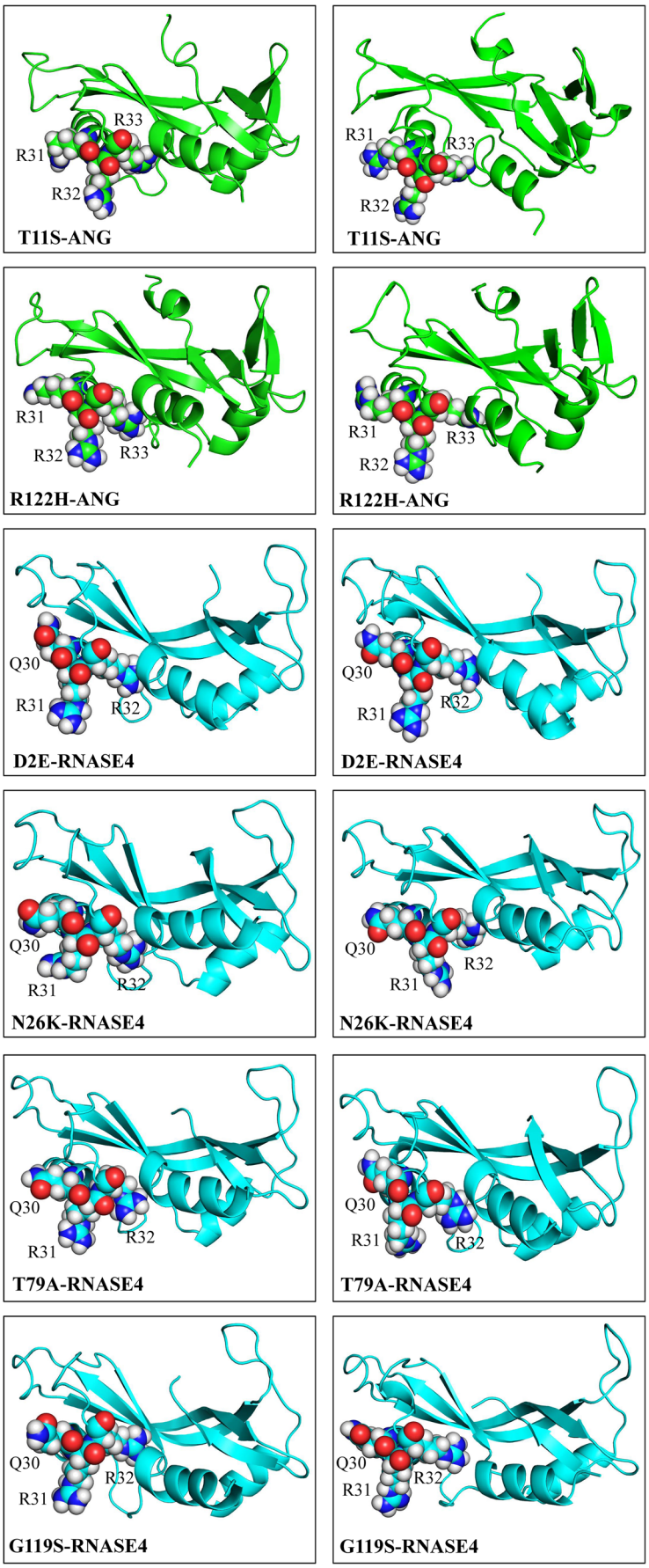

Fig. 6 Conformation of ^{31}RRR^{33} and ^{30}QRR^{32} residues from MD simulations of ANG and RNASE4 rare variants. Observed loose packing and open conformation of ^{31}RRR^{33} and ^{30}QRR^{32} residues during MD simulations of ANG and RNASE4 variants resulting in higher solvent accessible surface areas

