

**RESEARCH ARTICLE**

## **A Comparative Study of Homology Modeling Algorithms for NPTX2 Structure Prediction**

**Sowmya H**

Department of Biotechnology, School of Life Sciences, Vels University, Pallavaram, Chennai, Tamil Nadu, India.

\*Corresponding Author E-mail: [sowmya.se@velsuniv.ac.in](mailto:sowmya.se@velsuniv.ac.in)

### **ABSTRACT:**

Alzheimer's Disease (AD) is a very prevalent neurological disorder that results in loss of memory due to the weakening of synapses. Down-regulation of neuronal Pentraxin (*NPTX2*), a secretory protein is one of the causes for AD. The structure of a protein is very important to predict the protein's function. The experimental structure of NTPX2 is not available yet. Hence in this study, three structures for NPTX2 were generated using Geno3D, Modeller9.20 and Swiss Model. The quality of the protein was validated using PROCHECK and ERRAT. The PROCHECK results for the structures modeled using Geno3D, Modeller, and Swiss Model showed 63%, 87.3% and 88.2% residues in the most favoured regions and 2.5%, 0.0%, 0.00% residues in disallowed regions respectively. The ERRAT results showed an overall quality factor of 90.27, 60.476, 77.54 respectively. The model generated from Swiss Model can be considered as best model because in PROCHECK the model showed high number of residues in the most favoured region and no residues were in the disallowed region. The ERRAT result also showed an overall quality factor greater than 75. Though Geno3D showed a good overall quality factor, it showed about 2.5% residues in the disallowed region. Structure Modeled using Modeller showed good results for Ramachandran plot with no residues in the disallowed region. However, its overall quality factor was only 60.476. These structures predicted can lay a foundation for discovering new drugs for the treatment of AD.

**KEYWORDS:** Alzheimer's Disease, Neuronal pentraxin, Geno3D, Modeller, Swiss Model

### **INTRODUCTION:**

AD is a common progressive brain disorder that leads to loss of memory and thinking skills. Mostly people aged in their mid 60s are highly affected with AD. AD causes loss of cognitive functioning—thinking, remembering, and reasoning. Formation of Amyloid  $\beta$  plaques and tau tangles in the brain are the common cause of AD. Many other proteins have been reported to be up regulated or down regulated in AD<sup>1</sup>. One such protein whose down regulation is a major cause for AD and a new biomarker identified in AD is NPTX2 also called NP2 or NARP (neuronal activity regulated pentraxin) that belongs to neuronal pentraxin family.

Neuronal pentraxins include two secreted proteins NPTX1 and NPTX2 and a type II transmembrane protein that exist in cleaved, soluble form, NPTXR (Neuronal Pentraxin Receptor)<sup>2</sup>.

The major function of NPTX2 is to regulate AMPA type glutamate receptor GluA4. In addition, it also promotes the formation, maturation and plasticity of synapses in brain, and regulates axon outgrowth in cortical explants<sup>1</sup>. Presynaptic expression of NPTX2 is necessary to regulate GluA4, an AMPA receptor subunit. In AD, a significant down regulation of NPTX2 is reported. This reduction is associated with a reduction of the AMPA type glutamate receptor GluA4, a major reason for loss of synapses and cognitive impairment in AD patients<sup>2</sup>.

based on a related template protein with known experimental structure. The similar residues in the query and the template are organized based on their corresponding topology in the template<sup>3</sup>. The experimental study of NPTX2 is not available. This study involves the prediction of 3D structure of NPTX2 using the principles of homology modeling. The predicted models were validated using PROCHECK, ERRAT. Structure of a protein is very important to know the protein's function. The predicted structure can be used for carrying out further studies on NPTX2, to find suitable targets that can up regulate NPTX2 and hence can lay a foundation for the treatment of AD.

**MATERIALS AND METHODS:**

**Human NPTX2 Protein Sequence Retrieval:**

The human NPTX2 protein sequence (ID: P47972.2) was retrieved from NCBI. NCBI is a database, part of National Institute of Health that contains all the protein and gene information<sup>4</sup>.

