Predicting the impact of missense mutations on an unresolved protein's stability, structure, and function: A case study of Alzheimer's disease-associated TREM2 R47H variant

Joshua Pillai [a,b] , Kijung Sung [b] , Chengbiao Wu [b,c,*]

[a] School of Biological Sciences, University of California, 9301 S Scholars Dr, La Jolla, San Diego, CA 92093, United States
[b] Department of Neurosciences, University of California San Diego, Medical Teaching Facility, 9500 Gilman Drive, La Jolla, CA 92093-0624, USA
[c] Biophotonics Technology Center, Institute of Engineering in Medicine, University of California, 9500 Gilman Dr, La Jolla, San Diego, CA 92093, USA

ARTICLE INFO

ABSTRACT

AlphaFold2 (AF2) has spurred a revolution in predicting unresolved structures of wild-type proteins with high accuracy. However, AF2 falls short of predicting the effects of missense mutations on unresolved protein structures that may be informative to efforts in personalized medicine. Over the last decade, countless in-silico methods have been developed to predict the pathogenicity of point mutations on resolved structures, but no studies have evaluated their capabilities on unresolved protein structures predicted by AF2. Herein, we investigated Alzheimer's disease (AD)-causing coding variants of the triggering receptor expressed on myeloid cells 2 (TREM2) receptor using in-silico mutagenesis techniques on the AF2-predicted structure. We first demonstrated that the predicted structure retained a high accuracy in critical regions of the extracellular domain and subsequently validated the in-silico mutagenesis methods by evaluating the effects of the strongest risk variant R47H of TREM2. After validation of the R47H variant, we predicted the molecular basis and effects on protein stability and ligand-binding affinity of the R62H and D87N variants that remain unknown in current literature. By comparing it with the R47H variant, our analysis reveals that R62H and D87N variants exert a much less pronounced effect on the structural stability of TREM2. These in-silico findings show the possibility that the R62H and D87N mutations are likely less pathogenic than the R47H AD. Lastly, we investigated the Nasu-Hakola (NHD)-causing Y38C and V126G TREM2 as a comparison and found that they imposed greater destabilization compared to AD-causing variants. We believe that the in-silico mutagenesis methods described here can be applied broadly to evaluate the ever-growing numbers of protein mutations/variants discovered in human genetics study for their potential in diseases, ultimately facilitating personalized medicine.

1. Introduction

Resolving protein structures with high accuracy has been one of the long-standing challenges in bioinformatics, as it is critical not only for mechanistic understanding of protein functions but also for novel drug design and discovery signs. With the revolution spurred by AlphaFold2 (AF2), a state-of-the-art deep learning model that can predict wild-type (WT) protein structures with high accuracy [1], this issue has been largely resolved [2,3]. Although a landmark achievement, AF2 falls short of predicting the impact of mutations on the three-dimensional (3D) structure of proteins [4–7]. Yet, accurate prediction of structural impacts by missense variant is crucial for assessing their disease-causing

potential to aid the development of personalized medicine, given that many variants are responsible for causing Mendelian diseases.

Enormous efforts over the past decade in computational techniques, significant developments have been made to decipher the pathogenicity, stability, and stereochemical effects of missense variants in resolved protein structures [8–11]. However, to the best of our knowledge, there has been no study or in-silico method that has systematically demonstrated these techniques on coding variants for unresolved protein structures. Herein, we developed a working protocol to address this issue. We investigated Alzheimer's disease (AD)-causing coding variants of the triggering receptor expressed on myeloid cells 2 (TREM2) using in-silico mutagenesis techniques on the AF2-predicted structure. TREM2

was chosen as the protein of interest in this study largely due to extensive literature on its missense variants and the high structure confidence prediction by AF2. Our primary objective was to first systematically validate the known effects of the strongest AD risk-modifying variant, R47H, and later to evaluate other rare coding variants associated of TREM2 associated with AD that are poorly understood in current literature.

TREM2 is an innate immune receptor that is upregulated in the cell surface of microglia associated with amyloid-beta (Aβ) plaques in AD. In 2013, genome-wide association studies (GWAS) identified TREM2 R47H homozygous loss-of-function mutation as a factor that increases the risk of AD [12,13]. Numerous meta-analyses have since validated the association of the R47H variant with AD [14–16]. Given the importance of the R47H variant in TREM2 biology, studies have since identified experimentally that the mutation causes reduced protein stability [17] and decreased binding-affinity to Aβ [18–20]. These studies are further validated with the reporting of the partially resolved structures from the WT and R47H TREM2 variant [21]. Sudom et al. identified that the R47H variant alters the conformation of the complementarity-determining region 2 (CDR2) loop, thus reducing critical hydrogen bonding in the C-terminal region of the CDR1 loop. The partially resolved WT and mutant structures reported by Sudom et al. [21] are essential to this study as they are a validation to the AF2 TREM2 structure and the in-silico mutagenesis techniques reported in this study.

After correctly predicting the effects of the R47H variant, we next investigated other rare coding variants. In particular, the R62H [16] and the D87N [22] variants that both have a strong association with AD-risk. However, other GWAS did not identify an association of these two variants with AD [23,24]. Nevertheless, both variants have been confirmed to decrease ligand-binding affinity [25–27]. But the effects of R62H, D87N on protein stability nor their molecular basis have yet been investigated. Here, using in-silico mutagenesis techniques, we predicted the stereochemical, stability, and functional alterations from both variants in comparison to the R47H variant. Our study has pointed to the possibility that both R62H and D87N have a less pronounced impact on the TREM2 structure and stability, likely to be less pathogenic than R47H.

Lastly, as a comparison to the AD-causing variants, we were interested in exploring variants that lead to the Nasu-Hakola disease (NHD). The Y38C variant of TREM2 was first identified by [28], where a Turkish man was observed with symptoms consistent with NHD, but the key symptom of basal ganglia calcification was not observed. However, two heterozygotes demonstrated these symptoms, consistent with NHD [28]. From biological studies, Y38C TREM2 was accumulated in the endoplasmic reticulum, suggesting evidence of misfolding [29,30]. The molecular basis of this variant remains unknown, and no association has been made with AD. Conversely, the V126G variant was reported in two unrelated patients with NHD [31,32], where its neuropathology remains unknown, and experimental studies have found that the mutation leads to poor surface expression and defective glycosylation in the Golgi apparatus [30]. However, structural studies have identified that NHD-causing variants are buried in the structure of TREM2 while AD-causing tends to be on the surface of the protein [26].

In summary, the protocol that we have developed here can accurately predict impacts of missense variants on the protein stability and structure. The method can potentially facilitate prioritization of our research efforts in understanding the consequences of ever-increasing number of missense variants discovered in human genetics.

2. Materials & methods

2.1. Acquisition of 3D structures

As of December 20, 2024, the experimental model of TREM2 has not been resolved, but the structure has been previously predicted by AF2 that is available from the Structure Database (Uniprot: Q9NZC2) with an overall predicted local distance difference test (pLDDT) score of 83.88. Our analyses primarily focused on Arg47 but also included Arg62, Asp87, Tyr38, and Val126 mutation sites. The pLDDT of Arg47 was 98.17, Arg62 was 97.06, Tyr38 was 97.50, and Val126 was 98.26, respectively, allowing for further investigation to be conducted. Additionally, it must be noted that the localized region of interest in the ectodomain, including residues 20–133, were of high confidence (pLDDT > 70), so identifications of possible molecular interactions would also retain high confidence. Given that the R47H variant is prominently reported in prior literature, the WT (PDB: 5UD7, Chain A-F) and variant (PDB: 5UD8, Chain A-B) structures have been partially resolved and were used for investigation in this study. For additional comparison, we also used the lower-resolution structure (PDB: 5ELI, Chain A-B) resolved by Kober et al., [26].