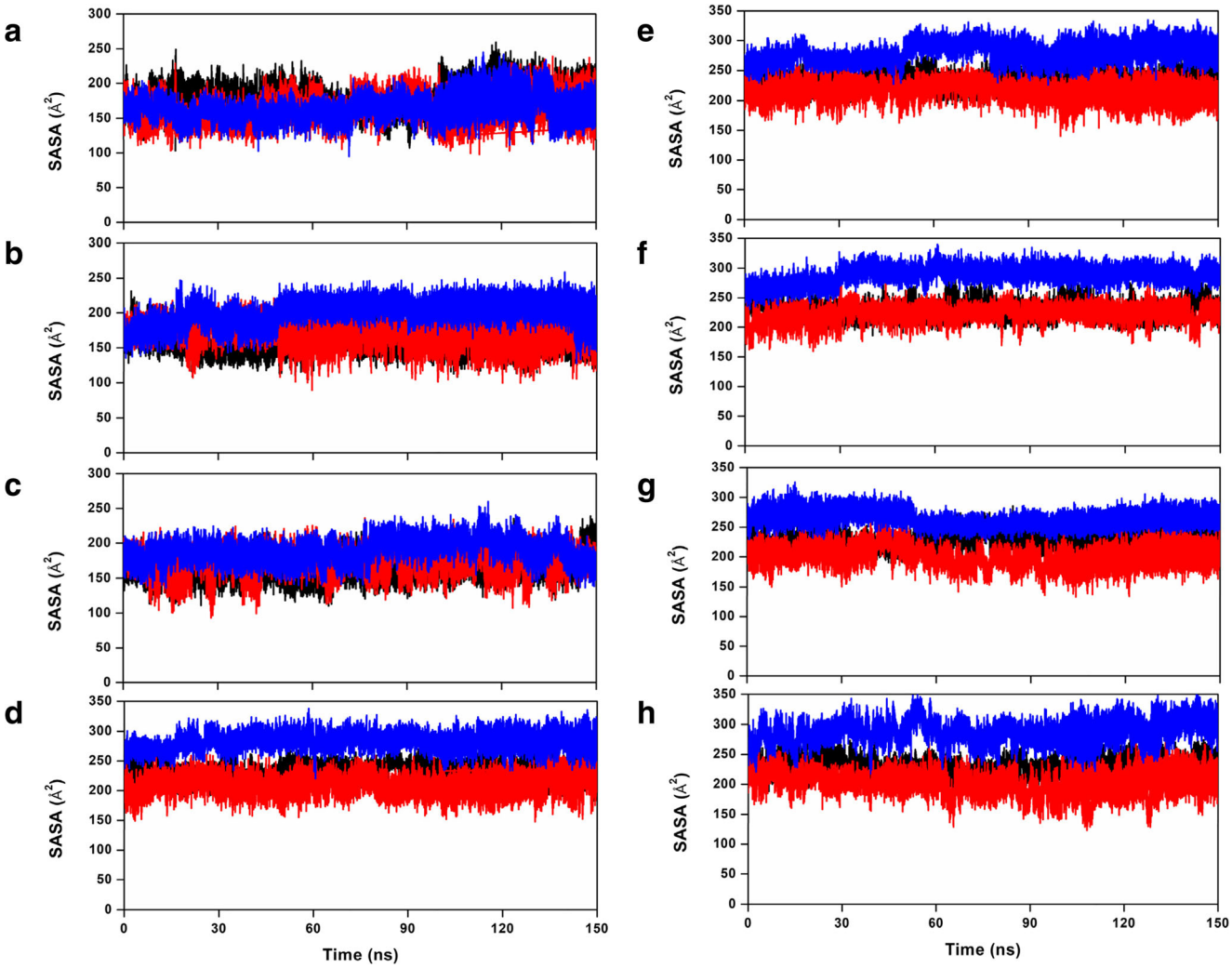

Fig. 7 Solvent accessible surface areas for ANG and RNASE4 rare variants. Computed solvent accessible surface areas of ^{31}RRR^{33} and ^{30}QRR^{32} residues from explicit-solvent MD simulations of ANG and RNASE4 rare variants. All the ANG variants (B) T11S-ANG and (C) R122H-ANG possessed solvent accessible surface areas comparable to (A) wild-type ANG and the RNASE4 rare variants (E) D2E-RNASE4, (F) N26K-RNASE4, (G) T79A-RNASE4 and (H) G119S-RNASE4 possessed accessible surface areas comparable to (D) wild-type RNASE4, suggesting they would retain their nuclear translocation activities. Solvent accessible surface areas of ^{31}RRR^{33} and ^{30}QRR^{32} residues are represented in black, red and blue colour respectively

Ribonucleolytic activity of rare ANG and RNASE4 variants

As the catalytic triad residues His13, Lys40 and His114 in ANG and His12, Lys40 and His116 are responsible for ribonucleolytic activity in ANG and RNASE4 respectively, we initially visualized their conformational dynamics from MD simulated trajectories and observed that among the catalytic triad residues in ANG and RNASE4, only His114 and His116 undergoes a characteristic conformational switching in ANG and RNASE4 during the simulations. This behaviour was found in the T11S variant of ANG and N26K, T79A variants of RNASE4 (Fig. 3). The remaining variants such as D2E-RNASE4, G119S-RNASE4 and R122H-ANG behaved like the wild-type proteins. Our atomic level quantification of the HA-CA-CB-CG dihedral angle of His114 and His116 clearly reflected the conformational switching where dihedral angle of His114 and His116 shifted from the −80° to −179° for the T11S-ANG, N26K-RNASE4 and T79A-RNASE4 variants (Fig. 4). This gave us an initial impression that probably these variants would retain only partial ribonucleolytic activity as certain ANG variants have been previously shown similar dihedral angle profile resulting in partial loss of activity (Padhi et al. 2014a, b, 2018; Wu et al. 2007). A stable and wild-type like behaviour of D2E-RNASE4, G119S-RNASE4 and R122H-ANG variants however suggested that these substitutions are probably not strong enough to induce a conformational change of catalytic Histidine and in loss of ribonucleolytic activity.

