ACS OMEGA

Open Access

http://pubs.acs.org/journal/acsodf

Article

Transcript-Level In Silico Analysis of Alzheimer's Disease-Related Gene Biomarkers and Their Evaluation with Bioactive Flavonoids to Explore Therapeutic Interactions

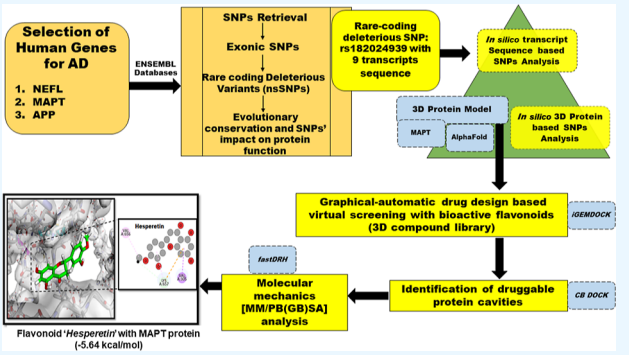
Muhammad Bilal Azmi,* Affan Ahmed, Tehniat Faraz Ahmed, Fauzia Imtiaz, Uzma Asif, Uzma Zaman, Khalid Ali Khan, and Asif Khan Sherwani

Cite This: *ACS Omega* 2023, 8, 40695−40712

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Alzheimer's disease (AD) is a progressive brain disorder that can significantly affect the quality of life. We used a variety of in silico tools to investigate the transcript-level mutational impact of exonic missense rare variations (single nucleotide polymorphisms, SNPs) on protein function and to identify potential druggable protein cavities that correspond to potential therapeutic targets for the management of AD. According to the NIA-AA (National Institute on Aging-Alzheimer's Association) framework, we selected three AD biomarker genes (APP, NEFL, and MAPT). We systematically screened transcript-level exonic rare SNPs from these genes with a minor allele frequency of 1% in 1KGD (1000 Genomes Project Database) and gnomAD (Genome Aggregation Database). With downstream functional effect predictions, a single variation (rs182024939: K > N) of the MAPT gene with nine transcript SNPs was identified as the most pathogenic variation from the large dataset of mutations. The machine learning consensus classifier predictor categorized these transcript-level SNPs as the most deleterious variations, resulting in a large decrease in protein structural stability ($\Delta\Delta G$ kcal/mol). The bioactive flavonoid library was screened for drug-likeness and toxicity risk. Virtual screening of eligible flavonoids was performed using the MAPT protein. Identification of druggable protein-binding cavities showed VAL305, GLU655, and LYS657 as consensus-interacting residues present in the MAPT-docked top-ranked flavonoid compounds. The MM/PB(GB)SA analysis indicated hesperetin (−5.64 kcal/mol), eriodictyol (−5.63 kcal/mol), and sakuranetin (−5.60 kcal/mol) as the best docked flavonoids with the near-native binding pose. The findings of this study provide important insights into the potential of hesperetin as a promising flavonoid that can be utilized for further rational drug design and lead optimization to open new gateways in the field of AD therapeutics.



1. INTRODUCTION

Alzheimer's disease (AD) is the most common type of dementia, characterized by a gradual and progressive decline in memory and other cognitive functions, ultimately leading to significant impairment of daily activities.[1] Brain changes associated with AD lead to the destruction of synapses in parts of the brain involved in memory, including the entorhinal cortex and hippocampus, culminating in brain atrophy.[2]

In 2019, AD was established as the sixth-leading cause of death in the USA and the fifth-leading cause of death for people aged 65 and older.[3] Global statistics show that AD affected 6.2 million Americans aged 65 and older in 2021. Without medical advancements to manage or cure the disease, this figure is projected to nearly double to 13.8 million by 2060.[3,4] In Southeast Asia, dementia of the Alzheimer's type accounted for 59.8% of all dementia cases in 2021.[5] Due to a rapidly growing elderly population, an estimated two million people in Pakistan are currently living with AD, and this number is steadily rising.[6]

AD is characterized by numerous molecular and cellular changes in the brain involving a complex interplay of genetic, environmental, and lifestyle factors that contribute to the disease's pathogenesis.[2] AD is biologically defined by the presence of β-amyloid-containing plaques and tau-containing neurofibrillary tangles (NFTs).[7] The anomalous processing of amyloid precursor protein (APP) by β and γ-secretases leads to the production of Aβ40 and Aβ42 monomers, which further oligomerize and aggregate into senile plaques.[8] The hyperphosphorylation of tau causes it to lose its normal function of stabilizing microtubules, leading to the formation of paired

Received: August 6, 2023
Accepted: September 25, 2023
Published: October 18, 2023

helical filaments which make up the NFTs.[9] These NFTs interfere with the transport mechanism of the cell, impairing the synaptic transmission between neurons. The formation of NFTs is a hallmark of AD and is used as a diagnostic criterion for the disease.[10] The accumulation of Aβ42 and p-tau in the brain triggers a series of events, including inflammation, oxidative stress, and impaired energy metabolism that eventually lead to neuronal death.[9,10] Emerging research suggests that aberrant tau and β-amyloid proteins, along with other variables, may interact intricately to cause the brain abnormalities associated with AD.[11]

AD has been designated as a global public health priority due to its expected tripled prevalence from 2015 to 2050, with the greatest impact expected in low- and middle-income countries.[3,4] Mixed etiology, overlapping symptoms, variation in presentation, and a lack of standardized accessible biomarkers leave AD still difficult to diagnose.[12] Even after more than a century of rigorous research, no disease-modifying treatment yet exists, and only symptomatic management is provided to AD patients.[13] This high failure rate of clinical trials can be attributed to pharmacological intervention late in the disease course. Since it is strongly suggested that treatment is rendered ineffective after a certain pathological threshold, there is a pressing need for early diagnostic markers of AD.[12,13] The research framework articulated by the NIA-AA (National Institute on Aging-Alzheimer's Association) envisions AD as a molecular rather than a clinical construct.[14]

Recently, following the NIA-AA research framework guidelines,[14,15] our group obtained encouraging results on investigating the potential use of plasma amyloid β 42 (Aβ42), phosphorylated tau (P-tau), and neurofilament light chain (NEFL) as biomarkers for AD.[16] To expand our biochemical analysis, we conducted this in silico analysis of the genes encoding the core proteins involved in AD pathogenesis. This approach allowed us to investigate both pathogenic and rare-coding mutations and their impact on the structure of the respective proteins. By examining the structural consequences of these mutations, we gained valuable insights into their potential contributions to AD development and progression.

Flavonoids, ubiquitous plant compounds comprising up to 60% of dietary polyphenols, have been extensively studied as potential drugs or food supplements owing to their biological functions.[68] They consist of a diphenyl propane−flavonoid skeleton and are divided into subclasses, such as anthocyanins, chalcones, flavanols, flavanones, flavones, and isoflavones. Flavonoids exhibit antioxidant properties by scavenging free radicals, regulating oxidative stress, and chelating transition metals. Different flavonoids such as myricetin, hesperetin, genistein, xanthohumol, and isoliquiritigenin, have been reported to have antioxidant and anti-inflammatory effects.[68,69] They also regulate carcinogenesis signaling pathways, interact with proteins, and modulate wingless-related integration site (Wnt) signaling pathways, affecting all cancer stages.[68,69]

Oxidative stress was initially proposed as a major factor in the development of AD in 1986.[68,69] Overwhelming evidence exists that the cells in the Alzheimer's brain undergo abnormally high levels of oxidative stress.[69] In AD, the brain appears to sustain more oxidative damage than normal with low levels of antioxidants. Recent research on secondary metabolites of plants, such as polyphenols, has demonstrated that they may slow the progression of AD. The mechanisms of action of flavonoids in AD involve the inhibition of acetylcholinesterase, butyrylcholinesterase, tau protein aggregation, β-secretase, oxidative stress, inflammation, and apoptosis through the modulation of signaling pathways implicated in cognitive and neuroprotective functions.[69] In addition, flavonoids, including epicatechin-3-gallate, gossypetin, naringenin, quercetin, and myricetin, have been reported to block β-amyloid and tau aggregation, scavenge free radicals, and sequester metal ions at clinically low concentrations.[69] The reported possible mechanisms of action of these flavonoids include decreasing the hyperphosphorylation of tau protein, oxidative stress, and inflammation, which in general result in the modulation of metabolic pathways as well as key enzymes to manage the progression of AD.[69] We further explored the therapeutic potential of flavonoids, known for their neuroprotective effects by investigating their binding interactions with the identified druggable protein cavities in AD-related proteins.[17] This investigation aimed to elucidate the potential therapeutic effects of flavonoids by targeting specific sites within AD-related proteins.

