Comparative Modelling and Prediction of Mutant Structures in TREM2 Protein using Computational Tools

K Mounika¹, K. Anbarasu²*

¹Research Scholar, Department of Bioinformatics, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamil Nadu, India, Pincode: 602105.
²Project Guide, Department of Bioinformatics, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamil Nadu, India, Pincode: 602105.

Abstract

Aim. Comparative modelling of mutant structures of TREM2 associated with Alzheimer’s disease using Homology modelling. Fold recognition and Novel Ab Initio methods. Materials and Methods. The sample size of 10 most deleterious mutations of TREM2 resulted previously by PROVEAN and PredictSNP was modelled using Bioinformatic computational tools such as SWISS model, Pyre2, tr Rosetta. Results. The best mutant model with 99.36% of identity and 99.72% ramachandran plot resulted by SWISS model is S116C. Most accurate mutants modelled by Pyre 2 are S116C,R98W,D39G,R6C,A105V all with 100% confidence in modelled structure and 50% of coverage. The perfect match score (i.e. Estimated TM score) between two structures was shown by S31F which was the result of tr Rosetta. Conclusion. The best modelled mutant structures has been predicted by using Comparative modelling, Fold recognition, Novel Ab initio methods.

Keywords. Alzheimer’s, Comparative modelling, mutant structures, Fold recognition, Homology modelling, Novel Ab Initio


INTRODUCTION

Alzheimer’s disease is a neurodegenerative disorder which is a common form of irreversible dementia. Alzheimer's disease progresses from episodic memory issues to a steady global decline in cognitive function, leaving individuals with end-stage dementia. (Citron 2010). Previously, the highly pathogenic mutations in the genes APP, PSEN1, PSEN2, MAPT, TREM2, and GRN have been associated with familial early onset dementia. These genes were chosen to screen for mutants in extremely modest groups or in single families correlated with late-onset AD (Cruchaga et al. 2012). The point mutations in a protein sequence can induce a structural change or divert the functional behaviour of native confirmations, which leads to directly or indirectly influence the function and distinct phenotypes as a result (Feyfant, Sali, and Fiser 2007).

The number of articles published since 2016 on this topic are nearly 951 in Google Scholar. Nearly 7 articles in Pubmed over variants associated with TREM2 causing Alzheimer’s disease. Many publications that focus on various applications were produced. For example, The TREM2 ECD domain's tertiary structure is mostly made up of nine-strands (A - F), which comprise three key complementarity-determining regions (called CDR loops), namely CDR1, CDR2, and CDR3 ligands bind to TREM2 ECD near apical CDR loops CDR2 maintains a stable conformation in normal conditions via maintaining H-bonding via the CDR1 loop, which is required for ligand interactions, according to previous research genetic variations result in the destabilization of these loops, and thus, by impairing ligand binding may have deleterious effects (Dash, Choi, and Moon 2020). Different variant structures that are found deleterious and associated with Alzheimer’s disease are R47H, R62C (Guerreiro et al. 2013); (Jonsson et al. 2013). Site-Directed mutagenesis is also used to introduce binding sites in heavy atoms in preparation for X Ray crystallography techniques (Heidelberger et al. 2009) When all of the side chains must be modelled on a single backbone, rotamers with a fixed backbone are frequently utilised. Some comparative
Comparative modelling and protein design approaches use this approach to avoid the combinatorial explosion that comes with a full conformational search of multiple sidechains function (Kognole et al. 2022) has been used for modelling point mutation in protein using main component CHARMM force field (Feyfant, Sali, and Fiser 2007).

Our institution is passionate about high quality evidence based research and has excelled in various fields (Parakh et al. 2020; Pham et al. 2021; Perunal, Antony, and Muthuramalingam 2021; Sathiyamoorthi et al. 2021; Devarajan et al. 2021; Dhanraj and Rajeshkumar 2021; Uganya, Radhika, and Vijayaraj 2021; Tesfaye Jule et al. 2021; Nandhini, Ezhilarasan, and Rajeshkumar 2020; Kamath et al. 2020). The lacunae of the study is mutations in TREM2 gene are reported to be associated with Alzeihmer’s disease due to lack of its 3D structure. Since Experimental exploration such as x ray crystallography, site-directed mutagenesis of different positions in protein structure is time consuming and expensive process which models the side chain. Although modelling a single sidechain in a given atomic environment appears to be one of the most straightforward of all protein structure prediction issues, it remains unsolved (Fiser 2004). Our aim of the research is to comparative modelling of mutant structure using Homology modelling, Fold recognition and novel Ab Initio.