2.2. Computational tools

In order to predict the effects of missense variants with high accuracy and confidence, we exclusively used computational tools that have been rigorously peer-reviewed and independently validated by numerous studies. To comprehensively predict the stereochemical and structural effects of variants, we relied on Missense3D [33]. To determine the alterations in stability of mutations in single protein chains, we used PremPS, DDMUT, DynaMut2, MAESTRO, DeepDDG, and INPS-MD [34–39]. Lastly, to determine the functional effects of variants, we used AlphaMissense, PolyPhen-2, SNP&GO, and SIFT for prediction [40–42]; (Kumar et al., 2018). All of the previously described tools are freely accessible as on public web servers, making for straightforward validation of the findings presented in this study. Our protocol is illustrated in Fig. 1. We have provided an expanded discussion of these computational tools in the Supplementary Material 1 in the Supplementary Material. For all molecular interactions, we first determined the precise atoms using Missense3D and conducted follow-up visualization with PyMOL (Schrödinger LLC, Portland, OR 97204, USA).

2.3. Statistical analyses

All statistical analyses and data plotting was performed using GraphPad Prism (version 5.04) (GraphPad Software, La Jolla, CA, USA). With all experiments, the raw data obtained from all computational tools was stored in Microsoft Word and the processed data in Excel. For continuous datasets, a Mann-Whitney U-test was performed due to the moderate sample size across experiments. To evaluate difference from an expected value (i.e. 0 Kcal/mol for stability alterations), we used the Wilcoxon Signed-Rank test to evaluate significance. For evaluation of multiple variables, a Kruskal-Wallis test was performed and subsequently post-hoc Dunn's Multiple Comparison test. All data has been presented as the mean $\pm$ standard error of the mean (SEM). Unless stated otherwise, a *p*-value less than 0.05 was statistically significant. All data and *p*-values are indicated in the corresponding figures and tables. ns: non significance; *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; ****$p < 0.0001$.

3. Results

3.1. AlphaFold2 predictions are highly consistent to the WT and R47H structures

Before beginning experimental procedures with the AF2 to predict TREM2 structure, we sought to validate it by comparing it with the partially resolved WT and R47H structures [21]. First, we superimposed the resolved WT TREM2 (residues 21–129) to the adjacent AF structures. The root means square deviation (r.m.s.d.) between the WT structures was 0.407 ± 0.01754 Å ($n = 6$, Chain A-F), suggesting that the two structures are nearly identical. Next, we visualized the molecular
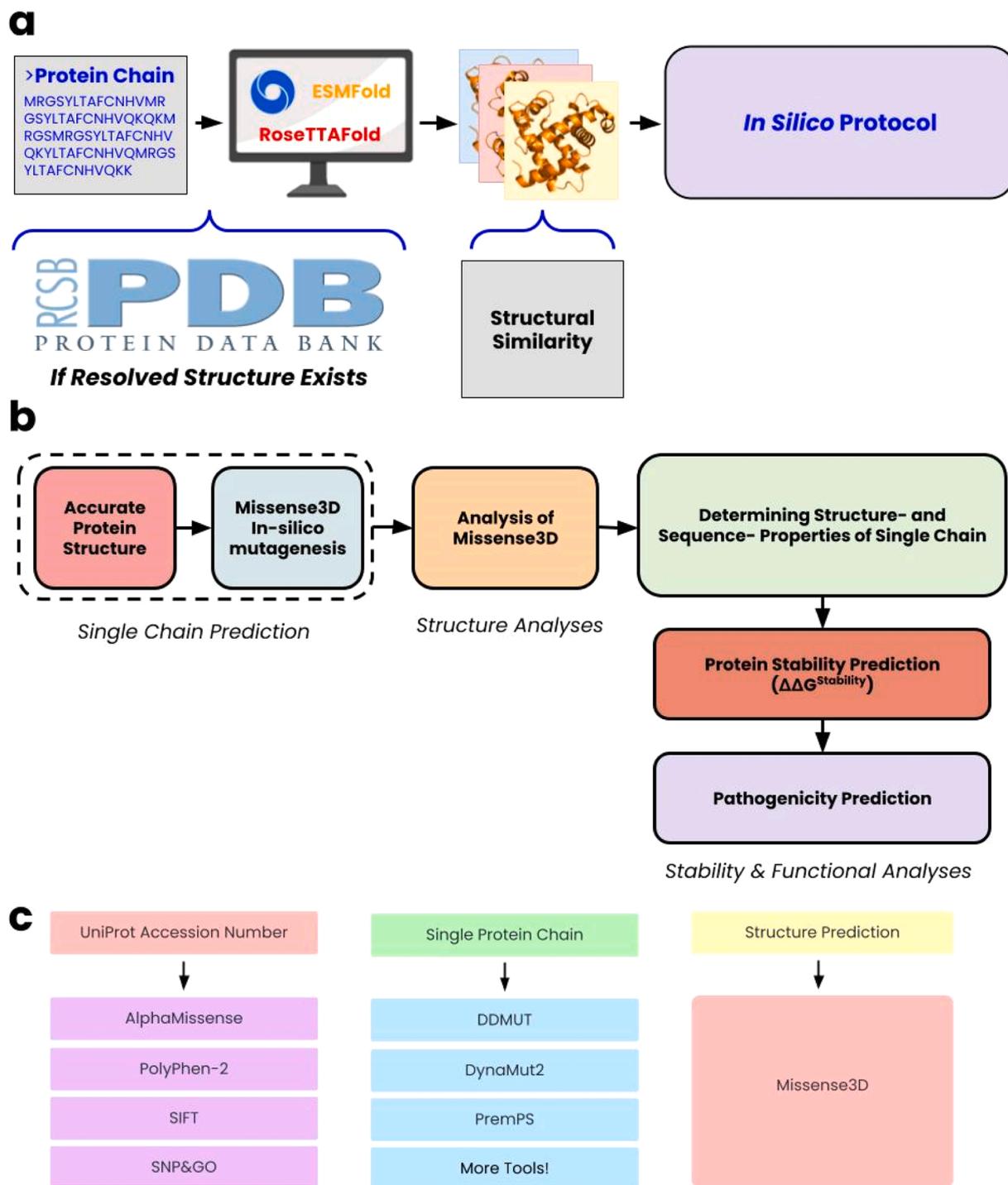
Fig. 1. Protocol for predicting effects of missense variants on unresolved proteins. **(a)** A simplified workflow of obtaining the unresolved or resolved protein structures. **(b)** The sequential protocol for predicting structural, stability, and pathogenicity predictions. **(c)** Tools used for pathogenicity, stability, and structural predictions.

interactions of Arg47 on the predicted structure (Fig. 2**A-B**), where a hydrogen bond network with the side chain of Ser65, carbonyl oxygens of Thr66 and His67, and Oδ1 Asn68 occurs. As a suitable comparison, we have also visualized the molecular interactions of Arg47 in the WT structures from Sodom et al. (2018) and Kober et al., [26], where the observed interactions were remarkably consistent to those provided in their respective report (Fig. 2**C-F**). Given the highly accurate WT prediction, we next evaluated the critical R47H variant.