Table 4 A comprehensive list of human ANG and RNASE4 variants reported in ALS patients. Variants for which the functional assay results are available are presented with their corresponding MD simulation derived attributes (i) conformational switching of catalytic His114/His116 responsible for loss of ribonucleolytic activity and (ii) folding of ^{31}RRR^{33}/^{30}QRR^{32} residues responsible for loss of nuclear translocation activity

Human ANG/ RNASE4 variant	His116/His114 conformational switching (MD simulations)	Ribonucleolytic activity (MD simulations)	Ribonucleolytic activity (Experimental reports)	Folding of ^{30}QRR^{32}/^{31}RRR^{33} (MD simulations)	Nuclear translocation activity (MD simulations)	Nuclear translocation activity (Experimental reports)
E48D-RNASE4	Yes (Padhi et al. 2019)	Loss (Padhi et al. 2019)	Loss (Padhi et al. 2019)	No (Padhi et al. 2019)	No loss (Padhi et al. 2019)	No loss (Padhi et al. 2019)
V75I-RNASE4	No (Padhi et al. 2019)	No loss (Padhi et al. 2019)	No loss (Padhi et al. 2019)	No (Padhi et al. 2019)	No loss (Padhi et al. 2019)	No loss (Padhi et al. 2019)
D22G-ANG	Yes (Padhi et al. 2014a, b)	Loss (Padhi et al. 2014a, b)	Loss (Padhi et al. 2014a, b)	No (Padhi et al. 2014a, b)	No loss (Padhi et al. 2014a, b)	No loss (Padhi et al. 2014a, b)
L35P-ANG	Yes (Padhi et al. 2014a, b)	Loss (Padhi et al. 2014a, b)	Loss (Padhi et al. 2014a, b)	Yes (Padhi et al. 2014a, b)	Loss (Padhi et al. 2014a, b)	Loss (Padhi et al. 2014a, b)
K17I-ANG	Yes (Padhi et al. 2012)	Loss (Padhi et al. 2012)	Loss (Wu et al. 2007; Thiyagarajan et al. 2012; Crabtree et al. 2007)	No (Padhi et al. 2012)	No loss (Padhi et al. 2012)	No loss (Wu et al. 2007; Thiyagarajan et al. 2012)
S28 N-ANG	Yes (Padhi et al. 2012)	Loss (Padhi et al. 2012)	Loss (Wu et al. 2007; Thiyagarajan et al. 2012)	Yes (Padhi et al. 2012)	Loss (Padhi et al. 2012)	Loss (Wu et al. 2007; Thiyagarajan et al. 2012)
P112L-ANG	Yes (Padhi et al. 2012)	Loss (Padhi et al. 2012)	Loss (Wu et al. 2007; Thiyagarajan et al. 2012)	Yes (Padhi et al. 2012)	Loss (Padhi et al. 2012)	Loss (Wu et al. 2007; Thiyagarajan et al. 2012)
I46V-ANG	Yes (Padhi et al. 2013a, b)	Loss (Padhi et al. 2013a, b)	Loss (Thiyagarajan et al. 2012; Crabtree et al. 2007)	No (Padhi et al. 2013a, b)	No loss (Padhi et al. 2013a, b)	No loss (Thiyagarajan et al. 2012)
K17E-ANG	Yes (Padhi et al. 2013a, b)	Loss (Padhi et al. 2013a, b)	Loss (Thiyagarajan et al. 2012; Crabtree et al. 2007)	No (Padhi et al. 2013a, b)	No loss (Padhi et al. 2013a, b)	No loss (Thiyagarajan et al. 2012)
R31K-ANG	Yes (Padhi et al. 2013a, b)	Loss (Padhi et al. 2013a, b)	Loss (Thiyagarajan et al. 2012; Crabtree et al. 2007)	No (Padhi et al. 2013a, b)	No loss (Padhi et al. 2013a, b)	No loss (Thiyagarajan et al. 2012)
V113I-ANG	No (Padhi et al. 2012)	No (Padhi et al. 2012)	Minor loss (Thiyagarajan et al. 2012)	Loss (Padhi et al. 2012)	Loss (Padhi et al. 2012)	Transient activity (Thiyagarajan et al. 2012)
R121H-ANG	No (Padhi et al. 2013a, b)	No loss (Padhi et al. 2013a, b)	No loss (Thiyagarajan et al. 2012)	Loss (Padhi et al. 2013a, b)	Loss (Padhi et al. 2013a, b)	Transient activity (Thiyagarajan et al. 2012)
K54E-ANG	No (Padhi et al. 2013a, b)	No loss (Padhi et al. 2013a, b)	No loss (Thiyagarajan et al. 2012)	Loss (Padhi et al. 2013a, b)	Loss (Padhi et al. 2013a, b)	Not reported
R33W-ANG	Yes (Tripolszki et al. 2019)	Loss (Tripolszki et al. 2019)	Loss (Tripolszki et al. 2019)	Yes (Tripolszki et al. 2019)	Loss (Tripolszki et al. 2019)	Loss (Tripolszki et al. 2019)