2. MATERIALS AND METHODS

2.1. Retrieving Genes, SNPs' Selection, and Extraction of Rare-Coding Exonic Mutations (nsSNP). As a continuation of our earlier work reported in 2023 regarding these biomarker gene targets, the main rationale for the selection of potential AD genes for biomarker analysis also follows the NIA-AA research framework guide.[15,16] We chose Aβ 42 protein from category A and P-tau from the T category of the AT(N) classification of biomarkers. Additionally, we included NEFL as a candidate biomarker gene as several studies have shown NEFL to be indicative of disease severity, although it is not specific to AD.[15,16] APP, MAPT, and NEFL genes were extracted from publicly accessible databases. The relevant genomic information was extracted through a database-associated protein search. The strategy for the retrieval of AD-associated genes/proteins was done from the NCBI (National Center for Biotechnology Information) databases[18] and the HUGO Gene Nomenclature Committee.[19] All essential genomic information was further retrieved from NCBI,[18] ENSEMBL,[20] UniProt,[21] and GeneCards.[22] Data related to the core single nucleotide polymorphisms (SNPs) of the aforementioned human genes (potential biomarker genes/proteins for AD) and their protein sequences (FASTA format) were downloaded from the NCBI[18] and ENSEMBL (https://asia.ensembl.org/index.html) databases (Human 127 Genome Assembly, GRCh38.p13 build 38).[20] Subsequently, the retrieved SNPs were further validated through the dbSNP database. The variant effect predictor (VEP) module was used to predict the downstream functional effects (DSFE) of variants, including SNPs, deletions, insertions, and single nucleotide variants (SNVs) on selected genes, transcripts, and protein sequences as well as regulatory regions. The rare-coding variants were evidenced in the 1000 Genome Project Database (1KGD)[23] and the Genome Aggregation Database (gnomAD).[24] The minor allele frequency (MAF) cutoff was set less than or equal to 0.01 (1%). The relevant information for all SNPs, including SNP ID, chromosomal location, nucleic acid changes, amino acid changes, and reported transcript sequence ID was also extracted. After the retrieval of all SNPs from the ENSEMBL database, the extraction of exonic variants was performed by applying a filter in the VEP module of ENSEMBL. The extracted details were downloaded and stored as a default comma-separated values (CSV) Excel file for further analysis.

https://doi.org/10.1021/acsomega.3c05769
ACS Omega 2023, 8, 40695−40712

2.2. Extraction of Deleterious SNPs through the DSFE Prediction Approach. To screen the deleterious exonic nsSNP in the genomic mutation dataset, a combination of various computational tools based on different machine learning and deep learning algorithms was employed. The details of the algorithm used to compute deleterious mutations include the following:

1. SIFT: SIFT (https://sift.bii.a-star.edu.sg/) is a tool for distinguishing pathogenic mutations with significant disease involvement from neutral polymorphisms. SIFT predictions make it easy to examine how amino acid substitutions affect protein function.[25] The SIFT score, which ranges from 0 to 1, is a normalized likelihood value of discovering the new amino acid at that site.
2. Polymorphism phenotyping (PolyPhen2): This tool (Polymorphism phenotyping, http://genetics.bwh.harvard.edu/pph2/dbsearch.shtml) was used to evaluate the functional impact of an amino acid substitution on a protein's stability and overall function. With regard to specificity and sensitivity for a mutation, the prediction (score) output is classed as "most likely deleterious", "may be deleterious", and "benign", which was employed as a basis for gene annotation.[26]
3. Combined annotation-dependent depletion (CADD) score: In numerous genomic datasets of humans, CADD is most often employed to filter the deleteriousness of SNVs as insertion/deletion variations. Because of the substantial correlation between the CADD score and allelic diversity, variant pathogenicity, and experimentally measured regulatory effects within individual genome sequences,[27] CADD scores of 20 or higher were used to select nsSNPs.
4. The rare exome variant ensemble learner (REVEL): REVEL is used to evaluate the pathogenicity of missense variations in order to determine the possible pathogenicity of SNV. This tool is not intended to be used as a descriptive prediction method; rather, it simply indicates that scores larger than "0.5" are "likely disease causing" and scores below "0.5" are "likely benign". Only 10.9% of neutral variants, but an estimated 75.4% of disease mutations, have a score above 0.5.[28]
5. MetaLR: This logistic regression-based method called MetaLR is used to classify SNV as either "damaging" or "tolerable". A number between 0 and 1 is also given in this instance and variants with higher values are more likely to be harmful.[29]
6. Mutation assessor class: The functional effects of substituting amino acids are predicted by this technique. The classification of the prediction is "high", which represents any variations with a higher likelihood of being harmful in genomic datasets.[30]

2.3. Transcript-Sequence-Based Consensus Classification of Deleterious nsSNPs. We separately validated the transcript-sequence-based nsSNPs with PhD-SNP,[31] PANTHER,[32] MutPred2,[33] SNAP,[34] MetaSNP,[35] SNAP2,[36] PMut,[37] and SNAP&Go[38] for their deleterious assessment. Further, to investigate the consensus classification of the transcript-sequence-based SNPs' impact on protein function, PredictSNP1.0 (http://loschmidt.chemi.muni.cz/predictsnp1/) tool was used to determine the aggregate predicted accuracy of nsSNP. This consensus classifier was implemented through nine of the top-performing prediction tools: PhD-SNP (predictor of human deleterious single nucleotide polymorphisms), SNAP, PANTHER (protein analysis through evolutionary relationships), PredictSNP, nsSNPAnalyzer, SIFT (sorting intolerant from tolerant), PolyPhen-1 (polymorphism phenotyping v1), PolyPhen-2 (polymorphism phenotyping v2), and MAPP (multivariate analysis of protein polymorphism).[39]

2.4. Characterizing the Effects of Rare nsSNPs on Protein Stability and the Evolutionary Conservation of Protein Structural Characteristics Using In Silico Supervised Learning-Based Methods. The in silico tools used to characterize the effects of rare nsSNPs on protein stability and functions include the following web servers/tools:

1. iPTree-STAB: The iPTREE-STAB/iStable v.2.0 is an interpretable decision tree-based approach that can distinguish between proteins that become stabilized or destabilized and predict their stability changes ($\Delta\Delta G$) upon single amino acid variation. Three nearby residues of the mutant site along the N- and C-terminals are used in classification and regression trees, respectively, as the major methods for discrimination and predictions (http://ncblab.nchu.edu.tw/iStable2/seqsubmit.html). Additionally, this database calculates the surface accessibility and secondary structural information on the protein sequences with mutations.[40]
2. MUpro (http://mupro.proteomics.ics.uci.edu/), this tool was created using support vector machines (SVM) and neural networks, employing machine learning techniques (NN). The benefit of this in silico approach was that it could anticipate changes in protein stability without the need for tertiary structures.[41]
3. I Mutant: I Mutant is a predictor of changes in protein stability brought on by mutations. I Mutant v3.0 is a SVM-based tool for the automatic prediction of protein stability alterations in response to single-point mutations at pH 7 and a temperature of 25 °C. The predictions are made from the protein sequence and can be used as a regression estimator to predict the related $\Delta\Delta G$ values as well as a classifier to predict the sign of the protein stability change upon mutation (http://gpcr2.biocomp.unibo.it/emidio/I-Mutant3.0/I-MutantSuiteHelp.html). The ternary classification includes the neutral class, which refers to modest $\Delta\Delta G$ value changes caused by single-point protein mutations (WT/new). I Mutant v3.0 classifies the prediction in three classes: neutral mutation ($-0.5 < \Delta\Delta$ G < 0.5), large decrease (< -0.5) and large increase (> 0.5).[42]
4. INPS-MD: Using sequences and structures, the INPS-MD approach (impact of nonsynonymous mutations on protein stability-multi-dimension) predicts the stability of protein variations. The simplified support vector used by the libsvm package, which was only evaluated with linear and radial basis function kernels, is the foundation of the INPS-MD predictor employing sequences. To determine stabilizing ($\Delta\Delta G > 0$) and destabilizing ($\Delta\Delta G < 0$) variations, we evaluated INPS-MD predictions on 9 transcript-level nsSNPs.[43]
5. ConSurf and SOPMA analysis: By computing the conservation score using a special technique, the ConSurf software (http://consurf.tau.ac.il) was utilized to assess the evolutionary conservation of amino acids' variation in protein. The technique applies the Bayesian

https://doi.org/10.1021/acsomega.3c05769
ACS Omega 2023, 8, 40695−40712

method to compute normalized conservative scores (NMC) with a confidence interval (CI) for each of the estimated evolutionary conservation scores. Evolutionarily conservative amino acids were those with scores between 7 and 9.[44] The secondary structure of amino acid residues (helix, turn, and coil) was collaboratively predicted by the SOPMA neural network approach (PHD). To calculate the secondary structural information on proteins, this software (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html) employed five distinct methods.[45]

2.5. Computing the Impact of Transcript-Level Structural Modifications of Rare nsSNPs on the Secondary Structure, Torsional Angles, Energetic Potentials, and Protein Dynamics. To examine the effects of the substitution of an amino acid on protein sequence, protein secondary structure, surface accessibility, percent disordered residue, and torsional angles, phi (φ) and psi (ψ), were computed from NetSurfP-2.0 web server of the Technical University of Denmark's (DTU).[46] The Cologne University Protein Stability Analysis Tool (CUPSAT) webserver was used to look into any potential 3D structural variation associated with the protein of interest (http://cupsat.tu-bs.de/). With the use of this web tool, transformations in the potential of the targeted atom(s) and torsion angle potential were also obtained.[47] The free energy difference of the unfolding conformational state between the wild type (WT)/native protein and the nsSNP-based modified protein 3D structure was calculated using this sequence-based approach.