First, we generated mutations on the AF-predicted structure using Missense3D to formulate the R47H structure. Upon superimposition, the

generated structure was highly similar to those reported from Sodom et al. (2018) with an r.m.s.d. of 0.464 ± 0.056 Å (*n* = 2, Chain A-B). In visualization and tabular form, His47 does not retain most of the hydrogen bond network, but observed interactions include the side chain hydroxyl of Thr66 both as an acceptor and donor, and carbonyl oxygen on Asn68 (Fig. 3**A-B**). From Sodom et al., (2018), this predicted result is consistent for Thr66, but no interactions are observed with Asn68, and instead with imidazole ring on His67 (Fig. 3**C-D**). While the AF structure does provide a strong prediction, it failed to detect the interaction of His67 because the residue undergoes a change in

a
AF2 WT TREM2

b
Missense3D AF2 WT TREM2

Wild type

Donor	Acceptor	Distance (Å)	Type
A0047-ARG NE	A0066-THR O	3.52	SM
A0047-ARG NE	A0067-HIS O	2.94	SM
A0047-ARG NH1	A0065-SER OG	3.15	SS
A0047-ARG NH1	A0066-THR O	2.95	SM
A0047-ARG NH1	A0068-ASN OD1	2.81	SS
A0047-ARG NH2	A0068-ASN OD1	3.25	SS

c
Sodom et al. WT TREM2 (Chain A)

d
Missense3D WT TREM2

Wild type

Donor	Acceptor	Distance (Å)	Type
A0047-ARG NE	A0066-THR O	3.19	SM
A0047-ARG NE	A0067-HIS O	2.92	SM
A0047-ARG NH1	A0065-SER OG	2.73	SS
A0047-ARG NH1	A0066-THR O	2.75	SM

Failed to detect ASN68 Oδl and R47 NH2 on Chains A-F.

e
Kober et al. WT TREM2 (Chain B)

f
Missense3D WT TREM2

Wild type

Donor	Acceptor	Distance (Å)	Type
B0047-ARG NE	B0066-THR O	3.70	SM
B0047-ARG NH1	B0065-SER OG	2.96	SS
B0047-ARG NH1	B0066-THR O	3.49	SM
B0047-ARG NH2	B0068-ASN O	3.60	SM

Failed to detect HIS67 Nε and R47 O on Chains A-B.

Fig. 2. Predicted molecular interactions of Arg47 on TREM2. **(a)** Position of residues of Arg47 forming the hydrogen bond network with Ser65, Thr66, His67, and Asp68 on the AF2 predicted structure (UniProt: Q9NZC2), and in **(b)** tabular form. **(c)** Partially resolved WT structure from Sodom et al. (2016), focused on Arg47 and in **(d)** tabular form, but missing the interaction Asn68 that was denoted in their report. **(e)** Partially resolved structure from Kober et al. [26], focused on Arg47 and in **(f)** tabular form. The interaction with His67 is not noted in this structure.

conformation by 180° towards His47, allowing for π–π stacking interactions. In-silico tools are currently unable to formulate this conformational change because the alteration is made rigidly and not dynamic in nature. While this remains a current limitation in the field, we still believe that the overall prediction is reasonably close to that of the resolved structure and could be further compared in stability and stereochemical effects.

Using the generated R47H structure, we then evaluated the accuracy of the computational tools in predicting the known stability, structural, and functional effects of the R47H variant. Firstly, we calculated the Gibbs free energy of stability ($\Delta\Delta G^{stability}$), where it was estimated that the variant would destabilize the structure by -0.7934 ± 0.06429 Kcal/mol significantly different from no stability change ($p = 0.0312$)

(Fig. 3**E**). To confirm this data, we performed structural analysis using Missense3D, where it was identified that no significant structural damages were caused from the mutation except minor alterations. Specifically, there were increases in local clash scores ($+2.67$) and relative solvent accessibility (RSA) ($+1.3$ %) from WT to mutant, but not significant enough to detect structural damage despite breakages in the hydrogen bond network. Lastly, for studying the functional effects of R47H, most tools predicted the mutation was benign, with Alpha-Missense predicting a score of 0.187 (benign), SIFT predicting a score of 0.06 (benign), SNP&GO predicting a score of 4 (benign), but only PolyPhen-2 predicting 1 (pathogenic) (Table 1). The functional results of this variant will further be compared and discussed with others in Section 4. Overall, the findings of the stability analyses are consistent with
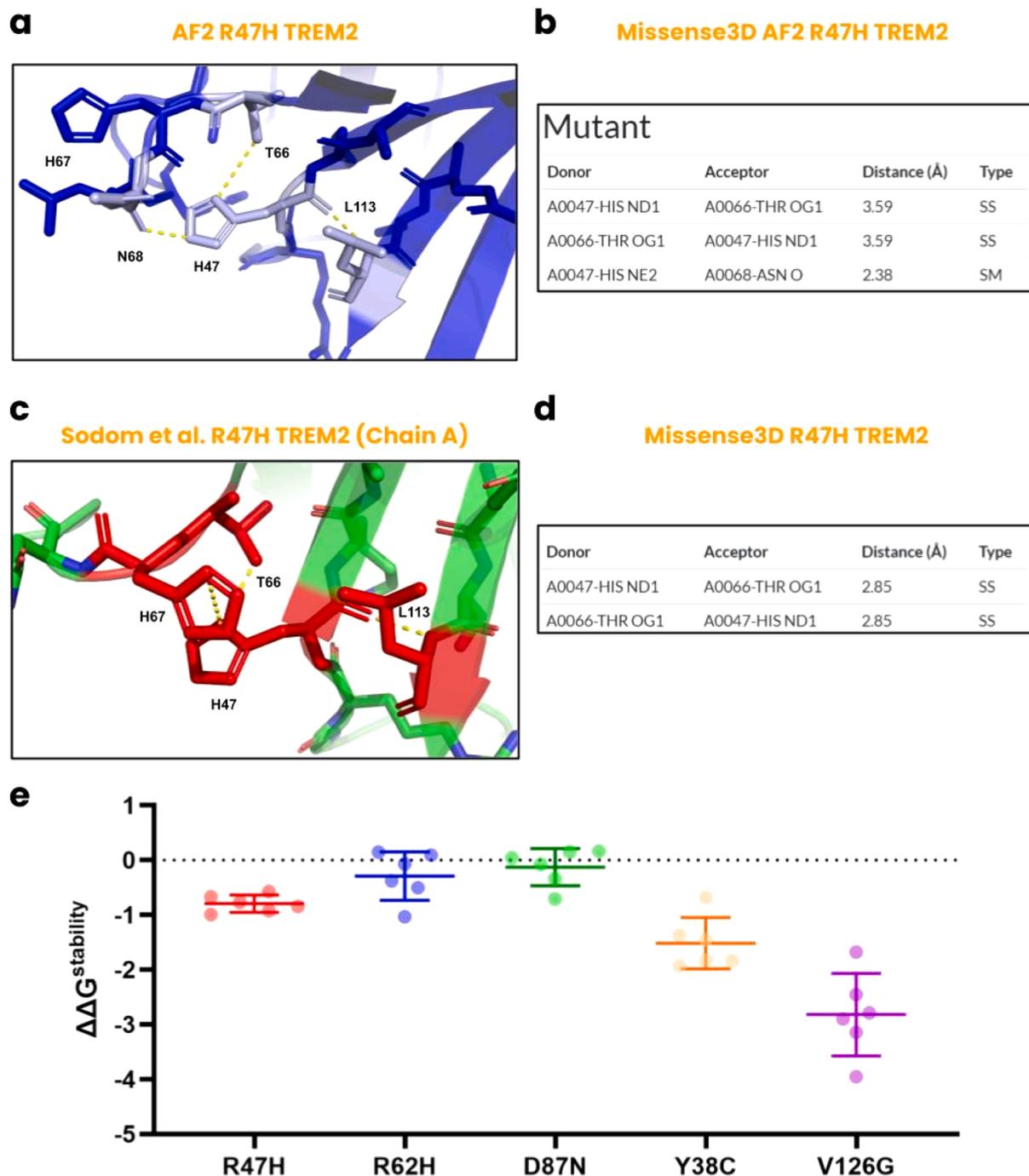
Fig. 3. Predicted molecular interactions of the R47H variant on TREM2. **(a)** Position of residues after mutagenesis of Arg47, where an interaction with Thr66 and Asn68 is observed. **(b)** The interaction is also presented in tabular form. **(c)** The resolved R47H structure (PDB: 5ud8) displays the same interaction with Thr66, but His67 has altered conformations, and there is no interaction that occurs with Asn68. **(d)** This is also displayed in tabular form. **(e)** The change in protein stability of all variants evaluated in this study has been displayed as boxplots ($n = 6$). All data has been presented as the mean ± SEM.

