# PHYLOGENETICS AND HOMOLOGY MODELING STUDIES OF BETA-SECRETASE 1- AN *IN SILICO* APPROACH TO THERAPEUTIC TARGET FOR ALZHEIMER'S DISEASE

**Soujanya Ketha[1], Samuel David Raj[1], Godi Sudhakar[2*]**

[1] Department of Biotechnology, Dr.V.S. Krishna College, Visakhapatnam- 530022, A.P. India
[2] Department of Human Genetics, Andhra University, Visakhapatnam-530003. A.P. India.

## *ABSTRACT*

Alzheimer's disease (AD) is a genetically complicated, heterogeneous neurodegenerative disorder that affects the brain's neurons due to the accumulation of plaques formed by an amyloid-β peptide. Beta-Secretase 1 (BACE1) is a membrane-associated aspartic protease that cleaves the amyloid and is a prime target for developing anti-AD pharmaceuticals to reduce the amyloid burden. In this study, the MEGA software was used to know the phylogenetic relationship among secretase, and the Modeller tool was utilized to generate three-dimensional structures of BACE1. The 3D model was verified using PROCHECK, VERIFY 3D, and ERRAT programs, followed by binding site prediction by DoGSiteScorer. The model predicted by Modeller in this study showed 93.5% residues in the most favored regions and 0% of residues in the disallowed regions as identified in the Ramachandran plot, which indicates the obtained model's good quality. Furthermore, the mode of inhibitions and proving certain drugs' activity can be studied with this predicted model using *in silico* approach as it is the therapeutic target for AD.

**Keywords:** Homology Modeling, Beta Secretase 1, Alzheimer's disease, Amyloid beta-peptide

## INTRODUCTION

Alzheimer's disease (AD) is the most common neurodegenerative disorder in the elderly. It affects the neurons in the brain due to the accumulation of plaques and tangles formed by the amyloid beta-peptide (Aβ) (Hong et al., 2004). AD is a genetically complicated and heterogeneous disorder that primarily involves the brain's parts that control thought, memory, and language (Braak H & Braak E. 1991; Bertram et al., 2007; Scodellaro and Pin, 2011). The symptoms typically start with memory loss and slowly affect other cognitive domains like visual

skills, motor skills, executive function, language, and daily activities (McKhann et al., 2011; Huang and Mucke, 2012).

Epidemiological studies predicted that this AD affects more than 35 million people worldwide (Querfurth and LaFerla, 2010; Kumar et al., 2015). Furthermore, this incidence will rise to 65 million by 2030 and about 115 million in 2050 (Prince et al., 2013; Wimo et al., 2017). Despite the recent progress in understanding the pathophysiology and neurobiology of AD, no therapeutic strategies can effectively cure or prevent it. Several proteins are responsible for this disease condition, but the real causatives are yet to be discovered. Autopsy of the brains of AD patients demonstrated an extensive neuronal loss in the brain region that controls cognition and memory and the presence of extracellular aggregates of proteinaceous debris, known as "amyloid plaques." (Selkoe, D J. 1991; Walsh, D.M., & Selkoe, D, J. 2007). These plaques are composed of aggregates of a peptide, known as an amyloid-β peptide. It is generated due to abnormal proteolysis of a more significant membrane-embedded protein called amyloid precursor protein (APP) (Murphy, M. P., & LeVine, H., 2010). Family history of patients suffering from early-onset AD reveals that AD is inherited from parents to progeny, indicating the involvement of few genetic risk factors in the early-onset form of AD (Mattson, 2004).