**Physico-chemical characteristics:**

To analyze the physic chemical characteristics, Expasy-Protpram was used<sup>5</sup>. ProtParam allows the computation of various physical and chemical parameters like molecular weight, theoretical pI<sup>6</sup>, amino acid composition, estimated half-life, aliphatic index<sup>7</sup> and grand average of hydropathicity (GRAVY)<sup>8</sup>.

**Functional Characterization:**

To identify the transmembrane regions, SOSUI server performed<sup>9</sup>. To understand the functional linkages disulphide bonds in the protein should be known. The disulfide bonds in the protein were identified using the tool CYS\_REC<sup>10</sup>. CYS\_REC identifies the number and position of cysteine amino acid in the protein. The information about motif regions was determined using motiffinder (<https://www.genome.jp/tools/motif/>).

**Identification of Protein Secondary Structure:**

Secondary structure of a protein comprises mainly  $\alpha$ -helix,  $\beta$  s-sheets, turns and coils. The knowledge about protein secondary structure is important for knowing the tertiary and quaternary structures. The secondary structure of the protein was determined using Self-Optimized Prediction Method with Alignment (SOPMA)<sup>11</sup>.

**Model building and evaluation:**

Three software Modeller<sup>12</sup>, Geno 3D<sup>13</sup> and Swissmodel<sup>14</sup> were used to predict the three dimensional structure of the NPTX2. The models were energy minimized using

swiss PDB viewer. Finally the models were validated using Saves server PROCHECK- Ramchandran plot analysis<sup>15</sup>. Further verification of models was performed using ERRAT<sup>16</sup>.

**RESULTS AND DISCUSSION:**

AD is the most common cause of dementia with about 10% people above the age of 50 is commonly affected. NPTX are a family of proteins that are involved in maintaining the synaptic plasticity in neurons, NPTXs bind to glutamate receptor and can act as a presynaptic factor to cause post synaptic induction. Down-regulation of NPTX2 causes a cognitive impairment and can be a target in AD. The protein sequence of NPTX2 was retrieved from the NCBI, a public domain database. The FASTA format of protein sequences were retrieved and used for further analysis. Physico-chemical characteristics are determined using Expasy's ProtParam tool was used to compute various protein parameters (Table1). The results showed pI (isoelectric Point) value and molecular weight of 5.45 and 47041.55 g/mol. The isoelectric point is the point at which the pH of the protein is zero. At the pI, minimum solubility of protein occurs, thus the pI value determination can be significant during purification of protein. A cycle of synthesis and degradation of proteins occurs within the body. Proteins regularly get degraded and are again replaced with other new copies. Half life is the time taken to degrade half the quantity of initial quantity of a protein. The half life determines how long a protein remains stable. The half life of the NPTX2 was 30 hrs, indicating that the protein remains functionally stable for about 30 hrs after the protein gets degraded and new copies are synthesized. Hydrophilic and hydrophobic proteins can be identified using GRAVY. The protein had a GRAVY value -0.247. Negative GRAVY score indicate hydrophilic protein<sup>17</sup>. The number of negative and positive residues is 53 and 43 respectively. The relative volume occupied by alanine, valine, leucine and isoleucine amino acids determines the aliphatic index of a protein. A higher positive value indicates that the protein is highly thermostable. The aliphatic index of NPTX2 was 91.93. This very high aliphatic index value indicates that NPTX2 is thermally stable for wide range of temperatures. The nonpolar amino acid of protein forms the transmembrane domain that traverse the phospholipid bilayer. The transmembrane domain of NPTX2 was determined using SOSUI server. The transmembrane domain in NTPX2 was about 7 amino acids long (Table2).