Materials and Methods
The proposed work is done in the Bioinformatic laboratory, Department of Bioinformatics at Saveetha School of Engineering, Saveetha Institute of Medical And Technical Sciences. There is no ethical approval as human samples are not involved. The number of groups is 1 for three alignment algorithms called Comparative modelling, HH search, Rosetta.

The sample size is 10 per group. The bioinformatic tools named SWISS MODEL, Pyre, tr rosetta are used. The dataset required for study were selected as most deleterious mutations that were resulted by various computational tools named PROVEAN, Predict SNP. The most deleterious mutations that were selected for study are R47H, R47C, S116C, S16F, M31F, A105V, R62C, R136W, D39G, R98W. Initially the sequence of Triggering receptor expression on myeloid cells 2 were retrieved from uniprotkb(id:Q9NZC2). The mutations are included manually in the wildtype sequence and then submitted to the modelling tools for the mutant modelled structures. Finally the modelled mutations are visualized under Pymol inorder to make changes like removing water molecules, ligands etc., in case required.

Comparative modelling or homology modelling (Used in SWISS model)
Homology modelling is the most precise computational tool for creating dependable structural models, and it is employed in a wide range of biological applications. Homology modelling uses template protein sequence alignment to predict the 3D structure of a query protein.

Step 1: The mutant sequence for each sample was uploaded

Step 2: The sequence from step one is used to run a query on the SWISS-MODEL template library to look for evolutionary similar protein structures.

Step 3: When the template search is finished, the templates are ordered according on the expected quality of the generated models, as determined by the Global Model Quality Estimate (GMQE) and Quaternary Structure Quality Estimate algorithms (QSQE).

Step 4: A 3D protein model is automatically built for each specified template by first transferring conserved atom coordinates as indicated by the target-template alignment.

Step 5: SWISS-MODEL uses the QMEAN scoring algorithm to measure modelling flaws and provide estimates for predicted model correctness. To obtain global and per-residue quality estimations, QMEAN uses statistical potentials of mean force. Pairwise distance constraints, which provide ensemble information from all template structures discovered, help to improve local quality estimation.

HH search (Used in Pyre 2)
HH-suite is a popular open source software suite for finding sensitive sequence similarity and recognising protein folds. It is based on profile Hidden Markov models (HMMs), which represent multiple sequence alignments of homologous proteins and are aligned pairwise.

Step 1: The mutant sequence for each sample was uploaded
Step 2: The sequence from the first step is used as query sequence
Step 3: Multiple sequence alignment
Step 4: Secondary structure prediction
Step 5: Produce query Hidden Markov model
Step 6: HH search through HH database of known structures
Step 7: Alignment between query sequence and Template
Step 8: Produces crude backbone perform loop modelling add side chains using rotamers
Step 9: Final model

**Network prediction (tr Rosetta)**

trRosetta is an algorithm for rapid and reliable protein structure prediction. It constructs the protein structure on the basis of direct energy minimizations with a restrained Rosetta.

**Statistical Analysis**

The statistical analysis used for performing analysis in the SWISS model was QMEANDisCo and Qmean. QMEANDisCo is a composite scoring function to derive the absolute quality estimates of single model and to check the errors. Qmean is also a scoring function with combination of Global and Local scores where each one is combination of four statistical potential terms. By default these scores are transformed into Z scores to relate with what we would expect from a good resolution x ray structures. In pyre2 the prediction accuracy of protein has increased to 80% when compared to 1990’s where the accuracy was only around 65% (Bogachev et al. 2016)

**Results**

An overview of results by SWISS model was tabulated in Table 1. It demonstrates the sequence identity, globalscore, Z scores, Ramachandran favored by each deleterious mutation of TREM2. The Ramachandran plot of most accurate model S16F was shown, the chart is plotted by taking pie angle on x-axis and psi angle on y-axis as shown in Fig. 1. The visualized PDB structure of the most destabilized mutant structure of the S16F model was shown in Fig. 2. An overview of results by Pyre was tabulated in Table 2. The confidence and coverage score of each mutation with the template that was previously recorded in the HH database was represented. In Fig. 3 The pdb structures of S116C and R98W which are having high confidence with templates has been visualized using pymol of version 2. An overview of mutants resulted by tr rosetta that are having high Estimated TM scores are tabulated in Table 3. The TM score indicates the perfect match between two proteins and tells us which mutant is modeled more accurately. In Fig. 4, the pdb structure of S16F has been shown which was visualized by pymol version 2.