A positive correlation is observed between mutations on chromosome 21 and the frequency of early-onset of AD. Mutations on chromosome 21 cause abnormality in amyloid precursor protein (APP) processing (Goate, A et al. 1991; Borchelt et al. 1997). Aβ is generated due to consecutive proteolysis of APP by two proteases, namely β and γ-secretase. β-Secretase 1 (BACE1) is a membrane-associated aspartic protease that cleaves the amyloid precursor protein (APP) at the N-terminus of the amyloid-beta (Aβ) peptide domain, thereby catalyzing the first step in A peptide generation. BACE1 is the key enzyme for producing the causative agent in AD. BACE1 is a compact globular protein with the molecular weight of 43722.54 Daltons, and isoelectric point (pI) is 5.04, formed by two domains specifically, residues 47-146 and residues 146-385 (Artimo et al., 2012). It appears to be an important drug target upstream of Aβ generation. Therefore, β-Secretase is considered the prime target for developing anti-AD pharmaceuticals, and BACE1 inhibitors can reduce the amyloid burden (Selkoe DJ.2001; Rochette MJ & Murphy MP 2002).

There are several methods like crystallography and nuclear magnetic resonance spectroscopy (NMR) to determine protein structure, which is more expensive and tedious than *In silico* methods. On the other hand, homology/comparative modeling is the best method for sequences with the highest template similarity to predict the structure and identify errors in the structure for further studies like molecular docking and molecular simulations. Hence, this *In silico* study on structural analysis of BACE1 can lay a foundation for drug discovery and provide more insights into the molecular mechanisms of AD.

## MATERIALS AND METHODS

**Sequence Analysis and Phylogenetic Studies:** The protein sequences of secretase from homosapiens were aligned in FASTA format and then analyzed with ClustalW (Thompson *et al.,* 1994) using MEGA software (Kumar, Stecher, Li, Knyaz, and Tamura 2018). The program was then used to calculate the best match for selected sequences; similarities and differences were identified by viewing the Cladogram based on the UPGMA (unweighted pair group method with arithmetic mean) approach.

**Homology Search:** The protein sequence of Beta-secretase 1 (P56817-BACE1_HUMAN) in FASTA format was obtained from the UniProtKB (http://www.uniprot.org/). BLASTP (Altshul *et al.,* 1990) was used with the default parameters to find suitable templates to identify the most appropriate templates based on sequence identity and score for homology modeling of BACE1. The homology modeling was carried out using MODELLER (http://www.salilab.org/modeler) (Sali and Blundell, 1993), is a computer program that models the three-dimensional (3D) structure of proteins and their assemblies by satisfying spatial restraints. The core modeling procedure begins with aligning the sequence to be modeled (target) with a related known 3D structure (template). Then, the spatial features, such as Cα-Cα distance, hydrogen bonds, and main chain and side-chain dihedral angles, are transferred from the templates to the target (Vyas, V. K., 2012).

**Validation of Modeled Structure:** The 3D model of BACE1 was verified using the PROCHECK; VERIFY 3D and ERRAT programs available from the Structural Analysis and Verification Server (SAVES) (http://nihserver.mbi.ucla.edu/SAVES). Out of the models generated by MODELLER, the one with the best G-score of PROCHECK (Laskoswki *et al.,*

1993) to evaluate the quality of the modeled 3D structure of BACE1, and VERIFY3D (Eisenberg *et al.,*1997) was used to check the residue profiles of the obtained 3D models. In addition, to assess the stereochemical quality of the 3D models, Quality evaluation of the model for the environment profile was also predicted using ERRAT (http://nihserver.mbi.ucla.edu/ERRATv2/).

**Binding site Prediction:** To find out the binding affinities, the modeled structure of BACE1 from homosapiens and the default binding sites were predicted by submitting the structure to DoGSiteScorer, an active site prediction and analysis server (Andrea Volkamer 2012).

## RESULTS AND DISCUSSION

**Phylogenetic analysis of BACE-1:** Evolution is the process where the population is altered with respect to time, and these changes in the population can be depicted in a Phylogenetic tree. The basic theory behind Phylogenetics is that members of a group share a common evolutionary history and are more related to each other than to the other group members. In this research study, the hierarchical clustering method indicates and infers that the Cladogram, as shown in Figure 1; the Clade of four Secretase sequences found in the Protein Data Bank (PDB) (https://www.rcsb.org/) were selected with its PDB ID's; BACE-1 (6EJ2), BACE-2 (2EWY), and gamma-secretase (5A63) from homosapiens respectively. Furthermore, we successfully constructed phylogenetic relationships of Secretase sequences and found that the average supporting values of nodes on each branch were higher than 50%. However, beta-secretase one and beta-secretase two are related, but the gamma-secretase is distinct; Tree shows a high similarity between BACE1 and BACE2.

