scientific reports

OPEN

Dynamic insights into the effects of nonsynonymous polymorphisms (nsSNPs) on loss of TREM2 function

Raju Dash[1], Yeasmin Akter Munni[1], Sarmistha Mitra[1], Ho Jin Choi[1], Sultana Israt Jahan[2], Apusi Chowdhury[3], Tae Jung Jang[4] & Il Soo Moon[1]✉

Single nucleotide variations in Triggering Receptor Expressed on Myeloid Cells 2 (TREM2) are associated with many neurodegenerative diseases, including Nasu-Hakola disease (NHD), frontotemporal dementia (FTD), and late-onset Alzheimer's disease because they disrupt ligand binding to the extracellular domain of TREM2. However, the effects of nonsynonymous single nucleotide polymorphisms (nsSNPs) in TREM2 on disease progression remain unknown. In this study, we identified several high-risk nsSNPs in the *TREM2* gene using various deleterious SNP predicting algorithms and analyzed their destabilizing effects on the ligand recognizing region of the TREM2 immunoglobulin (Ig) domain by molecular dynamics (MD) simulation. Cumulative prediction by all tools employed suggested the three most deleterious nsSNPs involved in loss of TREM2 function are rs549402254 (W50S), rs749358844 (R52C), and rs1409131974 (D104G). MD simulation showed that these three variants cause substantial structural alterations and conformational remodeling of the apical loops of the TREM2 Ig domain, which is responsible for ligand recognition. Detailed analysis revealed that these variants substantially increased distances between apical loops and induced conformation remodeling by changing inter-loop nonbonded contacts. Moreover, all nsSNPs changed the electrostatic potentials near the putative ligand-interacting region (PLIR), which suggested they might reduce specificity or loss of binding affinity for TREM2 ligands. Overall, this study identifies three potential high-risk nsSNPs in the *TREM2* gene. We propose further studies on the molecular mechanisms responsible for loss of TREM2 function and the associations between TREM2 nsSNPs and neurodegenerative diseases.

Microglial cells in brain express an innate immune cell surface receptor called Triggering Receptor Expressed in Myeloid cells 2 (TREM2), which controls a wide range of microglial immune functions, including chemotaxis, phagocytosis, autophagy, survival and proliferation, proinflammatory cytokine production, and lipid metabolism[1–7]. TREM2 is activated when its ligand-binding domain interacts with a broad range of putative ligands, such as oligomeric amyloid-β[8], bacterial lipopolysaccharides, lipidated apolipoproteins like ApoA, ApoE, and CLU[9], various sphingomyelin and phospholipids[9,10], or nucleic acids[11], though it shows a preference for anionic substrates[12,13]. Ligand binding initiates downstream signaling via the recruitments of co-receptor DNAX-activation protein 12 (DAP12) and SYK, ERK, PLCG2, or NFAT[14,15]. In cerebrospinal fluid (CSF), TREM2 is also detected in a soluble form (sTREM2)[9], which is produced by the proteolytic actions of ADAM10 and ADAM17[16,17] or alternative transcription[18]. TREM2 plays a critical neuroprotective during early and mid-term Alzheimer's disease (AD), as it suppresses the Aβ diffusion and accumulation by regulating microglial activation around amyloid plaques[3,19,20]. Furthermore, TREM2 overexpression is also associated with the clearance of soluble and insoluble Aβ42 aggregates[21,22]. Studies suggest sTREM2 physically binds to Aβ42 and inhibits its polymerization and have reported that sTREM2 overexpression ameliorates AD in a mouse model[23] and protects cognitive reserve[24].

Current genome-wide association studies indicate loss of TREM2 function due to the presence of homozygous or heterozygous variants is highly associated with the progression of many neurodegenerative disorders, including AD, Nasu-Hakola disease (NHD), Frontotemporal dementia (FTD), and Parkinson's disease (PD)[25–29].

[1]Department of Anatomy, Dongguk University College of Medicine, Gyeongju 38066, Republic of Korea. [2]Department of Biotechnology and Genetic Engineering, Noakhali Science and Technology University, Noakhali 3814, Bangladesh. [3]Department of Pharmaceutical Science, North-South University, Dhaka 1229, Bangladesh. [4]Department of Pathology, Dongguk University College of Medicine, Gyeongju 38066, Republic of Korea. ✉email: moonis@dongguk.ac.kr

Molecular modeling of the TREM2 ectodomain showed that AD-associated variants locate on the extracellular surface near the putative ligand-interacting region (PLIR), while NHD-associated variants are grossly damaging by frameshift, truncation, or unfolding. The most common variants associated with AD risk are R47H and R62H[25,26], which reportedly disrupt ligand interaction but do not markedly affect protein structure or stability[13,30,31]. Dynamics studies conducted by molecular simulation showed that AD-associated variants induced flexibility in the three loops of the complementarity-determining region (CDR) adjacent to the PLIR, which is responsible for ligand binding and recognition. This flexibility ultimately disrupts the structural integrity of the core structure of the Ig domain and between these CDRs, and results in the exposure of negatively charged buried residues[32]. These structural consequences are most severe when the structure contains FTD and NHD mutations and are associated with complete loss of function[7,33–35]. Other studies suggest that TREM2 variants at Q33X and W191X might influence loss of TREM2 function[28,36]. Although many biophysical and molecular dynamics-based studies have characterized the impacts of several known risk-associated variants[13,15,31,32], the disease triggering potentiality of TREM2 nonsynonymous single nucleotide polymorphisms (nsSNPs) remains uncharacterized.

In silico methods provide an effective means of finding deleterious nsSNPs in specific genes[15,37–40]. Thus, we utilized bioinformatic prediction-based tools to identify high-risk SNPs and then all-atom MD simulation to analyze the magnitudes of their effects on TREM2 structure. The MD simulation findings obtained identified three high-risk nsSNPs that alter TREM2 Ig domain structure and destabilize ligand-binding regions and concurred with previous reports[15,31,32].

Results

Identification of deleterious nsSNPs in _TREM2_. Missense variants play vital roles in many complex diseases by modulating in vivo protein functions[41,42]. Available missense SNPs in the dbSNP database (a total of 228 SNPs) were subjected to analysis to determine their effects on TREM2 structural stability and dynamics (Supplementary file 1). Using structure and sequence-based approaches (Table S1, Supplementary file 2), a total of 17 in silico nsSNP prediction algorithms were used to analyze these SNPs. In silico nsSNP prediction algorithms such as SIFT, PolyPhen, Condel, CADD, DANN, FATHMM, M-CAP, MetaLR, MutPred, MutationAssessor, PROVEAN, VEST3, fathmm-MKL, MuPro, iStable, PhD-SNP, and SNAP2 were used in this study (Table S1). The DANN algorithm predicted the highest number of deleterious SNPs, while FATHMM and MetaLR predicted the lowest number (Fig. 1A). The predictions of all algorithms were found to correlate significantly with one another. However, two algorithms, FATHMM and VEST3, produced negative correlations with other tools (Fig. 1B). SNPs that have been identified as deleterious by the most of the tools are often more likely to be deleterious than other SNPs [43,44]. Three SNPs rs549402254 (W50S), rs749358844 (R52C), and rs1409131974 (D104G) were predicted by all algorithms, that is, by at least 15 tools, were considered high-risk nsSNPs and subjected to further analysis (Table S2, Supplementary file 2).

Structural consequences by molecular dynamics (MD) simulation

MD simulations were performed for variants and wild-type structures to obtain detailed structural and dynamic insights of the roles played by the identified nsSNPs in _TREM2_. Previous biophysical studies revealed that the TREM2 Ig domain consists of a hydrophobic patch and contributes to ligand binding (Fig. 2A). This region, which is also known as the complementarity-determining region (CDR), is composed of three major loops (CDR1 to CDR3) (Fig. 2B). A putative ligand-interacting region (PLIR) is also located in the Ig domain and displays a positively charged patch of surface-exposed residues that includes residues from the CDR2 and βC″ strands. Figure 2 shows that the variants W50S and R52C are located near the PLIR in the βC strand (Fig. 2Ca&b) and are buried, whereas D104G is located in the loop between βE and βF (Fig. 2Cc).

MD simulation was used to analyze the structural stabilities and dynamics of the three TREM2 variants. As described in "Methods", three independent simulations were run for each variant and wild-type for 500 ns (a total of 1.5 μs) and subjected to root mean square deviation (RMSD) analysis to determine equilibrated trajectories by considering initial protein backbone structures (Supplemental Video 1, 2, 3 & 4). As shown in Figure S1 (Supplementary file 2), all variants and wild-type achieved equilibration after around ~ 100 ns of simulation and remained stable thereafter with a maximum RMSD of < 3.0 Å. To improve conformational sampling efficiency in the simulated trajectory analysis, the last 300 ns of trajectory from each run was extracted and concatenated to a sub-trajectory of 900 ns for further analysis. To confirm the sufficiency of conformational sampling, we analyzed sub-trajectory cosine contents, which have been shown to be good indicators of sufficient trajectory sampling to achieve convergence[45–48]. When the cosine content value is high (i.e., close to 1), protein dynamics resemble random diffusion, indicating inadequate sampling. In contrast, a cosine content near 0 indicates convergent sampling. Cosine content analysis showed that all trajectories had cosine values < 0.1 (Table S3, Supplementary file 2)[47], which indicated trajectories were convergent and appropriate for detailed analysis[49].

Alterations in conformational stabilities. To examine the variations in the conformational stabilities of variant structures, RMSD values of the equilibrated trajectories, which indicate geometrical similarities between structures, were calculated for TREM2 variants and the wild-type[50,51]. As shown in Figure S2A (Supplementary file 2), TREM2 W50S, R52C, and D104G variants (Figures S2Aa, b & c, Supplementary file 2) had larger RMSD deviations than the wild-type (Figure S2dAd, Supplementary file 2), and D104G exhibited a large RMSD distribution in the density plot shifted right of wild-type distribution (Figures S2Ac, Supplementary file 2) with a maximum at ~ 1.4 Å. To confirm these deviation changes, we calculated radius of gyration (Rg), as they are indicators of protein compactness and lack of flexibility (Figures S2B Supplementary file 2). Rg analysis showed that TREM2 W50S, R52C, and D104G variants had greater Rg values than the wild-type (Figures S2Ba,b&c, Supplementary file 2), and the difference was marked for R52C but slight for W50S and D104G. Since all vari-
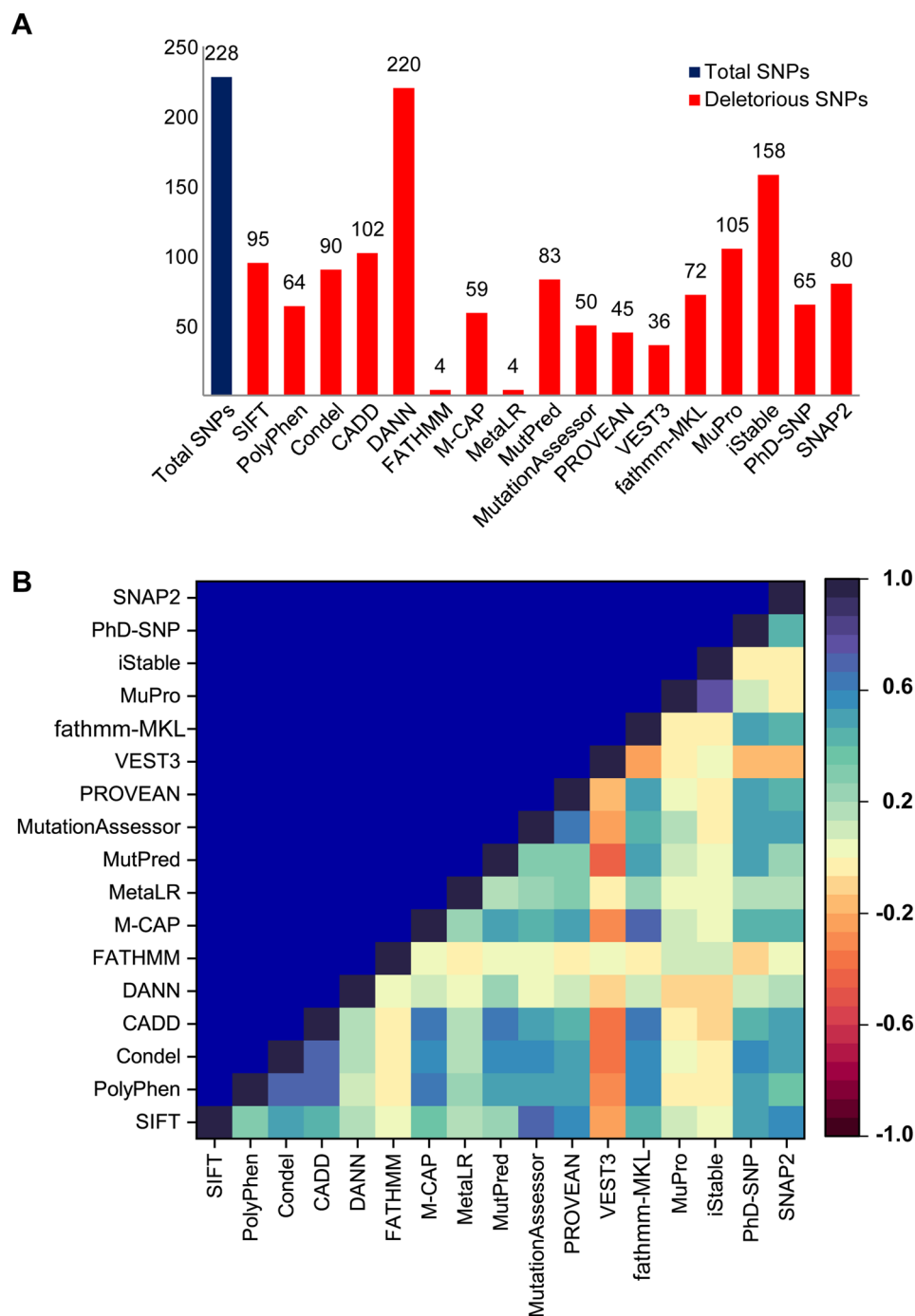
Figure 1. Identification of deleterious SNPs in *TREM2* using different computational algorithms. (**A**) Numbers of total and deleterious SNPs predicted by each tool. (**B**) Pairwise correlations between deleterious SNP predicting algorithms presented as a color-coded heatmap, which shows a positive correlation (+) when algorithmic predictions agreed and a negative correlation (−) when they disagreed.