**Table 1: Results of physicochemical parameters of NTPX2 determined from Expasy's ProtParam tool.**

Molecular Weight (g/mol)	pI	Sequence Length	Half life (hrs)	GRAVY	-R	+R	AI
47041.55	5.45	431	30	-0.247	53	43	91.93

Physicochemical properties of NPTX2 protein (M.wt.: Molecular weight; pI: Isoelectric point; -R: Number of negative residues; +R: Number of positive residues; AI: Aliphatic index; GRAVY: Grand Average Hydropathy)

**Table 2: Transmembrane regions identified from SOSUI server.**

Transmembrane region	Length
MLALLAASVALAVAAG	16

The protein structure is stabilized by the formation of disulfide bonds. The disulfide bonds are formed between cysteine amino acids in a protein. Hence knowledge about cysteine amino acid can provide useful information about the disulfide bond. The number of cysteine amino acids was determined using CYS\_REC tool (Table3). A total of 7 cysteine residues were present in NTPX2. Motifs are structural patterns of a protein. The motif in NPTX2 determined using MotifFinder is listed in Table 4. Intermediate secondary structures in protein are formed by hydrogen bonding between the amino and carboxyl group of amino acids. Secondary structure residues in NPTX2 analyzed using SOPMA<sup>18</sup> (Table 5), showed 45.01%, 36.43% and 12.76 % of alpha helix, random coils, and extended strands formation respectively. Default parameters (Window width: 17, similarity threshold: 8 and number of states: 4) are used for secondary structure analysis.

**Table 3: Number of Cysteine Residues Identified using CYS\_REC**

No of cysteines	Position of Cysteine
7	Cys 29, cys 41, cys 94, cys 253, cys 313, cys 394, cys 424

**Table4: Results of Motif determined using Motiffinder**

Motif	Position in the Protein	Description
Pentaxin	232 to 416	Pentaxin family
Laminin_G_3	244 to 388	Concanavalin A-like lectin/glucanases superfamily
GOLGA2L5	64 to 190	Putative golgin subfamily A member 2-like protein 5
GrpE	132 to 226	GrpE
LIAS_N	200 to 253	N-terminal domain of lipoyl synthase of Radical_SAM family
DUF2408	125 to 212	Protein of unknown function
CorA	123 to 213	CorA-like Mg2+ transporter protein
Laminin_G_2	285 to 377	Laminin G domain
HOOK	135 to 200	HOOK protein
Exonuc_VII_L	121 to 208	Exonuclease VII, large subunit
Sun2_CC2	162 to 182	SUN2 coiled coil domain 2

The three dimensional structure prediction of NPTX2 was performed using Modeller9.20, Geno 3D, Swiss Model. These software works based on the principles of homology modeling. Homology modeling constructs model based on a homologous template protein with a known experimental structure available. To find the

structurally homologous proteins BLAST was performed against PDB. Protein with >30% identity can be chosen as a template<sup>19</sup>. Two template structures and the query sequence were aligned using Multialign (fig. 1). MultiAlign is an effective tool for obtaining multiple sequence alignment<sup>20</sup>. Protein with maximum alignment in MultiAlign was chosen as the template. The template structure was retrieved from PDB. Any bad geometry in the PDB structures were fixed using What if server. The final template structure was then downloaded and used for further query structure prediction. The constructed models were energy minimized using SWISS PDB viewer<sup>21</sup>. Through energy minimization any bad angle or bad length formed between the amino acids can be corrected. The validation of models performed using PROCHECK-Ramachandran plot and ERRAT. Dihedral angles  $\psi$  against  $\phi$  of amino acid residues in protein structure can be visualized using Ramachandran plot. PROCHECK and ERRAT determines the quality of the predicted structures (Table 6). The number of amino acids in the allowed and the disallowed region can determine the protein quality. The ramachandran plot from Geno3D showed the results with 63% in most favoured region and 2.5% residues are in the disallowed region. ERRAT results for structure modeled using Geno3D showed a quality factor of 90.27. Ramachandran plot for model structured using Modeller showed 87.3% residues in most favoured region and 0.0% residues in the disallowed region (Figure 1). The ERRAT results showed a quality factor of 60.476. The Swiss Model structure showed 88.2% residues in the most favoured region and 0.00% amino acid in the disallowed region. The ERRAT result for Swiss Model structure showed a quality factor of 77.54. A comparison of results from Modeller, Geno3D and Swiss Model are listed in Table 6 (Figure. 2). Finally the modeled structures were viewed using Pymol (Figure 3).