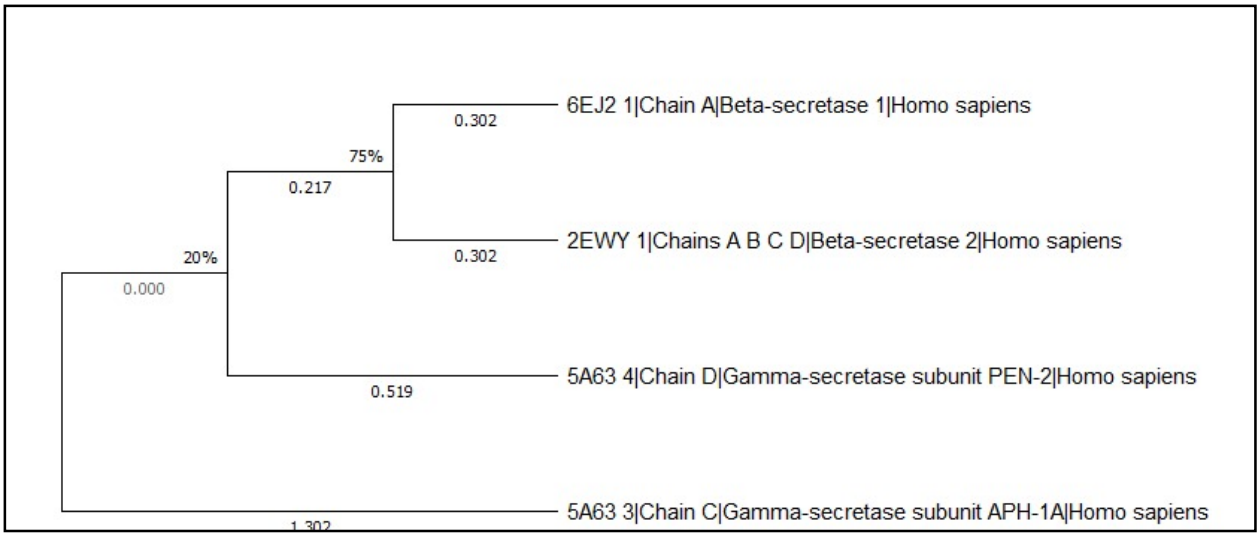



**Figure 1: The phylogenetic analysis of Secretase proteins from homosapiens:** Phylogenetic tree of the closest homolog for the protein, Beta-secretase with known 3D structure obtained with the BLAST server against PDB.

**Homology modeling and validation of the structure**: The structure of BACE1 has been determined by using a homology modeling protocol. Based on the maximum identity and high score, the results were categorized and showed homology with the BACE-1 in the complex of compound #32 from Homo sapiens (PDB code 6UWP) with 99 % identity. This structure was selected as a template for further conformational studies by Structural Analysis and Verification Server (SAVES) (http://nihserver.mbi.ucla.edu/SAVES). These verification programs rely on the Ramachandran plot that determines the quality of the predicted structure by assessing different parameters such as lengths, angles, and planarity of the peptide bonds, the geometry of the hydrogen bonds, and side-chain conformations of protein structures as a function of atomic resolution (Hollingsworth, S. A., & Karplus, P. A. 2010). The model that satisfied all the validation criteria based on PROCHECK is presented in the Ramachandran plot (Fig 2 and Table 1) showed that the Φ/Ψ (Phi/Psi) angles of modeled protein generated from the BACE1 sequence were (93.5%) residues are in the most favored regions, 6.5% residues in the additional allowed region, 0% residues in generously allowed region, and 0 % residues are in the disallowed region, confirming that the predicted model is of good quality. The Verify 3D server indicated that 97.71% of the residues in modeled structure had an average 3D-1D score > 0.2, thereby verifying the model. For the present 3D model, the overall quality factor predicted by the ERRAT server

was 85.195. ERRAT is a so-called overall quality factor for nonbonded atomic interactions, with higher scores indicating higher quality. The generally accepted range is >50 for a high-quality model (Colovos 1993).