prior experimental results by [17] that the mutation imparts lower stability. Our stereochemical and functional results also suggest that while there is increased exposure of the residue to the surrounding solvent that could potentially alter ligand-binding affinity, the changes induced by the mutation was not significant enough to cause complete loss-of-function of TREM2. These conclusions are consistent with prior literature [26].

3.2. Predicted molecular basis and effects of R62H and D87N structures

Using the validated AF2-predicted WT structure, we performed the same experimental procedures used to formulate the R47H variant for the R62H variant. Using Missense3D, the R62H structure was generated, and follow-up stability, structure, and functional analyses were performed. From the AF-structure, Arg62 is predicted to form a hydrogen bond with the side chain carbonyl oxygen on Gln111, and the mutation results in His62 forming a bond with a hydroxyl on Ser65 (Fig. 4**A-B**). As an additional confirmation of this prediction, mutagenesis was also performed on the WT structures provided by Sudom et al., [21] and Kober et al., [26]. From the WT structure provided by Sodom et al., there is no hydrogen bonding that occurs on the WT structure, but the bond with Ser65 is observed after mutagenesis (Fig. 4**C**). However, no

Table 1

Predicted pathogenicity of AD- and NHD-causing TREM2 variants. For the NHD-causing variants, almost all metrics except for SNP&GO predicted that the variants are pathogenic. In contrast, most tools suggested that the AD-causing variants are benign except for PolyPhen-2. The location of these mutations are the most critical factor in determining pathogenicity, whether being on the surface or buried in the structure of TREM2. * This prediction was not available on the AlphaFold Structure Database and was instead obtained from the Hedge Group [43]. ** Predicted CADD GRCh38-v1.7 for Y38C (Chr6:41161541 A>G), R47H (Chr6:41161514 G>A), R62H (Chr6:41161469 G>A), D87N (Chr6:41161395 G>A), and V126G (Chr6:41161277 T > G) calculated from the CADD web-server (https://cadd.gs.washington.edu/). R47H and R62H are risk modifiers, while D87N is a potential risk modifier for AD. Y38C and V126G are pathogenic for NHD.

MT	PolyPhen-2	AlphaMissense	SIFT	SNP&GO	CADD**	Disease
Y38C	1 (Pathogenic)	0.9 (Pathogenic)*	0.01 (Pathogenic)	2 (Benign)	26.8	NHD
R47H	1 (Pathogenic)	0.187 (Benign)	0.06 (Benign)	4 (Benign)	22.5	AD
R62H	0.084 (Benign)	0.112 (Benign)	0.1 (Benign)	7 (Benign)	16.24	AD
D87N	1 (Pathogenic)	0.129 (Benign)	0.06 (Benign)	7 (Benign)	22.8	AD
V126G	1 (Pathogenic)	0.683 (Pathogenic)	0.01 (Pathogenic)	4 (Benign)	27.3	NHD
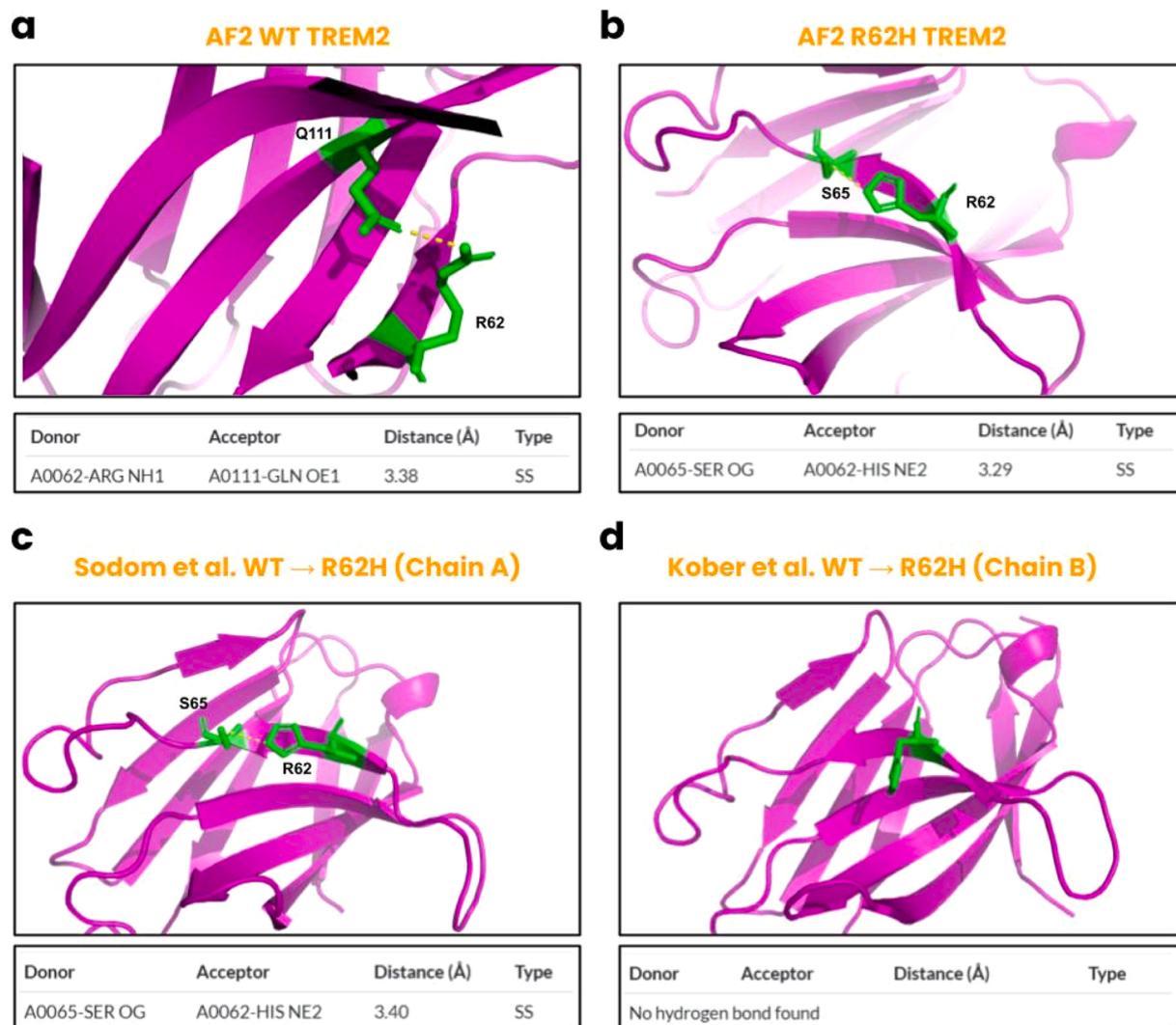
Fig. 4. Predicted molecular interactions of WT Arg62 and R62H variants. **(a)** Position of Arg62 with hydrogen bonding with Gln111 on the AF2-predicted WT structure, and shown in tabular form. **(b)** His62 forms an hydrogen bonding interaction with Ser65 after mutagenesis of the AF2- predicted structure. **(c)** This interaction was confirmed on the structure produced from mutagenesis of the WT structure from Sodom et al. (2018). **(d)** However, this was not observed on the lower-resolution WT structure form Kober et al., [26].

