Analysis of mutations in APP1 protein associated with development and protection against Alzheimer's disease - an *In silico* approach

Análise de mutações na proteína APP1 associadas ao desenvolvimento e proteção contra a doença de Alzheimer - uma abordagem *In silico*

DOI:10.34117/bjdv8n6-278

Recebimento dos originais: 21/04/2022
Aceitação para publicação: 31/05/2022

Gabriel Rodrigues Coutinho Pereira
M.Sc
Instituição: Universidade Federal do Estado do Rio de Janeiro (UNIRIO)
Endereço: Avenida Pasteur, 296, CEP: 22290-250, Urca, Rio de Janeiro - RJ
E-mail: gabrielkytz@hotmail.com

Maíra de Oliveira Torres
M.Sc
Instituição: Universidade Federal do Estado do Rio de Janeiro (UNIRIO)
Endereço: Avenida Pasteur, 296, CEP: 22290-250, Urca, Rio de Janeiro - RJ
E-mail: torres.biomed@gmail.com

Vicente Salgado Pires
M.Sc
Instituição: Universidade Federal do Estado do Rio de Janeiro (UNIRIO)
Endereço: Avenida Pasteur, 296, CEP: 22290-250, Urca, Rio de Janeiro - RJ
E-mail: visalpi1@hotmail.com

Joelma Freire de Mesquita
Ph.D
Instituição: Universidade Federal do Estado do Rio de Janeiro (UNIRIO)
Endereço: Avenida Pasteur, 296, CEP: 22290-250, Urca, Rio de Janeiro - RJ
E-mail: joelma.mesquita@unirio.br

ABSTRACT

Introduction: Alzheimer's disease (AD) is the dementia with the highest number of cases worldwide, causing great social and economic impact. The amyloid cascade is the most accepted hypothesis to explain the beginning of AD. According to it, neurodegeneration is caused by the accumulation of Aβ peptides and the formation of amyloid plaques in the brain. In familial cases of Alzheimer's, mutations in the APP1 protein lead to increased production of Aβ plaques. Objectives: Analyze *in silico* the structural and functional impact of missense mutations in APP1 and construct a complete model for the protein. Methods: We generated and validated a theoretical structure of APP1 protein using structural modeling and quality assessment algorithms. We further analyzed the effects of AD-related mutations on APP1 protein and also the neuroprotective mutation A673T by performing functional predictions, evolutionary conservation analysis, and molecular dynamics (MD). Results: The predictive analysis indicated that most mutations occur in conserved regions of APP1 and also present an elevated rate of deleterious predictions, pointing to their harmful effects. The computational modeling generated an

unprecedented, accurate and complete model of human APP1, whose quality was corroborated by validation algorithms and structural alignment. The MD simulations of codon 673 variants pointed to flexibility and essential dynamics alterations at the AICD and Aβ domains, which could have strong and non-intuitive consequences on APP1 interactions, including those involved in β-secretase cleavage and, consequently, aβ peptide formation. Conclusions: Flexibility and essential dynamics alterations upon codon 673 variants may have functional implications for APP1, influencing the generation of Aβ peptide, the main responsible for APP1 toxicity in AD.

Keywords: Alzheimer's disease, amyloid precursor protein 1, *In silico.*

RESUMO

Introdução: A doença de Alzheimer (AD) é a demência com o maior número de casos a nível mundial, causando um grande impacto social e económico. A cascata amilóide é a hipótese mais aceite para explicar o surgimento da doença de Alzheimer. Segundo ela, a neurodegeneração é causada pela acumulação de peptídeos Aβ e pela formação de placas amilóides no cérebro. Em casos familiares de doença de Alzheimer, as mutações na proteína APP1 levam a um aumento da produção de placas Aβ. Objectivos: Analisar em silico o impacto estrutural e funcional das mutações de missense na proteína APP1 e construir um modelo completo para a proteína. Métodos: Gerámos e validámos uma estrutura teórica da proteína APP1 utilizando modelos estruturais e algoritmos de avaliação da qualidade. Analisámos ainda os efeitos das mutações relacionadas com a AD na proteína APP1 e também a mutação neuroprotectora A673T através da realização de previsões funcionais, análise de conservação evolutiva, e dinâmica molecular (MD). Resultados: A análise preditiva indicou que a maioria das mutações ocorrem em regiões conservadas de APP1 e também mostram uma elevada taxa de previsões deletérias, apontando para os seus efeitos nocivos. A modelação computacional gerou um modelo de APP1 humano sem precedentes e completo, cuja qualidade foi corroborada por algoritmos de validação e alinhamento estrutural. As simulações MD das 673 variantes do códão apontaram para flexibilidade e mudanças essenciais na dinâmica nos domínios AICD e Aβ, o que poderia ter consequências fortes e não-intuitivas para as interacções APP1, incluindo as envolvidas na clivagem secretora de β e, consequentemente, a formação do peptídeo aβ. Conclusões: Flexibilidade e mudanças essenciais na dinâmica do códão 673 variantes podem ter implicações funcionais para o APP1, influenciando a geração do péptido Aβ, que é o principal responsável pela toxicidade do APP1 na AD.

Palavras-chave: doença de Alzheimer, proteína precursora amilóide 1, *In silico.*

1 INTRODUCTION

Alzheimer's disease (AD) represents about 60% of all dementias. It is estimated that more than 35 million people are affected by AD worldwide (FERRI *et al.*, 2005). Alzheimer's also has a major social and financial impact. In 2019, AD brought to the United States alone a total outlay of approximately $290 billion, which is estimated to rise to $1.1 trillion in just over thirty years (ALZHEIMER'S ASSOCIATION, 2021). At this time, Alzheimer's doesn't have a cure, and the anti-Alzheimer's drugs currently

approved by the FDA just aim to manage the symptoms and provide a better quality of life (LONG; HOLTZMAN, 2019).

The amyloid cascade is the most accepted hypothesis to explain the development of AD. According to it, neurodegeneration is caused by the accumulation of β-amyloid (Aβ) peptides and the formation of amyloid plaques in the brain (ROSENBERG *et al.*, 2016). In its normal processing, the Amyloid Precursor Protein 1 (APP1) undergoes proteolytic catalysis by α-secretase, generating non-significant levels of Aβ peptide. However, under some conditions, the cleavage of the C-terminal portion of APP1 by the enzymes aspartyl protease β-site APP cleaving enzyme 1 (BACE1) and subsequently γ-secretase leads to the formation of significant and, usually, toxic levels of Aβ peptide (LONG; HOLTZMAN, 2019). In familial cases of Alzheimer's, mutations in the APP lead to an increase in the production of beta-amyloid plaques (SELKOE, 2001).

Despite the central role of APP1 in the pathogenesis of AD, no complete three-dimensional structure has been experimentally determined for this protein. Important regions to APP1 function and pathogenicity are still unknown (BATEMAN *et al.*, 2017), including the Aβ domain, directly involved in the production of Aβ plaques (O'BRIEN; WONG, 2011). In this work, following the methodology previously established by our group (Pereira et al., 2021; Pereira et al., 2022), we generated *in silico* an unprecedented, accurate, and complete three-dimensional structure of human APP1, as well as analyzed the functional effects of seventy-one APP1 missense mutations most of them not yet characterized. We also performed a thoroughly MD analysis of codon 673 variants, which affect a key residue for the generation of Aβ peptide (JONSSON *et al.*, 2012).

2 MATERIALS AND METHODS

A workflow containing the methodology applied to this work is shown in S1 Fig.

2.2 DATASET

The native sequence of APP1 protein was obtained in the UNIPROT database (Uniprot ID: P05067) (BATEMAN, 2019). APP1 mutations were compiled from the OMIM (HAMOSH *et al.*, 2000) (ID: 104760), UNIPROT, Molgen (Cruts and Van-Broeckhoven, 2018), and ALZforum databases (KINOSHITA; CLARK, 2007).

2.3 FUNCTIONAL PREDICTION ANALYSIS

The native sequence of APP1 and its variants were submitted to prediction

algorithms. The algorithms MutPred, PhD-SNP, PMut, Provean, PredictSNP, MAPP, SIFT, PolyPhen2, SNAP, and SNPs & GO were used to predict the functional effects of APP1 mutations (PEREIRA; VIEIRA; MESQUITA, 2021).