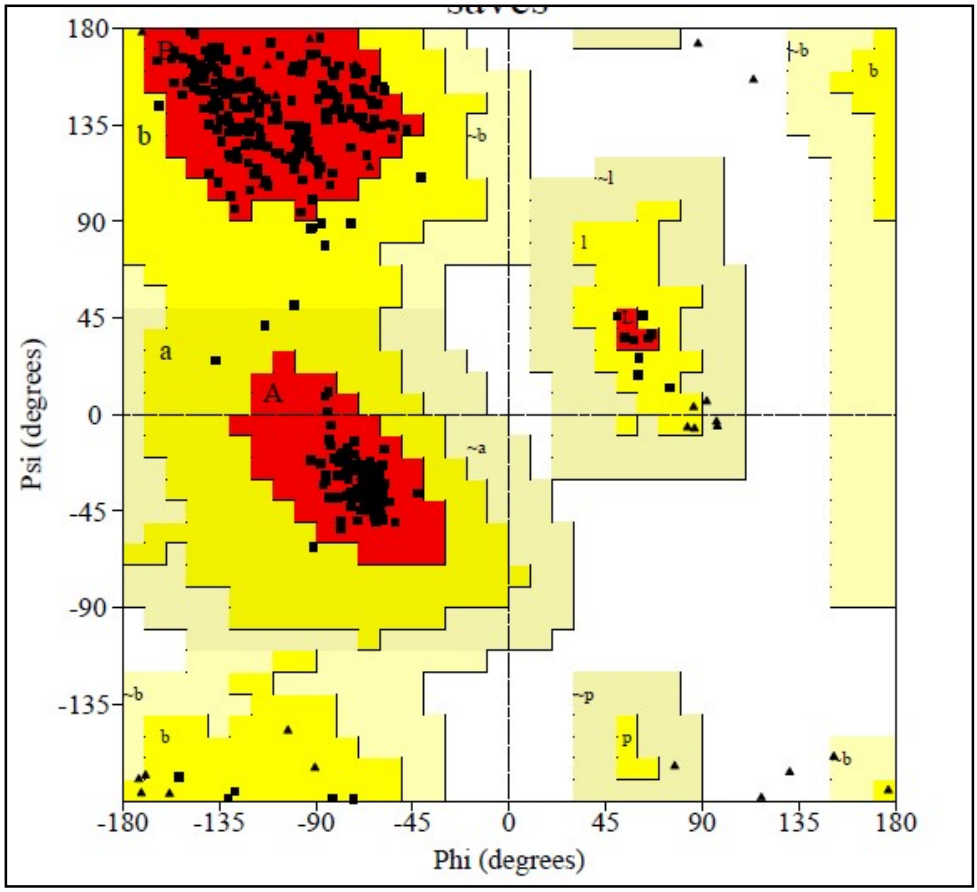



**Figure 2:** Ramachandran plot of modeled protein derived from PROCHECK.

Table 1: Ramachandran plot statistics of modeled protein.