hydrogen bonds were identified in either the WT or mutant structures for Kober et al., [26] (Fig. 4D). A stability analysis was then performed on the AF2 structure, where unfortunately the computational tools were not unanimous, estimating -0.2908 ± 0.1813 Kcal/mol ($n = 6$) not significant from no stability change ($p = 0.3125$) (Fig. 3E). From further structural analysis of the AF-structure, no significant structural damage was detected except small alterations. The local clash score ($+2.42$) and

RSA ($+6.9$ %) both increased. We then performed a test evaluating the functional effects of R62H and found that most algorithms predicted that the variant was benign. AlphaMissense predicted a score of 0.112 (benign), SIFT predicted a score of 0.01 (benign), SNP&GO predicted a score of 7 (benign), but only PolyPhen-2 predicted 0.084 (benign) in Table 1. Overall, with the in-silico orthogonal data of R62H, it appears that the mutation does increase steric hindrance while decreasing local

hydrophobic interactions in the localized structure, and could be impacting the binding affinity of ligands onto this or surrounding residues, leading to the pathology of AD. The molecular interactions of this residue also appear to be minimal in comparison to Arg47 that contained a strong hydrogen bond network, indicating a smaller effect size in identifying a change in protein stability.

Next, we evaluated the D87N variant from the predicted WT structure using the same experimental procedure used for prior variants. From a structural analysis of the WT AF structure, there are hydrogen bond networks consisting of a hydroxyl Tyr38 and carbonyl oxygen

a AF2 D87N TREM2

b Missense3D

Wild type

Donor	Acceptor	Distance (Å)	Type
A0038-TYR OH	A0087-ASP O	3.23	SM
A0089-LEU N	A0087-ASP OD1	3.13	MS
A0090-GLY N	A0087-ASP OD1	2.97	MS

Mutant

Donor	Acceptor	Distance (Å)	Type
A0038-TYR OH	A0087-ASN O	3.23	SM
A0089-LEU N	A0087-ASN OD1	3.07	MS
A0090-GLY N	A0087-ASN OD1	2.71	MS

c Sodom et al. WT → D87N (Chain A)

d

Wild type

Donor	Acceptor	Distance (Å)	Type
A0087-ASP N	A0086-ASP OD1	2.81	MS
A0089-LEU N	A0087-ASP OD1	3.23	MS
A0090-GLY N	A0087-ASP OD1	2.89	MS

Mutant

Donor	Acceptor	Distance (Å)	Type
A0087-ASN N	A0086-ASP OD1	2.81	MS
A0089-LEU N	A0087-ASN OD1	3.26	MS
A0090-GLY N	A0087-ASN OD1	2.81	MS

e Kober et al. WT → D87N (Chain B)

f

Wild type

Donor	Acceptor	Distance (Å)	Type
B0038-TYR OH	B0087-ASP O	3.06	SM
B0087-ASP N	B0086-ASP OD1	2.96	MS
B0089-LEU N	B0087-ASP OD1	3.59	MS
B0090-GLY N	B0087-ASP OD1	3.49	MS
B0092-THR OG1	B0087-ASP OD2	3.86	SS

Mutant

Donor	Acceptor	Distance (Å)	Type
B0038-TYR OH	B0087-ASN O	3.06	SM
B0087-ASN N	B0086-ASP OD1	2.96	MS
B0089-LEU N	B0087-ASN OD1	3.34	MS
B0090-GLY N	B0087-ASN OD1	3.03	MS

Fig. 5. Predicted molecular interactions of the D87N variant. **(a)** Asp87 is expected to form hydrogen bonds with Tyr38, Leu89, and Gly90, and **(b)** these interactions are retained after mutagenesis is performed. **(c)** Almost all of the interactions are observed on the altered structure from Sodom et al. (2018) except for excluding Tyr38 and instead Asp86, and **(d)** shown in tabular form. **(e)** The same structural components are also observed in the altered structure by Sodom et al. (2016), but including Tyr38, **(f)** displayed in tabular form.

Asp87, amine Leu89 and Oδ1Asp87, and amine Gly90 and Oδ1 Asp87 (Fig. 5A-B). There are no alterations in hydrogen bonding between residues after mutagenesis, although there is a decrease in distance for Leu89 and Gly90. As a confirmation of these interactions, mutagenesis was performed on the resolved structure by Sodom et al. (2018), where the interactions with Leu89 and Gly90 were observed, but Tyr38 was absent, and a new interaction was observed with the amine Asn86 (Fig. 5C-D). Likewise, with the structure by Kober et al., [26], the interactions with Leu89 and Gly90 were present, but Tyr38 and Asn86 were observed (Fig. 5E-F). We then performed a stability analysis, where the findings were inconclusive, estimating -0.1278 ± 0.1384 Kcal/mol and not significant from no stability change ($p = 0.6875$) (Fig. 3E). From structural analysis with Missense3D on the AF-structure, no damage was detected other than small alterations. The clash score decreased (-0.03) and the RSA increased (+5.85 %). For functional effects, just like R47H and R62H, a majority of predictions were benign. AlphaMissense predicted a score of 0.129 (benign), SIFT predicted a score of 0.06 (benign),

SNP&GO predicted a score of 7 (benign), but only PolyPhen-2 predicted 1 (pathogenic) in Table 1. Overall, from the orthogonal analyses, it appears that the majority of interactions at Asn87 are retained, demonstrating a smaller difference in stability. Similarly to the R62H variant, there is an increase in steric hindrance observed but likely retaining the original hydrophobic interactions, impacting ligand-binding affinity, and leading to AD-pathology.

3.3. Predicted molecular basis of NHD-causing Y38C and V126G structures

Finally, we predicted the basis of the NHD-causing Y38C and V126G TREM2 variants as a comparison to the AD-causing variants. Using the prior experimental procedures, the Y38C variant was created, and a structural analysis was performed. On the WT AF-structure, two hydrogen interactions were observed between the hydroxyl of Tyr38 and Oδ1 Asp86 and carbonyl oxygen on Asp87 (Fig. 6A). This was then

a AF2 WT TREM2

Donor	Acceptor	Distance (Å)	Type
A0038-TYR OH	A0086-ASP OD1	2.68	SS
A0038-TYR OH	A0087-ASP O	3.23	SM

b AF2 Y38C TREM2

Donor	Acceptor	Distance (Å)	Type
A0038-CYS SG	A0091-GLY O	3.51	SM

c Sodom et al. WT → Y38C (Chain A)

Donor	Acceptor	Distance (Å)	Type
A0048-LYS NZ	A0038-CYS SG	3.13	SS

d Kober et al. WT → Y38C (Chain B)

Donor	Acceptor	Distance (Å)	Type
B0048-LYS NZ	B0038-CYS SG	2.94	SS
B0038-CYS SG	B0091-GLY O	3.72	SM