2.4 STRUCTURAL MODELING

Complete three-dimensional models of wild-type APP1 protein were generated using the following algorithms: IntFold, RaptorX, SwissModel Rosetta, I-Tasser, and Phyre2 (SCHWEDE, 2013). The template structures were selected using the Protein Blast server based on sequence coverage and identity. The selected template structures (i.e. PDB IDs: 3KTM; 3L3T; 5BUO; 3DXC) were structurally aligned to the complete theoretical models of APP1 in the TM-Align Server (ALBÀ; CASTRESANA, 2007).

2.5 STRUCTURE VALIDATION

The selected model, which is the most structurally similar to the template structures (PDB IDs: 3KTM_A; 3L3T_E; 5BUO_A; 3DXC_B), was submitted to ERRAT, PROCHECK (WILLIAMS *et al.*, 2018), RAMPAGE, MolProbity (Lovell et al., 2003), ProSa, and QMEAN (WEBB; SALI, 2016) validation servers.

2.6 EVOLUTIONARY CONSERVATION ANALYSIS

The evolutionary conservation of each amino acid of APP1 was analyzed using the default parameters in the ConSurf server (ASHKENAZY *et al.*, 2016).

2.7 MOLECULAR DYNAMICS

We performed Molecular Dynamics (MD) simulations of APP1 wild-type and its 673 codon variants (i.e. A673T, A673V, and A673Y). MutatorPlugin1.3, available in the VMD 1.9.3 software (HUMPHREY; DALKE; SCHULTEN, 1996) was used to induce the A673T, A673V, and A673Y substitutions on the validated model of APP1.

Following the methodology previously established by our group (PEREIRA; VIEIRA; MESQUITA, 2021), MD simulation of APP1 wild-type and its variants was performed using the GROMACS 2018.8 package (HESS; SPOEL; LINDAHL, 2014), and the AMBER99SB-ILDN force field. The structures were solvated in a triclinic box using TIP3P water. The systems were neutralized by adding Na+ Cl- ions and minimized using the steepest descent method. After system minimization, NVT and NPT ensembles were performed at 1atm and 300K for the duration of 100 ps using the V-rescale

thermostat and Parrinello-Rahman barostat.

The production simulations were then performed at 300K for the duration of 150ns using the PME and LINCS algorithms (PEREIRA; VIEIRA; MESQUITA, 2021). The following parameters were analyzed from the MD trajectories using GROMACS distribution programs: root-mean-square deviation (RMSD), root-mean-square fluctuation (RMSF), and B-factor. The MD trajectories were also analyzed using the Bio3D library (GRANT *et al.*, 2006) implemented in R software to perform a Principal Component Analysis (PCA).

3 RESULTS AND DISCUSSION

3.1 DATASET

The FASTA sequence of the APP1 protein has 770 amino acid residues, which cover its complete sequence length (BATEMAN *et al.*, 2017). APP1 functional domains are shown in Fig 1. The E1 domain (residues 18 – 190) binds to heparin, Cu, and Zn, and can stimulate neurite outgrowth. The E2 domain (residues 295 – 500) is involved in APP1 dimerization (REINHARD; HÉBERT; DE STROOPER, 2005).

The E1 and E2 domains are connected through the acidic domain (AcD) (228 – 294), a flexible and negatively charged region (COBURGER *et al.*, 2013). Between the E1 and AcD domains, there is a partially flexible region known as the extension domain (residues 191 – 227) (REINHARD; HÉBERT; DE STROOPER, 2005). The Aβ domain, directly involved in the formation of Aβ peptides, is located between the residues 672 and 720. At last, the C-terminal domain (AICD), which participates in several APP1 interactions, is located between residues 721 and 770 (O'BRIEN; WONG, 2011).

Seventy-one mutations in APP1 protein were compiled from the databases and literature, as shown in Fig 1. Detailed information in mutations compiled is shown in the S1 Table.

3.2 FUNCTIONAL PREDICTION

As shown in S2 Fig, 10% of the mutations were predicted as deleterious by all the functional prediction algorithms used (A500T, V710G, V717A, L720S, M722K, L723R, and K724M), indicating that these mutations are the most likely to present harmful effects on protein function. Moreover, 45% percent of the analyzed mutations were predicted as deleterious by at least seven algorithms (S2 Fig), which shows the harmful potential of APP1 mutations. The individual results of these algorithms are shown in the S2 Table.

A group of frequent Alzheimer's-related mutations in APP1 occurring at the Aβ domain (residues 672 and 720) have been already characterized. The mutations A673T and A673V are already known to be neuroprotective and deleterious, respectively (JONSSON *et al.*, 2012). Both mutations were only predicted as deleterious by MutPred. A673Y, in turn, also affects codon 673 but is not yet characterized. A673Y was predicted as deleterious for five functional prediction algorithms. The most frequent APP1 mutation, i.e. V717I (London mutation) increases Aβ levels, being AD-related (MURATORE *et al.*, 2014). Nonetheless, the deleterious potential of V717I was only detected by four functional prediction algorithms. Three other codon 717 mutations (V717F, V717G, and V717L) were predicted as deleterious by at least eight prediction algorithms (S2 Table), which is in agreement with the literature consulted (S1 Table). At last, Mutation A713T, already associated with the development of familial AD (LANOISELÉE *et al.*, 2017), was predicted as deleterious by nine algorithms (S2 Table).

Our functional prediction results presented high accuracy in predicting the functional effect of variants A673T, V717F, V717G, V717L, and A713T. Nonetheless, they presented low accuracy in detecting the deleterious effect of A673V, which highlights the importance of using an ensemble of functional prediction algorithms, as previously shown by our group, particularly regarding the different prediction methods they use (PEREIRA; VIEIRA; MESQUITA, 2021).

3.3 STRUCTURAL MODELLING

No complete three-dimensional structure of human APP1 protein has been experimentally determined (BATEMAN, 2019), which impairs the analysis of mutations located in undetermined regions and hamper the development of anti-Alzheimer drugs targeting APP1 (NAQVI *et al.*, 2019). The structural modeling of wild-type APP1 returned 22 models, from which seven were complete. Among the complete models, five were generated by the Robetta (comparative and *ab initio*), and two were generated by the I-Tasser (*threading*). Only complete models were selected for the structural alignment.

A protein BLAST was performed to search for potentially related sequences of known structures for APP1, which pointed to four potential template structures in the PDB: 3KTM, 5BUO, 3L3T, and 3DXC. All of them are crystallographic fragments of the APP1 protein that present high sequence identity and variable sequence coverage, as shown in S3 Table. The computational models of APP1 were aligned on the TM-Align

server to the selected template structures, returning RMSD and TM-score values (S4 Table). TM-score values higher than 0.5 and close to 1, as well as RMSD values lower than 2Å and close to 0, indicate structural similarity between the generated models and template structures (ZHANG; SKOLNICK, 2005).

As shown in S4 Table, Model 1 generated using the Rosetta algorithm presented the lowest RMSD and the highest TM-score. This model was then selected for validation analysis due to its greater similarity to the template structures. The visual inspection of the alignment, shown in S3 Fig, further confirms the high similarity between the selected model and template structures.

3.4 STRUCTURE VALIDATION

The selected model was submitted to validation servers, which are algorithms widely used to assess model quality through stereochemical analyses and comparisons with experimentally determined protein structures (Pereira et al., 2019).

Procheck and Rampage use a Ramachandran plot to perform stereochemical analysis. Structures with high stereochemical quality should have more than 90% of the residues in favorable regions for PROCHECK, and around 98% for Rampage (Lovell et al., 2003; Williams et al., 2018). As shown in S4 Fig, the theoretical model of APP1 presented 90.5% of its residues in favorable regions on PROCHECK and 97.8% on RAMPAGE, indicating high stereochemical quality. Errat estimates the probability of error for each amino acid to generate an overall quality factor. The theoretical model of APP1 presented an ERRAT score of 96.23%, which is comparable to that of high-resolution protein structures (~95%) (S4 Fig) (WILLIAMS *et al.*, 2018).