ants had higher Rg values than the wild-type, we calculated solvent accessible surface areas (SASAs) (Figure S2C, Supplementary file 2), which reflect surface areas accessible for biomolecular interactions. Figures S2Ca,b&c (Supplementary file 2) show that the TREM2 structure containing all three variants had a greater SASA distribution than the wild-type (> 64 nm^2) (Figure S2Cd, Supplementary file 2). D104G had a higher average SASA than W50S or R52C and exhibited a wider SASA distribution (64 to 67 nm^2) (Figure S2Cc, Supplementary file 2). RMSD and Rg calculations and SASA analysis indicated all variants exhibited conformational changes in the structure of TREM2, which suggested substantial domain or regional fluctuations. Moreover, Rg and SASA calculations collectively suggested that the presence of variants in TREM2 reduces structural compactness, and thus, increases protein flexibility.
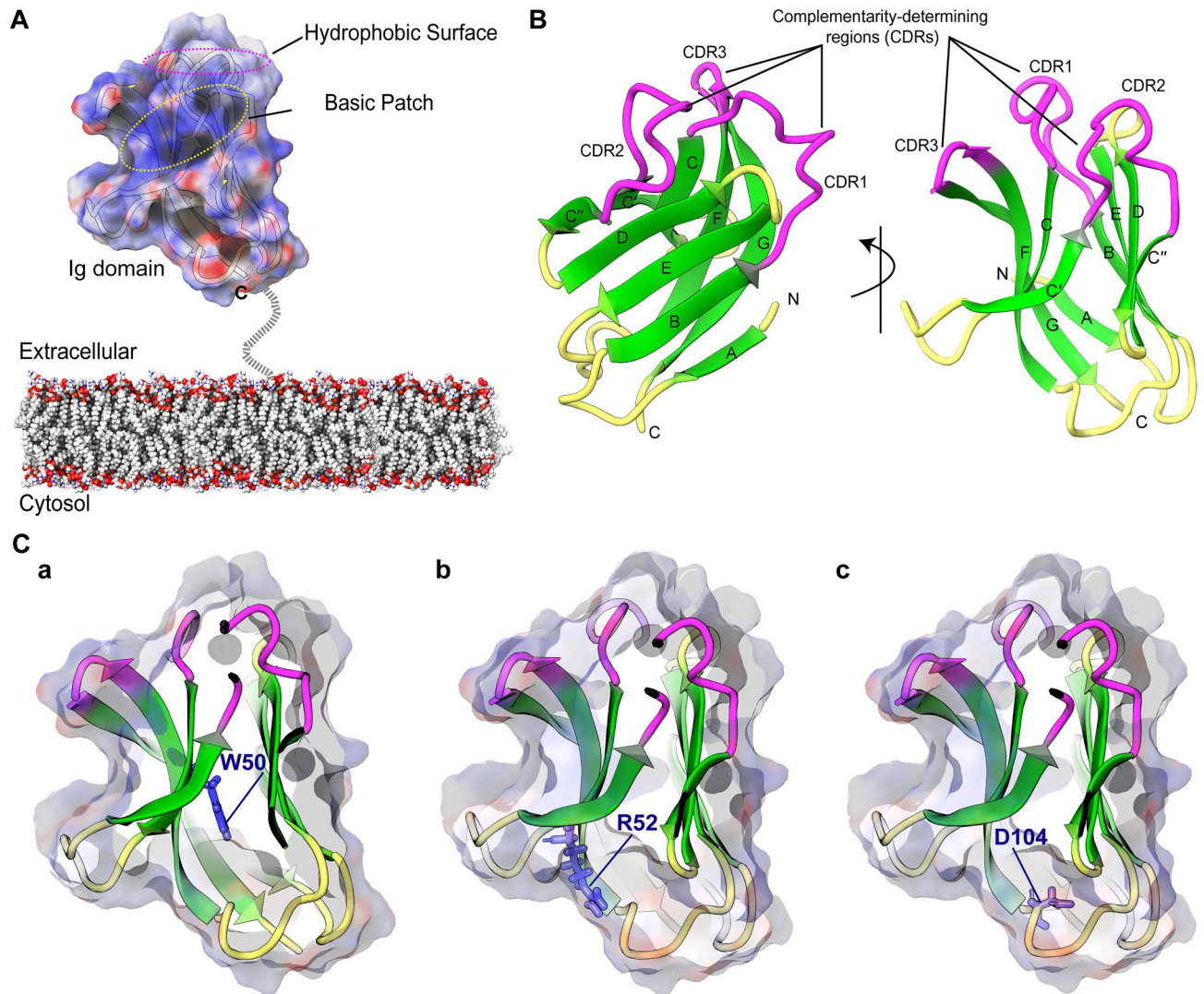
Figure 2. Molecular architecture showing the domain organization of TREM2 and the locations of SNPs. (**A**) In the cell membrane, TREM2 presents its immunoglobulin (Ig) domain (ectodomain) in surface view mode. The putative ligand-interacting region (PLIR) is indicated by the dotted yellow circle. (**B**) Three-dimensional representation of TREM2 Ig domain, CDRs are highlighted in pink and labeled with β-strands. (**C**) Locations of the identified SNPs in the 3-D structure of wild-type TREM2 Ig, where W50S (**a**), R52C (**b**), and D104G (**c**) are represented by blue sticks.

Changes in regional flexibility and correlative motion.

Root mean square fluctuation (RMSF) analysis was used to quantify TREM residual flexibility in the equilibrium state and identify regions that most influence conformational motion and stability[52]. Results are plotted in Fig. 3, which shows D104G increased residual fluctuation as compared with the wild-type (Fig. 3Ac) and that these effects were more pronounced in the CDR1 loop and the βE region. W50S and R52C also increased fluctuation in the CDR1 loop to a level higher than that of wild-type (Fig. 3Ab). W50S also showed substantially greater fluctuations in residues of the βC-βC' loop than the wild-type (Fig. 3B). These observations suggest that the identified variants might interrupt intramolecular networks essential for the stability of functional regions, like apical CDRs.

Dynamic cross-correlation maps (DCCMs) were constructed based on equilibrated trajectories to investigate pairwise relative motion amid all residue pairs in the TREM2 Ig domain and enable correlations between these regional changes (Fig. 3) and dynamic motions as determined using color-coded maps. In accord with RMSF analysis, DCCM also revealed notable changes between the variants and the wild-type, and a substantial reduction in relative motion was observed for D104G (Fig. 4A). W50S increased anticorrelated motion between the βC-βC' loop and the N-terminal segment of TREM2 (residues, 100 to 129), and this loop (βC-βC') also exhibited high flexibility by RMSF analysis. In addition, all CDR loops in R52C showed slight increases in anticorrelated motions relative to each other (Fig. 4B). On the other hand, correlative motion was substantially reduced in the core domain of D104G.
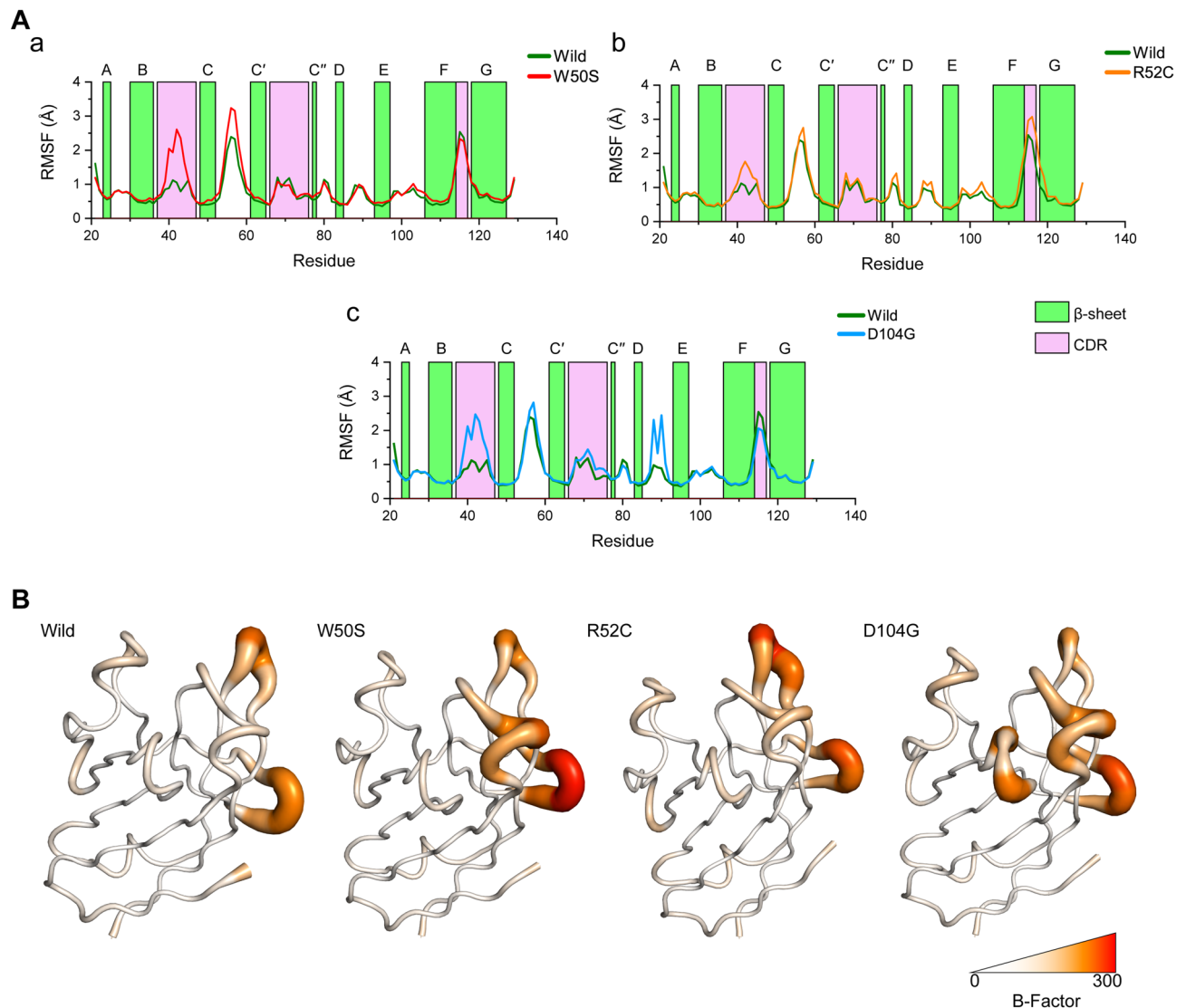
Figure 3. Changes in residual flexibilities during simulation. (**A**) Changes are represented by root mean square fluctuation (RMSF) plots for W50S (**a**), R52C (**b**), and D104G (**c**) versus the wild-type structure. Color bars represents β-sheet propensity and CDRs. (**B**) Comparison of TREM2 Ig domain fluctuations as demonstrated by B-factor coloring and thickness. RMSF values were converted to B-factors to represent structures in colored putty representations . Coloring was made by white-red spectrum, white means B-factor with a value 0, while red means 300. Tube thicknesses also indicate levels of fluctuation, i.e., tube thicknesses represent levels of fluctuation.