**Table5: Percentage of available secondary structure from SOPMA**

Secondary structure	Percentage of Secondary structure
Alpha helix	45.01 %
310 helix	0.00 %
Pi helix	0.00 %
Beta bridge	0.00 %
Extended strand	12.76 %
Beta turn	5.80 %
Bend region	0.00 %
Random coil	36.43 %
Ambiguous states	0.00 %
Other states	0.00 %

**Table 6: Comparison of Values of PROCHECK and ERRAT for NPTX2 Structures Modeled using Geno3D, Modeller9.20, Swiss Model**

Homology Modeling Tools	PROCHECK		ERRAT
	No. of residues in most favoured region	No of residues in disallowed region	
Geno3D	63%	2.5%	90.27
Modeller9.20	87.3%	0.0%	60.476
Swiss Model	88.2%	0.00%	77.54

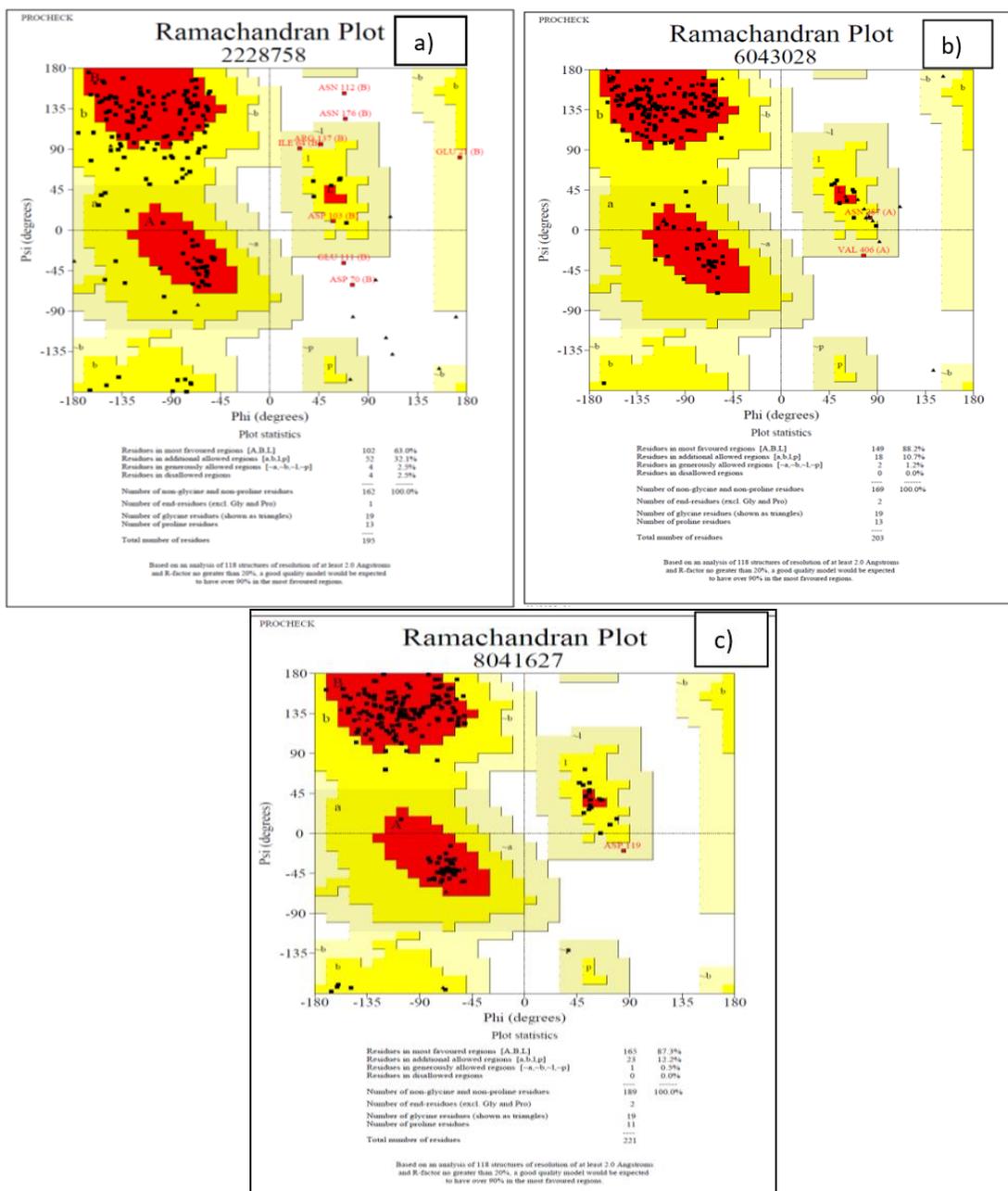
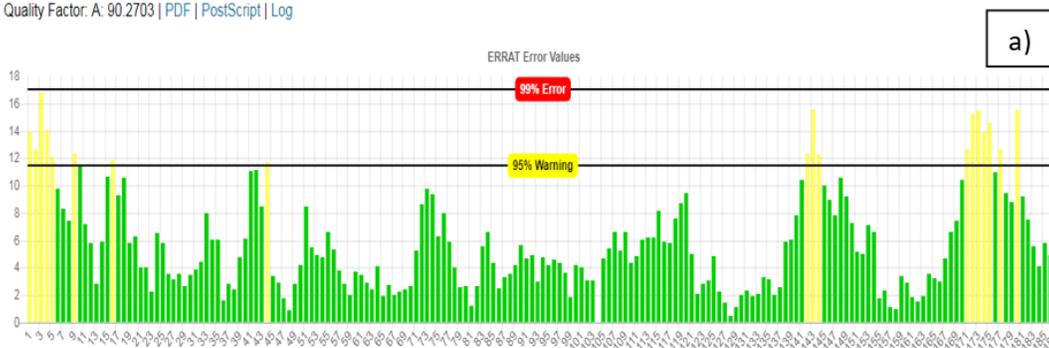
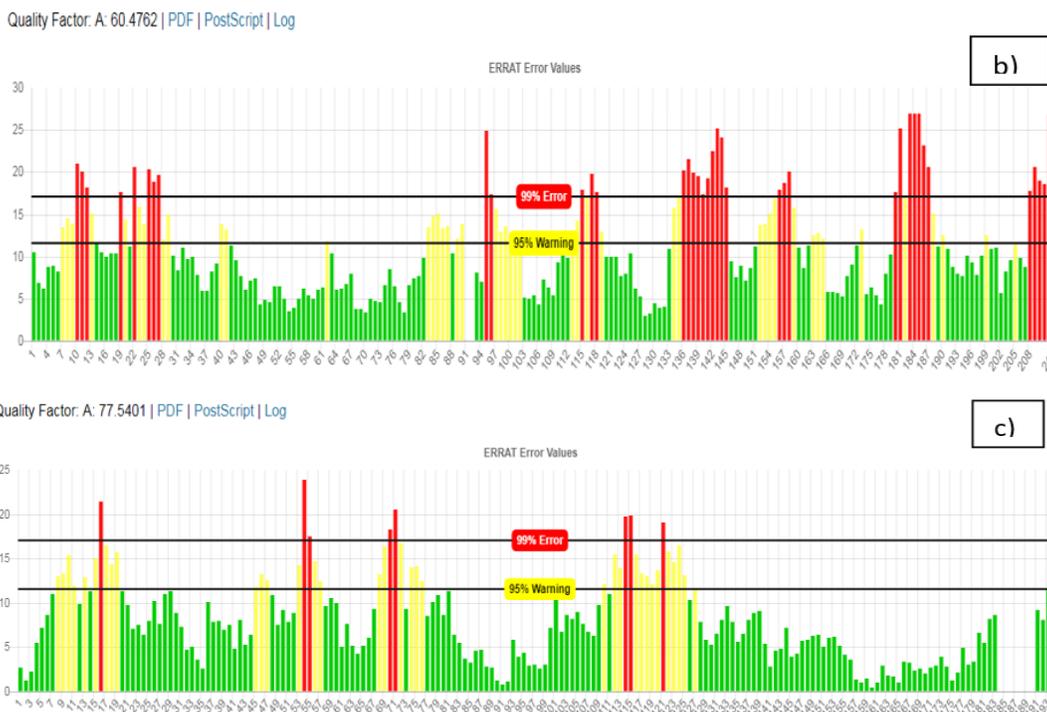