ProSa analyzes the total free energy of a given model and calculates its overall quality, which is plotted in a graph containing the scores computed for experimentally-determined structures in the PDB (WEBB; SALI, 2016). As shown in Fig S5, the APP1 model has a Z-score of -9.8, which is comparable to that of NMR structures. QMEAN assesses geometric aspects and torsion angles to estimate the overall quality of a given model (QMEAN-score), which is plotted on a graph containing the QMEAN-scores of high-resolution crystallographic structures (WEBB; SALI, 2016). The theoretical model has a QMEAN-score of 1.00, which is comparable to high-resolution structures (Fig S5). MolProbity assesses interatomic contacts, main-chain torsion angles, and side-chain rotamer angles to estimate an overall quality structure score (MolProbity-score). The

theoretical model of APP1 presented a MolProbity-score of 1.02 (Fig S5), which is comparable to high-quality experimental structures (≤2) (CHEN *et al.*, 2010).

The theoretical model of APP1, shown in Fig 2, is accurate since it is structurally similar to the experimentally determined PDB templates (S4 Table) and had its overall quality affirmed by the algorithms ProSa, QMEAN, MolProbity, PROCHECK, RAMPAGE, and ERRAT. The model's coordinates were stored in a PDB file, which is available in S1 File.

3.5 EVOLUTIONARY CONSERVATION ANALYSIS

The ConSurf algorithm performs multiple sequence alignments to estimate the evolutionary conservation of each amino acid (ASHKENAZY *et al.*, 2016). Residues with functional importance for protein are generally conserved due to the higher selective pressure. Thus, mutations at conserved positions are most likely to lead to deleterious effects (PEREIRA; VIEIRA; MESQUITA, 2021). The APP1 structure was colored according to the ConSurf colorimetric scale (Fig 3). Detailed information on ConSurf predictions is shown in S6 Fig.

As shown in Fig3, most APP1 mutations (53%) occur at conserved positions. *i.e.* ConSurf-score ≥7. Furthermore, twenty-nine percent of all mutations occur at positions that received the maximum conservation score (ConSurf-score=9). Among them, the variant K724M (S6 Fig) was also predicted as deleterious by all the functional prediction algorithms (S2 Table). Mutations A500T, V710G, V717A, L720S, M722K, and L723R, which were predicted as deleterious by all the functional prediction algorithms (S2 Table), occur at conserved and, possibly, important positions for APP1 function (ASHKENAZY *et al.*, 2016). V717A and M722K are pathogenic, while the other variants have not yet been classified (S1 Table). Our findings then suggest that the variants A500T, V710G, L720S, L723R, and particularly K724M may be harmful to APP1 function and, consequently, also valuable targets for future investigation.

3.6 MOLECULAR DYNAMICS

Molecular dynamics is an *in silico* method of solving Newtonian equations of motions for a group of atoms applied to reproduce the behavior of macromolecules in their biological environment (PEREIRA; VIEIRA; MESQUITA, 2021). The impact of codon 673 variants on protein structure were investigated through molecular dynamics. A673T is also known as the neuroprotective mutation, and it is the only APP1 mutation

reported with this effect. This mutation inhibits BACE1 cleavage, leading to a decrease in Aβ production. The substitution of alanine A673V, on the other hand, increases Aβ production and leads to the development of Alzheimer's (JONSSON *et al.*, 2012). In turn, the effects of mutation A673Y have not yet been described in the literature.

RMSD is useful to analyze overall structural displacement and assess system equilibration throughout the simulation. As shown in Fig 4, a sudden increase in the RMSD values was observed at the beginning of all simulations, followed by a moment of structural instability. This behavior might be attributed to the initial kinetic shock that occurs in MD simulation systems (AGRAHARI *et al.*, 2018). The establishment of plateaus in the RMSD values is observed after approximately 75ns in all simulations, which indicates that the structures fluctuate around average conformations (PEREIRA; VIEIRA; MESQUITA, 2021). The initial effects observed in the MD trajectories, *i.e*. those occurring before 75ns, were then not considered in further analyses for meaningful comparison.

The RMSF and B-factor are useful measures of local structural flexibility. The B-factor values computed during the simulations for each amino acid of APP1 were projected on the protein surface, which provides an interesting three-dimensional representation of structural flexibility (Pereira et al., 2019). The RMSF and B-factor analyses, shown in Fig 5 and S7 Fig, pointed to a drastic flexibility increase at the Aβ and AICD domains of the A673T variant. Variant A673T also presented decreased flexibility at the AcD and E2 domains. These analyses pointed to flexibility alterations at the E2, Aβ, and AICD domains of variant A673V and A673Y.

PCA is a statistical technique useful to reduce the total number of dimensions needed to describe protein dynamics. PCA allows separating the essential, probable biologically relevant movements to proteins from the remaining irrelevant local fluctuations. The essential protein motions are usually confined to the first two principal component modes (PC1 and PC2) (OLIVEIRA *et al.*, 2019). The PCA analysis suggested that the first two principal components capture the dominant motions, accounting for at least 50% of the total variance. As shown in Fig 6, the variants A673T, A673V, and A673Y presented changes in the direction of the clusters and the regions explored in the conformational space when compared to the wild-type APP1, which indicates alterations in the overall essential dynamics of codon 673 variants.

We also analyzed the RMSF contribution of each amino acid to PC1 and PC2, which are shown in Fig 7 and S8 Fig. This analysis pointed to essential mobility

alterations at the AICD, acidic, E2, and Aβ domains of variant A673T and A673V when compared to the wild-type. The variant A673Y, in turn, presented essential mobility alterations at the acidic and E2 domains. These essential mobility alterations were observed to a greater and lesser extent in the analyzed variants, but the A673T variant presented the greatest alterations when compared to the wild-type.

The deposition of Aβ peptide in the brain is the main responsible for the pathogenesis of AD. The events underlying Aβ peptide generation from the proteolytic cleavage of the Aβ domain are summarized in S9 Fig. The proteolytic cleavage process by α-secretase or β-secretase, which occur at the Aβ domain, directly influences the generation of Aβ peptides. The study of Jonsson *et.al*. 2012 suggested that A673 is a key APP1 position in regulating β-cleavage site selection (JONSSON *et al.*, 2012). Similarly, the study of Kimura *et. al.* 2016 indicated that the selectivity of BACE1-APP1 interaction is heavily dependent on the Aβ substrate sequence; thus, amino acid substitutions induced by mutations, such as those observed at position 673, have a direct impact on BACE-1 selectivity (KIMURA; HATA; SUZUKI, 2016).

Alterations in protein flexibility and essential dynamics, as those observed during the MD simulations for codon 673 variants, may have important and non-intuitive consequences on protein binding properties (PEREIRA; VIEIRA; MESQUITA, 2021). According to Forrey *et. al.* 2012, even small changes in protein flexibility can considerably impact protein binding affinity and specificity (FORREY; DOUGLAS; GILSON, 2012). Our MD findings then suggest that the mutations A673T, A673V, and A673Y may favor or disfavor APP1 interactions at the acidic, E2, AICD, and Aβ domains, which were the regions most affected by flexibility and essential dynamics changes over the simulations when compared to the wild-type. It possibly includes APP1-BACE-1 interactions at the Aβ domain, previously described in the literature as affected by A673T and A673V variants. Together with the wet-lab experiments discussed above, our findings further suggest that the APP1-BACE1 interactions may be affected by codon 673 variants, differentially influencing β-cleavage and Aβ peptide formation (JONSSON *et al.*, 2012; KIMURA; HATA; SUZUKI, 2016).

4 CONCLUSION

Seventy-one missense mutations in APP1 were compiled from the databases and literature. The predictive and evolutionary analyses indicated an elevated rate of deleterious predictions for the APP1 mutations, which most affect conserved and possibly

functionally important amino acids, pointing to their harmful effects. The computational modeling generated an unprecedented, accurate and complete model of APP1 protein. Alterations in protein flexibility and essential dynamics during the MD simulations of codon 673 variants may have strong and non-intuitive consequences on APP1 binding. It could favor or disfavor APP1 interactions in the affected regions, possibly including those at the Aβ domain involved in β-secretase cleavage and, consequently, Aβ peptide formation. These alterations may have functional implications for APP1, influencing the generation of Aβ peptide, the main responsible for APP1 toxicity in AD. The results presented in this paper are available at the SNPMOL (http://www.snpmol.org/) a free human curated database developed by our group.