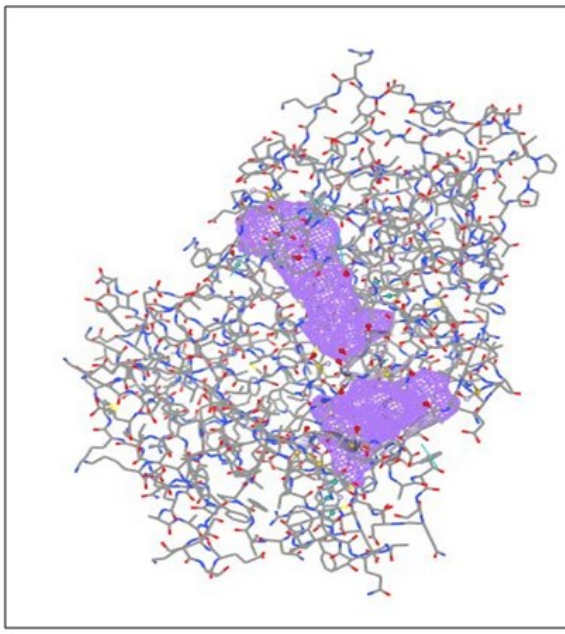
| PLOT STATISTICS | |
|---|---|
| **RESIDUES** | **PERCENTAGE (%)** |
| Residues in the most favored regions (A, B, L) | 93.5 |
| Residues in additional allowed regions (a, b, l, p]) | 6.5 |
| Residues in generously allowed regions (-a, -b, -l, -p) | 0 |
| Residues in disallowed regions | 0 |

**Superimposition of Template and Modeled structures**: This work studied the superimposition of protein structures at the residue level with dynamically weighted RMSD (root-mean-square

deviation). Superimpositions and plots of residue fluctuations of proteins identified protein domains and suggested possible protein motions. A significant problem in protein modeling is positioning a given group of protein structures in three-dimensional space once they are determined, i.e., protein structure alignment or superimposition (Wu, D. and Wu, Z. 2010). A conventional approach to superimposing a group of structures is to translate and rotate the structures so that the arithmetic average of the coordinate differences of the corresponding atoms in the structures, called the RMSD, is minimized. Here, the best superimposition of the structures is obtained when the minimal possible root-mean-square deviation is reached and is used to measure the similarity of the structures. When superimposed and aligned, i.e., the PDB template structure with modeled structure and the RMSD was observed (Fig 3). The RMSD value of 6UWP vs. modeled structure was 0.798 $A^0$.

**Figure 3:** Regular root mean squares deviation (RMSD) superimposition of 6UWP (Template) and modeled structure: The picture shows the superimposition of proteins 6UWP (red) and Modeled protein (green) visualized by Pymol.

**Binding site Prediction:** The modeled structure submitted to DoGSiteScorer, by default, detected two potential pockets on the surface of the modeled protein structure, listed in Table 2. These pockets are estimated to be druggable and were used for further docking studies. The properties of the searched pockets showed best with pocket two, such as hydrogen bond acceptors, hydrogen bond donors, and hydrophobic contacts, respectively. Furthermore, when compared with the surface volume ratio, pocket two also showed highest (0.54) than pocket one. (Figure 4). Further, as the target is not available for every protein-ligand interaction, using these physicochemical properties, the potential binding sites could be identified to analyze molecular interaction studies.



A

| Sno | Properties | Pocket 1 | Pocket 2 |
|---|---|---|---|
| 1 | Acceptors | 16 | 18 |
| 2 | Dogsite Simple Score | 0.07 | 0.08 |
| 3 | Donors | 9 | 12 |
| 4 | Hydrophobicity | 0.64 | 0.64 |
| 5 | Metal | 0 | 0 |
| 6 | Surface | 167.04 $A^0$ | 197.76 $A^0$ |
| 7 | Surface-Volume-Ratio | 0.43 | 0.54 |
| 8 | Volume | 307.20 $A^0$ | 294.91 $A^0$ |

B

**Figure 4: Detection of potential binding pockets in modeled structure by DoGSiteScorer**: A) Two binding pockets were identified by default B) The geometric and physicochemical properties of potential binding pockets.

## CONCLUSION

In conclusion, the three-dimensional structure of BACE1 was generated, which is helpful to study protein-ligand interactions to develop drugs to target AD. The model predicted by

Modeller in this study showed 93.5% residues in the most favored regions and 0% of residues in the disallowed regions, which indicates the obtained model's good quality. The superimposition suggested by the RMSD value may also have the potential to become a starting point for researchers when studying protein conformational transformation and molecular dynamic simulations. Furthermore, the mode of inhibitions and proving certain drugs' activity can be studied with this predicted model using *in silico* approach as it is the therapeutic target for AD.