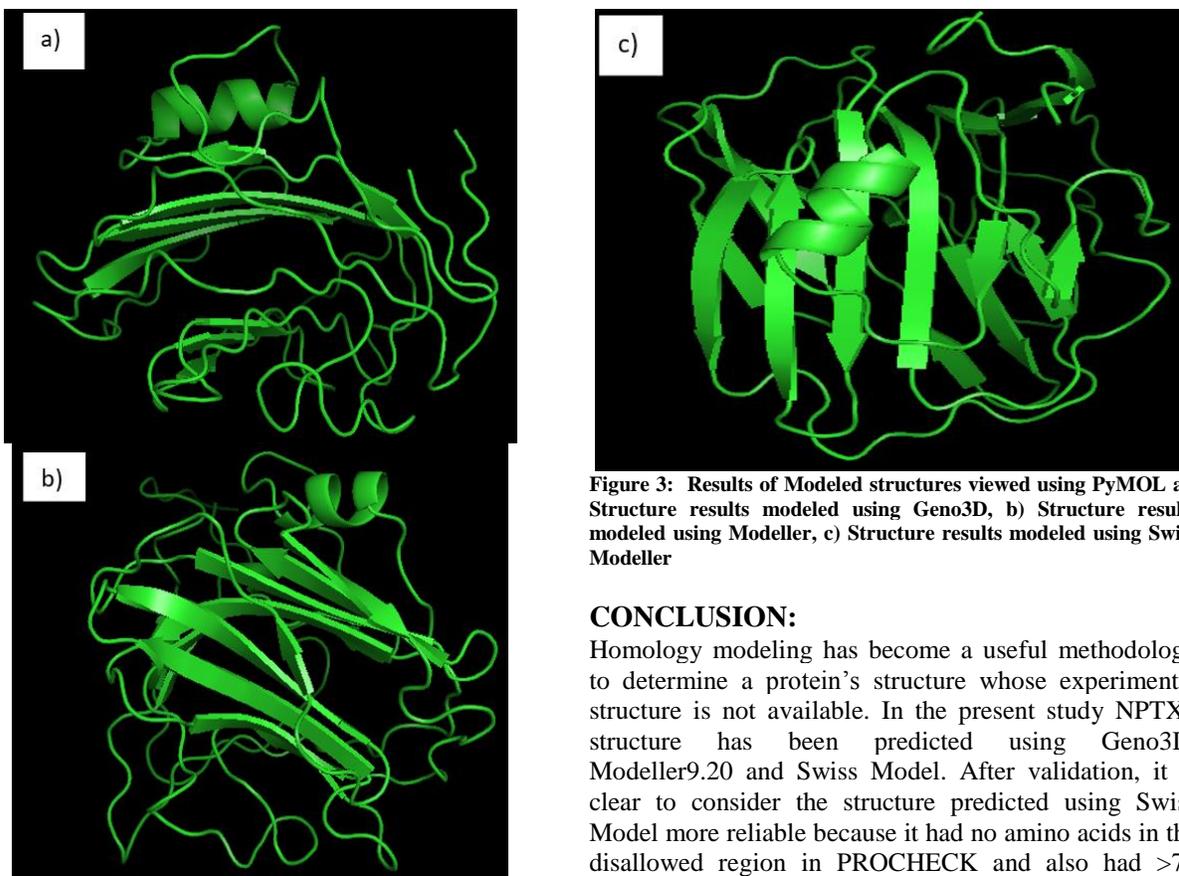
Figure 1: Results from PROCHECK. a) Ramachandran Plot for Structure Modeled using Geno3D, b) Ramachandran Plot for structure Modeled using Modeller9.20, c) Ramachandran Plot for structure Modeled using Swiss Modeller

Quality Factor: A: 90.2703 | PDF | PostScript | Log





**Figure 2: Results from ERRAT. A) ERRAT Plot for Structure Modeled using Geno3D, b) ERRAT Plot for structure Modeled using Modeller9.20, c) ERRAT Plot for structure Modeled using Swiss Modeller.**