Changes in collective motion.   Protein dynamics can be visualized by principal component analysis (PCA), which represents ensembles using a sequence of eigenvectors, whereby each eigenvector represents an aspect of protein motion by a phase space behavior[53]. PCA analysis (Fig. 5) showed the variances of variant structures differed from the wild-type. The similarities and differences between essential subspaces of wild-type and variant ensembles were highlighted by root mean square inner product (RMSIP) calculations[53–55]. Plots representing pairwise comparisons between the wild-type and variants suggested that the dynamic motions of the variants and wild-type differed, and substantial differences were observed for the first three principle components (PCs) (Fig. 5 Aa, b & c). In addition, PCA trace values of the wild-type, W50S, R52C, and D104G were 46.5, 70.492, 66.412, and 63.905 Å^2, respectively, which indicated the variants had higher collective flexibilities than the wild-type[37]. The conformational distributions of protein structure in each subspace were projected onto a 2D plot for the first three PCs of wild-type and variant trajectories (Fig. 5B). The wild-type had more concentrated, overlapping dots on PC 1/2 projections than the variants (Fig. 5Ba and Fig. 5Bb, c&d), and projected directions of dot distributions also differed, suggesting the change of conformational behavior. Similar patterns were also observed in projections of PC 1/3 and 2/3, though differences were more subtle than those in wild-type and variant projections (Figure S3, Supplementary file 2). Color-coded scattered dots in PCA plots represent conformational states; red dots denote the steady-state, blue dots unstable states, and white dots intermediate
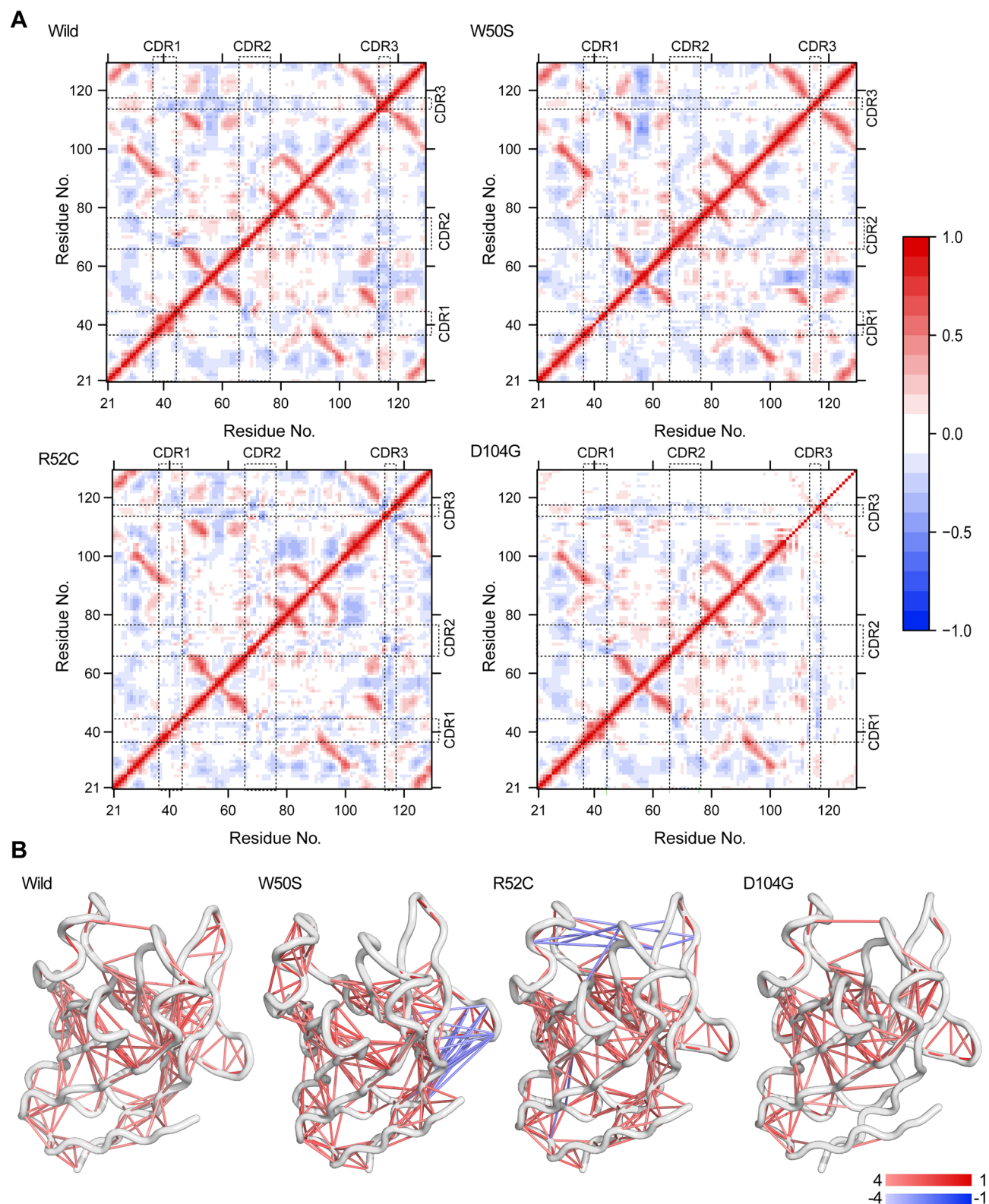
Figure 4. Alterations in the correlated motions in TREM2 variants. (**A**) The dynamic cross-correlation matrix (DCCM) plot shows anticorrelated and correlated motions between every two residues in the structure. Here a perfect correlated motion is denoted by red (+ 1), and an anticorrelated motion by blue (− 1). CDRs are indicated by dotted boxes. (**B**) DCCM representation in the structural view of TREM2, where lines denote correlations between two residues. A blue line indicates anticorrelated motions, and a red line denotes positive correlations. Color intensities indicate strengths of correlations.
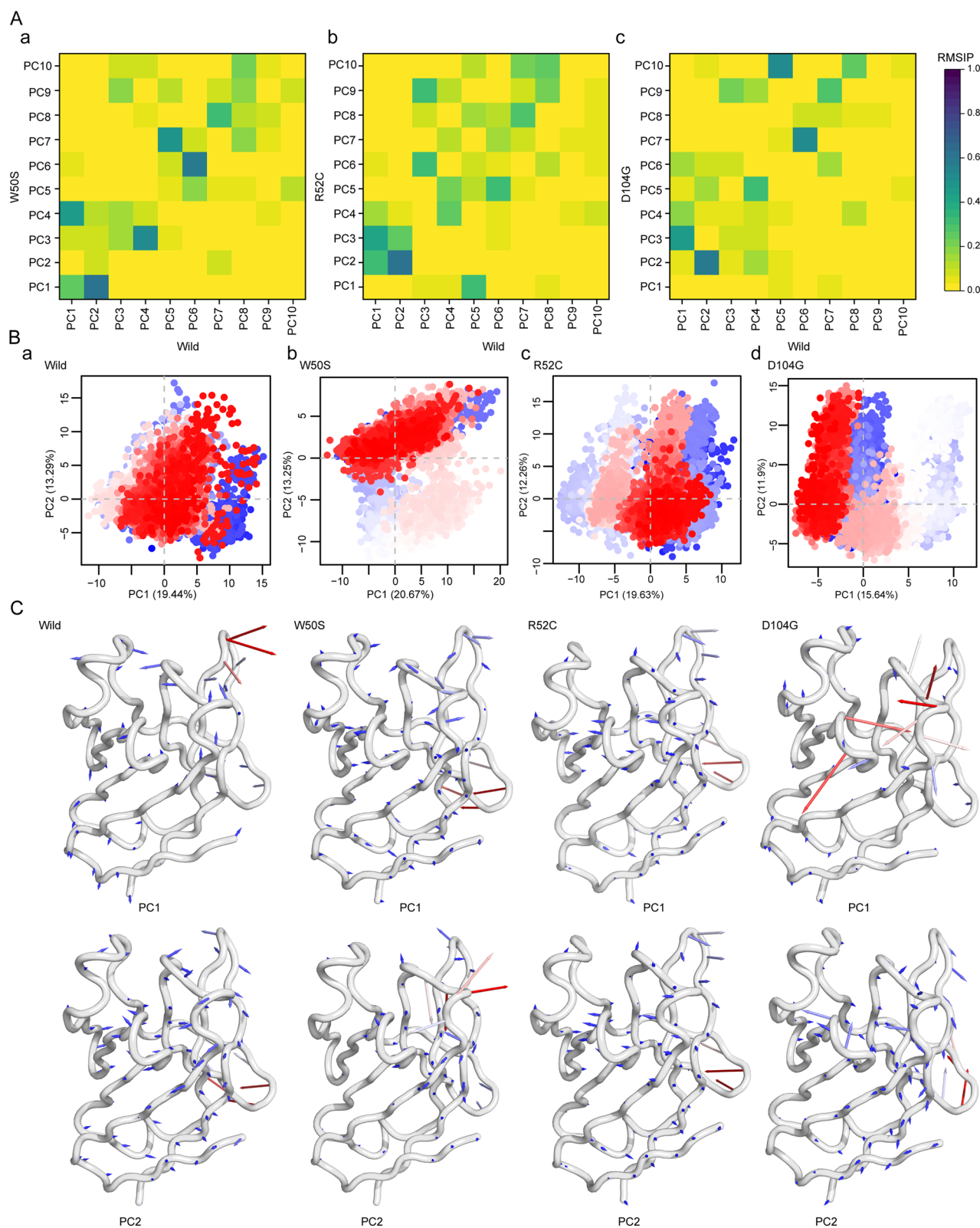
Figure 5. Changes protein dynamics induced by TREM2 variants were obtained by principal component analysis. (**A**) Root-Mean Square Inner Products (RMSIPs) of the first ten principal components showing similarities between the conformational spaces of the wild-type and variants. A gradient heat map indicating RMSIP values from yellow to dark blue. (**B**) The distribution of trajectory conformers was projected onto the principal planes of components 1 and 2, where dots represent the structure colored based on time of evolution (from blue to red). (**C**) Contributory movements in the first (upper panel) and second (lower panel) are represented by porcupine plots for all variants and wild-type.

states[56]. To visualize these changes in protein structure, a porcupine plot was generated for each variant and wild-type (Fig. 5B). Corresponding fluctuations are presented as line plots in Figure S4 (Supplementary file 2).