Our molecular docking results further supported these results where we observed that altered His114 or His116 conformation resulted in failure of H-bond or salt bridge interactions with the azo-group of NCI-65828 – a known inhibitor of ribonucleolytic activity (Fig. 5). Further, the T11S-ANG, N26K-RNASE4 and T79A-RNASE4 variants exhibited considerably lower binding energy compared to wild-type and other variants retaining the native His114 or His116 conformations (Table 3). A close inspection of the T11S-ANG, N26K-RNASE4 and T79A-RNASE4 variants further revealed that they possess conserved H-bond interaction paths Ser11-Arg33-Ile29-Tyr25-Asp15-Ile46-His13-(Thr44|Leu115)-Gln117-Asp116-His114, Lys26-Asp22-Arg95-Arg82-Asp80-Thr44-Asp118-His116 and Ala79-Phe45-Asn43-Thr44-Asp118-His116 that induce the His114 and His116 conformational switching in ANG and RNASE4 variants respectively (Supplementary Information Fig. 3).

Nuclear translocation activity of rare ANG and RNASE4 variants

Previous studies including recent reports from us have shown that ANG and RNASE4 use their nuclear localization signal ^{29}IMRRRGL^{35} and ^{28}MMQRRKM^{34} for translocation into nucleus in HUVEC and P19 cells (Wu et al. 2007; Thiyagarajan et al. 2012; Li et al. 2013). Biochemical and site directed mutagenesis experiments have effectively shown that among ^{29}IMRRRGL^{35}, ANG typically uses the three consecutive Arginine residues ^{31}RRR^{33} for nuclear translocation, where R33 is essential for the activity and ^{31}RR^{32} modulate it (Moroianu and Riordan 1994; Moroianu and Riordan 2006). Similar to ANG's ^{31}RRR^{33}, RNASE4 also possesses a signal sequence ^{30}QRR^{32}, present on the protein surface and largely accessible to solvent. Since our previous studies have effectively shown that ALS associated ANG and RNASE4 variants exhibit characteristic conformational changes in ^{31}RRR^{33} and ^{30}QRR^{32} residues respectively, we examined similar attributes for all the rare ANG and RNASE4 variants. We found that the six rare variant behaved similar to wild-type proteins and retained an open, loosely packed conformation of ^{31}RRR^{33} and ^{30}QRR^{32} residues during the MD simulations, resulting higher accessible surface area profiles (Figs. 6 and 7). It has been documented in previous reports that certain ANG variants although identified from ALS patients, do not lose their ribonucleolytic and nuclear translocation activities, thereby suggesting although they retain their neuroprotective activities, they may still be able to cause ALS due to the presence of different genetic mutations and other aspects of gene expression such as mRNA processing (Renton et al. 2014).

Further, we were curious to inspect how our MD simulation based results correlated with the in silico predictions obtained from the widely used methods for the pathogenicity of the rare ANG and RNASE4 variants. Our collective analyses from 8 different web-tools and algorithms based on sequence and structure based features suggested that the N26K-RNASE4 and T79A-RNASE4 variants could be possibly deleterious and may affect the structure and function (Table 2). Analyses based on physicochemical properties also pointed out the similar variants to be damaging to the ANG and RNASE4 proteins. Our simulation results showed that the not only N26K-RNASE4 and T79A-RNASE4, but the T11S-ANG variant would also be pathogenic as a result of loss of ribonucleolytic activity. Our MD simulation based method presented a comprehensive picture of loss-of-function mechanisms for ANG and RNASE4 rare variants and their plausible association with ALS.

Reliability of the MD simulation in predicting loss-of-function mechanisms of rare ANG and RNASE4 variants

In our previous studies along with a recent report based on MD simulations and examination of certain structural and dynamic attributes, we have effectively shown that ANG or RNASE4 variants lose their ribonucleolytic activity due to a characteristic conformational alteration of catalytic residue His114 or His116 (Padhi et al. 2014a, b, 2018). These results were later corroborated completely with functional assay experiments and demonstrated the strength and reliability of our MD simulation approach (Table 4). Here, our goal was to investigate if any of the Project MinE documented rare ANG and RNASE4 variants would lose their ribonucleolytic or nuclear translocation activity. Our comprehensive analyses of all the rare ANG and RNASE4 variants showed that T11S-ANG, N26K-RNASE4 and T79A-RNASE4 variants would lose their ribonucleolytic activity due to conformational switching of catalytic His114 and His116 residue. Interestingly, the MD simulation results are in good agreement with the predictions from several widely used web-tools and methods. In a nutshell, the close match between previously reported MD simulations results corroborated with experiments indicates that our method comprising physicochemical, structural and dynamic characterization are reliable to be accountable for loss of ribonucleolytic activity in ANG and RNASE4 rare variants.