To predict the 3D structure of each transcript-sequence, we used the Phyre2 web-based server for transcript-level protein 3D structure prediction (http://www.sbg.bio.ic.ac.uk/phyre2).[48] The Phyre2 services use the concepts and methods of homology modeling to forecast the 3D structure of a protein sequence. A protein sequence of interest (the target) can be modeled accurately on a very distantly related sequence of a known structure (the template), provided that the relationship between the target and template can be determined through sequence alignment as the structure of a protein is more conserved in evolution than its amino acid sequence. Profiles and hidden Markov models are currently the most effective and reliable techniques for identifying and aligning remotely related sequences (HMMs). These profiles/HMMs serve as an accurate representation of the mutational propensity of each location in an amino acid sequence based on reported mutations in similar sequences.[48]

SDM (site-directed mutator) was another computational method for analyzing the variation of amino acid replacements occurring at particular structural environments that were permitted within the family of homologous proteins with known 3D structures, turning them into substitution probability tables.[49] To analyze and visualize protein dynamics by sampling conformations and to evaluate the effects of residue modification/possible mutations on protein dynamics as well as the stability brought by changes in vibrational entropy, DynaMut, a web server that implemented two independent, well-established normal mode techniques, was employed.[50]

2.6. Protein's Homology Model, Bioactive Flavonoid Selection, and Compilation of 3D Compounds' Library. To prefer the complete 3D protein homology with comprehensive structural coverage of amino acid residues, the best homology model of canonical protein coding gene sequence (MAPT) was retrieved from the AlphaFold protein structure database available through the UniProt database (https://alphafold.ebi.ac.uk/search/text/MAPT).[21,51] This strategy was adopted because of the inaccessibility of a complete 3D protein model in the Protein Data Bank (PDB). Prior to this, we performed protein BLAST (basic local alignment search tool) using the PDB database. The sequence query coverage was less than 60%; therefore, our preference was to choose a 3D model with full coverage (amino acid length of 758) from the canonical sequence of MAPT.

The SAVES (structure validation server: https://saves.mbi.ucla.edu/) web server was used to know the probable residue properties, chi-plot analysis, side chain parameters, *G*-factor, and planar group analysis.[52] Bioactive flavonoid compounds were selected based on their previously proposed role as neuroprotective compounds, which can prevent the progression of age-related disorders, including AD, and can be used to design and develop new potential drugs that are effective in cognitive disorders.[53] According to Ayaz et al., 2019, we chose various flavonoid compound classes for the identification of therapeutic targets in the present protein.[17] A thorough literature search was used to get the basic information on these bioactive substances for establishing a 3D compound library including all bioactive flavonoids reported from various sources and classes. From the PubChem compound database (https://pubchem.ncbi.nlm.nih.gov), the 3D structures of all reported compounds were retrieved.[54] Following a visual examination of each 3D compound, the structures were initially downloaded in SDF (structure data format) and translated to PDB format using Open Babel software.[55] All the 3D compounds were subjected to Pfizer's rule of five for evaluating drug-likeness criteria.[56] Following the evaluation, only 3D bioactive flavonoids which satisfied all five of Pfizer's rule of five physicochemical requirements were chosen because any one infraction was treated as a failure or an elimination aspect.[57] Only eligible compounds were further screened with the help of OSIRIS Property Explorer (https://www.organic-chemistry.org/prog/peo/), which uses chemical structures in SMILE format for drug property prediction in terms of topological polar surface area, clog *P* calculation, log *S* calculation, molecular weight, and others. The drug-likeness properties were further computed and validated and only those 3D bioactive flavonoids were selected which have drug-score >0.8 (80%).[58]

2.7. Virtual Screening of Bioactive Flavonoid Compounds through the Molecular Docking-Based Graphical-Automatic Drug Design Approach. The structure of the MAPT protein from the AlphaFold protein structure database was chosen for this experiment. In our selected protein 3D model, the details of active site residues and pockets have not been well characterized because of the lack of complete 3D model details in the PDB. For this reason, virtual screening using a blind docking approach is a computational technique used in drug discovery when the information on the active site of a target protein is unknown or poorly characterized. These methods help to identify potential binding sites and candidate compounds for further experimental validation. Second, the main objective of applying this strategy was to significantly expedite the flavonoid screening process through blind docking-based virtual screening by narrowing the pool of compounds for further experimental validation. The basic reason for virtual screening was to dock a

https://doi.org/10.1021/acsomega.3c05769
ACS Omega 2023, 8, 40695−40712

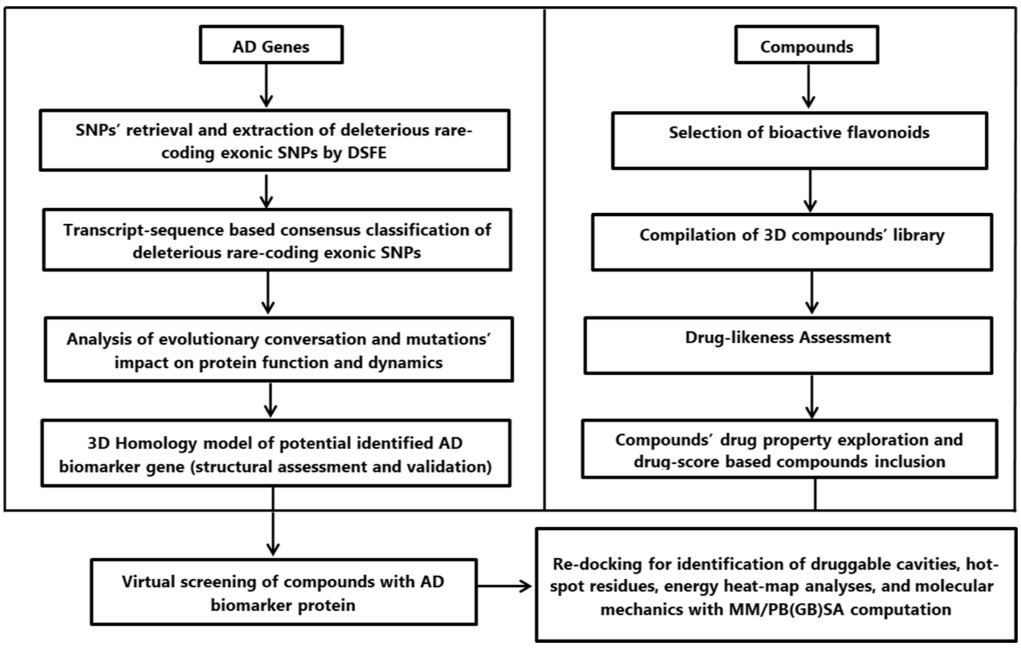


Figure 1. Research work-flow.

library of all bioactive flavonoid compounds into the binding pocket of the AD protein and then rank the small ligand molecules based on their binding affinity with the MAPT protein. The virtual screening approach resulted in some top-ranked compounds in terms of binding energy, which were used to be moved forward for testing toward hit identification. In this regard, we employed the iGEMDOCK graphic automatic drug design system to identify compound-protein associations via the fitness value. The corresponding formats of the ligands (flavonoids) and MAPT protein for this analysis were PDB files. iGEMDOCK provides post-analysis tools (i.e., atomic compositions) by providing a graphical-integrated environment for virtual screening (molecular docking), utilizing k-means and hierarchical clustering techniques based on docked poses (i.e., protein−ligand interactions) as well as compound characteristics. A novel approach for determining compound similarity was atomic composition (AC), which was comparable to the amino acid content of a protein sequence.[59]

2.8. Identifying Druggable Protein Cavities and Predicting Near-Native Binding Pose through Molecular Mechanic Computation with Poisson−Boltzmann/Generalized Born Surface Area Solvation. The binding cavities in the protein of interest were explored using the CB-dock (cavity-detection guided blind docking) technique.[60] We employed this method in order to find the best protein-binding (druggable) cavities in terms of their molecular interactions with bioactive flavonoid compounds. CB-dock is an automated protein−ligand docking method that locates binding cavities/sites. Utilizing the COACH prediction tool as a benchmark set, CB-dock compared and rated cavities using a technique known as "CurPocket" with cutting-edge algorithms for predicting protein−ligand binding sites. Using a novel curvature-based cavity detection approach, the CB-dock additionally determines the center and size of the docking box of the putative cavity as a critical process parameter. This prediction method was carefully optimized and achieved a ∼ 70% success rate for top-ranking poses whose root-mean-square deviation was within 2 Å from the X-ray pose.[60] All eligible screened bioactive compounds were CB-docked with a 3D model of the MAPT protein, and the binding affinities for each drug-like compound were documented. Ultimately, the highest grading cavities produced after each compound's interaction with the chosen protein model were used to assess for its further molecular visualization and interpretation of intervening therapeutic target with AD protein using a 2D ligand plot.[60]