## REFERENCES


Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., de Castro, E., Duvaud, S., Flegel, V., Fortier, A., Gasteiger, E., Grosdidier, A., Hernandez, C., Ioannidis, V., Kuznetsov, D., Liechti, R., Moretti, S., Mostaguir, K., Redaschi, N., Rossier, G., … Stockinger, H. (2012). Expasy: Sib Bioinformatics Resource Portal. *Nucleic Acids Research*, *40*(W1). https://doi.org/10.1093/nar/gks400

Bertram, L., McQueen, M. B., Mullin, K., Blacker, D., & Tanzi, R. E. (2007). Systematic meta-analyses of Alzheimer Disease Genetic Association Studies: The alzgene database. *Nature Genetics*, *39*(1), 17–23. https://doi.org/10.1038/ng1934

Borchelt, D. R., Ratovitski, T., van Lare, J., Lee, M. K., Gonzales, V., Jenkins, N. A., Copeland, N. G., Price, D. L., & Sisodia, S. S. (1997). Accelerated amyloid deposition in the brains of transgenic mice coexpressing mutant presenilin 1 and amyloid precursor proteins. *Neuron*, *19*(4), 939–945. https://doi.org/10.1016/s0896-6273(00)80974-5

Braak, H., & Braak, E. (1991). Neuropathological stageing of alzheimer-related changes. *Acta Neuropathologica*, *82*(4), 239–259. https://doi.org/10.1007/bf00308809

Colovos, C., & Yeates, T. O. (1993). Verification of protein structures: Patterns of nonbonded atomic interactions. *Protein Science*, *2*(9), 1511–1519. https://doi.org/10.1002/pro.5560020916

Eisenberg, D., Lüthy, R., & Bowie, J. U. (1997). [20] VERIFY3D: Assessment of protein models with three-dimensional profiles. *Methods in Enzymology*, 396–404. https://doi.org/10.1016/s0076-6879(97)77022-8

Goate, A., Chartier-Harlin, M.-C., Mullan, M., Brown, J., Crawford, F., Fidani, L., Giuffra, L., Haynes, A., Irving, N., James, L., Mant, R., Newton, P., Rooke, K., Roques, P., Talbot, C., Pericak-Vance, M., Roses, A., Williamson, R., Rossor, M., … Hardy, J. (1991). Segregation of a missense mutation in the amyloid precursor protein gene with familial alzheimer's disease. *Nature*, *349*(6311), 704–706. https://doi.org/10.1038/349704a0

Hollingsworth, S. A., & Karplus, P. A. (2010). A fresh look at the Ramachandran plot and the occurrence of standard structures in proteins. *BioMolecular Concepts*, *1*(3-4), 271–283. https://doi.org/10.1515/bmc.2010.022

Hong, L., He, X., Huang, X., Chang, W., & Tang, J. (2004). Structural features of human memapsin 2 (β-secretase) and their biological and pathological implications. *Acta Biochimica Et Biophysica Sinica*, *36*(12), 787–792. https://doi.org/10.1093/abbs/36.12.787

Huang, Y., & Mucke, L. (2012). Alzheimer mechanisms and therapeutic strategies. *Cell*, *148*(6), 1204–1222. https://doi.org/10.1016/j.cell.2012.02.040

Kumar, A., Singh, A., & Ekavali. (2015). A review on alzheimer's disease pathophysiology and its management: An update. *Pharmacological Reports*, *67*(2), 195–203. https://doi.org/10.1016/j.pharep.2014.09.004

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). Mega X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, *35*(6), 1547–1549. https://doi.org/10.1093/molbev/msy096

Laskowski, R. A., MacArthur, M. W., Moss, D. S., & Thornton, J. M. (1993). Procheck: A program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography*, *26*(2), 283–291. https://doi.org/10.1107/s0021889892009944

Mattson, M. P. (2004). Pathways towards and away from alzheimer's disease. *Nature*, *430*(7000), 631–639. https://doi.org/10.1038/nature02621