Conclusions

It is established that low-frequency rare variants play an important role in disease susceptibility such as ALS as a result of rapid population growth and weak purifying selection. Although the structure-function-disease relationship of the rare variants needs to be established through a multidisciplinary approach involving in vitro and in vivo studies, extensive

in silico analyses and characterization by combining biologically-relevant long time-scale simulations will be immensely useful to shortlist potential rare variants for follow-up experiments and clinical studies. The computational approach presented here along with the one employed and tested in our previous studies serves as a unified methodology for high-throughput screening of novel variants from ribonuclease superfamily, followed by their association with ALS susceptibility. Work is under way to expand the computational methodology to other important family of proteins associated with ALS, which will be crucial in bridging the gap between variant identification, characterization and ALS development through an understanding of molecular mechanisms of disease pathophysiology.

Acknowledgements Aditya K. Padhi acknowledges Council of Scientific and Industrial Research (CSIR), Government of India for research fellowship. Priyam Narain is thankful to IIT Delhi for Senior Research Fellowship. Helpful suggestions from Prof B. Jayaram (IIT Delhi) are gratefully acknowledged.

Author's contributions AKP; Collection of variants: PN; Analyzed the data: AKP, PN and JG; Contributed to the writing of the manuscript: AKP, PN and JG.

Compliance with ethical standards

Conflict of interest The authors declare no conflicts of interest.

Research involving in human and animal rights This article does not contain any study with human or animal subjects performed by any of the authors.

References

Abel O, Powell JF, Andersen PM, Al-Chalabi A (2012) ALSoD: a user-friendly online bioinformatics tool for amyotrophic lateral sclerosis genetics. Hum Mutat 33:1345–1351. https://doi.org/10.1002/humu.22157

Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR (2010) A method and server for predicting damaging missense mutations. Nat Methods 7:248–249. https://doi.org/10.1038/nmeth0410-248

Andersen PM, Al-Chalabi A (2011) Clinical genetics of amyotrophic lateral sclerosis: what do we really know? Nat Rev Neurol

Berendsen HJC, Postma JPM, Van Gunsteren WF et al (1984) Molecular dynamics with coupling to an external bath. J Chem Phys 81:3684–3690. https://doi.org/10.1063/1.448118

Brown RH, Al-Chalabi A (2017) Amyotrophic lateral sclerosis. N Engl J Med 377:162–172. https://doi.org/10.1056/NEJMra1603471

Cady J, Allred P, Bali T, Pestronk A, Goate A, Miller TM, Mitra RD, Ravits J, Harms MB, Baloh RH (2015) Amyotrophic lateral sclerosis onset is influenced by the burden of rare variants in known amyotrophic lateral sclerosis genes. Ann Neurol 77:100–113. https://doi.org/10.1002/ana.24306

Case DA, Babin V, Berryman JT et al (2014) Amber 14. University of California, San Francisco, CA

Choi Y, Chan AP (2015) PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. Bioinformatics. 31:2745–2747. https://doi.org/10.1093/bioinformatics/btv195

Chow CY, Landers JE, Bergren SK, Sapp PC, Grant AE, Jones JM, Everett L, Lenk GM, McKenna-Yasek DM, Weisman LS, Figlewicz D, Brown RH, Meisler MH (2009) Deleterious variants of FIG4, a phosphoinositide phosphatase, in patients with ALS. Am J Hum Genet 84:85–88. https://doi.org/10.1016/j.ajhg.2008.12.010

Chun S, Fay JC (2009) Identification of deleterious mutations within three human genomes. Genome Res 19:1553–1561. https://doi.org/10.1101/gr.092619.109

Crabtree B, Thiyagarajan N, Prior SH, Wilson P, Iyer S, Ferns T, Shapiro R, Brew K, Subramanian V, Acharya KR (2007) Characterization of human angiogenin variants implicated in amyotrophic lateral sclerosis. Biochemistry. 46:11810–11818. https://doi.org/10.1021/bi701333h