In order to identify the near-native mode of binding of MAPT protein with ligand through rescoring docking poses, we used the fastDRH algorithm. This tool generated protein−ligand binding poses by the molecular docking approach, structure-truncated MM/PB(GB)SA free energy calculation, and predicting the hotspot residues by per-residue free energy decomposition analysis. In structure-based drug design, predicting the native binding pose of a small molecule within a protein-binding pocket is considered a crucial challenge, particularly in the hit-to-lead and lead optimization phases. The fastDRH server integrates AutoDock Vina and AutoDock-GPU docking engines, structure-truncated MM/PB(GB)SA free energy calculation methods, and several poses based on per-residue energy decomposition analysis into an accessible and multifaceted online platform. The combined structure-truncated MM/PB(GB)SA rescoring processes demonstrated a success rate of >80% in the benchmark, which was significantly higher than that of AutoDock Vina (70%) in terms of protein−ligand binding mode prediction. The per-residue energy decomposition analysis technique based on several poses is a more reliable solution for hotspot residue identification than the one employing only a single pose, and the effectiveness of the outcomes has been empirically proven in various drug discovery projects. The molecular mechanics energies in combination with the MM/PBSA and MM/GBSA methods were used to calculate the free energy of small ligands binding to biological macromolecules, such as proteins. This approach is an efficient and reliable free energy simulation method for modeling molecular recognition.[61]

https://doi.org/10.1021/acsomega.3c05769
ACS Omega 2023, 8, 40695−40712

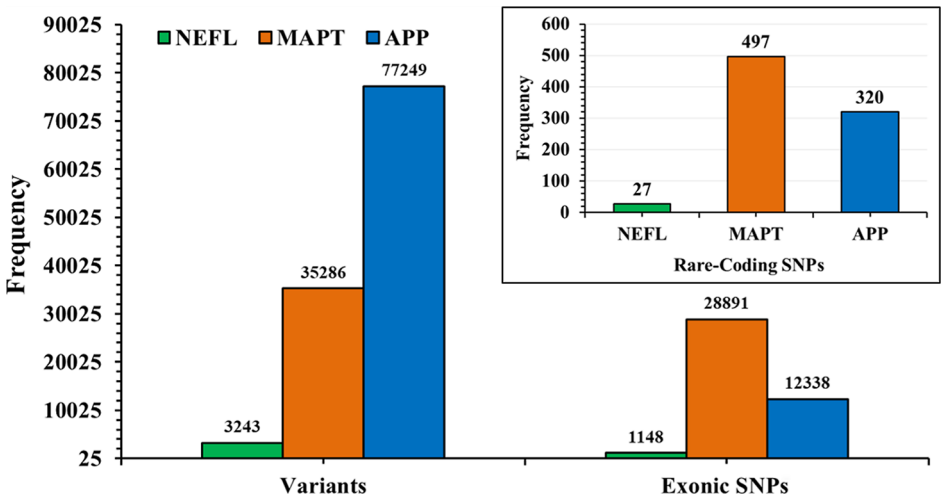


Figure 2. Mutational (SNPs) estimation of selected AD gene biomarkers.

3. RESULTS

3.1. Selection of the AD-Associated Genes and Proteins. Figure 1 depicts the research scheme employed in this study. According to the prescribed criteria of NIA-AA, we extracted three human genes encoding a particular protein biomarker. The first gene was NEFL (NCBI Gene ID: 4747), which has a four-exon count, two transcripts, and the following genomic location on chromosome 8:24,950,955−24,956,721, reverse strand. The protein coded by NEFL is NEFL polypeptide (UniProt accession no. P07196). The second gene was microtubule-associated protein tau (*MAPT*; NCBI Gene ID: 4137), which has 16 exons and 30 transcripts and the genomic location of chromosome 17:45,894,527−46,028,334 on the forward strand. The protein coded by the *MAPT* gene is MAPT (UniProt accession no. P10636). The third biomarker gene was amyloid β precursor protein (*APP*; NCBI Gene ID: 351), which has 20 exons and 20 transcripts and the genomic location of chromosome 21:25,880,535−26,171,128 reverse strand. The protein coded by the *APP* gene is an APP (UniProt accession no. P05067).

3.2. Genes Mutational Estimation and Extraction of Rare-Coding Deleterious SNPs through DSFE Prediction. The total number of mutations estimated was 3243 in the *NEFL*, 35286 in the *MAPT*, and 77249 in the *APP* gene. The overall mutations (SNPs) from these three biomarker genes were 115,778. These mutations were further filtered in terms of exonic SNPs, which reduced the count to 1148 in *NEFL*, 28891 in *MAPT*, and 12338 in *APP* genes, with an overall count of 42,377 SNPs. On the basis of MAF criteria of 1%, we further filtered the rare mutation numbers in all three genes and found the estimates of 27 in *NEFL*, 497 in *MAPT*, and 320 in the *APP* gene (Figure 2 and Table S1). By applying the set MAF criteria and DSFE prediction, we obtained only a single exonic rare-coding SNP, *rs182024939*, with nine (9) different transcripts of the MAPT gene (Table 1). Hence, because all other SNPs were excluded, this variant (*rs182024939*) from MAPT was characterized as a rare-coding high-risk mutation with nine different transcripts and was considered for further analysis.

The SNV was sourced from dbSNP, and it exhibits a G > A substitution at position 17:46010327 in the GRCh38 reference genome. The supporting evidence for the variant includes a MAF of T = 0.000005/1 in gnomAD exomes, as well as a MAF of T = 0.000156/1 in the 1KGD. The nine transcript sequences of the extracted rare-coding deleterious SNP (*rs182024939*) by DSFE prediction include *MAPT*-201 (K672N: *ENST00000262410*.10), *MAPT*-204 (K280N: *ENST00000351559*.10), *MAPT*-205 (K615N: *ENST00000415613*.6), *MAPT*-206 (K251N: *ENST00000420682*.6), *MAPT*-208 (K222N: *ENST00000446361*.7), *MAPT*-212 (K597N: *ENST00000571987*.5), *MAPT*-214 (K280N: *ENST00000574436*.5), *MAPT*-217 (K251N: *ENST00000680542*.1), and *MAPT*-218 (K222N: *ENST00000680674*.1).

All nine filtered transcript sequences of MAPT showed similar chromosomal locations, alleles, and global MAF status. However, in all nine transcripts, the amino acid positions of the mutations were different. The SIFT, PolyPhen, and MetaLR statuses of all nine transcript-level mutations were deleterious, probably damaging, and damaging, respectively. The remaining three prediction tools, REVEL, CADD, and mutation assessor, classified these nine transcript-level mutations of MAPT as likely disease-causing, damaging, and medium, respectively (Table 1).

3.3. Consensus Classification-Based Computation of All Transcript Sequences of Identified Rare-Coding Pathogenic Variant of the MAPT Gene. By applying six different machine learning algorithms to compute the sequence-based pathogenicity of all nine transcript-based variants (rs182024939) of the MAPT gene, we found the most pathogenic effect in the MAPT-212 (K597N) transcript sequence. Only the PANTHER tool showed a potentially damaging effect. The remaining algorithms (PhD SNP, SNAP, *Meta*-SNP, SNAP2, and PMut) showed a disease impact. Besides this, MAPT-201 (K672N) and MAPT-205 (K615N) also showed some disease-linked impact in a few prediction tools (Table 2). The evolutionary conservation scores computed by the PANTHER tool for MAPT transcripts 201, 205, 206, 2012, and 217 showed possibly damaging effects. As per the findings reported in Table 2, the transcripts computed with deleterious impacts were MAPT 201 and 212. Interestingly, it was observed that MAPT 204, 208, 212, and 214 were disease-causing mutations, specifically indicated by PMut (Table 2).

https://doi.org/10.1021/acsomega.3c05769
ACS Omega 2023, 8, 40695−40712

Table 1. Extraction of Rare-Coding Deleterious SNPs through DSFE Predictions on Genomic Variants of Selected AD Genes

Gene transcript ID	Variant ID[a]	Location	Alleles	Global MAF	AA	SIFT class	PolyPhen class	CADD class	REVEL class	MetaLR class	Mutation assessor class
MAPT-201	rs182024939	17:46010327	G/A/T	<0.001	K672N	deleterious	probably damaging	likely deleterious	likely disease-causing	damaging	medium
MAPT-204	rs182024939	17:46010327	G/A/T	<0.001	K280N	deleterious	probably damaging	likely deleterious	likely disease-causing	damaging	medium
MAPT-205	rs182024939	17:46010327	G/A/T	<0.001	K615N	deleterious	probably damaging	likely deleterious	likely disease-causing	damaging	medium
MAPT-206	rs182024939	17:46010327	G/A/T	<0.001	K251N	deleterious	probably damaging	likely deleterious	likely disease-causing	damaging	medium
MAPT-208	rs182024939	17:46010327	G/A/T	<0.001	K222N	deleterious	probably damaging	likely deleterious	likely disease-causing	damaging	medium
MAPT-212	rs182024939	17:46010327	G/A/T	<0.001	K597N	deleterious	probably damaging	likely deleterious	likely disease-causing	damaging	medium
MAPT-214	rs182024939	17:46010327	G/A/T	<0.001	K280N	deleterious	probably damaging	likely deleterious	likely disease-causing	damaging	medium
MAPT-217	rs182024939	17:46010327	G/A/T	<0.001	K251N	deleterious	probably damaging	likely deleterious	likely disease-causing	damaging	medium
MAPT-218	rs182024939	17:46010327	G/A/T	<0.001	K222N	deleterious	probably damaging	likely deleterious	likely disease-causing	damaging	medium

[a]Missense variants reported in 1KGD and gnomAD projects. SIFT = 0.0 (deleterious) to 1.0 (tolerated); PolyPhen = 0.0 (benign) to 1.0 (probably damaging); CADD = ≥30 (likely deleterious); REVEL = >0.5 (likely disease causing); MetaLR = >0.9 (damaging); and mutation assessor = >0.8.