McKhann, G. M., Knopman, D. S., Chertkow, H., Hyman, B. T., Jack, C. R., Kawas, C. H., Klunk, W. E., Koroshetz, W. J., Manly, J. J., Mayeux, R., Mohs, R. C., Morris, J. C.,

Rossor, M. N., Scheltens, P., Carrillo, M. C., Thies, B., Weintraub, S., & Phelps, C. H. (2011). The diagnosis of dementia due to alzheimer's disease: Recommendations from the National Institute on aging-alzheimer's association workgroups on diagnostic guidelines for alzheimer's disease. *Alzheimer's & Dementia*, *7*(3), 263–269. https://doi.org/10.1016/j.jalz.2011.03.005

Murphy, M. P., & LeVine, H. (2010). Alzheimer's disease and the amyloid-β peptide. *Journal of Alzheimer's Disease*, *19*(1), 311–323. https://doi.org/10.3233/jad-2010-1221

Prince, M., Bryce, R., Albanese, E., Wimo, A., Ribeiro, W., & Ferri, C. P. (2013). The global prevalence of dementia: A systematic review and metaanalysis. *Alzheimer's & Dementia*, *9*(1), 63. https://doi.org/10.1016/j.jalz.2012.11.007

Querfurth, H. W., & LaFerla, F. M. (2010). Alzheimer's disease. *New England Journal of Medicine*, *362*(4), 329–344. https://doi.org/10.1056/nejmra0909142

Rochette, M. J., & Murphy, M. P. (2002). Γ-secretase: Substrates and inhibitors. *Molecular Neurobiology*, *26*(1), 081–096. https://doi.org/10.1385/mn:26:1:081

Šali, A., & Blundell, T. L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *Journal of Molecular Biology*, *234*(3), 779–815. https://doi.org/10.1006/jmbi.1993.1626

Scodellaro, C., & Pin, S. (2011). The ambiguous relationships between aging and alzheimer's disease: A critical literature review. *Dementia*, *12*(1), 137–151. https://doi.org/10.1177/1471301211421230

Selkoe, D. J. (1991). The molecular pathology of alzheimer's disease. *Neuron*, *6*(4), 487–498. https://doi.org/10.1016/0896-6273(91)90052-2

Selkoe, D. J. (2001). Alzheimer's disease: Genes, proteins, and therapy. *Physiological Reviews*, *81*(2), 741–766. https://doi.org/10.1152/physrev.2001.81.2.741

Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). Clustal W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, *22*(22), 4673–4680. https://doi.org/10.1093/nar/22.22.4673

Volkamer, A., Kuhn, D., Rippmann, F., & Rarey, M. (2012). Dogsitescorer: A web server for automatic binding site prediction, analysis and Druggability Assessment. *Bioinformatics*, *28*(15), 2074–2075. https://doi.org/10.1093/bioinformatics/bts310

Vyas, V. K., Ukawala, R. D., Chintha, C., & Ghate, M. (2012). Homology modeling a fast tool for drug discovery: Current perspectives. *Indian Journal of Pharmaceutical Sciences*, *74*(1), 1. https://doi.org/10.4103/0250-474x.102537

Walsh, D. M., & Selkoe, D. J. (2007). A? oligomers ? A decade of Discovery. *Journal of Neurochemistry*, *101*(5), 1172–1184. https://doi.org/10.1111/j.1471-4159.2006.04426.x

Wimo, A., Guerchet, M., Ali, G. C., Wu, Y. T., Prina, A. M., Winblad, B., Jönsson, L., Liu, Z., & Prince, M. (2016). The worldwide costs of dementia 2015 and comparisons with 2010. *Alzheimer's & Dementia*, *13*(1), 1–7. https://doi.org/10.1016/j.jalz.2016.07.150

Wu, D., & Wu, Z. (2009). Superimposition of protein structures with dynamically weighted RMSD. *Journal of Molecular Modeling*, *16*(2), 211–222. https://doi.org/10.1007/s00894-009-0538-6