FIGURES

Fig 1. Graphical representation of APP1 mutations. The APP1 functional domains are highlighted in the figure. The extension domain is colored yellow. Amino acids affected by mutations are indicated in the graphical representation.

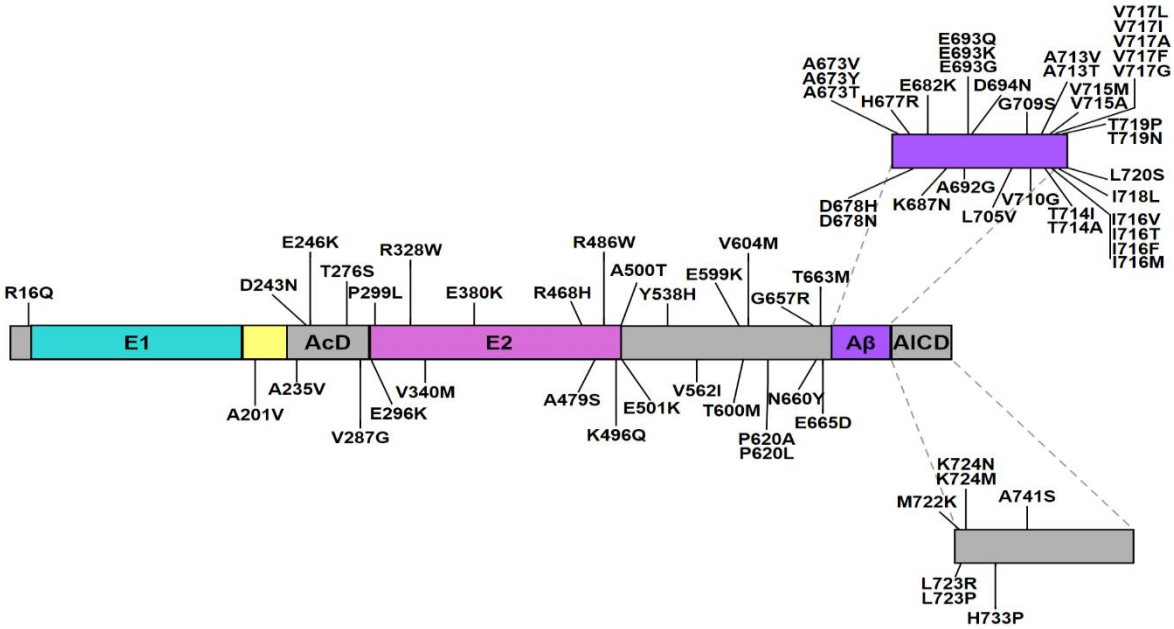


Fig 2. Validated theoretical model of human APP1. Three-dimensional representation for the theoretical model of APP1 generated by Rosetta and validated by MolProbity, Prosa, QMEAN, PROCHECK, Rampage, and ERRAT. The E1 domain, extension domain, E2 domain, and Aβ domain are represented in blue, yellow, pink, and purple, respectively. The acidic domain is represented in light gray, while the C-terminal domain is represented in dark gray.

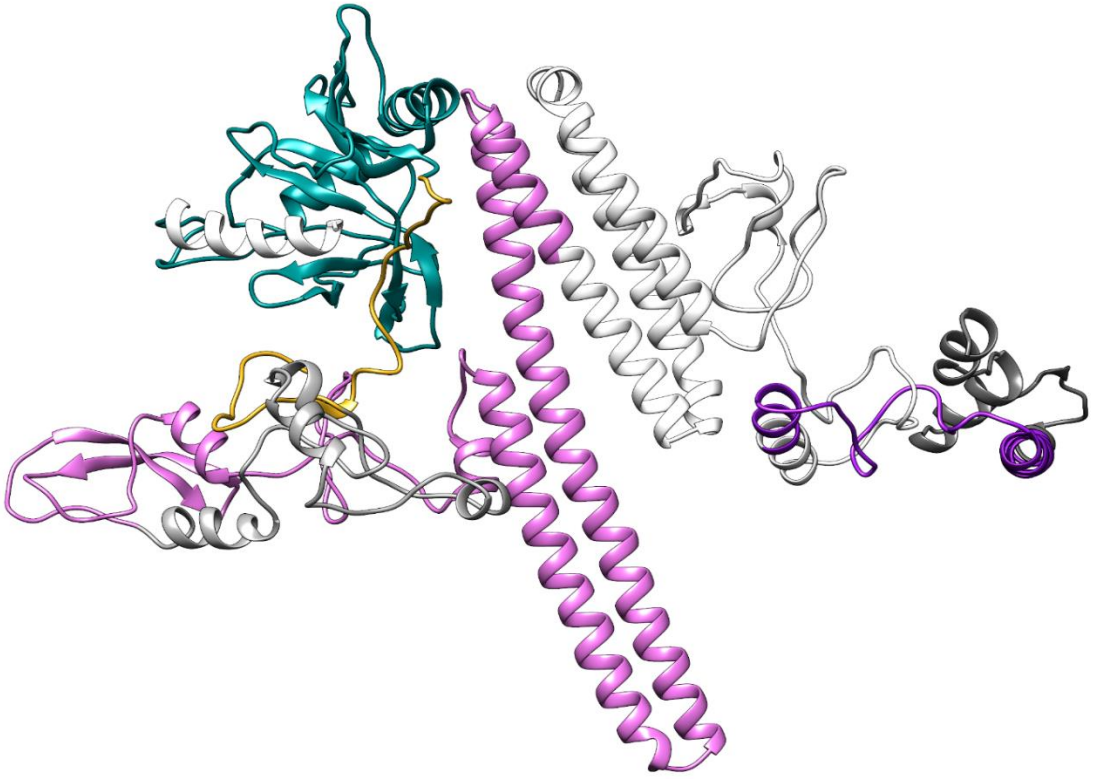


Fig 3. Evolutionary conservation analysis. (A) The conservation profile of APP1 is displayed at three angles, rotating 90 degrees from each other. Each amino acid of APP1 is represented as a space-filling model and colored according to the ConSurf conservation score, varying from cyan and highly variable to maroon and highly conserved. (B) The circle chart indicates the percentage of mutations classified according to the nine conservation levels.