As shown in Figs. 5B and S3A (Supplementary file 2), dynamic changes in variant structures largely occurred in CDRs, as was shown by RMSF analysis. Furthermore, D104G caused a substantial fluctuation in the CDR1 region and in the βD-βE loop but reduced fluctuation in the CDR3 loop in PC1 (Figure S4A, Supplementary file 2). Notably, the wild-type structure conferred a stable transition in the CDR3 loop, and this was disrupted in the variants. W50S showed a large amplitude at the βC-βC' loop in PC1 but a lower amplitude in PC2 (Figure S4B, Supplementary file 2) and exhibited marked fluctuations in the CDR1 loop. R52C showed notable changes in the CDR2 and CDR3 regions of PC1 and PC2 as compared with the wild-type.

Stability of secondary structures in the TREM2 Ig domain.

Define Secondary Structure of Proteins (DSSP) is frequently used to identify changes in protein secondary structures during MD simulation. Since RMSF, DCCM, and PCA analyses (Figs. 3, 4 & 5) suggested variants induced significant alteration in CDR dynamics, we examined the total occupancy of essential secondary structures, including helix and strand formation, and contributions of residues in equilibrated trajectories. As shown in Fig. 6A, in the wild-type structure α-helix formation was observed in CDR1 and CDR2 loops with occupancies of ~ 20% and 40%, respectively. However, TREM2 variants exhibited less A-helix formation and more 3_{10}-helix formation in CDRs and other regions, including residues 55–58, 87 to 91, and 101 to 106, respectively. For example, A-helix formation in W50S was 11% lower, but 3_{10}-helix formation was 45%. In addition, a marked increase in 3_{10}-helix occupancy was observed in the CDR2 region of R52C, while its C-terminal region (110 to 129) lost most of its β-strand structure. On the other hand, D104G exhibited increased β-strand structure formation, especially in 90 to 129.

Stability of CDRs in the TREM2 Ig domain.

Previous studies have reported that stability in CDR regions is critically maintained by intra-residual communication, which is typically altered by mutation[15,31,32]. As substantial alterations in the stability and structural organization of CDRs had been noted in the variants containing TREM2 structures, we monitored and plotted inter-loop distances among the α-carbons of CDR1, CDR2, and CDR3 (Fig. 6B). As shown by Fig. 6B, the distances between CDR1 and CDR2 (Fig. 6Bb) and between CDR1 and CDR3 (Fig. 6Bc) in all variants differed substantially from the wild-type. The distance between CDR2 and CDR3 in R52C showed a wide distribution, whereas D104G showed a narrow distribution (Fig. 6Bd).

All variants exhibited a greater average distance between CDR1 and CDR2 and between CDR2 and CDR3 than the wild-type (Fig. 6Be&g). As regards, average distances between CDR1 and CDR3 in W50S and D104G were markedly greater than in R52C and the wild-type (Fig. 6Bf), which supported our fluctuation and motion analyses findings. An increase in average inter-loop distance suggests that increased motion caused CDRs to spread and disrupt electrostatics potentials near the PLIR, including CDR2 and the βC" strand.

Since the stability of CDR loops can be altered or achieved by inter-loop communication, total contact between CDR loops was counted and plotted in heatmap (Figure S5, Supplementary file 2). Total contact between any two residues of different loops was only considered when contact occupancy exceeded 10%. The results obtained showed inter-loop interaction between CDR1 and CDR2 (Figure S5A, Supplementary file 2) and CDR1 and CDR3 (FigureS4B, Supplementary file 2), and no interaction between CDR2 and CDR3. As can be seen in the wild-type, Arg47 of CDR1 maintained contact at ~ 100% with Thr66 to Asn68 residues of CDR2 (Figure S5Aa, Supplementary file 2), but different interaction patterns with Leu69 and Trp70 in variant containing structures. All three variants exhibited reduced contact formation between Asn68 and Gly45, but substantial inter-variant differences were observed for contacts between Trp70 and Lys42 or Arg47 (Figure S5A b, c & d, Supplementary file 2). As regards contact formation between the CDR1 and CDR3, the Arg46 residue of the CDR loop of R52C exhibited contact formation with Ser116 to Glu117, whereas the wild-type showed interactions only with His114 and Gly115. All variants exhibited substantially less contact between His114 and His43 or Lys42 than the wild-type, whereas R52C showed less contact with His43 (< 20% vs. > 40% for the wild-type).

Changes in electrostatic potentials near the PLIR.

Electrostatic potentials over the solvated protein surface play an essential role in the recognition and binding of macromolecules. TREM2 disease-associated mutants have been reported to change electrostatic potentials near the PLIR[57]. We used free energy landscape analysis (FEL) to surface map electrostatic potentials near this region for all variants and wild-type structures. FEL represents conformational distributions from high to low energy minima[58,59]. In Fig. 7A, purple and red colors represent minima from low to high energy. Substantial changes in the conformational space region were observed in FEL maps of variants as compared with the wild-type (Fig. 7A). As shown in Fig. 7B, the wild-type structure showed narrow deeper purple grooves than the variant structures, whereas variant structures showed wider grooves, suggesting that variants exhibited conformational and structural transitions. Representative conformers with stable lowest energy states were extracted using FEL to draw electrostatic potential surface maps for each variant (Fig. 7C), and these maps showed subtle changes in wild-type electrostatic potentials (Figs. 7Ca, b, c &d). W50S (Fig. 7Cb) and R52C (Fig. 7Cc) variants increased electrostatic potential near the PLIR, which was absent in D104G (Fig. 7Cd). Furthermore, all variants had CDRs with helical conformations, indicating altered conformational stability in the ligand-binding region (Figure S6, Supplementary file 2).

Discussion

Using a series of bioinformatic algorithms, the present study identified three potential deleterious SNPs, namely, W50S, R52C, and D104G, in TREM2 and confirmed their deleteriousness by dynamic behavior analysis. It is worth noting that the prediction of deleterious SNPs using a single tool can result in false positives, and thus, several algorithms, including sequence and structure-based approaches, were utilized to maximize prediction
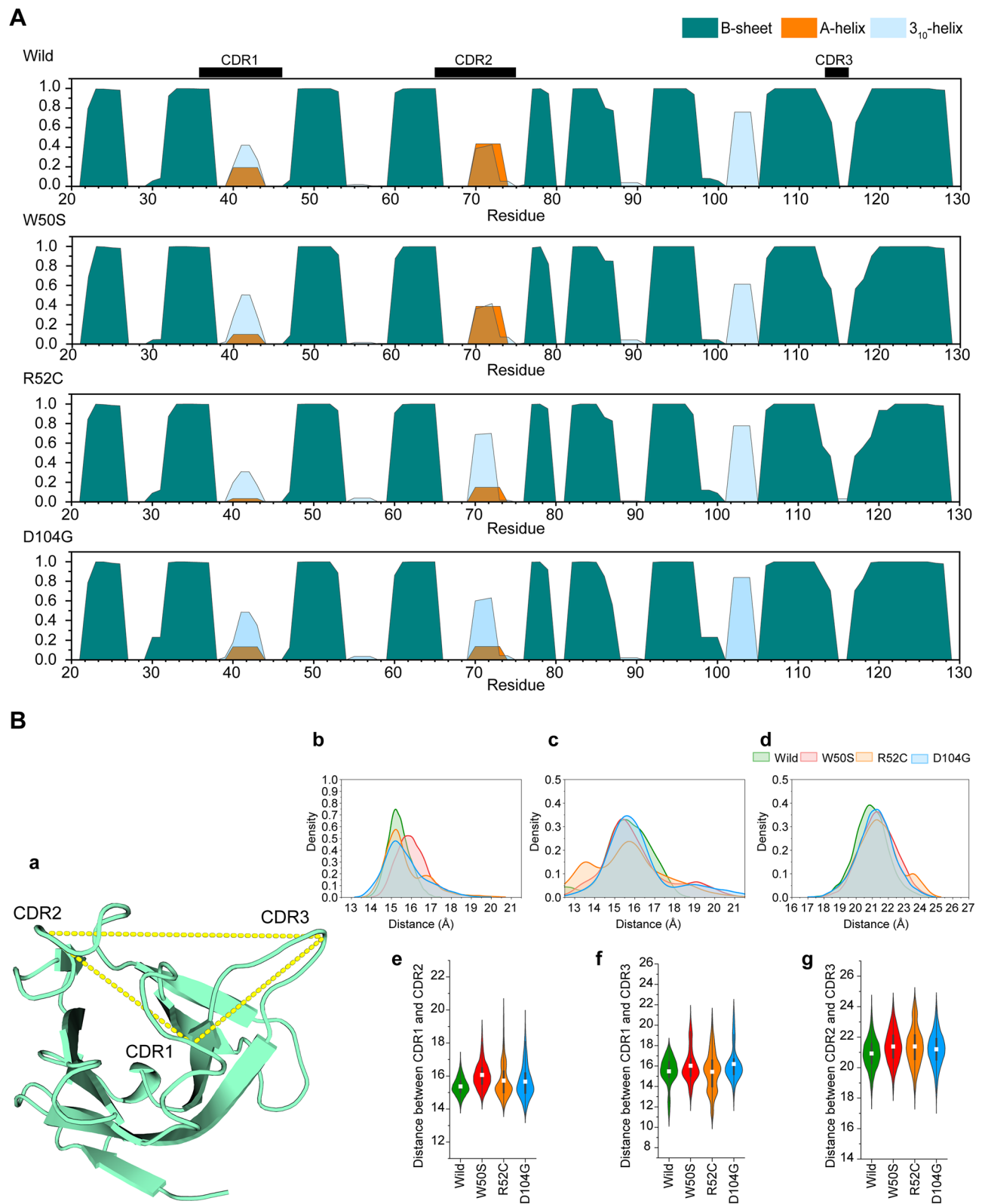
Figure 6. Structural alterations at CDRs due to the presence of nsSNPs. (**A**) Occupancies of secondary structures represented as fractions and percentages were calculated for α-helices (A-helix), β-sheets (B-sheet), and 3_{10}-helices (3_{10}-helix). (**B**) Changes in the inter-loop distances between CDRs. Distance calculations were performed on α-carbons of His43 (CDR1), Leu72 (CDR2), and Ser116 (CDR3) based on considerations of equilibrated trajectories (**a**). Changes in inter-loop distances represented as probability density functions for CDR1 and CDR2 (**b**), CDR1 and CDR3 (**c**), and CDR2 and CDR3 (**d**). Mean differences in distances between CDR1 and CDR2 (**e**), CDR1 and CDR3 (**f**), and CDR2 and CDR3 (**g**) are shown as violin plots.
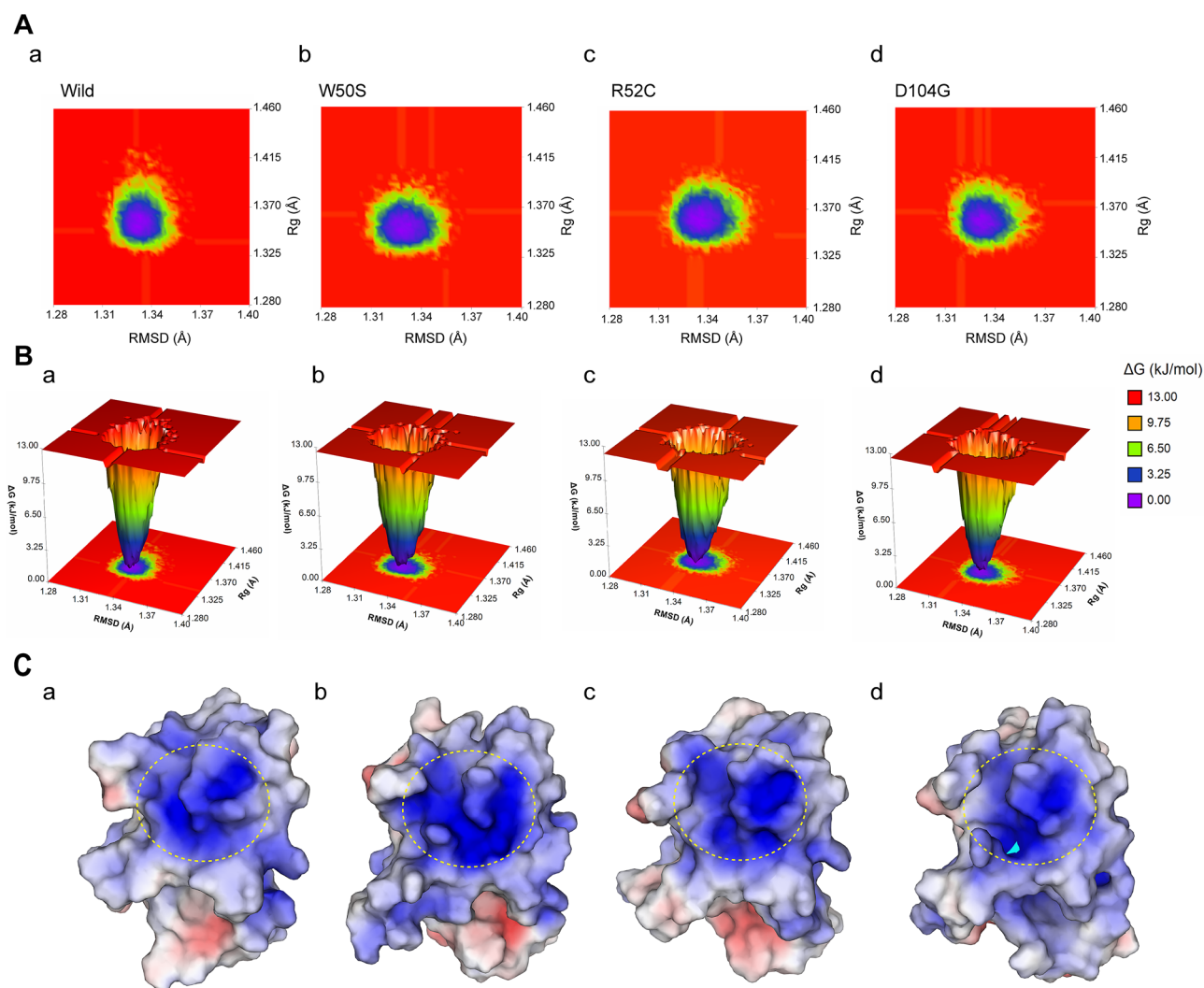
Figure 7. Effect of variants on the positive electrostatic potential near CDRs. (**A**) Free energy contour maps showing conformational distributions according to depth of energy minima for the wild-type (**a**), W50S (**b**), R52C (**c**), and D104G (**d**). Energy states as demonstrated by a color-coded map, where purple indicates a protein conformer with a lower energy minimum, and red a higher energy minimum then the wild-type. (**B**) Three-dimensional representation of the free energy landscapes (FEL) or the wild-type (**a**), W50S (**b**), R52C (**c**), and D104G (**d**). (**C**) Surface map of positive electrostatic potentials near the PLIR for the wild-type (**a**), W50S (**b**), R52C (**c**), and D104G (**d**) variants. The map was constructed using the lowest energy conformers retrieved from FEL, at corresponding timestamps of 292.5 (wild-type), 639.1 (W50S), 827.6 (R52C), and 868.1 ns (D104G). Darker blue regions indicate electropositive areas (> 6 kT/e), and red regions electronegative areas (< -6 kT/e). The yellow dotted line indicates the PLIR.