**Figure 3: Results of Modeled structures viewed using PyMOL a): Structure results modeled using Geno3D, b) Structure results modeled using Modeller, c) Structure results modeled using Swiss Modeller**

**CONCLUSION:**

Homology modeling has become a useful methodology to determine a protein’s structure whose experimental structure is not available. In the present study NPTX2 structure has been predicted using Geno3D, Modeller9.20 and Swiss Model. After validation, it is clear to consider the structure predicted using Swiss Model more reliable because it had no amino acids in the disallowed region in PROCHECK and also had >75 overall quality factor. Modeller structure also showed a

good result in PROCHECK with no amino acids in the disallowed region; however the overall quality factor determined using ERRAT was only 60. Geno3D showed an overall quality factor of 90.27 however had 2.5% amino acids in the disallowed region. The predicted structure can be used to carry out further studies about the protein like functional analysis of the protein and also can lay a foundation for drug discovery.

## REFERENCES:

1. Kelley BJ, Petersen RC. Alzheimer's disease and mild cognitive impairment. *Neurol Clin.* 2007;25(3):577-609.
2. Xiao MF, Xu D, Craig MT, et al. NPTX2 and cognitive dysfunction in Alzheimer's Disease. *Elife.* 2017;6:1-27.
3. Vyas VK, Ukawala RD, Ghate M, Chintha C. Homology modeling a fast tool for drug discovery: current perspectives. *Indian J Pharm Sci.* 2012;74(1):1-17.
4. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* 2015;44(D1):D7-19.
5. Gasteiger E. Protein Identification and Analysis Tools on the ExPASy Server. In: John M. Walker ed, *The Proteomics Protocols Handbook*, Humana Pres. 2005: 571-607.
6. Gill SC, Von Hippel PH. Calculation of protein extinction coefficients from amino acid sequence data. *Anal Biochem.* 1989;182: 319- 326.
7. Ikai AJ. Thermo stability and aliphatic index of globular proteins. *J Biochem.* 1980; 88: 1895-1898
8. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. *J Mol Biol.* 1982; 157: 105-132.
9. Hirokawa T, Boon-Chieng S, Mitaku S. SOSUI: classification and secondary structure prediction system for membrane proteins. *Bioinformatics.* 1998;14(4): 378-379.
10. CYS\_REC. [http://sun1.softberry.com/berry.phtml?topic=cys\\_rec&group=help&subgroup=propt](http://sun1.softberry.com/berry.phtml?topic=cys_rec&group=help&subgroup=propt). (27/10/2006)
11. Geourjon C, Deléage G. SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Bioinformatics.* 1995; 11(6): 681-684.
12. Sali A, Blundell TL. Comparative protein modeling by satisfaction of spatial restraints. *J Mol Biol.* 1993; 234: 779-815
13. Combet C, Jambon M, Deleage G, Geourjon C. Geno3D: Automatic comparative molecular modelling of protein. *Bioinformatics.* 2002; 18: 213-214.
14. Arnold K, Bordoli L, Kopp J, Schwede T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics.* 2006; 22: 195-201.
15. Ramachandran GN, Ramakrishnan C, Sasisekharan V. Stereochemistry of polypeptide chain configurations. *J Mol Biol.* 1963; 7: 95-99.
16. Colovos C, Yeates TO. Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci.* 1993; 9 :1511-1519.
17. Yang YH, Dai L, Xia HC, Zhu KM, Liu HJ, Chen KP: Protein profile of rice (*Oryza sativa*) seeds. *Genet Mol Biol.* 2013, 36 (1): 87-92.
18. Geourjon C, Deléage G. SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Comput Appl Biosci.* 1995;11; 681-684.
19. Eswar N, Webb B, Marti-Renom MA, et al. Comparative protein structure modeling using Modeller. *Curr Protoc Bioinformatics.* 2006; Chapter 5:Unit-5.6.
20. Essoussi N, Boujenfa K, Limam M. A comparison of MSA tools. *Bioinformatics.* 2008;2(10):452-5.
21. Nicolas G, Peitsch MC. SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis* 1997; 18: 2714-2723.