To get a consensus classification on these results, we used a classifier predictor tool to establish the pathogenic effects of these transcript-associated mutations. Conclusively, with the aid of a consensus classifier predictor, we computed that all nine transcript-based variants (*rs182024939*) of the MAPT gene have been characterized with potential pathogenic effects (Table 3). The consensus classification pattern indicated that the nine filtered transcripts of MAPT exhibited a pathogenic impact. Interestingly, the most pathogenic transcripts of MAPT computed using this classifier were 204, 205, and 214, respectively (Table 3). However, very few exceptions were also noticed in the different algorithms jointly covered under the consensus classification pattern (Table 3).

3.4. Categorizing the Transcript-Based Effects of Rare-Coding Deleterious Variations on Function, Stability, and Evolutionary Conservation of MAPT Protein. The computational investigations from various algorithms used to compute the effects of all nine transcript-based variants (rs182024939) of the MAPT gene on protein function and stability showed a marked decrease in the stability of proteins upon this mutation. The tools used to calculate the $\Delta\Delta G$ (kcal/mol) difference in energies showing the impact of the mutation on protein functions include iPTREE-STAB, MUpro, I Mutant, and INPS (Table 4). To investigate the impact of mutations on protein function, our findings indicated that the use of the algorithms mentioned above resulted in a uniform pattern of protein destabilization with very little variation in computed scores (Table 4). In addition, the findings in Table 4 strengthen and validate the outcomes of the previously used consensus classification experiments on nine filtered transcripts of MAPT.

The computed evolutionary conservation scores (BNC-Bayesian normalized conserved score) showed that the three transcript sequences, that is, MAPT-201 (K672N), MAPT-205 (K615N), and MAPT-212 (K597N), were highly conserved as well as exposed, and the mutations present in these sequences were on the extended strand, β-turn, and α helix, respectively. Additionally, the protein sequences of these transcripts of the MAPT gene have a higher proportion of exposed residues, which were computed using an algorithm that calculates the NMC using the Bayesian method with a CI for each of the inferred evolutionary conservation scores. Specifically, the findings from this study indicated that MAPT transcripts 201, 205, and 212 were highly conserved (evolutionary conservative amino acids: residues with scores between 7 and 8). The rest of the transcript sequences also showed good, conserved scores, and were exposed in the transcript-sequence function determination of entire proteins (Table 5).

3.5. Impact of Transcript-Sequence-Based Modifications of Rare-Coding Deleterious Variations on Secondary Structure, Torsional Angles, Energetic Potentials, and Protein Dynamics. By computing the impact of transcript-sequence-based modifications of rare-coding deleterious variation (nsSNPs) on secondary structure and torsional angles, energetic potentials showed the overall destabilization of protein structure upon mutation (Table 6). Similarly, the predicted energy difference because of mutations in nine different transcript sequences of the MAPT gene showed validation of the entire structural destabilization impact (Table 6). In particular, the computed higher RSA and ASA values observed in MAPT transcripts 204 and 214 also showed lower $\Delta\Delta G$ energetic values than the others. However, this analysis at the protein's secondary structural level helps to relate the

https://doi.org/10.1021/acsomega.3c05769
ACS Omega 2023, 8, 40695−40712

Table 2. Transcript-Sequence-Based Evolutionary Conservation Analyses and Prediction of nsSNPs' Deleterious Impact on the MAPT Gene

Variant ID	Gene transcript ID	Amino acid change	PANTHER	PhD SNP	SNAP	*Meta*-SNP	SNAP2	PMut
rs182024939	MAPT-201	K672N	0.5[a]	0.518[c]	0.685[c]	0.671[c]	57 (75%)	NC
rs182024939	MAPT-204	K280N	0.366[b]	0.489	NC	0.469	49 (71%)	0.83[c]
rs182024939	MAPT-205	K615N	0.5[a]	0.523[c]	0.560[c]	0.644[c]	34 (66%)	NC
rs182024939	MAPT-206	K251N	0.5[a]	0.484[c]	NC	0.466	39 (66%)	NC
rs182024939	MAPT-208	K222N	0.366[b]	0.462	0.475	0.422	39 (66%)	0.83[c]
rs182024939	MAPT-212	K597N	0.5[a]	0.521[c]	0.620[c]	0.611[c]	90 (95%)	0.65[c]
rs182024939	MAPT-214	K280N	0.366[b]	0.489	NC	0.469	49 (71%)	0.83[c]
rs182024939	MAPT-217	K251N	0.5[a]	0.484	NC	0.466	39 (66%)	NC
rs182024939	MAPT-218	K222N	0.366[b]	0.549[c]	0.470	0.485	37 (66%)	NC

[a]Possibly damaging. [b]Neutral. [c]Disease causing SNP; NC = not computed. PANTHER, PhD-SNP, SNAP, *Meta*-SNP = output normalized between 0 and 1 (if > 0.5 mutation is predicted disease causing); SNAP2 = A high score (>50 indicates strong signal for effect), weak signals (−50 < score <50), and a low score (score <−50), strong signal for neutral/no effect. The prediction scores of PMut are from 0 to 1 and the cutoff value is set to 0.5 (neutral, 0−0.5; pathological, 0.5−1).

Table 3. Transcript-Based Computation of Percent Expected Accuracy through Consensus Classifier Prediction of the Rare-Coding Deleterious nsSNPs of the MAPT Gene

Gene transcripts ID	AA	Chromosome: base pair	PredictSNP	PhD-SNP	PolyPhen-1	PolyPhen-2	SIFT	SNAP	PANTHER	MAPP
MAPT-201	K672N	17:46010327	72[a]	82[a]	59[a]	59[a]	79[a]	62[a]	71[a]	64[b]
MAPT-204	K280N	17:46010327	87[a]	82[a]	59[a]	54[a]	79[a]	62[a]	71[a]	57[a]
MAPT-205	K615N	17:46010327	87[a]	82[a]	59[a]	63[a]	79[a]	56[a]	71[a]	43[a]
MAPT-206	K251N	17:46010327	76[a]	82[a]	59[a]	54[a]	79[a]	50[b]	71[a]	46[a]
MAPT-208	K222N	17:46010327	51[a]	82[a]	59[a]	74[b]	79[a]	50[b]	71[a]	64[b]
MAPT-212	K597N	17:46010327	61[a]	82[a]	59[a]	59[a]	79[a]	50[b]	71[a]	70[b]
MAPT-214	K280N	17:46010327	87[a]	82[a]	59[a]	54[a]	79[a]	62[a]	71[a]	57[a]
MAPT-217	K251N	17:46010327	76[a]	82[a]	59[a]	54[a]	79[a]	50[b]	NC	46[a]
MAPT-218	K222N	17:46010327	72[a]	86[a]	59[a]	63[a]	79[a]	61[b]	71[a]	48[a]

[a]Deleterious and disease-causing. [b]Neutral; NC = not computed; AA = amino acid.

Table 4. Transcript-Based Computation of Changes in Free Energy of Amino Acid Substitution in Rare nsSNPs of the MAPT Gene[a]

Gene transcript ID	variant ID	Amino acid variation	iPTREE-STAB $\Delta\Delta G$ (kcal/mol)	MUpro $\Delta\Delta G$ (kcal/mol)	MUpro_SVM $\Delta\Delta G$ (kcal/mol)	MUpro_NN $\Delta\Delta G$ (kcal/mol)	I mutant 2.0 $\Delta\Delta G$ (kcal/mol)	INPS (kcal/mol)
MAPT-201	rs182024939	K672N	−1.38	−1.364	−1	−0.827	−0.55	−0.897982
MAPT-204	rs182024939	K280N	−1.38	−1.364	−1	−0.827	−0.55	−0.873567
MAPT-205	rs182024939	K615N	−1.38	−1.364	−1	−0.827	−0.55	−0.873567
MAPT-206	rs182024939	K251N	−1.38	−1.364	−1	−0.827	−0.55	−0.873567
MAPT-208	rs182024939	K222N	−1.38	−1.364	−1	−0.827	−0.55	−0.847001
MAPT-212	rs182024939	K597N	−1.43	−1.364	−1	−0.827	−0.55	−0.847001
MAPT-214	rs182024939	K280N	−1.38	−1.364	−1	−0.827	−0.55	−0.873567
MAPT-217	rs182024939	K251N	−1.38	−1.364	−1	−0.827	−0.55	−0.873567
MAPT-218	rs182024939	K222N	−1.38	−1.364	−1	−0.827	−0.55	−0.847001

[a]The computed negative values of $\Delta\Delta G$ indicate a decrease in the stability of the protein. The above predictions can be interpreted to identify stabilizing ($\Delta\Delta G > 0$) and destabilizing ($\Delta\Delta G < 0$) variations.

previous computational prediction as well as to better understand the inclusive destabilization impact in almost all transcripts of MAPT (Table 6).