accuracy[43,44]. Accordingly, W50S, R52C, and D104G, were classified as high-risk SNPs, by consensus prediction. Interestingly, these three variants exhibited substantial changes in the CDR loops as compared TREM2. More specifically, they showed changes in dynamic motion, secondary structure, and increased inter-loop distances. Previous studies have reported a short α-helix in the CDR2 loop of AD-associated variants[15,31] and suggested it might be associated with disease severity[32]. MD simulation showed that the CDR2 loop of the wild-type structure maintained an α-helix conformation in around 40% trajectories (Fig. 6A), which suggested the α-helix in the CDR2 loop is included the dynamic properties of the wild-type. However, in the TREM2 variants, the helical pattern was of the 3_{10}-helix type (Fig. 6A), which indicates that this structural conversion might be associated with disease severity.

Several studies have provided evidence that disease-associated variants of TREM2 exhibit conformational remodeling of CDR[32]. This region comprises a hydrophobic patch mainly composed of residues 40–47 (CDR1), 67–78 (CDR2), and 115–120, which are essential for ligand binding specificity and strength. Furthermore, it has been suggested conformational remodeling depends on inter-loop communication and that nonbonded interactions between these loops modulate secondary conformation. CDR2 conformational stability is critically maintained by interactions between Ser65, Thr66, His67, and Asn68 and Arg47 (in the CDR1 loop)[31]. In the present study, these interactions differed in variants during simulation, except for interactions involving Leu71

(Figure S5, Supplementary file 2). Interestingly, distances between CDRs were greater in all variants, especially in W50S. It has been predicted, alterations of positive electrostatic potentials near the PLIR can increase CDR motion[15,60], and thus, greater distances between CDRs might expose a buried patch of negative electrostatic potential and influence ligand binding. Analysis of a single protein structure obtained by FEL analysis revealed significant changes in secondary organizations and electrostatic potential surface area expansion (Fig. 7), indicating variants exposed residues typically buried in the wild-type structure. In the case of R52C, a marked change was observed, whereby increased flexibility induced significant correlated motion of CDRs (Fig. 4B), which increased nonbonded contacts between loops and facilitated conformational remodeling in CDR1 and CDR3 (Fig. 4B). Although these helix formations did not grossly change electrostatic potential surface areas (Fig. 7C), a substantial reduction in β-sheet formation was also observed near the CDR3 loop region (Fig. 6A). In a previous study, it was demonstrated that reductions in CDR stability reduced ligand binding, which it was argued would induce immune inactivation and possibly underlie the neurodegenerative effects of TREM2 variants[40] (Table S4, Supplementary file 2).

Replacement of residues in proteins can change inter- and intra-molecular interactions and communications by altering protein flexibility and causing steric clashes and unfavorable interactions[38,61,62]. In the current study, we like others[31,32,37], identified major changes in the inter-loop interactions of variants (Fig. 6B and Figure S5), which have been posited to underlie loss of CDR stability and function[39,40]. In the W50S, R52C, and D104G variants, substituted residues differed from those of the wild-type in hydrophobicity, charge, and size. Briefly, serine is smaller and more hydrophilic than tryptophan at 50, while cysteine in R52C is small, neutral in terms of charge, and more hydrophobic than the wild-type residue. Similarly, glycine is more hydrophobic, negatively charged, and more flexible than aspartic acid at 104. Increased hydrophilicity in W50S results in the loss of hydrophobic interactions with neighboring residues, whereas increased hydrophobicity in R52C and D104G disrupts ionic interactions and hydrogen bonding. Furthermore, W50S is buried in the core of the Ig domain, and the smaller size of serine might disturb core structure.

The findings of the present study agree well with previously published MD simulation findings and experimental data regarding loss of ligand binding in TREM2 variants due to the conformational remodeling of CDRs driven by alterations in inter-residue contacts, and thus, TREM2 binding site alterations (Table S4, Supplementary file 2). Actually, the in silico deleterious SNP prediction methods also identified several rare disease-associated TREM2 variants. The previously reported rare AD-associated missense variants R47H, H157Y and D87N (Supplementary file 1)[63] were also identified as deleterious by more than five tools in the present study. R47H has deleterious characteristics similar to W50S, R52C, and D104G, including CDR conformational remodeling and loss of ligand binding in functional assays. In addition, the NHD and FTD associated missense variants Y38C, T66M, S31F, and R47C[64] were identified as deleterious by most of the in silico tools used in this study, and Y38C and T66M were found experimentally to cause loss of ligand binding. In a molecular dynamics simulation study, it was suggested that Y38C and T66M cause severe conformational remodeling in CDRs loops by changing the interloop nonbonded interaction network and structural dynamics[15]. Interestingly, our MD simulation results for identified variants concurred with previously reported MD simulation results and experimental, which indicates that the methods used for physicochemical, structural, and dynamic characterizations reliably explained loss of ligand binding by TREM2 variants.

Although this in silico study provides detailed insight into the disruptive effects of nsSNPs, further biochemical and structural comparative studies on known variants are required to confirm our findings. In addition, long-timescale atomistic simulations and dynamic studies using replica exchange or other extensive sampling techniques are required to allow firm conclusions to be drawn as to whether nsSNPs functionally disrupt TREM2 in a background of neurodegeneration.

Conclusion

The present study was conducted using a comprehensive bioinformatics design and identified three nsSNPs, namely, rs549402254 (W50S), rs749358844 (R52C), and rs1409131974 (D104G), in the *TREM2* from nsSNPs contained in the NCBI database, which induces structural alterations in the TREM2 Ig domain revealed by MD simulation. Detailed characterizations of the simulated trajectories of these variants demonstrated they exhibited increased loop motion and instability, particularly in CDRs. Although further experimental validation is required to confirm these variants cause reduced TREM2 ligand binding, the altered structural dynamics of TREM2 variants found in the present study concur with those reported for previously identified disease-associated TREM2 variants. We believe our findings reveal that in silico studies provide another means of revealing links between genetic-based studies and neurodegenerative disorders. Furthermore, the study provides new information on the dynamics of CDR regulations in wild-type TREM2 and its deleterious variants, and thus, provides clues regarding drug design and TREM2 gene therapy for the treatment of neurodegenerative diseases.

Methods

Data collection and identification of deleterious SNPs of TREM2. Information regarding *TREM2* SNPs was retrieved using the NCBI SNP database with their corresponding rs IDs. As we focused exclusively on the identification of deleterious SNPs, 17 widely accepted in silico tools were used to identify high-risk SNPs in the *TREM2* gene. The prediction tools used were: SIFT, PolyPhen, Condel, FATHMM, DANN, CADD, M-CAP, Meta-LR, MutPred, MutationAssessor, PROVEAN, VEST3, fathmm-MKL, MuPRO, iStable, PhD-SNP, and SNAP2. Sorting Intolerant from Tolerant or SIFT is an analysis program based on the PSI-Blast algorithm[65] that differentiates neutral and deleterious SNPs[66] and predicts deleterious effects by applying the sequence homology approach[67]. Polyphen calculates the absolute value of differences between the profile scores of allelic variants in their polymorphic positions[68]. The number of aligned sequences is shown at the query position, and these

aid evaluations of the reliabilities of profile score calculations[68]. Condel is an in silico tool used to validate the outcomes of nonsynonymous single nucleotide variants SNVs based on the ensemble scores of multiple prediction tools (SIFT, Polyphen2, MutationAssessor, and FATHMM)[69]. The prediction result is denoted by a score between 0 to 1, which are considered to be deleterious or neutral, respectively[70]. CADD provides a comprehensive evaluation of variants using C-scores or "Phred" scores[71]. CADD score is described as a "meta-annotation" tool because it utilizes data from many other functional annotation tools[72] and can effectively predict a variant's effect on a protein[73]. DANN utilizes a DNN (deep neural network) algorithm that captures non-linear relationships among Boolean features defined for each variant[74]. This allows DANN to annotate non-coding variants and prioritize putative causal variants[74]. FATHMM provides an efficient species-independent method and utilizes large-scale genome sequencing projects with functional, molecular, and phenotypic associations[75]. fathmm-MKL is an integrated algorithm based on the Hidden Markov model and predicts the functional effects of non-coding and coding sequence variants[76]. The pathogenicity classifier tool M-CAP exhibits 95% sensitivity at dismissing 60% of rare missense variants of uncertain significance[77]. MutPred can predict the pathogenicity of amino acid substitution and the disease mechanism by employing Random forest depending on the amino acid sequence, functional properties, calculated protein structure and dynamics, and evolutionary information[78]. The prediction is expressed as a probability of deleterious effect, and a probability of > 50% denotes pathogenicity[79]. MutationAssessor predicts based on the evolutionary conservancy of affected residues and predicts the possible role of a mutation on phenotype using multiple sequence alignment[80,81]. Functional impact scores are generated from this evaluation of evolutionary information (FIS) and categorize nsSNP as neutral, low, medium, or high[81]. PROVEAN[82] is a high throughput online prediction tool that utilizes an alignment-based approach for single and multiple amino acid insertions, deletions, and substitutions to identify disease causing variants[83]. VEST3 uses a machine-learning classifier, which ranks rare missense variants by probability of disease association[84]. VEST3 is based on a random forest algorithm that identifies functional missense mutations[85] and provides p-values for false discovery rates (FDR)[86]. MuPro (a web tool) was used to predict nsSNP-induced alterations in protein stability from energy change values[87], which are quantified using confidence scores that range between − 1 and 1, where a score of < 0 indicates a decrease in protein stability, and a score of > 0 indicates an increase in protein stability[88,89]. The degree of protein destabilization and free energy variation was calculated using iStable[90] (http://predictor.nchu.edu.tw/istable/indexSeq.php) and a support vector machine (SVM). iStable predicts changes in protein stability caused by a single amino acid residue mutation[91]. PhD-SNP[92] is another tool that utilizes SVM to predict neutral or disease-associated SNPs[93]. Based on a neural network classifier, SNAP2 predicts nsSNP-induced changes in secondary structure[94], and distinguishes between the effects of neutral to deleterious SNPs by considering the solvent accessibility effect, secondary structure, and evolutionary conservation[95].