**Discussion**

In this article the protein structure prediction and analytical methods were used such as comparative modelling, folding recognition and Ab initio were utilised to design the structures for potential mutations of Trem2 which are considered to be main cause for accumulation of toxic proteins such as Tau, Amyloid in the microglia of brain cells and leads to Alzheimer’s disease. Extensive experimental studies such as X-ray crystallography would be a time-consuming and inconvenient process. The application of bioinformatics tools might be easy to proceed and also results most accurate modelled structures.

There are some similar findings related to protein structure prediction methods such as Homology modelling has evolved into a critical tool in structural biology facilitating to bridge the void between known protein sequences and empirically established structures. Fully automated workflows and servers simplify and streamline the homology modelling process, allowing even non-computational experts to build accurate protein models (Waterhouse et al. 2018). By aligning profile HMMs instead of simple sequence profiles we are able to improve both sensitivity and alignment quality and can produce much precise structures (Söding 2004). Novel Ab-initio method (Zanchet et al. 2016) is extremely used to determine the model and acquire protein structures without all of the resources like known templates based strategies involved in other more traditional biophysical methods (Söding 2004; Murthy 2020)

The limitations faced in our research study is that the sample structures generated by the novel Ab initio and homology modelling are only 50 each and in those 20 destabilized structures of TREM2 are modelled. The future scope of this project is to work on atleast 1000 structures of TREM2 protein with the help of an advanced computer system so that the possibilities of obtaining better protein structures will be high.

**Conclusion**

TREM2 is one of many immune-related genes expressed in microglia that have been connected to Alzheimer's disease aetiology. It's likely that some of these genes are functionally linked and belong to common pathways that
could be targeted for future therapeutic intervention. Bioinformatics tools were utilised to assess the depleted effects of these pathogenic single nucleotide polymorphism (SNPs) on the inflicted protein because experimental studies for every mutation are time-consuming. To improve forecast accuracy, a combination of a few tools was applied. The best modelled mutant structures has been predicted by using Comparative modelling, Fold recognition, Novel Ab initio methods.

**Declarations**

**Conflict of Interest**

No conflict of interest in this manuscript.

**Author contribution**

Author KM was involved in data collection, data analysis, manuscript writing. Author KA was involved in Conceptualization, guidance and critical review of manuscript.

**Acknowledgements**

The authors would like to express their gratitude towards Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (Formerly known as Saveetha University) for providing the necessary infrastructure to carry out this work successfully.

**Funding:** We thank the following organisations for providing financial support that enabled us to complete the study

1. Qbiogen LLP, Chennai
2. Saveetha University
3. Saveetha Institute of Medical and Technical Sciences
4. Saveetha School of Engineering

**References**

K. Anbarasu, et.al: Comparative Modelling and Prediction of Mutant Structures in TREM2 Protein using Computational Tools


Table 1. Results of Sequence identity and ramachandran favored by mutant structures of TREM2 modelled by SWISS model

<table>
<thead>
<tr>
<th>Mutation</th>
<th>Sequenc Identity</th>
<th>GMQ E</th>
<th>QMEADisC o</th>
<th>Global Qmea n</th>
<th>C beta</th>
<th>All atom Solvatio n</th>
<th>Torsio n</th>
<th>Ramachandra n Favoured</th>
</tr>
</thead>
<tbody>
<tr>
<td>S116C</td>
<td>99.35%</td>
<td>0.47</td>
<td>0.78</td>
<td>0.08</td>
<td>-1.8</td>
<td>0.56</td>
<td>-0.61</td>
<td>-0.59</td>
</tr>
<tr>
<td>A105V</td>
<td>98.72%</td>
<td>0.48</td>
<td>0.8</td>
<td>0.08</td>
<td>-1.41</td>
<td>0.49</td>
<td>-0.63</td>
<td>-0.07</td>
</tr>
<tr>
<td>R98W</td>
<td>98.72%</td>
<td>0.48</td>
<td>0.8</td>
<td>0.08</td>
<td>-1.66</td>
<td>1.02</td>
<td>-0.66</td>
<td>-0.41</td>
</tr>
<tr>
<td>R62C</td>
<td>98.72%</td>
<td>0.48</td>
<td>0.8</td>
<td>0.08</td>
<td>-1.62</td>
<td>0.82</td>
<td>-0.53</td>
<td>-0.67</td>
</tr>
<tr>
<td>S31F</td>
<td>98.72%</td>
<td>0.48</td>
<td>0.8</td>
<td>0.08</td>
<td>-1.42</td>
<td>-0.4</td>
<td>-0.71</td>
<td>-0.1</td>
</tr>
<tr>
<td>D39G</td>
<td>98.72%</td>
<td>0.48</td>
<td>0.8</td>
<td>0.08</td>
<td>-1.87</td>
<td>0.56</td>
<td>-0.58</td>
<td>-0.42</td>
</tr>
<tr>
<td>R47H</td>
<td>98.72%</td>
<td>0.48</td>
<td>0.8</td>
<td>0.08</td>
<td>-1.48</td>
<td>0.36</td>
<td>-0.57</td>
<td>-0.44</td>
</tr>
<tr>
<td>S16F</td>
<td>99.36%</td>
<td>0.48</td>
<td>0.8</td>
<td>0.08</td>
<td>-1.6</td>
<td>0.83</td>
<td>-0.59</td>
<td>-0.18</td>
</tr>
<tr>
<td>R136W</td>
<td>98.72%</td>
<td>0.48</td>
<td>0.8</td>
<td>0.08</td>
<td>-1.57</td>
<td>0.51</td>
<td>-0.65</td>
<td>-0.4</td>
</tr>
<tr>
<td>R47C</td>
<td>99.35%</td>
<td>0.48</td>
<td>0.8</td>
<td>0.08</td>
<td>-1.44</td>
<td>0.45</td>
<td>-0.65</td>
<td>-0.42</td>
</tr>
</tbody>
</table>
Table 2: Results of confidence percentage of mutants modelled by Pyre 2