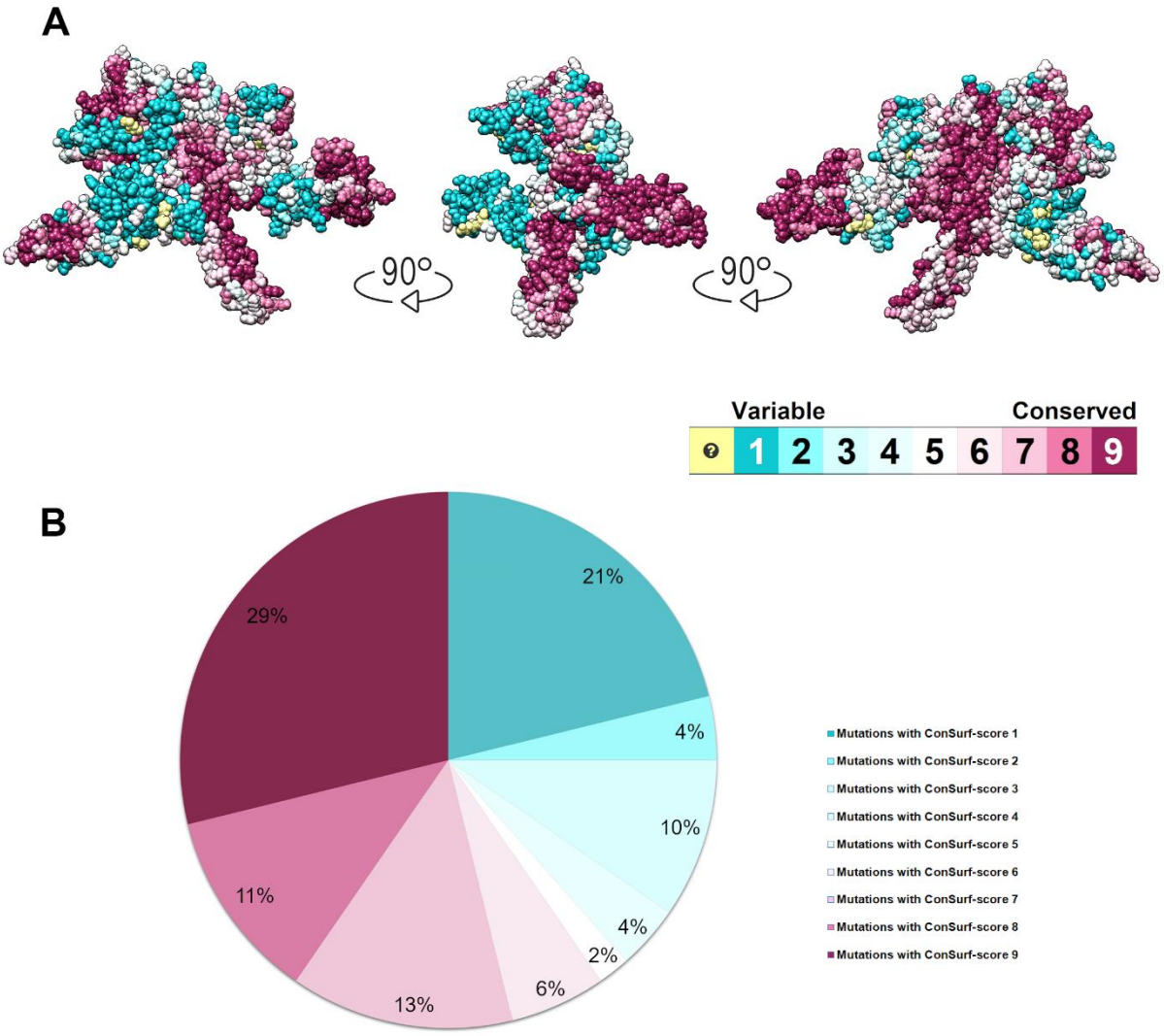


Fig 4. RMSD analysis. The RMSD values computed for the backbone atoms of wild-type APP1 and its variants are shown over time. The wild-type is represented in black, variants A673T, A673V, and A673Y are represented in orange, purple, and cyan, respectively.

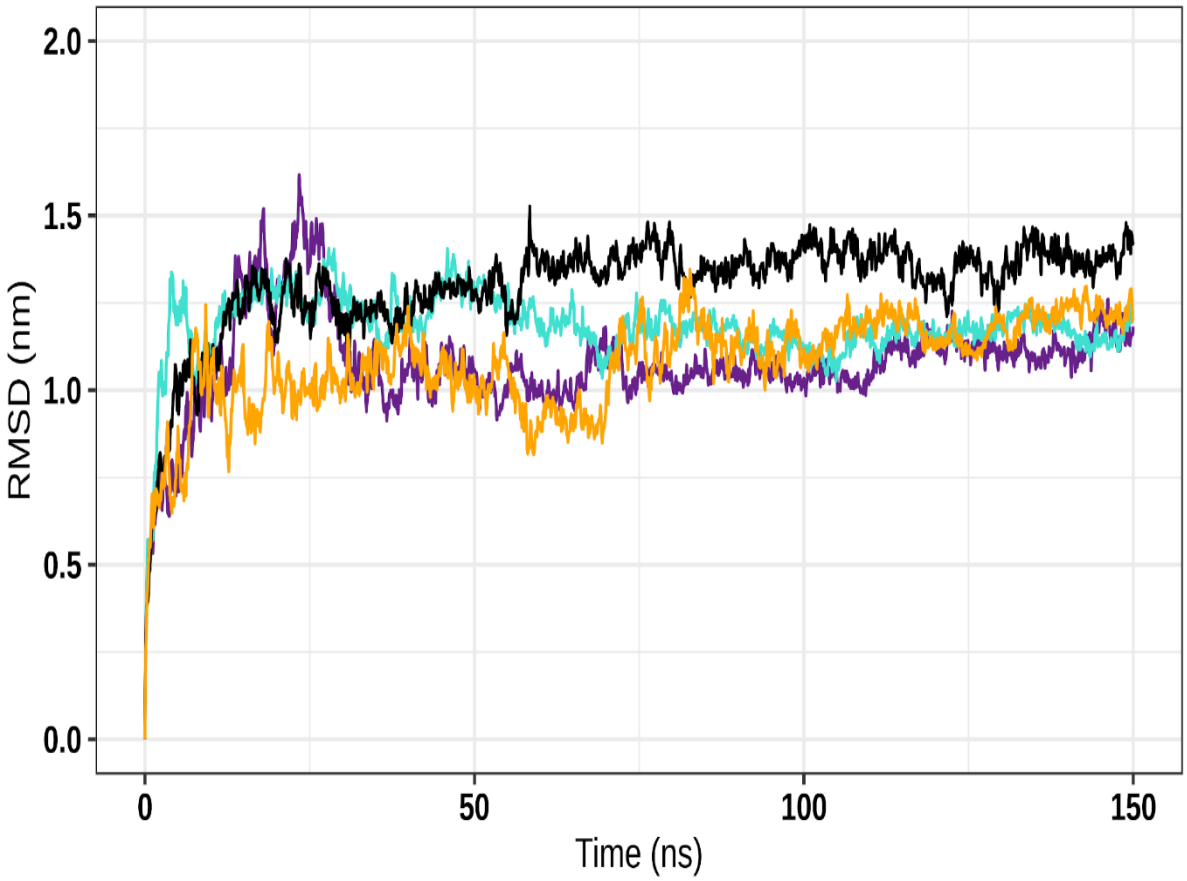


Fig 5. RMSF analysis. The RMSF values for each residue of wild-type APP1 and its variants are shown as a line plot. Schematic representations of the APP1 functional regions are also shown to further comparison. The wild-type is represented in black, variants A673T, A673V, and A673Y are represented in orange, purple, and cyan, respectively.

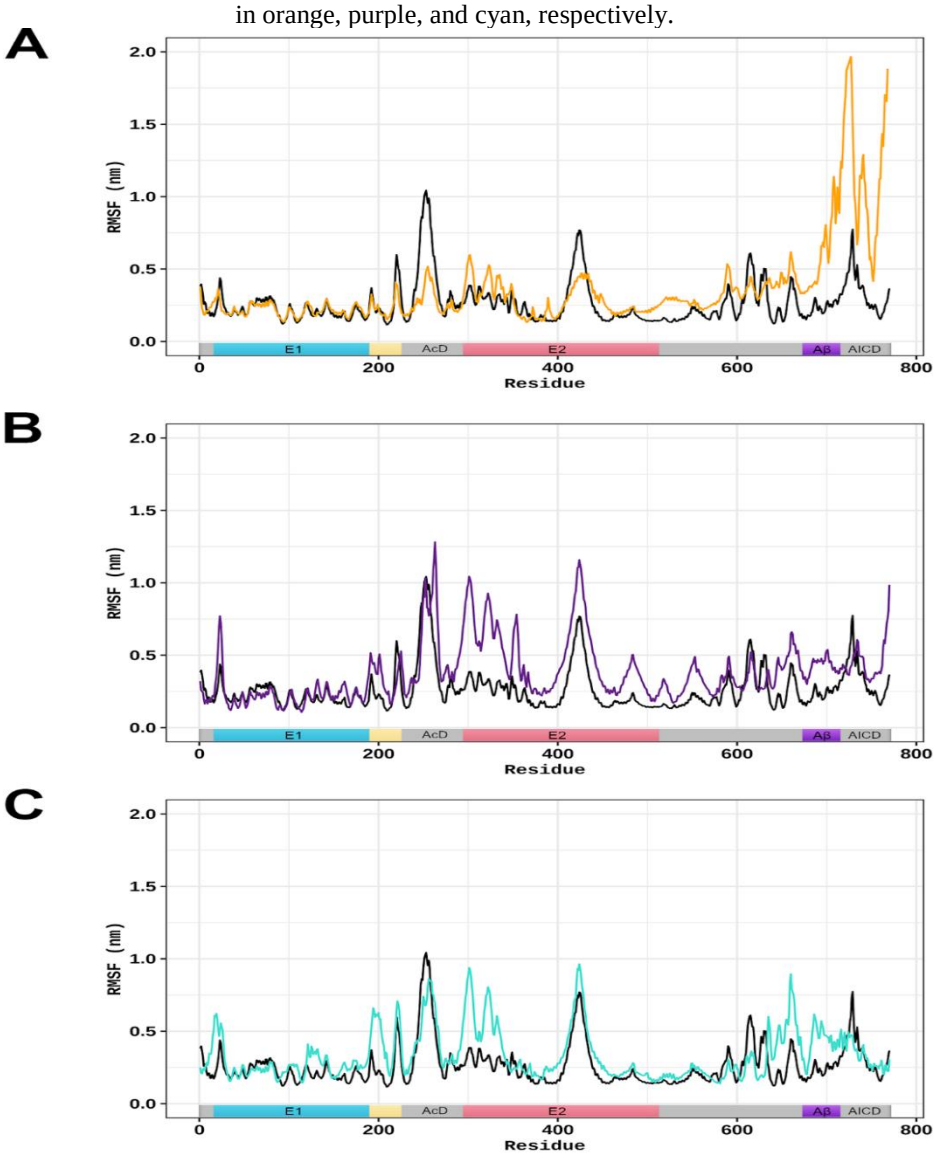


Fig 6. Principal Component Analysis. Projections of the first two principal components (PC1 and PC2) extracted from the MD trajectories of wild-type APP1 (black), variant A673T (orange), variant A673V (purple), and variant A673Y (cyan).