Fig. 6. Predicted molecular interactions of the Y38C variant. **(a)** Position of residues Tyr38 on the AF2-predicted WT structure, where interactions are observed with Asp86 and Asp87. **(b)** However, after mutagenesis, there appears to be an interaction with Gly91 on the structure after mutagenesis. **(c)** This hydrogen bond does not appear on the structure by Sodom et al. (2018), and instead has an interaction with Lys48. **(d)** Yet, with the altered structure by Kober et al., [26], both residues Gly91 and Lys48 appear to interact.

altered to a single interaction with the carbonyl oxygen on Gly91 (Fig. 6**B**). We also performed the same procedure on the WT structure of Sodom et al. (2018), where the interaction with Oδ1 Asp86 was only present in the WT and altered to interact with the terminal amine on Lys48 (Fig. 6**C**). On the structure of Kober et al., [26], the same WT interaction observed in the AF-structure was observed, and after mutagenesis was with Gly91 and Lys48 (Fig. 6**D**). From a stability analysis, the variant was significantly destabilized at -1.516 ± 0.1903 Kcal/mol that was significant from no change ($p = 0.0312$). A Kruskal-Wallis test was then conducted, where a difference was observed between the 5 mutants ($p = 0.0001$), and post-hoc revealed that the Y38C was more destabilized than R47H although not significant ($p > 0.9999$). From further structural analysis with Missense3D on the AF-structure, there were increases in local clash scores ($+2.46$) and RSA ($+8.8$ %), and an expansion of cavity volume by 45.144 Å^3. For functional results, AlphaMissense predicted 0.9 (pathogenic), SIFT predicted a score of 0.01 (pathogenic), PolyPhen-2 predicted 1 (pathogenic), but SNP&GO predicted a score of 2 (benign) in Table 1.

Lastly, the V126G variant was determined using the prior protocol. After mutagenesis of the WT AF-structure, there were no hydrogen bonds identified in either structure, and similar results with the respective structures by Sodom et al. (2018) and Kober et al., [26] (Fig. 7**A-C**). Upon structural analysis of the AF-structure, Missense3D identified significant structural damage. The local clash scores decreased (-0.7) and RSA increased ($+9.3$), and most importantly the residue was altered from buried to exposed. From a stability analysis, there was significant destabilization estimated at -2.818 ± 0.3074 Kcal/mol ($p = 0.0312$). After a one-way ANOVA, post-hoc revealed that the destabilization was significantly greater compared to D87N ($p = 0.0004$) and R62H ($p = 0.0027$), and greater for R47H and Y38C although not significant. For functional results, AlphaMissense predicted

0.683 (pathogenic), SIFT predicted a score of 0.01 (pathogenic), PolyPhen-2 predicted 1 (pathogenic), but SNP&GO predicted a score of 4 (benign) in Table 1. Overall, the NHD-causing missense variants do cause larger effects on stability of TREM2 compared to AD-causing variants. This is a strong confirmation of prior literature, especially from Kober et al. [26] that NHD-causing variants are buried while AD-causing are located on the protein surface and therefore impose a larger effect on the overall stability.

4. Discussion

In this short communication, we sought to use the AF predicted structure of TREM2 to accurately predict alterations in the global structure of this critical receptor from AD-associated missense mutations. After validating that the predicted structure was nearly identical to partially resolved critical regions of TREM2, we used state-of-the-art computational tools to analyze the stability, structural, and functional effects of the prominent R47H variant and compared our results with experimental evidence from prior studies as a validation of these tools. From our analyses, we correctly identified that the protein would destabilize [17] and identified almost all breakages in hydrogen bonding from WT to mutant. These experiments essentially proved the validity of the computational methods used in this study and provided grounds to evaluate the R62H and D87N variants with confidence.

For the R62H mutation, our data for protein stability was largely inconclusive because the stereochemical effects were not as abundantly clear as those from the R47H variant. A single hydrogen bond alteration would be much more challenging to confidently determine effects compared to multiple interactions and displays that current computational tools still can be improved in their precision. Likewise, the same conclusions can be made about the D87N variant that still retained its
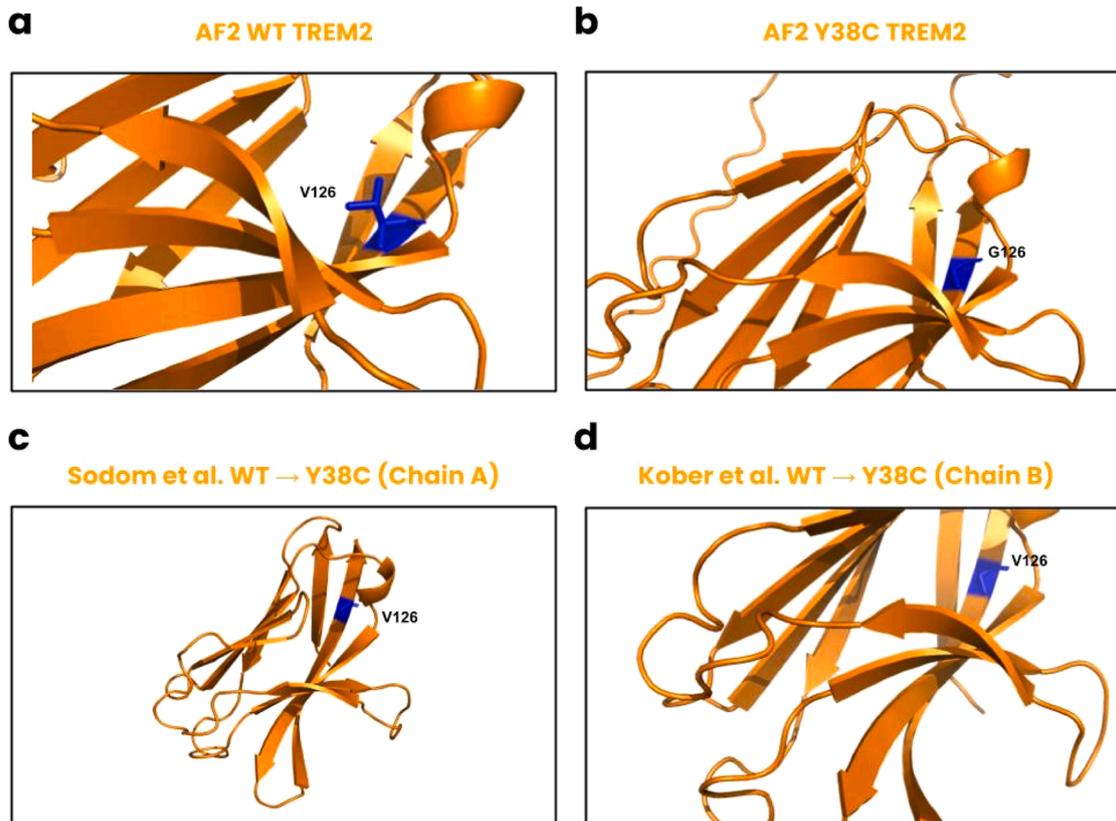
Fig. 7. Structural components of V126G TREM2 variant. **(a)** Position of Val126 with no hydrogen bonding interactions with other residues but is buried in the AF2-predicted WT structure. **(b)** Mutagenesis is performed on the structure, where the residue appears to now be exposed. **(c)** Similar patterns are also observed for the altered structures from Sodom et al. (2018) and **(d)** Kober et al., [26].

three original hydrogen bonding. From our data, the D87N and R62H variant caused significantly less stability and structural alterations to TREM2, and its regulation of ligand-binding compared to those of the R47H variant.