DeJesus-Hernandez M, Mackenzie IR, Boeve BF, Boxer AL, Baker M, Rutherford NJ, Nicholson AM, Finch NCA, Flynn H, Adamson J, Kouri N, Wojtas A, Sengdy P, Hsiung GYR, Karydas A, Seeley WW, Josephs KA, Coppola G, Geschwind DH, Wszolek ZK, Feldman H, Knopman DS, Petersen RC, Miller BL, Dickson DW, Boylan KB, Graff-Radford NR, Rademakers R (2011) Expanded GGGGCC Hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. Neuron. 72:245–256. https://doi.org/10.1016/j.neuron.2011.09.011

Du Y, Wen Y, Guo X et al (2018) A genome-wide expression association analysis identifies genes and pathways associated with amyotrophic lateral sclerosis. Cell Mol Neurobiol 38:635–639. https://doi.org/10.1007/s10571-017-0512-2

Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG (1995) A smooth particle mesh Ewald method. J Chem Phys 103:8577–8593. https://doi.org/10.1063/1.470117

Ferraiuolo L, Kirby J, Grierson AJ, et al (2011) Molecular pathways of motor neuron injury in amyotrophic lateral sclerosis. Nat Rev Neurol

González-Pérez A, López-Bigas N (2011) Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. Am J Hum Genet doi 88:440–449. https://doi.org/10.1016/j.ajhg.2011.03.004

Greenway MJ, Alexander MD, Ennis S et al (2004) A novel candidate region for ALS on chromosome 14q11.2. Neurology. https://doi.org/10.1212/01.WNL.0000144344.39103.F6

Greenway MJ, Andersen PM, Russ G et al (2006) ANG mutations segregate with familial and "sporadic" amyotrophic lateral sclerosis. Nat Genet 38:411–413. https://doi.org/10.1038/ng1742

Hardiman O, Al-Chalabi A, Chio A, Corr Giancarlo Logroscino EM et al (2019) Amyotrophic lateral sclerosis. Prog Med Chem. https://doi.org/10.1016/bs.pmch.2018.12.001

Horn HW, Swope WC, Pitera JW, Madura JD, Dick TJ, Hura GL, Head-Gordon T (2004) Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. J Chem Phys 120:9665–9678. https://doi.org/10.1063/1.1683075

Huey R, Morris GM, Olson AJ, Goodsell DS (2007) Software news and update a semiempirical free energy force field with charge-based desolvation. J Comput Chem 28:1145–1152. https://doi.org/10.1002/jcc.20634

Humphrey W, Dalke A, Schulten K (1996) VMD: Visual molecular dynamics. J Mol Graph 14:33–38. https://doi.org/10.1016/0263-7855(96)00018-5

Krüger S, Battke F, Sprecher A, Munz M, Synofzik M, Schöls L, Gasser T, Grehl T, Prudlo J, Biskup S (2016) Rare variants in neurodegeneration associated genes revealed by targeted panel sequencing in a German ALS cohort. Front Mol Neurosci 9. https://doi.org/10.3389/fnmol.2016.00092

Leonidas DD, Shapiro R, Allen SC, Subbarao GV, Veluraja K, Acharya KR (1999) Refined crystal structures of native human angiogenin and two active site variants: implications for the unique functional

properties of an enzyme involved in neovascularisation during tumour growth. J Mol Biol 285:1209–1233. https://doi.org/10.1006/jmbi.1998.2378

Li S, Sheng J, Hu JK, Yu W, Kishikawa H, Hu MG, Shima K, Wu D, Xu Z, Xin W, Sims KB, Landers JE, Brown RH, Hu GF (2013) Ribonuclease 4 protects neuron degeneration by promoting angiogenesis, neurogenesis, and neuronal survival under stress. Angiogenesis. 16:387–404. https://doi.org/10.1007/s10456-012-9322-9

Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C (2015) ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. J Chem Theory Comput 11:3696–3713. https://doi.org/10.1021/acs.jctc.5b00255

Marangi G, Traynor BJ (2015) Genetic causes of amyotrophic lateral sclerosis: new genetic analysis methodologies entailing new opportunities and challenges. Brain Res 1607:75–93

McLaren W, Pritchard B, Rios D, Chen Y, Flicek P, Cunningham F (2010) Deriving the consequences of genomic variants with the Ensembl API and SNP effect predictor. Bioinformatics. 26:2069–2070. https://doi.org/10.1093/bioinformatics/btq330

Moroianu J, Riordan JF (1994) Identification of the nucleolar targeting signal of human angiogenin. Biochem Biophys Res Commun 203:1765–1772. https://doi.org/10.1006/bbrc.1994.2391

Moroianu J, Riordan JF (2006) Nuclear translocation of angiogenin in proliferating endothelial cells is essential to its angiogenic activity. Proc Natl Acad Sci 91:1677–1681. https://doi.org/10.1073/pnas.91.5.1677

Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ (1998) Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. J Comput Chem 19:1639–1662. https://doi.org/10.1002/(SICI)1096-987X(19981115)19:14<1639::AID-JCC10>3.0.CO;2-B