Molecular dynamics (MD) simulation. The crystal structure of the TREM2 Ig domain was downloaded from the RCSB protein databank (PDB ID: 5UD7) and then prepared in Schrödinger 2017–1 (Schrödinger, LLC, New York, NY, USA, 2017), as previously described[15,59,96–99]. After preparing the structure, Schrödinger 2017–1 (Schrödinger, LLC, New York, NY, USA, 2017) was used to include the three variants (W50S, R52C, and D104G) in the structure using mutant residue script. This was followed by a short MD simulation to refine and minimize structure energies using YAMBER3 force field[100,101] in YASARA Dynamics software (YASARA Biosciences GmBH, Vienna, Austria), as previously described[44,52,96]. Further analyses were then conducted using the lowest energy MD conformer.

MD simulation was conducted using YASARA (YASARA Biosciences GmBH, Vienna, Austria) dynamics software, hydrogen bond network optimization[96,98,99,102,103], and AMBER14 to apply force fields[100,101]. A simulation cell (Cubic box) was prepared in a periodic boundary condition and extended on each side such that it was 10 Å larger than the selected protein. The water model TIP3P (transferable intermolecular potential3 points) was added to solvate the whole system using a solvent density of 0.997 gL^{-1}[104]. pKa (acid dissociation constant) values of protein residues were evaluated in the solvated state. Neutralization was performed by adding Na$^+$ or Cl$^-$ ions to the cubic cell to maintain a physiological concentration of 0.9% (0.15 M NaCl). The SCWRL algorithm and H-bonding network optimization were used to carry on the protonation state accurately for each specific amino acid[105]. A simulated a nnealing method was followed by minimizing the energy of each simulated system using the steepest gradient approach for 5000 cycles. Time step interval was fixed at 2.00 fs for all simulations using a multiple time-step (MTS) algorithm. MD simulation was then run at a NaCl concentration of 0.9%, pH 7.4, and 298 K using PME (particle-mesh Ewald), and an 8 Å cut-off distance was used to calculate long-range electrostatic interactions while maintaining the periodic boundary condition[106]. The Berendsen thermostat was applied at constant pressure to maintain the temperature of the simulation system, and time was incremented at 2.00 fs[107–109]. Three independent 500 ns MD runs were performed for each system, and results were obtained using a time interval of 100 ps for all simulated trajectories. Results were analyzed using RMSF and RMSD. We also analyzed SASAs of protein backbones and Rg values using VMD (Version 1.9.3) software[110,111] developed from the default script of YASARA[109].

Dynamic cross-correlation maps (DCCMs). DCCMs were calculated to explore the linked motions of all Cα atoms in equilibrated trajectories for all variants and wild-type. The Bio3D[112] package integrated into the R program was used to analyze DCCMs and provided Pearson's covariance matrix correlation coefficients, known as "cov2dccm" coefficients. C_{ij} (the cross-correlation ratio) was determined for Cα electrons[113] using the following equation,

$$C_{ij} = \frac{\langle \Delta r_i \, . \, \Delta r_j \rangle}{(\langle \Delta r_i^2 \rangle \langle \Delta r_j^2 \rangle)^{1/2}}$$

where Δr_i and Δr_j are average location s of the ith and jth atoms, respectively. C_{ij} values range from -1 to $+1$, where positive values indicate degrees of correlated motion between residues *i* and *j*, and negative values indicate degrees of anticorrelated motion.

Principal component analysis (PCA).

To understand atomic movements and protein loop dynamics, PCA was implemented by calculating and diagonalization of the atomic coordinates of eigenvectors and positional covariance matrices based on eigenvectors and equal eigenvalues, which amplify the displacements of atoms in MD simulation trajectories[114,115]. The Bio3D package[112] was used for PCA as we previously described[96].

In addition, root mean square inner products (RMSIPs) were calculated using the first three principal components to obtain similarities between two sets of modes from normal modes or principal components. Here, the range of RMSIP values was set from 0 to 1, where 0 indicates orthogonal directionality and 1 indicates identical directionalities of sample subspaces[47,116]; values were calculated using Bio3D[54]. Also, the essential dynamics program of GROMACS was used to calculate the cosine content of principal components obtained from each simulation[117].

Free energy landscape (FEL).

Gibb's free energies, a mapping system, and free energy landscape (FEL) analysis were used to identify most stable states from protein conformations; FEL explains energy distributions of protein folding throughout MD simulations[118,119]. FEL was also used to obtain protein enthalpy and entropy functions[97]. The following equation was used to determine Gibb's free energy landscapes,

$$G_i = -K_B T \ln (N_i / N_{max})$$

Here, K_B is Boltzmann's constant, and the temperature (T) was set at 300 K. N_i represents the population of bin I, and N_{max} represents the most occupied bin. Different color codes were generated to depict maximum and minimum energy levels. Based on radius of gyration and RMSD values, free energy contour maps were constructed for the most stable energy conformers.

References

1. Takahashi, K., Rochford, C. D. P. & Neumann, H. Clearance of apoptotic neurons without inflammation by microglial triggering receptor expressed on myeloid cells-2. *J. Exp. Med.* **201**, 647–657. https://doi.org/10.1084/jem.20041611 (2005).
2. Koth, L. L. *et al.* DAP12 is required for macrophage recruitment to the lung in response to cigarette smoke and chemotaxis toward CCL2. *J. Immunol.* **184**, 6522–6528. https://doi.org/10.4049/jimmunol.0901171 (2010).
3. Wang, Y. *et al.* TREM2-mediated early microglial response limits diffusion and toxicity of amyloid plaques. *J. Exp. Med.* **213**, 667–675. https://doi.org/10.1084/jem.20151948 (2016).
4. Otero, K. *et al.* TREM2 and β-catenin regulate bone homeostasis by controlling the rate of osteoclastogenesis. *J. Immunol.* **188**, 2612–2621. https://doi.org/10.4049/jimmunol.1102836 (2012).
5. Ulland, T. K. *et al.* TREM2 maintains microglial metabolic fitness in Alzheimer's disease. *Cell* **170**, 649-663.e613. https://doi.org/10.1016/j.cell.2017.07.023 (2017).
6. Andreone, B. J. *et al.* Alzheimer's-associated PLCγ2 is a signaling node required for both TREM2 function and the inflammatory response in human microglia. *Nat. Neurosci.* **23**, 927–938. https://doi.org/10.1038/s41593-020-0650-6 (2020).
7. Bouchon, A., Hernández-Munain, C., Cella, M. & Colonna, M. A DAP12-mediated pathway regulates expression of CC chemokine receptor 7 and maturation of human dendritic cells. *J. Exp. Med.* **194**, 1111–1122. https://doi.org/10.1084/jem.194.8.1111 (2001).
8. Zhong, L. *et al.* Amyloid-beta modulates microglial responses by binding to the triggering receptor expressed on myeloid cells 2 (TREM2). *Mol. Neurodegen.* **13**, 15. https://doi.org/10.1186/s13024-018-0247-7 (2018).
9. Yeh, F. L., Wang, Y., Tom, I., Gonzalez, L. C. & Sheng, M. TREM2 binds to apolipoproteins, including APOE and CLU/APOJ, and thereby facilitates uptake of amyloid-beta by microglia. *Neuron* **91**, 328–340. https://doi.org/10.1016/j.neuron.2016.06.015 (2016).
10. Wang, Y. *et al.* TREM2 lipid sensing sustains the microglial response in an Alzheimer's disease model. *Cell* **160**, 1061–1071. https://doi.org/10.1016/j.cell.2015.01.049 (2015).
11. Kawabori, M. *et al.* Triggering receptor expressed on myeloid cells 2 (TREM2) deficiency attenuates phagocytic activities of microglia and exacerbates ischemic damage in experimental stroke. *J. Neurosci.* **35**, 3384–3396. https://doi.org/10.1523/jneurosci.2620-14.2015 (2015).
12. Berner, D. K. *et al.* Meprin β cleaves TREM2 and controls its phagocytic activity on macrophages. *Faseb J.* **34**, 6675–6687. https://doi.org/10.1096/fj.201902183R (2020).
13. Kober, D. L. *et al.* Neurodegenerative disease mutations in TREM2 reveal a functional surface and distinct loss-of-function mechanisms. *Elife* **5**, https://doi.org/10.7554/eLife.20391 (2016).
14. Hall-Roberts, H. *et al.* TREM2 Alzheimer's variant R47H causes similar transcriptional dysregulation to knockout, yet only subtle functional phenotypes in human iPSC-derived macrophages. *Alzheimers Res. Ther.* **12**, 151–151. https://doi.org/10.1186/s13195-020-00709-z (2020).
15. Dash, R., Choi, H. J. & Moon, I. S. Mechanistic insights into the deleterious roles of Nasu-Hakola disease associated TREM2 variants. *Sci. Rep.* **10**, 3663. https://doi.org/10.1038/s41598-020-60561-x (2020).
16. Feuerbach, D. *et al.* ADAM17 is the main sheddase for the generation of human triggering receptor expressed in myeloid cells (hTREM2) ectodomain and cleaves TREM2 after Histidine 157. *Neurosci. Lett.* **660**, 109–114. https://doi.org/10.1016/j.neulet.2017.09.034 (2017).
17. Thornton, P. *et al.* TREM2 shedding by cleavage at the H157–S158 bond is accelerated for the Alzheimer's disease-associated H157Y variant. *EMBO Mol. Med.* **9**, 1366–1378. https://doi.org/10.15252/emmm.201707673 (2017).

18. Del-Aguila, J. L. *et al.* TREM2 brain transcript-specific studies in AD and TREM2 mutation carriers. *Mol. Neurodegen.* **14**, 18. https://doi.org/10.1186/s13024-019-0319-3 (2019).

19. Yuan, P. *et al.* TREM2 haplodeficiency in mice and humans impairs the microglia barrier function leading to decreased amyloid compaction and severe axonal dystrophy. *Neuron* **90**, 724–739. https://doi.org/10.1016/j.neuron.2016.05.003 (2016).

20. Keren-Shaul, H. *et al.* A unique microglia type associated with restricting development of Alzheimer's disease. *Cell* **169**, 1276–1290.e1217. https://doi.org/10.1016/j.cell.2017.05.018 (2017).

21. Jiang, T. *et al.* Upregulation of TREM2 ameliorates neuropathology and rescues spatial cognitive impairment in a transgenic mouse model of Alzheimer's disease. *Neuropsychopharmacology* **39**, 2949–2962. https://doi.org/10.1038/npp.2014.164 (2014).