Additionally, to contextualize our findings of the AD-causing variants, we conducted our computational method on variants that lead to Nasu-Hakola disease (NHD). The primary difference in variants that lead to NHD is that they are buried deep in the fold of TREM2, where they largely impact the folding and stability of the protein [26]. Within this study, we performed supplemental analyses of the variants Y38C and V126G, where they were both found to cause significantly more destabilization of TREM2 compared to the AD-causing variants. From structural analyses, the Y38C was not detected to cause severe structural damages, but there were many more alterations compared to the AD-causing variants, such as increased cavity volume. Likewise, for the V126G variant, structural damage was detected, where the substitution causes the residue to become significantly more exposed. Lastly, from functional analyses, most tools predicted that the variants are pathogenic (Table 1). All of these findings are expected because of the location of the mutation site: AD-causing variants are on the surface while NHD-causing are buried.

Given the functional predictions for all variants, we believe that the AD-causing variants were expected to be benign because these variants do not lead to complete loss-of-function of TREM2 compared to NHD. Each computational tool differs in their approach to predicting functional effects and has its own strengths and weaknesses in its prediction. For instance, AlphaMissense has been found to perform exceptionally well in predicting protein function in vitro given its training data derived from protein structures, sequences, and other biological contexts. The AD-causing variants lie on the surface of TREM2, impairing regulation of ligand-binding, but not entirely impacting the function of the protein, compared to the NHD-causing variants that result in complete loss-of-function. Multiple studies have indirectly confirmed that the R47H variant leads to partial functional loss of TREM2 [44,45], and our study formally validates these conclusions.

Overall, this study has demonstrated that the AF predicted structure of TREM2 was highly similar to previously resolved regions, allowing for us to subsequently validate the reliability of in-silico techniques in predicting AD-causing variants through the well-known R47H mutation. We provide molecular details of the stereochemical effects of the R62H and D87N variants, which have not been reported previously to the best of our knowledge. The R47H variant caused greater stability, structural, and functional effects than either variant. And for the first time, we have broadly validated the conclusions of prior studies for the effects of AD-causing TREM2 variants. Future studies should focus on determining the molecular details of these variants individually as well as their impacts in critical regulations of ligand-binding. Additionally, we believe that validated AF predicted structure of TREM2 may be of particular use in evaluating global alterations of this critical receptor. Future studies on many other targets will be needed to further validate the utility of our methods. The significant implication of our in-silico mutagenesis methods described here is that the protocol can be applied broadly to evaluate the large number of mutations discovered in human genetics study to guide our efforts in research and treatment development.

5. Limitations

There are multiple limitations of this study that must be noted in context of the in-silico approaches used in this study. First, while we employed stringent state-of-the-art tools to obtain our preliminary structural results, it is important to note that the structures of all variants are simply informed predictions of unresolved structures. We solely relied on known experimental data of TREM2 from the prior decade of research to support the reported findings. The research reported in this work is meant to serve as a baseline for future studies attempting to identify the mechanisms of lesser-known TREM2 variants that lead to

AD or NHD, especially R62H and D87N. With the R47H variant, we were able to identify all molecular interactions of the WT structure and almost all of those of the mutant except for His67 that had an altered conformation. This is a major limitation of current in-silico mutagenesis studies that must be addressed. Our methods used in this study relied on the residues remaining in a rigid conformation and were not dynamic in nature. While rigid-structure predictions still provide reasonable predictions, they currently lack high precision and can have slight inaccuracies. Though, we strongly believe that our protocol can help guide experimental efforts and identify the key molecular mechanisms of these variants that remain unknown.

Another limitation that must be discussed is the importance of testing multiple structure predictions before evaluating missense variants. For instance, in this study, we originally considered using the AlphaFold 3 (AF3) model for providing a WT prediction of TREM2 [46] but found that it failed to identify the hydrogen bond network that was broken from the R47H variant. Additionally, the top predictions from AF3 had a higher r.m.s.d. value and therefore were not considered in this study. When making predictions of unresolved structures, it is valuable to have multiple algorithms for comparison along with supporting structure data. AF2 in the case of this study provided the strongest WT prediction, and subsequently the best candidate for mutagenesis studies. In the future, we hope to expand our protocol and to make it more simplified for broad usage.

There are also a few smaller limitations that must be stated in this study. Firstly, there are localized regions that are disordered within the AF2 structure, but since the critical regions of the ectodomain were of high confidence, we still relied on the entire global structure. Additionally, as per the benchmarks provided by AF, the structure was still of high confidence. Until the complete structure of TREM2 has been resolved, structure predictions currently remain the only current approach for estimating global structural alterations from missense variants. Additionally, we had hoped to utilize a greater sample size of computational tools in gaining mechanistic insights into the variants since there is low power in this study. For variants that cause significant alterations, as demonstrated with R47H, Y38C, and V126G, it is easier to decipher their trends compared to more subtle variants like R62H and D87N. This is due to the low power in all statistical tests and the lack of data normality. While we have focused on using stringent statistical analyses in this study, we encourage future studies deciphering the mechanisms of TREM2 variants to be attentive to the overall trends in the observed data. Ultimately, while there are many limitations in this study, they largely represent the current challenges in computational techniques available. Overall, the use of protein structure predictions, in-silico mutagenesis techniques, and experimental efforts in our method provide for a reasonable baseline prediction of missense variants for unknown protein structures.

Author Contributions

J.P. and C.W. jointly conceived the study. All authors planned the entire experimental procedure presented. J.P. performed all experiments. C.W. and K.S. analyzed the source data. All authors wrote the manuscript.

CRediT authorship contribution statement

Wu Chengbiao: Writing – review & editing, Writing – original draft, Project administration, Conceptualization. **Sung Kijung:** Writing – original draft, Supervision, Project administration, Investigation, Conceptualization. **Pillai Joshua:** Writing – review & editing, Writing – original draft, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Declaration of Competing Interest

Acknowledgements

Supplementary Material

Supplementary Material 1 - Additional materials and sources on the protocol used in this study.

Supplementary Material 2 - Source data from outputs given by Missense3D of all structures used or generated in this study.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.csbj.2025.01.024.

Data Availability

All source data files are provided at https://github.com/Joshua-Pillai/AlphaFold2_TREM2.