The SDM impact analysis further confirmed the energetic variations as well as the characteristic attribute difference because of residue modification (genetic variation-nsSNPs), which could serve as a basis for understanding the appearance of protein destabilization upon such mutations (Table 7). Through SDM analysis, we further validated that the MAPT transcripts 204 and 214 had higher RSA and MTDEPTH status (compared to their WTs) as well as a higher protein destabilization impact (Table 7). Additionally, this was also observed in our previous experiment (Table 6), which showed the destabilization impact of these transcript-level mutations on the protein structure.

Moreover, the DynaMut prediction of protein stability changes in nine different transcript sequences of the MAPT gene showed an increase in vibrational entropy due to mutations. This increase in the vibrational entropy of all nine transcript-level mutant models of MAPT showed denaturation of the protein state due to amino acid substitutions (Table 8). We also found some mixed effects showing a decrease in the molecular stabilization pattern, especially in all filtered transcripts between WT and mutant models (nsSNPs) of the MAPT transcript sequence compared (Table 8).

https://doi.org/10.1021/acsomega.3c05769
ACS Omega 2023, 8, 40695−40712

Table 5. Evolutionary Conservation Analyses and Protein Structure Prediction of Rare-Coding Deleterious SNPs of the *MAPT* Gene

Gene transcript ID	Variant ID	Amino acid change	ConSurf score/BNC score (CI)[a]	Evolutionary analysis	SOPMA prediction	Secondary structure information	Surface accessibility
MAPT-201	rs182024939	K672N	8/−0.71 (9:−0.937 and 7:−0.563)	highly conserved and exposed	extended strand	α-helix = 1.9%, β-strand = 1.7%, and coil = 96.4%	buried = 5% and exposed = 95%
MAPT-204	rs182024939	K280N	7/−0.422 (7:−0.628 and 6:−0.315)	exposed	β turn	α-helix = 4.3%, β-strand = 4.1%, and coil = 91.6%	buried = 8.6% and exposed = 91.4%
MAPT-205	rs182024939	K615N	8/−0.677 (8:−0.898, 7:−0.536)	highly conserved and exposed	β turn	α-helix = 2.4%, β-strand = 2.4%, and coil = 95.1%	buried = 5% and exposed = 95%
MAPT-206	rs182024939	K251N	6/−0.417 (7:−0.654, 6:−0.260)	exposed	random coil	α-helix = 4.9%, β-strand = 3.4%, and coil = 91.7%	buried = 10.7% and exposed = 89.3%
MAPT-208	rs182024939	K222N	6/−0.440 (7:−0.677, 6:−0.281)	exposed	random coil	α-helix = 3.1%, β-strand = 2.9%, and coil = 94%	buried = 10.2% and exposed = 89.8%
MAPT-212	rs182024939	K597N	8/−0.816 (9:−1.049, 7:−0.648)	highly conserved and exposed	α helix	α-helix = 2.5% and coil = 97.5%	buried = 2.9% and exposed = 97.1%
MAPT-214	rs182024939	K280N	7/−0.422 (7:−0.628, 6:−0.315)	exposed	β turn	α-helix = 4.3%, β-strand = 4.1%, and coil = 91.6%	buried = 8.6% and exposed = 91.4%
MAPT-217	rs182024939	K251N	6/−0.417 (7:−0.654, 6:−0.260)	exposed	random coil	α-helix = 4.9%, β-strand = 3.4%, and coil = 91.7%	buried = 10.7% and exposed = 89.3%
MAPT-218	rs182024939	K222N	6/−0.306 (9:−1.189, 4:0.174)	exposed	random coil	α-helix = 4.2%, β-strand = 6.6%, and coil = 89.2%	buried = 26.2% and exposed = 73.8%

[a]BNC = Bayesian method for calculating rates with a CI to each of the inferred evolutionary conservation scores. The amino acids with scores between 7 and 9 were evolutionary conservative amino acids.

Table 6. Impact of Structural Modifications through Transcript-Level Homology Models of MAPT Protein on the Secondary Structure, Torsional Angles, and Energetic Potentials of Protein[a]

MAPT gene transcript	Mutations	RSA (%)	ASA (Å)	SS3	SS8	Phi (ϕ)	Psi (ψ)	Intrinsically disordered regions (%)	Torsion angle potentials	Overall stability	Predicted $\Delta\Delta G$ (kcal/mol)
201	K672N	50	117	strand	B-sheet	−122	138	4	favorable	destabilizing	−0.34
204	K280N	52	123	strand	B-sheet	−123	139	3	unfavorable	destabilizing	−0.82
205	K615N	49	117	strand	B-sheet	−124	138	4	favorable	destabilizing	−0.34
206	K251N	50	117	strand	B-sheet	−123	139	3	favorable	destabilizing	−0.34
208	K222N	51	119	strand	B-sheet	−125	139	3	favorable	destabilizing	−0.34
212	K597N	49	117	strand	B-sheet	−123	139	3	favorable	destabilizing	−0.34
214	K280N	52	123	strand	B-sheet	−123	139	3	unfavorable	destabilizing	−0.82
217	K251N	50	117	strand	B-sheet	−123	139	3	favorable	destabilizing	−0.34
218	K222N	51	120	strand	B-sheet	−124	138	3	favorable	destabilizing	−0.32

[a]RSA = relative solvent accessible area and ASA = accessible surface area. The computed negative values of $\Delta\Delta G$ indicate a decrease in the stability of the protein. The above predictions can be interpreted to identify stabilizing ($\Delta\Delta G > 0$) and destabilizing ($\Delta\Delta G < 0$) variations.

Table 7. SDM Analyses to Investigate the Impact of Amino Acid Substitution on 3D Protein Models and on the Overall Structural Environment of MAPT Protein

MAPT gene transcript ID	Mutations	WT RSA[a] (%)	WT DEPTH[b] (Å)	WT OSP[c]	MT RSA[d] (%)	MT DEPTH[e] (Å)	MT OSP[f]	SDM predicted[g] $\Delta\Delta G$
201	K672N	60.9	3.5	0.13	76.8	3.4	0.13	0.24
204	K280N	51.8	3.4	0.17	57.5	3.5	0.13	−0.61
205	K615N	60.9	3.5	0.13	76.8	3.4	0.13	0.24
206	K251N	60.9	3.5	0.13	76.8	3.4	0.13	0.24
208	K222N	60.9	3.5	0.13	76.8	3.4	0.13	0.24
212	K597N	60.9	3.5	0.13	76.8	3.4	0.13	0.24
214	K280N	51.8	3.4	0.17	57.5	3.5	0.13	−0.61
217	K251N	60.9	3.5	0.13	76.8	3.4	0.13	0.24
218	K222N	60.9	3.3	0.13	74.9	3.3	0.13	0.25

[a]Residue relative solvent accessibility for WT. [b]Residue depth for WT. [c]Occluded surface packing value for WT residue. [d]Relative solvent accessibility for modified residue. [e]Depth for modified residue. [f]Occluded surface packing value for modified residue. [g]The computed negative values of $\Delta\Delta G$ indicate a decrease in the stability of the protein. The above predictions can be interpreted to identify stabilizing ($\Delta\Delta G > 0$) and destabilizing ($\Delta\Delta G < 0$) variations.