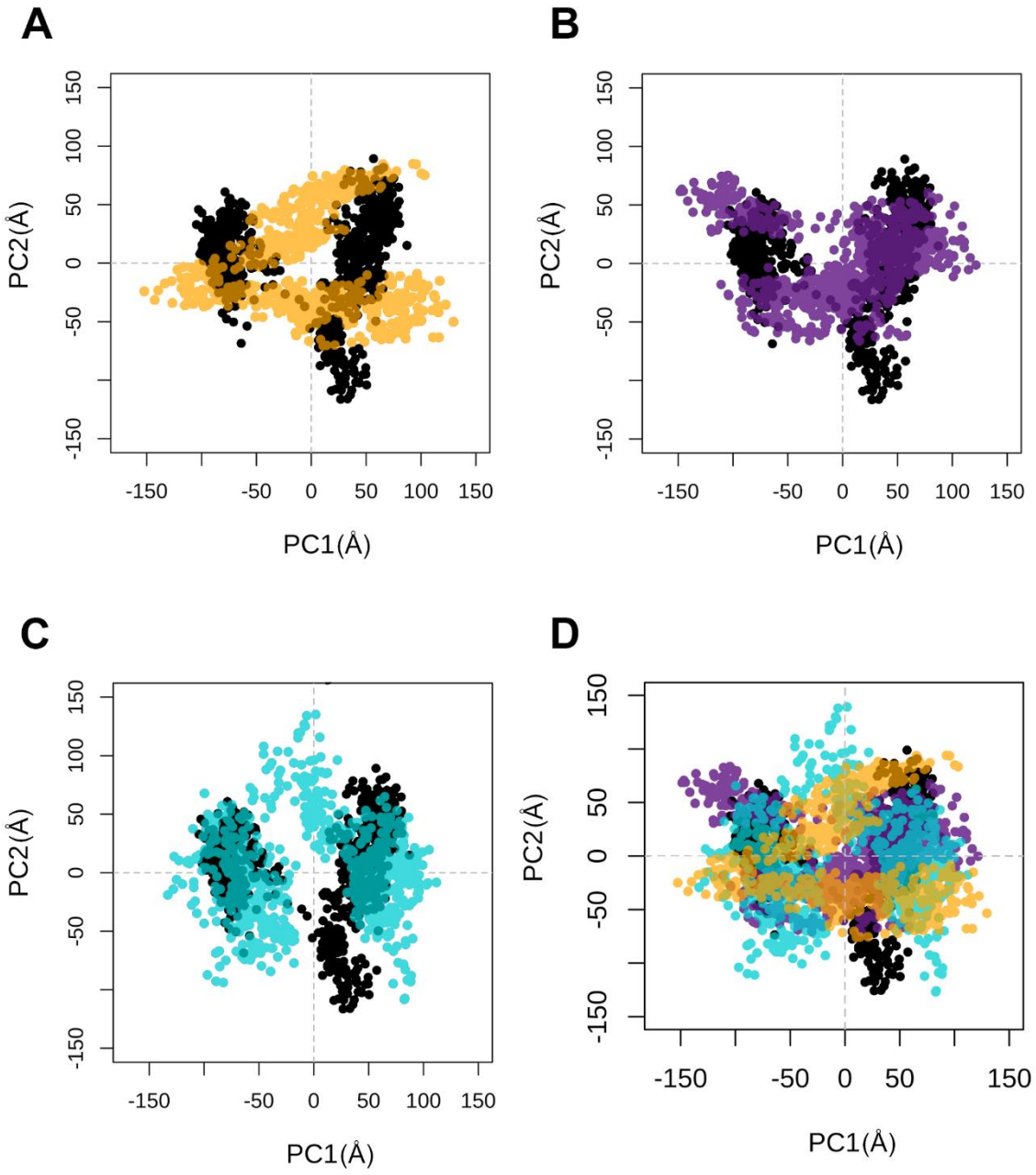
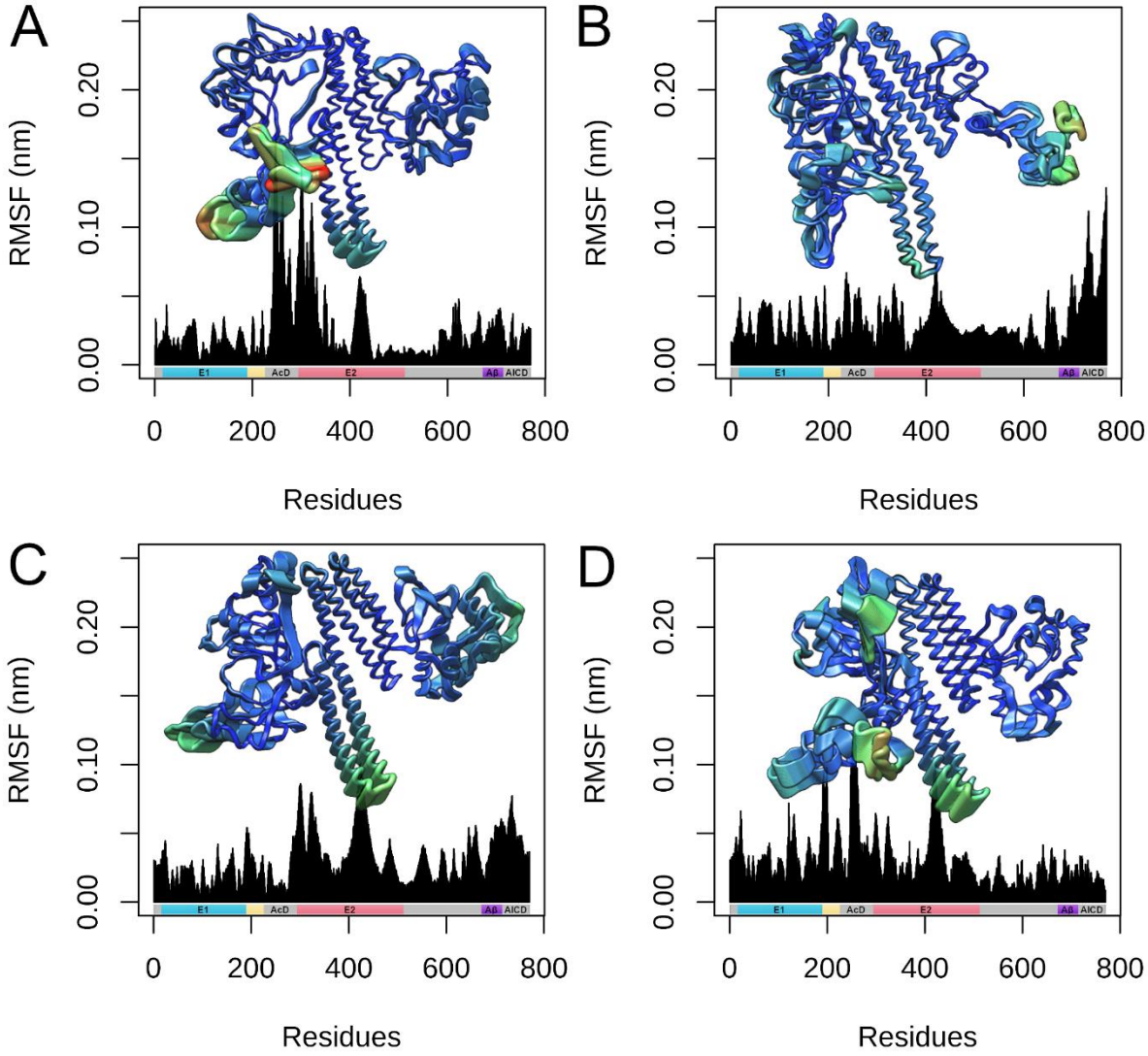


Fig 7. RMSF contribution to PC1. The RMSF contribution to PC1 for APP1 wild-type (A), variant A673T (B), variant A673V (C), and variant A673Y (C) are shown as line plots and projected on the corresponding protein structure. Schematic representations of the APP1 functional regions are also shown to further comparison. Each amino acid of APP1 wild-type and variants was colored and sized according to its RMSF contribution, which follows a coloring-thickness scale varying from blue and thin (low fluctuations) to red and thick (high fluctuations).