<table>
<thead>
<tr>
<th>Mutations</th>
<th>Template</th>
<th>Confidence</th>
<th>Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>R47C</td>
<td>C6xdsA</td>
<td>99.99%</td>
<td>53%</td>
</tr>
<tr>
<td>S31F</td>
<td>C6xdsA</td>
<td>99.99%</td>
<td>52%</td>
</tr>
<tr>
<td>S116C</td>
<td>C6xdsA</td>
<td>100.00%</td>
<td>50%</td>
</tr>
<tr>
<td>R136W</td>
<td>C6xdsA</td>
<td>99.99%</td>
<td>50%</td>
</tr>
<tr>
<td>R98W</td>
<td>C6xdsA</td>
<td>100%</td>
<td>50%</td>
</tr>
<tr>
<td>R47H</td>
<td>C6xdsA</td>
<td>99.99%</td>
<td>50%</td>
</tr>
<tr>
<td>D39G</td>
<td>C6xdsA</td>
<td>100%</td>
<td>50%</td>
</tr>
<tr>
<td>R62C</td>
<td>C6xdsA</td>
<td>100%</td>
<td>50%</td>
</tr>
<tr>
<td>A105V</td>
<td>C6xdsA</td>
<td>100%</td>
<td>50%</td>
</tr>
<tr>
<td>S16F</td>
<td>C6xdsA</td>
<td>99.99%</td>
<td>53%</td>
</tr>
</tbody>
</table>

Table 3: Estimated TM-Scores of mutations modelled by tr Rosetta

<table>
<thead>
<tr>
<th>Mutation</th>
<th>Estimated TM-Score</th>
<th>Disorder (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>S31F</td>
<td>0.631</td>
<td>47.40%</td>
</tr>
<tr>
<td>R47C</td>
<td>0.619</td>
<td>50%</td>
</tr>
<tr>
<td>S16F</td>
<td>0.622</td>
<td>49.60%</td>
</tr>
<tr>
<td>A105V</td>
<td>0.613</td>
<td>47.40%</td>
</tr>
<tr>
<td>R98W</td>
<td>0.608</td>
<td>48.30%</td>
</tr>
<tr>
<td>S116C</td>
<td>0.619</td>
<td>45.20%</td>
</tr>
<tr>
<td>D39G</td>
<td>0.616</td>
<td>49.60%</td>
</tr>
<tr>
<td>R62C</td>
<td>0.612</td>
<td>49.10%</td>
</tr>
<tr>
<td>R47H</td>
<td>0.619</td>
<td>50%</td>
</tr>
<tr>
<td>R136W</td>
<td>0.62</td>
<td>48.70%</td>
</tr>
</tbody>
</table>
Fig. 1. Ramachandran plot of S16F

Fig. 2. Visualized PDB Structure of S16F
K. Anbarasu, et.al: Comparative Modelling and Prediction of Mutant Structures in TREM2 Protein using Computational Tools

Fig. 3 Visualized pdb structure of S116C

Fig. 3 Visualized pdb structures of S31F
Fig. 4: PDB structure of S31F