3.6. Homology Model Selection, Validation, Bioactive (Flavonoids) Compounds' 3D Library Preparation, and Selection on the Basis of Drug-like Properties Assessment. We retrieved the WT 3D homology model of a selected MAPT protein from the AlphaFold database, and the PDB file of this 3D protein was accessed through UniProt accession no. P10636. The reason for the selection of this model was the availability of a complete 3D model according to the sequence length (1−758), with reference to the 3D protein model prediction strategy. The downloaded model in PDB format

https://doi.org/10.1021/acsomega.3c05769
ACS Omega 2023, 8, 40695−40712

Table 8. DynaMut Prediction of Protein Stability Changes of Amino Acid Substitution on 3D Protein Models through NMA of MAPT Protein[a,b,c]

MAPT gene transcript ID	Mutations	SAAMBE ΔΔG (kcal/mol)	DynaMut ΔΔG (kcal/mol)	NMA ENCoM ΔΔG (kcal/mol)	mCSM ΔΔG (kcal/mol)	SDM ΔΔG (kcal/mol)	DUET ΔΔG (kcal/mol)	ΔΔS Vib ENCoM (kcal.mol^{-1}.K^{-1})
201	K672N	−0.12^{D}	0.525^{S}	−1.017^{D}	0.088^{S}	0.100^{S}	0.327^{S}	1.271^{b}
204	K280N	−0.12^{D}	−0.266^{D}	−0.338^{D}	0.185^{S}	−0.040^{D}	0.395^{S}	0.423^{b}
205	K615N	−0.12^{D}	0.525^{S}	−1.017^{D}	0.088^{S}	0.100^{S}	0.327^{S}	1.271^{b}
206	K251N	−0.12^{D}	0.525^{S}	−1.017^{D}	0.088^{S}	0.100^{S}	0.327^{S}	1.271^{b}
208	K222N	−0.12^{D}	0.525^{S}	−1.017^{D}	0.088^{S}	0.100^{S}	0.327^{S}	1.271^{b}
212	K597N	−0.12^{D}	0.525^{S}	−1.017^{D}	0.088^{S}	0.100^{S}	0.327^{S}	1.271^{b}
214	K280N	−0.12^{D}	−0.266^{D}	−0.338^{D}	0.185^{S}	−0.040^{D}	0.395^{S}	0.423^{b}
217	K251N	−0.12^{D}	0.525^{S}	−1.017^{D}	0.088^{S}	0.100^{S}	0.327^{S}	1.271^{b}
218	K222N	−0.12^{D}	0.525^{S}	−0.809^{D}	0.082^{S}	0.100^{S}	0.322^{S}	1.011^{b}

[a]S = stabilizing and D = destabilizing. [b]Increase of molecular flexibility. [c]NMA = normal-mode analysis, ENCoM = elastic network contact model, mCSM = mutation cutoff scanning matrix, SDM = site-directed mutator, and ΔΔS Vib ENCoM = vibrational entropy energy between WT and mutant. The computed negative values of ΔΔG indicate a decrease in the stability of the protein. The above predictions can be interpreted to identify stabilizing (ΔΔG > 0) and destabilizing (ΔΔG < 0) variations.

Table 9. Virtual Screening through a Graphical-Automatic Drug Design (*iGEMDOCK*) Approach and Post-Screening Analysis of Top-Ranked Flavonoid Complexes with MAPT

Flavonoids[a]	PUBCHEM ID	Fitness value (kcal/mol)	Rank	Atom	H-bond	Ligand interactions
luteolin	5280445	−69.928722	6	21	LYS607, VAL626, TYR627, and LYS628	LYS607, VAL626, TYR627, LYS628, PRO629, VAL630, and LEU632
eriodictyol	440735	−73.27035	4	21	GLU306	VAL305, GLU306, and LYS657
hesperetin	72281	−74.098458	2	22	THR308, GLN653, and GLU655	VAL305, THR308, GLN653, and GLU655
alpinetin	154279	−73.332331	3	20	ILE614 and LYS615	ILE614, LYS615, and HIS616
7-*O*-methyleriodictyol	1268276	−77.020284	1	22	GLU306 and LYS657	VAL305, GLU306, ILE307, VAL656, and LYS657
sakuranetin	73571	−72.6986	5	21	LYS615 and HIS616	ASN613, ILE614, LYS615, and HIS616
pinocembrin	68071	−68.404283	7	19	GLU655	VAL305, ILE307, GLU655, and LYS657
aromadendrin	122850	−67.066335	8	21	MET31 and SER516	MET31, ASP34, SER515, and SER516

[a]These flavonoids were screened and passed drug-likeness-based inclusion criteria.

was further investigated to ensure that the structure was clean, as no heteroatoms, ligands, water molecules, or other unwanted ions or charges were present that might affect the accuracy and reliability of the desired protein 3D model. The Ramachandran plot analysis of the 3D model selected for further work showed that 41.3% of residues were in the core region, and 29.1 and 10.8% were in the generously allowed region. The 3D structural residue properties showed that the maximum deviation was 23.2, the bad-contact score was one, and the bond/length angle was 35.4. The 3D structural *G*-factor analysis through SAVES showed that dihedrals were −1.36, covalent ones were −1.33, and overall was −1.28. Furthermore, the planar group analysis showed that 94.8% of residues were within limits, and 5.2% were highlighted.

Besides this, we extracted sixty-one flavonoids previously reported as prospective compounds for the modulation of neurodegeneration by Ayaz et al., 2019. For all selected compounds (flavonoids), the downloaded 3D structures were visually inspected to ensure a proper 3D transformation. Subsequently, they were screened using the Lipinski rule of five criteria. We identified forty-one flavonoid compounds that met all five criteria, and the remaining compounds were excluded. On forty-one screened compounds, we applied the OSIRIS drug property explorer, among which we found twenty-three compounds that were retained as nonmutagenic, nontumorigenic, and nonirritating compounds. We found five more compounds that had reproductive risks or effects so we excluded them. From eighteen compounds, we examined each individual compound on the basis of its OSIRIS computed drug score, and we included only eight compounds that had a score of >0.8 (80%) (Table S2).

3.7. iGEMDOCK-Based Virtual Screening of Bioactive Flavonoids, Identification of Druggable Protein Cavities with a Cavity Detection-Guided Binding Approach, and MM/PB(GB)SA Computation for Near Native Docking Pose through Docking, Rescoring, and Hotspot Residue Analysis. Initially, the *iGEMDOCK* tool was used for virtual screening of the library of flavonoid compounds, and the experiment was distinct in two steps. In the first step, we allowed all sixty-one flavonoid bioactive compounds to dock with the MAPT protein to obtain significant molecular interactions with the fitness score (Table S3). In the second step, we inspected and filtered only those compounds (with binding interactions) that initially passed all steps of the inclusion criteria (drug-likeness) for compound assessment. Of the eight screened flavonoid (exhibiting drug-likeness properties) compounds, 7-*O*-methyleriodictyol (−77.02 kcal/mol) was the top-ranked compound with the highest fitness value with the MAPT protein. The second- and third-ranked compounds were hesperetin (−74.098458 kcal/mol) and alpinetin (−73.33 kcal/mol), respectively (Table 9 and Figure 3).

We employed the CB-dock tool to identify druggable protein cavities in MAPT using a cavity detection-guided binding approach with screened drug-like flavonoids. Five top-ranked docked flavonoid compounds identified through the