Mueller SC, Backes C, Haas J, INHERITANCE Study Group, Katus HA, Meder B, Meese E, Keller A (2014) Pathogenicity prediction of non-synonymous single nucleotide variants in dilated cardiomyopathy. Brief Bioinform 16:769–779. https://doi.org/10.1093/bib/bbu054

Narain P, Gomes J, Bhatia R, Singh I, Vivekanandan P (2017) C9orf72 hexanucleotide repeat expansions and Ataxin 2 intermediate length repeat expansions in Indian patients with amyotrophic lateral sclerosis. Neurobiol Aging 56:211.e9–211.e14. https://doi.org/10.1016/j.neurobiolaging.2017.04.011

Narain P, Pandey A, Gupta S, Gomes J, Bhatia R, Vivekanandan P (2018) Targeted next-generation sequencing reveals novel and rare variants in Indian patients with amyotrophic lateral sclerosis. Neurobiol Aging 71:265.e9–265.e14. https://doi.org/10.1016/j.neurobiolaging.2018.05.012

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

Padhi AK, Gomes J (2019) A molecular dynamics based investigation reveals the role of rare ribonuclease 4 variants in amyotrophic lateral sclerosis susceptibility. Mutat Res - Fundam Mol Mech Mutagen 813:1–12. https://doi.org/10.1016/j.mrfmmm.2018.11.002

Padhi AK, Hazra S (2019) Insights into the role of d-amino acid oxidase mutations in amyotrophic lateral sclerosis. J Cell Biochem 120:2180–2197. https://doi.org/10.1002/jcb.27529

Padhi AK, Kumar H, Vasaikar SV, Jayaram B, Gomes J (2012) Mechanisms of loss of functions of human angiogenin variants implicated in amyotrophic lateral sclerosis. PLoS One 7:e32479. https://doi.org/10.1371/journal.pone.0032479

Padhi AK, Jayaram B, Gomes J (2013a) Prediction of functional loss of human angiogenin mutants associated with ALS by molecular dynamics simulations. Sci Rep 3. https://doi.org/10.1038/srep01225

Padhi AK, Vasaikar SV, Jayaram B, Gomes J (2013b) Fast prediction of deleterious angiogenin mutations causing amyotrophic lateral sclerosis. FEBS Lett 587:1762–1766. https://doi.org/10.1016/j.febslet.2013.04.022

Padhi AK, Banerjee K, Gomes J, Banerjee M (2014a) Computational and functional characterization of angiogenin mutations, and correlation with amyotrophic lateral sclerosis. PLoS One 9:e111963. https://doi.org/10.1371/journal.pone.0111963

Padhi AK, Vasaikar S V, Jayaram B, Gomes J (2014b) ANGDelMut – a web-based tool for predicting and analyzing functional loss mechanisms of amyotrophic lateral sclerosis-associated angiogenin mutations. F1000Research. https://doi.org/10.12688/f1000research.2-227.v3

Padhi AK, Narain P, Dave U, Satija R, Patir A, Gomes J (2019) Insights into the role of ribonuclease 4 polymorphisms in amyotrophic lateral sclerosis. J Biomol Struct Dyn 37:116–130. https://doi.org/10.1080/07391102.2017.1419147

Pang SYY, Hsu JS, Teo KC, Li Y, Kung MHW, Cheah KSE, Chan D, Cheung KMC, Li M, Sham PC, Ho SL (2017) Burden of rare variants in ALS genes influences survival in familial and sporadic ALS. Neurobiol Aging 58:238.e9–238.e15. https://doi.org/10.1016/j.neurobiolaging.2017.06.007

Renton AE, Majounie E, Waite A, Simón-Sánchez J, Rollinson S, Gibbs JR, Schymick JC, Laaksovirta H, van Swieten JC, Myllykangas L, Kalimo H, Paetau A, Abramzon Y, Remes AM, Kaganovich A, Scholz SW, Duckworth J, Ding J, Harmer DW, Hernandez DG, Johnson JO, Mok K, Ryten M, Trabzuni D, Guerreiro RJ, Orrell RW, Neal J, Murray A, Pearson J, Jansen IE, Sondervan D, Seelaar H, Blake D, Young K, Halliwell N, Callister JB, Toulson G, Richardson A, Gerhard A, Snowden J, Mann D, Neary D, Nalls MA, Peuralinna T, Jansson L, Isoviita VM, Kaivorinne AL, Hölttä-Vuori M, Ikonen E, Sulkava R, Benatar M, Wuu J, Chiò A, Restagno G, Borghero G, Sabatelli M, Heckerman D, Rogaeva E, Zinman L, Rothstein JD, Sendtner M, Drepper C, Eichler EE, Alkan C, Abdullaev Z, Pack SD, Dutra A, Pak E, Hardy J, Singleton A, Williams NM, Heutink P, Pickering-Brown S, Morris HR, Tienari PJ, Traynor BJ (2011) A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. Neuron. 72:257–268. https://doi.org/10.1016/j.neuron.2011.09.010