22. Jiang, T. *et al.* TREM2 overexpression has no improvement on neuropathology and cognitive impairment in aging APPswe/PS1dE9 mice. *Mol. Neurobiol.* **54**, 855–865. https://doi.org/10.1007/s12035-016-9704-x (2017).

23. Zhong, L. *et al.* Soluble TREM2 ameliorates pathological phenotypes by modulating microglial functions in an Alzheimer's disease model. *Nat. Commun.* **10**, 1365. https://doi.org/10.1038/s41467-019-09118-9 (2019).

24. Ewers, M. *et al.* Increased soluble TREM2 in cerebrospinal fluid is associated with reduced cognitive and clinical decline in Alzheimer's disease. *Sci Transl Med* **11**, doi:https://doi.org/10.1126/scitranslmed.aav6221 (2019).

25. Guerreiro, R. *et al.* TREM2 variants in Alzheimer's disease. *N. Engl. J. Med.* **368**, 117–127. https://doi.org/10.1056/NEJMoa1211851 (2013).

26. Jonsson, T. *et al.* Variant of TREM2 associated with the risk of Alzheimer's disease. *N. Engl. J. Med.* **368**, 107–116. https://doi.org/10.1056/NEJMoa1211103 (2013).

27. Rayaprolu, S. *et al.* TREM2 in neurodegeneration: evidence for association of the p.R47H variant with frontotemporal dementia and Parkinson's disease. *Mol. Neurodegen.* **8**, 19, doi:https://doi.org/10.1186/1750-1326-8-19 (2013).

28. Borroni, B. *et al.* Heterozygous TREM2 mutations in frontotemporal dementia. *Neurobiol. Aging* **35**(934), e937–e910. https://doi.org/10.1016/j.neurobiolaging.2013.09.017 (2014).

29. Cuyvers, E. *et al.* Investigating the role of rare heterozygous TREM2 variants in Alzheimer's disease and frontotemporal dementia. *Neurobiol. Aging* **35**(726), e711-729. https://doi.org/10.1016/j.neurobiolaging.2013.09.009 (2014).

30. Jin, S. C. *et al.* Coding variants in TREM2 increase risk for Alzheimer's disease. *Hum. Mol. Genet.* **23**, 5838–5846. https://doi.org/10.1093/hmg/ddu277 (2014).

31. Sudom, A. *et al.* Molecular basis for the loss-of-function effects of the Alzheimer's disease-associated R47H variant of the immune receptor TREM2. *J. Biol. Chem.* **293**, 12634–12646. https://doi.org/10.1074/jbc.RA118.002352 (2018).

32. Dean, H. B., Roberson, E. D. & Song, Y. Neurodegenerative disease-associated variants in TREM2 destabilize the apical ligand-binding region of the immunoglobulin domain. *Front. Neurol.* **10**, 1252. https://doi.org/10.3389/fneur.2019.01252 (2019).

33. McQuade, A. *et al.* Gene expression and functional deficits underlie TREM2-knockout microglia responses in human models of Alzheimer's disease. *Nat. Commun.* **11**, 5370–5370. https://doi.org/10.1038/s41467-020-19227-5 (2020).

34. Parhizkar, S. *et al.* Loss of TREM2 function increases amyloid seeding but reduces plaque-associated ApoE. *Nat. Neurosci.* **22**, 191–204. https://doi.org/10.1038/s41593-018-0296-9 (2019).

35. Zheng, H. *et al.* Opposing roles of the triggering receptor expressed on myeloid cells 2 and triggering receptor expressed on myeloid cells-like transcript 2 in microglia activation. *Neurobiol. Aging* **42**, 132–141. https://doi.org/10.1016/j.neurobiolaging.2016.03.004 (2016).

36. Song, W. *et al.* Alzheimer's disease-associated TREM2 variants exhibit either decreased or increased ligand-dependent activation. *Alzheimers Dement* **13**, 381–387. https://doi.org/10.1016/j.jalz.2016.07.004 (2017).

37. Padhi, A. K. & Zhang, K. Y. J. Mechanistic insights into the loss-of-function mechanisms of rare human D-amino acid oxidase variants implicated in amyotrophic lateral sclerosis. *Sci. Rep.* **10**, 17146. https://doi.org/10.1038/s41598-020-74048-2 (2020).

38. Padhi, A. K. *et al.* An integrated computational pipeline for designing high-affinity nanobodies with expanded genetic codes. *Brief Bioinform.* **22**, doi:https://doi.org/10.1093/bib/bbab338 (2021).

39. Padhi, A. K., Jayaram, B. & Gomes, J. Prediction of functional loss of human angiogenin mutants associated with ALS by molecular dynamics simulations. *Sci. Rep.* **3**, 1225. https://doi.org/10.1038/srep01225 (2013).

40. Kumar, V., Pandey, P., Idrees, D., Prakash, A. & Lynn, A. M. Delineating the effect of mutations on the conformational dynamics of N-terminal domain of TDP-43. *Biophys. Chem.* **250**, 106174. https://doi.org/10.1016/j.bpc.2019.106174 (2019).

41. Pal, L. R. & Moult, J. Genetic Basis of common human disease: insight into the role of missense SNPs from genome-wide association studies. *J. Mol. Biol.* **427**, 2271–2289. https://doi.org/10.1016/j.jmb.2015.04.014 (2015).

42. Han, Y. *et al.* Genome-wide association study identifies a missense variant at APOA5 for coronary artery disease in Multi-Ethnic Cohorts from Southeast Asia. *Sci. Rep.* **7**, 17921. https://doi.org/10.1038/s41598-017-18214-z (2017).

43. Tanwar, H., Kumar, D. T., Doss, C. G. P. & Zayed, H. Bioinformatics classification of mutations in patients with Mucopolysaccharidosis IIIA. *Metab. Brain Dis.* **34**, 1577–1594. https://doi.org/10.1007/s11011-019-00465-6 (2019).

44. Arifuzzaman, M. *et al.* In silico analysis of nonsynonymous single-nucleotide polymorphisms (nsSNPs) of the SMPX gene. *Ann. Hum. Genet.* **84**, 54–71. https://doi.org/10.1111/ahg.12350 (2020).

45. Maisuradze, G. G. & Leitner, D. M. Free energy landscape of a biomolecule in dihedral principal component space: sampling convergence and correspondence between structures and minima. *Proteins* **67**, 569–578. https://doi.org/10.1002/prot.21344 (2007).

46. Maisuradze, G. G., Liwo, A. & Scheraga, H. A. Principal component analysis for protein folding dynamics. *J. Mol. Biol.* **385**, 312–329. https://doi.org/10.1016/j.jmb.2008.10.018 (2009).

47. Hess, B. Convergence of sampling in protein simulations. *Physi. Rev. E Stat. Non-linear, Soft Matter Phys.* **65**, 031910, doi:https://doi.org/10.1103/PhysRevE.65.031910 (2002).

48. Pandini, A. & Bonati, L. Conservation and specialization in PAS domain dynamics. *Protein Eng. Des. Sel.* **18**, 127–137. https://doi.org/10.1093/protein/gzi017 (2005).

49. Dash, R. *et al.* Computational insights into the deleterious impacts of missense variants on N-Acetyl-d-glucosamine kinase structure and function. *Int J Mol Sci* **22**, doi:https://doi.org/10.3390/ijms22158048 (2021).

50. Dash, R. *et al.* Computational analysis and binding site identification of type III secretion system ATPase from Pseudomonas aeruginosa. *Interdiscip. Sci.* **8**, 403–411. https://doi.org/10.1007/s12539-015-0121-z (2016).

51. Junaid, M. *et al.* Molecular simulation studies of 3,3'-diindolylmethane as a potent MicroRNA-21 antagonist. *J. Pharm. Bioallied Sci* **9**, 259–265. https://doi.org/10.4103/jpbs.JPBS_266_16 (2017).

52. Hosen, S. M. Z., Dash, R., Junaid, M., Mitra, S. & Absar, N. Identification and structural characterization of deleterious non-synonymous single nucleotide polymorphisms in the human SKP2 gene. *Comput. Biol. Chem.* **79**, 127–136. https://doi.org/10.1016/j.compbiolchem.2019.02.003 (2019).

53. Yazhini, A. & Srinivasan, N. How good are comparative models in the understanding of protein dynamics?. *Proteins* **88**, 874–888. https://doi.org/10.1002/prot.25879 (2020).

54. Skjærven, L., Yao, X. Q., Scarabelli, G. & Grant, B. J. Integrating protein structural dynamics and evolutionary analysis with Bio3D. *BMC Bioinformatics* **15**, 399. https://doi.org/10.1186/s12859-014-0399-6 (2014).

55. Xu, L. & Chen, L. Y. Molecular determinant of substrate binding and specificity of cytochrome P450 2J2. *Sci. Rep.* **10**, 22267. https://doi.org/10.1038/s41598-020-79284-0 (2020).

56. Li, H. L. *et al.* Exploring the effect of D61G mutation on SHP2 cause gain of function activity by a molecular dynamics study. *J. Biomol. Struct. Dyn.* **36**, 3856–3868. https://doi.org/10.1080/07391102.2017.1402709 (2018).

57. Kober, D. L. *et al.* Functional insights from biophysical study of TREM2 interactions with apoE and Aβ(1–42). *Alzheimers Dement* https://doi.org/10.1002/alz.12194 (2020).

58. Dash, R. *et al.* Computational SNP analysis and molecular simulation revealed the most deleterious missense variants in the NBD1 domain of human ABCA1 transporter. *Int. J. Mol. Sci.* **21**, doi:https://doi.org/10.3390/ijms21207606 (2020).

59. Ripon, M. K. H. *et al.* N-acetyl-D-glucosamine kinase binds dynein light chain roadblock 1 and promotes protein aggregate clearance. *Cell Death Dis* **11**, 619. https://doi.org/10.1038/s41419-020-02862-7 (2020).

60. Dean, H. B., Roberson, E. D. & Song, Y. Neurodegenerative disease–associated variants in TREM2 destabilize the apical ligand-binding region of the immunoglobulin domain. *Front. Neurol.* **10**, doi:https://doi.org/10.3389/fneur.2019.01252 (2019).

61. Kumar Ghosh, D., Nanaji Shrikondawar, A. & Ranjan, A. Local structural unfolding at the edge-strands of beta sheets is the molecular basis for instability and aggregation of G85R and G93A mutants of superoxide dismutase 1. *J. Biomol. Struct. Dyn.* **38**, 647–659, doi:https://doi.org/10.1080/07391102.2019.1584125 (2020).

62. Whitney, D. S., Volkman, B. F. & Prehoda, K. E. Evolution of a protein interaction domain family by tuning conformational flexibility. *J. Am. Chem. Soc.* **138**, 15150–15156. https://doi.org/10.1021/jacs.6b05954 (2016).

63. Li, R., Wang, X. & He, P. The most prevalent rare coding variants of TREM2 conferring risk of Alzheimer's disease: A systematic review and meta-analysis. *Exp. Ther. Med.* **21**, 347–347. https://doi.org/10.3892/etm.2021.9778 (2021).

64. Sirkis, D. W. *et al.* Rare TREM2 variants associated with Alzheimer's disease display reduced cell surface expression. *Acta Neuropathol. Commun.* **4**, 98–98. https://doi.org/10.1186/s40478-016-0367-7 (2016).

65. Sim, N.-L. *et al.* SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res.* **40**, W452–W457. https://doi.org/10.1093/nar/gks539 (2012).

66. Kumar, P., Henikoff, S. & Ng, P. C. Predicting the effects of coding nonsynonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.* **4**, 1073–1081. https://doi.org/10.1038/nprot.2009.86 (2009).

67. Ng, P. C. & Henikoff, S. Accounting for human polymorphisms predicted to affect protein function. *Genome Res* **12**, 436–446. https://doi.org/10.1101/gr.212802 (2002).

68. Ramensky, V., Bork, P. & Sunyaev, S. Human nonsynonymous SNPs: server and survey. *Nucleic Acids Res.* **30**, 3894–3900. https://doi.org/10.1093/nar/gkf493 (2002).