FUNDING

This study was supported by Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Financiadora de Estudos e Projetos, Conselho Nacional de Desenvolvimento Científico e Tecnológico, Universidade Federal do Estado do Rio de Janeiro, and NVIDIA Corporation. The funders had no role in study design and manuscript preparation.

COMPETING INTERESTS

The material received as support from NVIDIA for this study does not alter our adherence to the journal policies on sharing data and materials.

SUPPORTING INFORMATION

The supporting information is available at: https://doi.org/10.6084/m9.figshare.20020286

S1_Table. APP1 variants compiled from the databases (UNIPROT, OMIM, Molgen, and ALZFORUM).

S2_Table. Functional prediction analysis of APP1 protein variants.

S3_Table. Protein Blast results for APP1.

S4_Table: Structural alignment of the theoretical models of APP1 protein and its structural fragments (PDB ID: 3KTM, 5BUO, 3L3T, and 3DXC) in the TM-align server.

S1_Fig: Methodology workflow.

S2_Fig. Functional prediction of APP1 missense mutations. The seventy-one APP1 missense mutations compiled from the literature and databases were submitted to ten functional prediction algorithms. The circle chart shows the percentage of variants predicted as deleterious by zero to ten algorithms.

S3_Fig: Structural alignment of the theoretical models of APP1 protein and its structural fragments (PDB ID: 3KTM, 5BUO, 3L3T, and 3DXC) in the TM-align server. The validated model of wild-type APP1 is colored in light gray. The PDB fragments 3KTM (pink), 5BUO (yellow), 3L3T (green), and 3DXC (blue) were colored and aligned to the theoretical model.

S4_Fig. Structure validation of the theoretical model of APP1 using the algorithms PROCHECK, Rampage, and ERRAT. (A) PROCHECK's Ramachandran graph shows that

APP1's theoretical model has 90.5% of its residues in the most favored regions (yellow), 9.4% of its residues in allowed regions, and 0.0% in disallowed regions. (B) RAMPAGE's Ramachandran graph shows that APP1's theoretical model has 97.8% of its residues in favored regions, 2.1% of its residues in allowed regions, and only 0.1% in disallowed regions. (C) Structural validation in ERRAT. The bar plot shows the error probability of each residue of the APP1 modeled structure. The overall quality factor of the submitted structure is 96.23%.

S5_Fig. Structure validation of the theoretical model of APP1 using the algorithms ProSa, QMEAN, and MolProbity. (A) Structural validation in ProSa, which shows the Z-score for the APP1 modeled structure (black dot), when compared to all crystallographic (light blue dots) and NMR structures (dark blue dots) deposited in PDB. The theoretical model presents an overall quality score comparable to that of NMR structures. (B) Structural validation in QMEAN, which shows the overall quality score (QMEAN-score) for the APP1 modeled structure (red star), when compared to a set of 9766 high-resolution protein structures contained in PDB (gray and black dots). The theoretical model presents an overall quality score comparable to that of high-resolution protein structures. (C) Structural validation in MolProbity. This analysis shows the overall quality score (MolProbity-score) for the APP1 modeled structure. High-quality models usually present MolProbity-score values below 2.0. The theoretical model presented a MolProbity score of 1.02, further confirming its structural quality.

S6_Fig. Evolutionary conservation analysis of APP1 protein variants. The bar plot shows the ConSurf score for each amino acid of APP1 affected by mutations. The bar plot was colored according to the ConSurf color-coding scheme, varying from (1) cyan and variable to (9) maroon and conserved.

S7_Fig. B-Factor analysis of wild-type APP1 and its variants. The amino-acids residues of wild-type APP1 (A) and its variants A673T (B), A673V (C), and A673Y (D) are sized and colored according to their B-factor values. The B-factor coloring-thickness scale varies from blue and thin (rigid amino acids) to red and thick (flexible amino acids).

S8_Fig. RMSF contribution to PC2. The RMSF contribution to PC2 for APP1 wild-type (A), variant A673T (B), variant A673V (C), and variant A673Y (C) are shown as line plots and projected on the corresponding protein structure. Schematic representations of the APP1

functional regions are also shown to further comparison. Each amino acid of APP1 wild-type and variants was colored and sized according to its RMSF contribution, which follows a coloring-thickness scale varying from blue and thin (low fluctuations) to red and thick (high fluctuations).

S9_Fig. Schematic representation of APP1 amyloidogenic and non-amyloidogenic pathways. APP1 is initially cleaved by α-(non-amyloidogenic) or β-secretase (amyloidogenic). The Aβ domain of APP1 is cleaved into a 43 amino-acid Aβ peptide through proteolytic processing by β- (BACE-1), and γ-secretase enzymes. APP1 is initially cleaved by β-secretase, generating sAPPβ and the carboxy-terminal fragment CTF99. This undergoes a second cleavage by γ-secretase, which generates significant and, usually, toxic levels of Aβ peptide. This last cleavage event releases a C-terminal domain (AICD), which is involved in most of the APP1 interactions, including those related to signaling pathways and regulation of cell survival. Alternatively, the cleavage of APP1 by α-secretase produces a soluble molecule, sAPPα, and generates a carboxy-terminal fragment (CTF83) that is later cleaved by γ-secretase. This process does not generate significant levels of Aβ peptide.

S1_File. PDB file containing the atomic coordinates for the validated model of APP1 protein.

ABBREVIATION

Aβ: β-amyloid
AcD: acidic domain
AD: Alzheimer's disease
AICD: C-terminal domain
APP1: Amyloid Precursor Protein 1
BACE1: β-site APP cleaving enzyme 1
MD: Molecular Dynamics
PCA: Principal Component Analysis
RMSD: root-mean-square deviation
RMSF: root-mean-square fluctuation

REFERENCES

AGRAHARI, Ashish Kumar *et al.* A profound computational study to prioritize the disease-causing mutations in PRPS1 gene. **Metabolic Brain Disease**, v. 33, n. 2, p. 589–600, 2018. Available at: https://doi.org/10.1007/s11011-017-0121-2

ALBÀ, M. Mar; CASTRESANA, Jose. On homology searches by protein Blast and the characterization of the age of genes. **BMC Evolutionary Biology**, v. 7, p. 1–8, 2007. Available at: https://doi.org/10.1186/1471-2148-7-53

ALZHEIMER'S ASSOCIATION. **Global Prevalence**. 2021. Available at: https://www.alz.org/aaic/releases_2021/global-prevalence.asp. Acesso em: 14 mar. 2022.