Renton AE, Chiò A, Traynor BJ (2014) State of play in amyotrophic lateral sclerosis genetics. Nat Neurosci 17:17–23

Reva B, Antipin Y, Sander C (2011) Predicting the functional impact of protein mutations: application to cancer genomics. Nucleic Acids Res 39:e118. https://doi.org/10.1093/nar/gkr407

Ribeiro JV, Tamames JAC, Cerqueira NMFSA, Fernandes PA, Ramos MJ (2013) Volarea - a bioinformatics tool to calculate the surface area and the volume of molecular systems. Chem Biol Drug Des 82:743–755. https://doi.org/10.1111/cbdd.12197

Roe DR, Cheatham TE (2013) PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. J Chem Theory Comput 9:3084–3095. https://doi.org/10.1021/ct400341p

Rosen DR, Siddique T, Patterson D, Figlewicz DA, Sapp P, Hentati A, Donaldson D, Goto J, O'Regan JP, Deng HX, Rahmani Z, Krizus A, McKenna-Yasek D, Cayabyab A, Gaston SM, Berger R, Tanzi RE, Halperin JJ, Herzfeldt B, van den Bergh R, Hung WY, Bird T, Deng G, Mulder DW, Smyth C, Laing NG, Soriano E, Pericak–Vance MA, Haines J, Rouleau GA, Gusella JS, Horvitz HR, Brown RH (1993) Mutations in cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis. Nature. 362:59–62. https://doi.org/10.1038/362059a0

Ryckaert JP, Ciccotti G, Berendsen HJC (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J Comput Phys 23:327–341. https://doi.org/10.1016/0021-9991(77)90098-5

Sanner MF (1999) Python: a programming language for software integration and development. J Mol Graph Model

Schwarz JM, Cooper DN, Schuelke M, Seelow D (2014) MutationTaster2: mutation prediction for the deep-sequencing age. Nat Methods 11:361–362. https://doi.org/10.1038/nmeth.2890

Shannon P, Markiel A, Ozier O et al (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13:2498–2504. https://doi.org/10.1101/gr.1239303

Sreedharan J, Blair IP, Tripathi VB et al (2008) TDP-43 mutations in familial and sporadic amyotrophic lateral sclerosis. Science (80-). https://doi.org/10.1126/science.1154584

Terzyan SS, Peracaula R, De Llorens R et al (1999) The three-dimensional structure of human RNase 4, unliganded and complexed with d(Up), reveals the basis for its uridine selectivity. J Mol Biol 285:205–214. https://doi.org/10.1006/jmbi.1998.2288

Thiyagarajan N, Ferguson R, Subramanian V, Acharya KR (2012) Structural and molecular insights into the mechanism of action of human angiogenin-ALS variants in neurons. Nat Commun 3:1121. https://doi.org/10.1038/ncomms2126

Tripolszki K, Danis J, Padhi AK, Gomes J, Bozó R, Nagy ZF, Nagy D, Klivényi P, Engelhardt JI, Széll M (2019) Angiogenin mutations in Hungarian patients with amyotrophic lateral sclerosis: clinical, genetic, computational, and functional analyses. Brain Behav. https://doi.org/10.1002/brb3.1293

Van Blitterswijk M, Landers JE (2010) RNA processing pathways in amyotrophic lateral sclerosis. Neurogenetics

Van Rheenen W, Pulit SL, Dekker AM et al (2018) Project MinE: study design and pilot analyses of a large-scale whole-genome sequencing study in amyotrophic lateral sclerosis. Eur J Hum Genet 26:1537–1546. https://doi.org/10.1038/s41431-018-0177-4

Vance C, Rogelj B, Hortobágyi T et al (2009) Mutations in FUS, an RNA processing protein, cause familial amyotrophic lateral sclerosis type 6. Science (80-). https://doi.org/10.1126/science.1165942

Vasaikar SV, Padhi AK, Jayaram B, Gomes J (2013) NeuroDNet - an open source platform for constructing and analyzing neurodegenerative disease networks. BMC Neurosci 14:3. https://doi.org/10.1186/1471-2202-14-3

Venselaar H, te Beek TAH, Kuipers RKP, Hekkelman ML, Vriend G (2010) Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. BMC Bioinformatics https://doi.org/10.1186/1471-2105-11-548

Wu D, Yu W, Kishikawa H, Folkerth RD, Iafrate AJ, Shen Y, Xin W, Sims K, Hu GF (2007) Angiogenin loss-of-function mutations in amyotrophic lateral sclerosis. Ann Neurol 62:609–617. https://doi.org/10.1002/ana.21221