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

70. Mustafa, M. I., Murshed, N. S., Abdelmoneim, A. H. & Makhawi, A. M. In silico analysis of the functional and structural consequences of SNPs in human ARX gene associated with EIEE1. *Informat. Med. Unlocked* **21**, 100447. https://doi.org/10.1016/j.imu.2020.100447 (2020).

71. Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J. & Kircher, M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res.* **47**, D886–D894. https://doi.org/10.1093/nar/gky1016 (2019).

72. Bandaru, N., Lee, T., Zhang, P., Chen, Y. & Guo, C. In silico prediction of clinical pathogenicty by CADD scoring of exome variants found in genome of a Male belonging to the Chinese Dai Minority.

73. Rentzsch, P., Schubach, M., Shendure, J. & Kircher, M. CADD-Splice-improving genome-wide variant effect prediction using deep learning-derived splice scores. *Genome Med* **13**, 31. https://doi.org/10.1186/s13073-021-00835-9 (2021).

74. Quang, D., Chen, Y. & Xie, X. DANN: a deep learning approach for annotating the pathogenicity of genetic variants. *Bioinformatics* **31**, 761–763. https://doi.org/10.1093/bioinformatics/btu703 (2015).

75. Shihab, H. A. *et al.* Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. *Hum. Mutat.* **34**, 57–65. https://doi.org/10.1002/humu.22225 (2013).

76. Shihab, H. A. *et al.* An integrative approach to predicting the functional effects of non-coding and coding sequence variation. *Bioinformatics* **31**, 1536–1543. https://doi.org/10.1093/bioinformatics/btv009 (2015).

77. Jagadeesh, K. A. *et al.* M-CAP eliminates a majority of variants of uncertain significance in clinical exomes at high sensitivity. *Nat. Genet.* **48**, 1581–1586. https://doi.org/10.1038/ng.3703 (2016).

78. Hepp, D., Gonçalves, G. L. & Freitas, T. R. O. d. Prediction of the damage-associated nonsynonymous single nucleotide polymorphisms in the human MC1R gene. *PLOS ONE* **10**, e0121812, doi:https://doi.org/10.1371/journal.pone.0121812 (2015).

79. Khabou, B. *et al.* Comparison of in silico prediction and experimental assessment of ABCB4 variants identified in patients with biliary diseases. *Int. J. Biochem. Cell Biol.* **89**, 101–109. https://doi.org/10.1016/j.biocel.2017.05.028 (2017).

80. Dash, R. *et al.* Computational insights into the deleterious impacts of missense variants on N-acetyl-d-glucosamine kinase structure and function. *Int. J. Mol. Sci.* **22**, 8048 (2021).

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

82. Choi, Y., Sims, G. E., Murphy, S., Miller, J. R. & Chan, A. P. Predicting the functional effect of amino acid substitutions and indels. *PLoS ONE* **7**, e46688. https://doi.org/10.1371/journal.pone.0046688 (2012).

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

84. Douville, C. *et al.* Assessing the pathogenicity of insertion and deletion variants with the variant effect scoring tool (VEST-Indel). *Hum. Mutat.* **37**, 28–35. https://doi.org/10.1002/humu.22911 (2016).

85. Carter, H., Douville, C., Stenson, P. D., Cooper, D. N. & Karchin, R. Identifying Mendelian disease genes with the variant effect scoring tool. *BMC Genomics* **14**(Suppl 3), S3–S3. https://doi.org/10.1186/1471-2164-14-S3-S3 (2013).

86. Navapour, L. & Mogharrab, N. In silico screening and analysis of nonsynonymous SNPs in human CYP1A2 to assess possible associations with pathogenicity and cancer susceptibility. *Sci. Rep.* **11**, 4977. https://doi.org/10.1038/s41598-021-83696-x (2021).

87. Cheng, J., Randall, A. & Baldi, P. Prediction of protein stability changes for single-site mutations using support vector machines. *Proteins* **62**, 1125–1132. https://doi.org/10.1002/prot.20810 (2006).

88. Narayana Swamy, A., Valasala, H. & Kamma, S. In silico evaluation of nonsynonymous single nucleotide polymorphisms in the ADIPOQ gene associated with diabetes, obesity, and inflammation. *Avicenna J. Med. Biotechnol.* **7**, 121–127 (2015).

89. Yousefi, T. *et al.* In silico analysis of nonsynonymous single nucleotide polymorphism in a human KLK-2 gene associated with prostate cancer. *Meta Gene* **21**, 100578. https://doi.org/10.1016/j.mgene.2019.100578 (2019).

90. Chen, C.-W., Lin, J. & Chu, Y.-W. iStable: off-the-shelf predictor integration for predicting protein stability changes. *BMC Bioinformatics* **14**(Suppl 2), S5–S5. https://doi.org/10.1186/1471-2105-14-S2-S5 (2013).

91. Soltani, I. *et al.* Comprehensive in-silico analysis of damage associated SNPs in hOCT1 affecting Imatinib response in chronic myeloid leukemia. *Genomics* **113**, 755–766. https://doi.org/10.1016/j.ygeno.2020.10.007 (2021).

92. Capriotti, E., Calabrese, R. & Casadio, R. Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. *Bioinformatics* **22**, 2729–2734. https://doi.org/10.1093/bioinformatics/btl423 (2006).

93. Venkata Subbiah, H., Ramesh Babu, P. & Subbiah, U. In silico analysis of nonsynonymous single nucleotide polymorphisms of human DEFB1 gene. *Egpt. J. Med. Hum. Genet.* **21**, 66. https://doi.org/10.1186/s43042-020-00110-3 (2020).

94. Bromberg, Y., Yachdav, G. & Rost, B. SNAP predicts effect of mutations on protein function. *Bioinformatics* **24**, 2397–2398. https://doi.org/10.1093/bioinformatics/btn435 (2008).

95. Rozario, L. T., Sharker, T. & Nila, T. A. In silico analysis of deleterious SNPs of human MTUS1 gene and their impacts on subsequent protein structure and function. *PLoS ONE* **16**, e0252932. https://doi.org/10.1371/journal.pone.0252932 (2021).

96. Dash, R. *et al.* Structural and dynamic characterizations highlight the deleterious role of SULT1A1 R213H polymorphism in substrate binding. *Int. J. Mol. Sci.* **20**, doi:https://doi.org/10.3390/ijms20246256 (2019).
97. Dash, R. *et al.* Unveiling the structural insights into the selective inhibition of protein kinase D1. *Curr. Pharm. Des.* **25**, 1059–1074. https://doi.org/10.2174/1381612825666190527095510 (2019).
98. Dash, R., Junaid, M., Mitra, S., Arifuzzaman, M. & Hosen, S. M. Z. Structure-based identification of potent VEGFR-2 inhibitors from in vivo metabolites of a herbal ingredient. *J. Mol. Model* **25**, 98. https://doi.org/10.1007/s00894-019-3979-6 (2019).
99. Islam, M. A. *et al.* N-Acetyl-D-Glucosamine kinase interacts with NudC and lis1 in dynein motor complex and promotes cell migration. *Int. J. Mol. Sci.* **22**, doi:https://doi.org/10.3390/ijms22010129 (2020).
100. Land, H. & Humble, M. S. YASARA: a tool to obtain structural guidance in biocatalytic investigations. *Methods Mol. Biol.* **1685**, 43–67. https://doi.org/10.1007/978-1-4939-7366-8_4 (2018).
101. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **25**, 1157–1174. https://doi.org/10.1002/jcc.20035 (2004).
102. Ali, M. C. *et al.* In silico chemical profiling and identification of neuromodulators from Curcuma amada targeting acetylcholinesterase. *Netw. Model. Anal. Health Informat. Bioinformat.* **10**, 59. https://doi.org/10.1007/s13721-021-00334-2 (2021).
103. Mitra, S. & Dash, R. Structural dynamics and quantum mechanical aspects of shikonin derivatives as CREBBP bromodomain inhibitors. *J. Mol. Graph Model* **83**, 42–52. https://doi.org/10.1016/j.jmgm.2018.04.014 (2018).
104. Harrach, M. F. & Drossel, B. Structure and dynamics of TIP3P, TIP4P, and TIP5P water near smooth and atomistic walls of different hydroaffinity. *J. Chem. Phys.* **140**, 174501. https://doi.org/10.1063/1.4872239 (2014).
105. Krieger, E., Dunbrack, R. L. Jr., Hooft, R. W. & Krieger, B. Assignment of protonation states in proteins and ligands: combining pKa prediction with hydrogen bonding network optimization. *Methods Mol. Biol.* **819**, 405–421. https://doi.org/10.1007/978-1-61779-465-0_25 (2012).
106. Krieger, E., Darden, T., Nabuurs, S. B., Finkelstein, A. & Vriend, G. Making optimal use of empirical energy functions: force-field parameterization in crystal space. *Proteins* **57**, 678–683. https://doi.org/10.1002/prot.20251 (2004).
107. Essmann, U. *et al.* A smooth particle mesh Ewald method. *J. Chem. Phys.* **103**, 8577–8593. https://doi.org/10.1063/1.470117 (1995).
108. Krieger, E. & Vriend, G. New ways to boost molecular dynamics simulations. *J. Comput. Chem.* **36**, 996–1007. https://doi.org/10.1002/jcc.23899 (2015).
109. Krieger, E., Koraimann, G., Vriend, G. Increasing the precision of comparative models with YASARA NOVA—a self-parameterizing force field. *Proteins* **47**(3), 393–402 (2002).
110. Dash, R. *et al.* In silico-based vaccine design against Ebola virus glycoprotein. *Adv. Appl. Bioinform. Chem.* **10**, 11–28. https://doi.org/10.2147/aabc.S115859 (2017).
111. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph* **14**(33–38), 27–38. https://doi.org/10.1016/0263-7855(96)00018-5 (1996).
112. Grant, B. J., Rodrigues, A. P., ElSawy, K. M., McCammon, J. A. & Caves, L. S. Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics* **22**, 2695–2696. https://doi.org/10.1093/bioinformatics/btl461 (2006).
113. Ichiye, T. & Karplus, M. Collective motions in proteins: a covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations. *Proteins* **11**, 205–217. https://doi.org/10.1002/prot.340110305 (1991).
114. Salmas, R. E., Yurtsever, M. & Durdagi, S. Investigation of inhibition mechanism of chemokine receptor CCR5 by micro-second molecular dynamics simulations. *Sci. Rep.* **5**, 13180. https://doi.org/10.1038/srep13180 (2015).
115. David, C. C. & Jacobs, D. J. Principal component analysis: a method for determining the essential dynamics of proteins. *Methods Mol. Biol.* **1084**, 193–226. https://doi.org/10.1007/978-1-62703-658-0_11 (2014).
116. Hess, B. Similarities between principal components of protein dynamics and random diffusion. *Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.* **62**, 8438–8448. https://doi.org/10.1103/physreve.62.8438 (2000).
117. Merlino, A., Vitagliano, L., Ceruso, M. A. & Mazzarella, L. Subtle functional collective motions in pancreatic-like ribonucleases: from ribonuclease A to angiogenin. *Proteins* **53**, 101–110. https://doi.org/10.1002/prot.10466 (2003).
118. Frauenfelder, H., Sligar, S. G. & Wolynes, P. G. The energy landscapes and motions of proteins. *Science* **254**, 1598–1603. https://doi.org/10.1126/science.1749933 (1991).
119. Lange, O. F. & Grubmüller, H. Generalized correlation for biomolecular dynamics. *Proteins* **62**, 1053–1061. https://doi.org/10.1002/prot.20784 (2006).

Acknowledgements

Author contributions

R.D. and I.S.M. conceived the study and designed the experimental work. R.D., S.M., Y.A.M., A.C., S.I.J., and H.J.C. conducted the experiments and analyzed the data. R.D., S.M., and Y.A.M. wrote the manuscript. I.S.M. and T.J.J. contributed to study supervision, project administration, and the acquisition of funding. All authors read and agreed to the published version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-022-13120-5.

Correspondence and requests for materials should be addressed to I.S.M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.