References

[1] Jumper J, et al. Highly accurate protein structure prediction with AlphaFold. Nature 2021;596:583–9.
[2] Perrakis A, Sixma TK. AI revolutions in biology. EMBO Rep 2021;22.
[3] Zonta F, Pantano S. From sequence to mechanobiology? Promises and challenges for AlphaFold 3. Mechanobiol Med 2024;2:100083.
[4] Buel GR, Walters KJ. Can AlphaFold2 predict the impact of missense mutations on structure? Nat Struct Mol Biol 2022;29:1–2.
[5] David A, Sternberg MJE. Protein structure-based evaluation of missense variants: resources, challenges and future directions. Curr Opin Struct Biol 2023;80:102600.
[6] Karakoyun HK, et al. Evaluation of AlphaFold structure-based protein stability prediction on missense variations in cancer. Front Genet 2023;14.
[7] Schmidt A, et al. Predicting the pathogenicity of missense variants using features derived from AlphaFold2. Bioinformatics 2023;39.
[8] De Oliveira Garcia FA, De Andrade ES, Palmero EI. Insights on variant analysis in silico tools for pathogenicity prediction. Front Genet 2022;13.
[9] Yazar M, Özbek P. In silico tools and approaches for the prediction of functional and structural effects of single-nucleotide polymorphisms on proteins: an expert review. OMICS a J Integr Biol 2020;25:23–37.
[10] Chen J, Gu Z, Lai L, Pei J. In silico protein function prediction: the rise of machine learning-based approaches. Med Rev 2023;3:487–510.
[11] Sanavia T, et al. Limitations and challenges in protein stability prediction upon genome variations: towards future applications in precision medicine. Comput Struct Biotechnol J 2020;18:1968–79.
[12] Guerreiro R, et al. TREM2 variants in Alzheimer's disease. N Engl J Med 2012;368:117–27.
[13] Jonsson T, et al. Variant of TREM2 associated with the risk of Alzheimer's disease. N Engl J Med 2012;368:107–16.
[14] Benitez BA, et al. TREM2 is associated with the risk of Alzheimer's disease in Spanish population. Neurobiol Aging 2013;34:1711.e15–7.
[15] Ghani M, et al. Mutation analysis of the MS4A and TREM gene clusters in a case-control Alzheimer's disease data set. Neurobiol Aging 2016;42:217.e7–217.e13.
[16] Sims R, et al. Rare coding variants in PLCG2, ABI3, and TREM2 implicate microglial-mediated innate immunity in Alzheimer's disease. Nat Genet 2017;49:1373–84.
[17] Park J-S, Ji IJ, Kim D-H, An HJ, Yoon S-Y. The Alzheimer's Disease-Associated R47H variant of TREM2 has an altered glycosylation pattern and protein stability. Front Neurosci 2017;10.
[18] Zhao Y, et al. TREM2 Is a Receptor for β-Amyloid that Mediates Microglial Function. Neuron 2018;97:1023–1031.e7.
[19] Wang Y, et al. TREM2 lipid sensing sustains the microglial response in an Alzheimer's disease model. Cell 2015;160:1061–71.
[20] 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 2016;90:724–39.
[21] 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 2018;293:12634–46.
[22] Song W, et al. Alzheimer's disease-associated TREM2 variants exhibit either decreased or increased ligand-dependent activation. Alzheimer S Dement 2016;13:381–7.
[23] Borroni B, et al. Heterozygous TREM2 mutations in frontotemporal dementia. Neurobiol Aging 2013;35:934.e7–934.e10.
[24] Cuyvers E, et al. Investigating the role of rare heterozygous TREM2 variants in Alzheimer's disease and frontotemporal dementia. Neurobiol Aging 2013;35:726.e11–9.
[25] Ulrich JD, Ulland TK, Colonna M, Holtzman DM. Elucidating the role of TREM2 in Alzheimer's disease. Neuron 2017;94:237–48.
[26] Kober DL, et al. Neurodegenerative disease mutations in TREM2 reveal a functional surface and distinct loss-of-function mechanisms. eLife 2016;5.
[27] Yeh FL, Wang Y, Tom I, Gonzalez LC, Sheng M. TREM2 binds to apolipoproteins, including APOE and CLU/APOJ, and thereby facilitates uptake of Amyloid-Beta by Microglia. Neuron 2016;91:328–40.
[28] Guerreiro RJ, Lohmann E, Brás JM, Gibbs JR, Rohrer JD, Gurunlian N, Dursun B, Bilgic B, Hanagasi H, Gurvit H, Emre M, Singleton A, Hardy J. Using exome sequencing to reveal mutations in TREM2 presenting as a frontotemporal dementia–like syndrome without bone involvement. JAMA Neurol 2013;70(1):78. https://doi.org/10.1001/jamaneurol.2013.579.
[29] Park J, Ji IJ, An HJ, Kang M, Kang S, Kim D, Yoon S. Disease-Associated mutations of TREM2 alter the processing of N-Linked oligosaccharides in the golgi apparatus. Traffic 2015;16(5):510–8. https://doi.org/10.1111/tra.12264.
[30] Sirkis DW, Aparicio RE, Schekman R. Neurodegeneration-associated mutant TREM2 proteins abortively cycle between the ER and ER–Golgi intermediate compartment. Mol Biol Cell 2017;28(20):2723–33. https://doi.org/10.1091/mbc.e17-06-0423.
[31] Klünemann HH, Ridha BH, Magy L, Wherrett JR, Hemelsoet DM, Keen RW, De Bleecker JL, Rossor MN, Marienhagen J, Klein HE, Peltonen L, Paloneva J. The genetic causes of basal ganglia calcification, dementia, and bone cysts. Neurology 2005;64(9):1502–7. https://doi.org/10.1212/01.wnl.0000160304.00003.ca.
[32] Paloneva J, Autti T, Hakola P, Haltia MJ. Polycystic Lipomembranous Osteodysplasia with Sclerosing Leukoencephalopathy (PLOSL). In: Adam MP, Ardinger HH, Pagon RA, Wallace SE, Bean LJH, Mefford HC, Stephens K, Amemiya A, Ledbetter N, editors. SourceGeneReviews® [Internet]. Seattle (WA): University of Washington, Seattle; 2002. p. 24. 1993-2017.
[33] Ittisoponpisan S, et al. Can predicted protein 3D structures provide reliable insights into whether missense variants are disease associated? J Mol Biol 2019;431:2197–212.
[34] Chen Y, et al. PremPS: Predicting the impact of missense mutations on protein stability. PLoS Comput Biol 2020;16:e1008543.
[35] Zhou Y, Pan Q, Rodrigues DEV, Ascher DB. DDMut: predicting effects of mutations on protein stability using deep learning. Nucleic Acids Res 2023;51:W122–8.
[36] Rodrigues CHM, Pires DEV, Ascher DB. DynaMut2: assessing changes in stability and flexibility upon single and multiple point missense mutations. Protein Sci 2020;30:60–9.
[37] Laimer J, Hofer H, Fritz M, Wegenkittl S, Lackner P. MAESTRO - multi agent stability prediction upon point mutations. BMC Bioinforma 2015;16.
[38] Cao H, Wang J, He L, Qi Y, Zhang JZ. DeepDDG: predicting the stability change of protein point mutations using neural networks. J Chem Inf Model 2019;59:1508–14.
[39] Savojardo C, Fariselli P, Martelli PL, Casadio R. INPS-MD: a web server to predict stability of protein variants from sequence and structure. Bioinformatics 2016;32:2542–4.
[40] Cheng J, et al. Accurate proteome-wide missense variant effect prediction with AlphaMissense. Science 2023;381.
[41] Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. Curr Protoc Hum Genet 2013;76.
[42] Capriotti E, et al. WS-SNPs&GO: a web server for predicting the deleterious effect of human protein variants using functional annotation. BMC Genom 2013;14(S6).
[43] Tordai H, et al. Analysis of AlphaMissense data in different protein groups and structural context. Sci Data 2024;11.
[44] Gratuze M, et al. Impact of TREM2R47H variant on tau pathology–induced gliosis and neurodegeneration. J Clin Investig 2020;130:4954–68.
[45] Cheng-Hathaway PJ, et al. The Trem2 R47H variant confers loss-of-function-like phenotypes in Alzheimer's disease. Mol Neurodegener 2018;13.
[46] Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, Ronneberger O, Willmore L, Ballard AJ, Bambrick J, Bodenstein SW, Evans DA, Hung C, O'Neill M, Reiman D, Tunyasuvunakool K, Wu Z, Žemgulytė A, Arvaniti E, Jumper JM. Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature 2024;630(8016):493–500. https://doi.org/10.1038/s41586-024-07487-w.