ASHKENAZY, Haim *et al.* ConSurf 2016 : an improved methodology to estimate and visualize evolutionary conservation in macromolecules. **Nucleic Acids Research**, v. 44, n. May, p. 344–350, 2016. Available at: https://doi.org/10.1093/nar/gkw408

BATEMAN, Alex *et al.* UniProt: The universal protein knowledgebase. **Nucleic Acids Research**, *[S. l.]*, v. 45, n. D1, p. D158–D169, 2017. Available at: https://doi.org/10.1093/nar/gkw1099

BATEMAN, Alex. UniProt: A worldwide hub of protein knowledge. **Nucleic Acids Research**, *[S. l.]*, 2019. Available at: https://doi.org/10.1093/nar/gky1049

CHEN, Vincent B. *et al.* MolProbity: All-atom structure validation for macromolecular crystallography. **Acta Crystallographica Section D: Biological Crystallography**, v. 66, n. 1, p. 12–21, 2010. Available at: https://doi.org/10.1107/S0907444909042073

COBURGER, Ina *et al.* Analysis of the overall structure of the multi-domain amyloid precursor protein (APP). **PLoS ONE**, 2013. Available at: https://doi.org/10.1371/journal.pone.0081926

CRUTS, Marc; VAN BROECKHOVEN, Christine. Data mining: Applying the AD&FTD mutation database to progranulin. **Methods in Molecular Biology**, v. 1806, n. Chapter 6, p. 81–92, 2018. Available at: https://doi.org/10.1007/978-1-4939-8559-3_6

FERRI, Cleusa P. *et al.* Global prevalence of dementia: a Delphi consensus study. **The Lancet**, v. 366, n. 9503, p. 2112–2117, 2005. Available at: https://doi.org/10.1016/S0140-6736(05)67889-0

FORREY, Christopher; DOUGLAS, Jack F.; GILSON, Michael K. The fundamental role of flexibility on the strength of molecular binding. **Soft Matter**, v. 8, n. 23, p. 6385–6392, 2012. Available at: https://doi.org/10.1039/c2sm25160d

GRANT, Barry J. *et al.* Bio3d: An R package for the comparative analysis of protein structures. **Bioinformatics**, v. 22, n. 21, p. 2695–2696, 2006. Available at: https://doi.org/10.1093/bioinformatics/btl461

HAMOSH, Ada *et al.* Online Mendelian Inheritance in Man (OMIM). **Human Mutation**, v. 15, n. 1, p. 57–61, 2000. Available at: https://doi.org/10.1002/(SICI)1098-1004(200001)15:1<57::AID-HUMU12>3.0.CO;2-G

HESS, Berk; SPOEL, David van der; LINDAHL, Erik. Gromacs User Manual. **Department of Biophysical Chemistry,** 2014. Available at: https://doi.org/10.1006/imms.1993.1022

HUMPHREY, W.; DALKE, A.; SCHULTEN, K. VMD: visual molecular dynamics. **Journal of molecular graphics**, v. 14, n. 1, p. 33–38, 27–28, 1996. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8744570. Acesso em: 9 maio. 2016.

JONSSON, Thorlakur *et al.* A mutation in APP protects against Alzheimer‘s disease and age-related cognitive decline. **Nature**, v. 488, n. 7409, p. 96, 2012. Available at: https://doi.org/10.1038/nature11283

KIMURA, Ayano; HATA, Saori; SUZUKI, Toshiharu. Alternative selection of site APP-cleaving enzyme 1 (BACE1) cleavage sites in amyloid β-protein precursor (APP) harboring protective and pathogenic mutations within the Aβ sequence. **Journal of Biological Chemistry**, v. 291, n. 46, p. 24041–24053, 2016. Available at: https://doi.org/10.1074/jbc.M116.744722

KINOSHITA, June; CLARK, Timothy. **Alzforum.** Available at: https://doi.org/10.1007/978-1-59745-520-6_19

LANOISELÉE, Hélène Marie *et al.* APP, PSEN1, and PSEN2 mutations in early-onset Alzheimer disease: A genetic screening study of familial and sporadic cases. **PLoS Medicine**, 2017. Available at: https://doi.org/10.1371/journal.pmed.1002270

LONG, Justin M.; HOLTZMAN, David M. **Alzheimer Disease: An Update on Pathobiology and Treatment Strategies**. Available at: https://doi.org/10.1016/j.cell.2019.09.001

LOVELL, S. C. *et al.* Structure validation by Calpha geometry: Phi,psi and Cbeta deviation. **Proteins**, 2003 a.

LOVELL, Simon C. *et al.* Structure validation by C alpha geometry: phi,psi and C beta deviation. **Proteins-Structure Function and Genetics**, v. 50, n. August 2002, p. 437–450, 2003 b. Available at: https://doi.org/10.1002/prot.10286

MURATORE, Christina R. *et al.* The familial alzheimer’s disease APPV717I mutation alters APP processing and Tau expression in iPSC-derived neurons. **Human Molecular Genetics**, 2014. Available at: https://doi.org/10.1093/hmg/ddu064

NAQVI, Ahmad Abu Turab *et al.* Advancements in Docking and Molecular Dynamics Simulations Towards Ligand-receptor Interactions and Structure-function Relationships. **Current Topics in Medicinal Chemistry**, v. 18, n. 20, p. 1755–1768, 2019. Available at: https://doi.org/10.2174/1568026618666181025114157

O’BRIEN, Richard J.; WONG, Philip C. Amyloid Precursor Protein Processing and Alzheimer’s Disease. **Annu Rev Neurosci**, v. 34, p. 185–204, 2011. Available at: https://doi.org/10.1146/annurev-neuro-061010-113613.Amyloid

OLIVEIRA, Clara Carolina Silva De *et al.* In silico analysis of the V66M variant of human BDNF in psychiatric disorders : An approach to precision medicine. **PLoS ONE**, v. 14, n. 4, p. e0215508, 2019.

PEREIRA, G. R. C. *et al.* In silico analysis and molecular dynamics simulation of human superoxide dismutase 3 (SOD3) genetic variants. **Journal of Cellular Biochemistry**, v. 120, n. 3, p. 3583–3598, 2019 a. Available at: https://doi.org/10.1002/jcb.27636

PEREIRA, Gabriel Rodrigues Coutinho *et al.* In silico analysis and molecular dynamics simulation of human superoxide dismutase 3 (SOD3) genetic variants. **Journal of Cellular Biochemistry**, v. 120, n. 3, p. 3583–3598, 2019 b. Available at: https://doi.org/10.1002/jcb.27636

PEREIRA, Gabriel Rodrigues Coutinho *et al.* In silico analyses of acetylcholinesterase (AChE) and its genetic variants in interaction with the anti - Alzheimer drug Rivastigmine. **Journal of Cellular Biochemistry**, n. May, p. 1–19, 2022. Available at: https://doi.org/10.1002/jcb.30277

PEREIRA, Gabriel Rodrigues Coutinho; VIEIRA, Barbara de Azevedo Abrahim; MESQUITA, Joelma Freire De. Comprehensive in silico analysis and molecular dynamics of the superoxide dismutase 1 (SOD1) variants related to amyotrophic lateral sclerosis. **PLOS ONE**, v. 1, n. 0247841, p. 1–27, 2021. Available at: https://doi.org/10.1371/journal.pone.0247841

REINHARD, Constanze; HÉBERT, Sébastien S.; DE STROOPER, Bart. **The amyloid-β precursor protein: Integrating structure with biological function**. Available at: https://doi.org/10.1038/sj.emboj.7600860

ROSENBERG, Roger N. *et al.* **Genomics of Alzheimer disease: A review**. Available at: https://doi.org/10.1001/jamaneurol.2016.0301

SCHWEDE, Torsten. Protein modeling: What happened to the “protein structure gap”? **Structure**, v. 21, n. 9, p. 1531–1540, 2013. Available at: https://doi.org/10.1016/j.str.2013.08.007

SELKOE, D. J. Alzheimer’s disease: genes, proteins, and therapy. **Physiological reviews**, v. 81, n. 2, p. 741–766, 2001. Available at: https://doi.org/10.1152/physrev.2001.81.2.741

WEBB, Benjamin; SALI, Andrej. Comparative Protein Structure Modeling Using MODELLER. **Current Protocols in Bioinformatics**, v. 54, n. 5.6.1-5.6.3.7, 2016. Available at: https://doi.org/10.1002/cpbi.3.Comparative

WILLIAMS, Christopher J. *et al.* MolProbity: More and better reference data for improved all-atom structure validation. **Protein Science**, 2018. Available at: https://doi.org/10.1002/pro.3330

ZHANG, Yang; SKOLNICK, Jeffrey. TM-align: A protein structure alignment algorithm based on the TM-score. **Nucleic Acids Research**, v. 33, n. 7, p. 2302–2309, 2005. Available at: https://doi.org/10.1093/nar/gki524