https://doi.org/10.1021/acsomega.3c05769
ACS Omega 2023, 8, 40695−40712

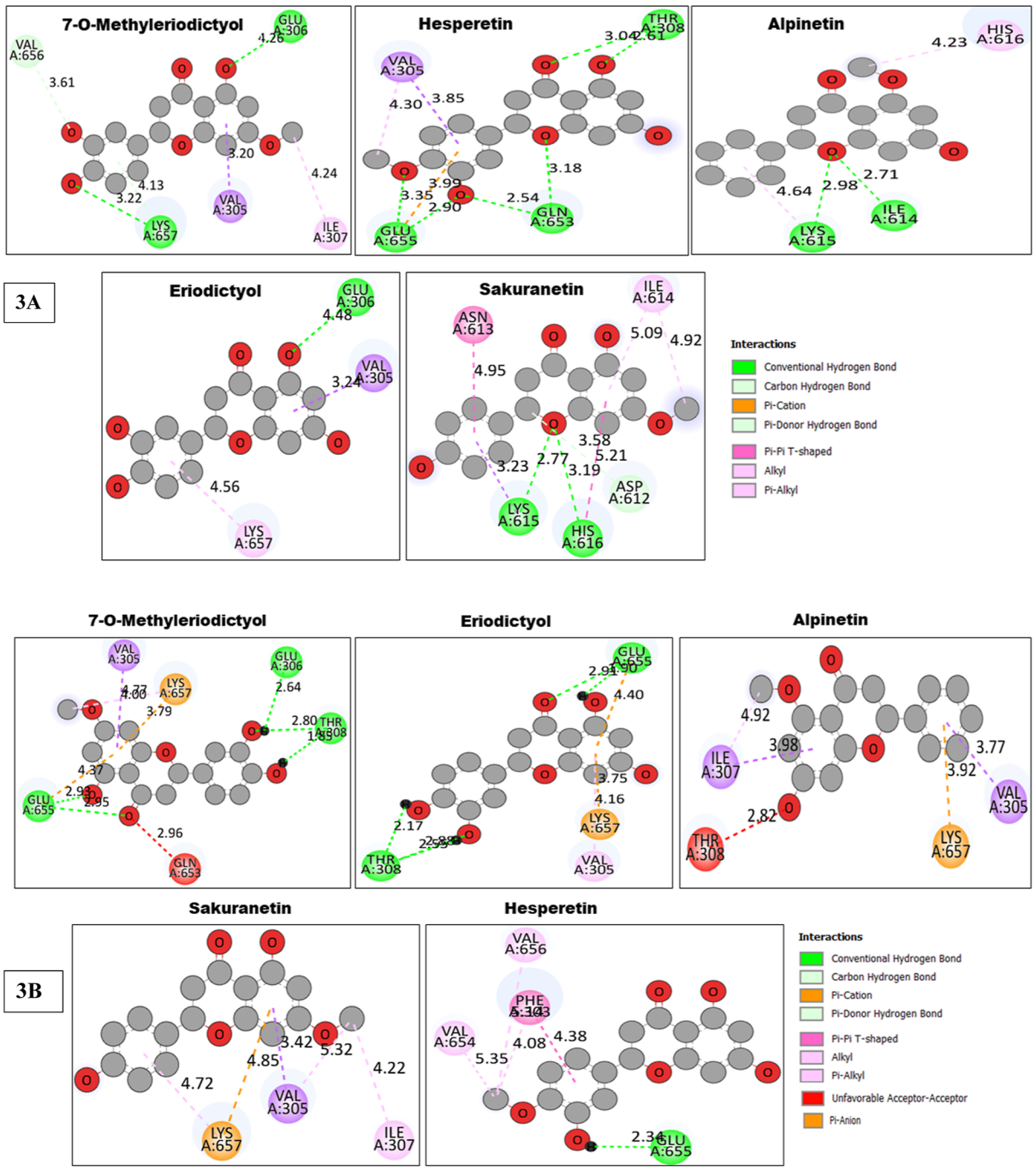


Figure 3. 2D ligand−protein interactions. (A) *iGEMDOCK*-based 2D ligand−protein interactions of the top five docked flavonoid compounds with the MAPT protein. (B) CB docking-based 2D ligand−protein interactions of the top five docked flavonoid compounds with the MAPT protein.

iGEMDOCK approach were used in this CB-dock experiment to identify druggable protein cavities that reside in the MAPT protein. In this case, the highest binding affinities were observed with 7-*O*-methyleriodictyol as well as eriodictyol; that is, −6.2 kcal/mol, which showed a prominent consistency of outcomes produced from both iGEMDOCK and CB-dock experimentations. The lowest binding affinity was observed for hesperetin, that is, −5.2 kcal/mol (Table 10 and Figure 3).

Besides this, we used the fastDRH-based prediction method for determination of the best native binding pose for the protein-compound (flavonoids) complex with the aid of MM/PB(GB)SA computation using rescoring functions. The outcomes showed hesperetin (−5.64 kcal/mol) as the best-docked compound on the basis of generated energetic estimates. Besides this, eriodictyol (−5.63 kcal/mol) and sakuranetin (−5.6 kcal/mol) were the other top-docked compounds (Table 11 and Figure 4). To provide a better understanding of these interactions, we included an energy plot of the top 10 residues (Figure 5) and a heat map of the top 30 residues (Figure 6), which were generated from the same above web server.

4. DISCUSSION

AD is a complex neurodegenerative disorder with multifactorial etiology, resulting from a combination of genetic, environmental, and lifestyle factors.[1] The genetics of AD are

https://doi.org/10.1021/acsomega.3c05769
ACS Omega 2023, 8, 40695−40712

Table 10. Identification of Druggable Protein Cavities with Molecular Residues Interactions of MAPT Protein with Top Five Identified Bioactive Flavonoids Using the Cavity-Binding (CB) Docking Approach

Compound	Protein cavity	Vina score (kcal/mol)	Cavity volume (Å^3)	Center (x, y, and z)	Docking size (x, y, and z)	Contact residues	Ligand interactions	H-bonds
7-O-methyleriodictyol	C1	−6.2	181	−38, −15, −19	21, 21, 21	HIS304, VAL305, GLU306, ILE307, THR308, PRO309, ASN310, GLN653, VAL654, GLU655, VAL656, and LYS657	VAL305, GLU306, THR308, GLN653, GLU655, and LYS657	GLU306, THR308, and GLU655
hesperetin	C1	−5.3	181	−38, −15, −19	20, 20, 20	PHE301, THR302, PHE303, HIS304, VAL305, VAL654, GLU655, VAL656, and LYS657	PHE303, VAL654, GLU655, and VAL656	GLU655
alpinetin	C1	−5.9	181	−38, −15, −19	20, 20, 20	HIS304, VAL305, GLU306, ILE307, THR308, PRO309, GLN653, GLU655, and LYS657	VAL305, ILE307, THR308, and LYS657	
eriodictyol	C1	−6.2	181	−38, −15, −19	21, 21, 21	HIS304, VAL305, GLU306, ILE307, THR308, PRO309, ASN310, GLN653, VAL654, GLU655, VAL656, and LYS657	VAL305, THR308, GLU655, and LYS657	THR308 and GLU655
sakuranetin	C1	−5.8	181	−38, −15, −19	21, 21, 21	PHE303, HIS304, VAL305, GLU306, ILE307, GLN653, GLU655, VAL656, and LYS657	VAL305, ILE307, and LYS657	

particularly complex, with multiple genes interacting to influence the disease's onset and progression.[2] The role of genetic variations which can impact protein functioning and set the basis for AD cannot be ignored.[62] Thus, to understand the impact of gene-specific mutations, annotation of potential AD biomarker gene variants is still needed. Identifying the genetic loci associated with biomarkers could aid in better understanding the specific pathophysiological components underpinning these biomarkers.[63] Therefore, we used the set NIA-AA criteria to investigate the functional impact of transcript-sequence-based deleterious exonic SNPs taken from the AD biomarker genes.[15]

We selected the large set of exonic mutations of APP, NEFL, and MAPT genes, applied the DSFE prediction tools, and found nine different transcripts of MAPT's single-gene variation (rs182024939: K > N) which emerged as the most damaging SNP among all of the mutations. By providing insights into the potential functional impact of a genetic variant, the DSFE predictions can help in categorizing promising gene-variant targets for further analysis as well as in benefiting the development of personalized treatments for individuals.[64] The obtained DSFE results were further investigated through a number of machine learning algorithms like PANTHER, PhD-SNP, SNAP, SNAP2, and PMut. Our analysis revealed that MAPT transcripts 201, 205, and 212 were found as the most pathogenic among the nine different transcripts of the rs182024939 SNP. The pathogenicity of all nine selected transcript-level SNPs of the MAPT gene was confirmed through a consensus classifier predictor algorithm. This finding is particularly significant given the critical role that MAPT plays in maintaining the structure and function of neurons in the brain.[8,9] Normally, tau helps stabilize microtubules, which are vital for maintaining neuronal shape and facilitating intracellular transport. However, in AD tau becomes abnormally phosphorylated, leading to the formation of tau tangles and the disruption of these important cellular processes.[9] Interestingly, our results suggest that all the transcript sequence-based MAPT mutations showed a significant extent of destabilization in terms of protein function, which might serve as the basis for understanding its possible role in AD.

Although the gene pathogenic variations that we found were relatively rare, their identification and characterization can provide valuable insights into the biological mechanisms underlying AD and pave the way for the development of new therapies. Previous studies have clearly reported that SNPs in the MAPT gene are associated with an increased risk of developing AD.[65] Interestingly, our study also evaluated the impact of missense mutations on the MAPT protein and found that the substitution of lysine with asparagine (K > N) in nine different transcript-level SNPs resulted in a generalized destabilizing effect on the protein's stability, function, and dynamics. This inference was supported by the calculation of the associated free energy change, indicating a significant impact on the protein's structural and functional properties. Additionally, our analysis of the impact of mutations on protein properties revealed that all nine transcript-level mutations of the MAPT protein had destabilizing energies. Mutations can affect a protein's structural modifications, secondary structure, torsional angles, and energetic potentials, which can ultimately affect its stability, function, and ability to interact with other molecules.[66] Further exploration of these

https://doi.org/10.1021/acsomega.3c05769
ACS Omega 2023, 8, 40695−40712

Table 11. Predicting the Near Native Binding Pose and Analysis for the MAPT Protein−Ligand Complexes through MM/PB(GB)SA Computation

Compounds	Docking score (kcal/mol)	PB1 (kcal/mol)	PB3 (kcal/mol)	PB4 (kcal/mol)	GB1 (kcal/mol)	GB2 (kcal/mol)	GB5 (kcal/mol)	GB6 (kcal/mol)	GB7 (kcal/mol)	GB8 (kcal/mol)
7-*O*-methyleriodictyol	−5.13	−3.4	−11.91	−11.74	−16.94	−15.02	−14.86	−11.06	−13.43	−14.29
hesperetin	−5.64	−2.33	−14.49	−18.97	−24.52	−21.55	−21.2	−16.73	−19.83	−20.2
alpinetin	−4.86	−4.79	−12.17	−12.99	−15.68	−14.26	−14.06	−11.9	−13.12	−12.73
eriodictyol	−5.63	0.6	−4.68	−5.58	−6.67	−5.3	−5.14	−2.77	−4.5	−5.45
sakuranetin	−5.6	−0.75	−12.75	−17.48	−23.38	−20.5	−20.14	−13.62	−18.15	−18.49

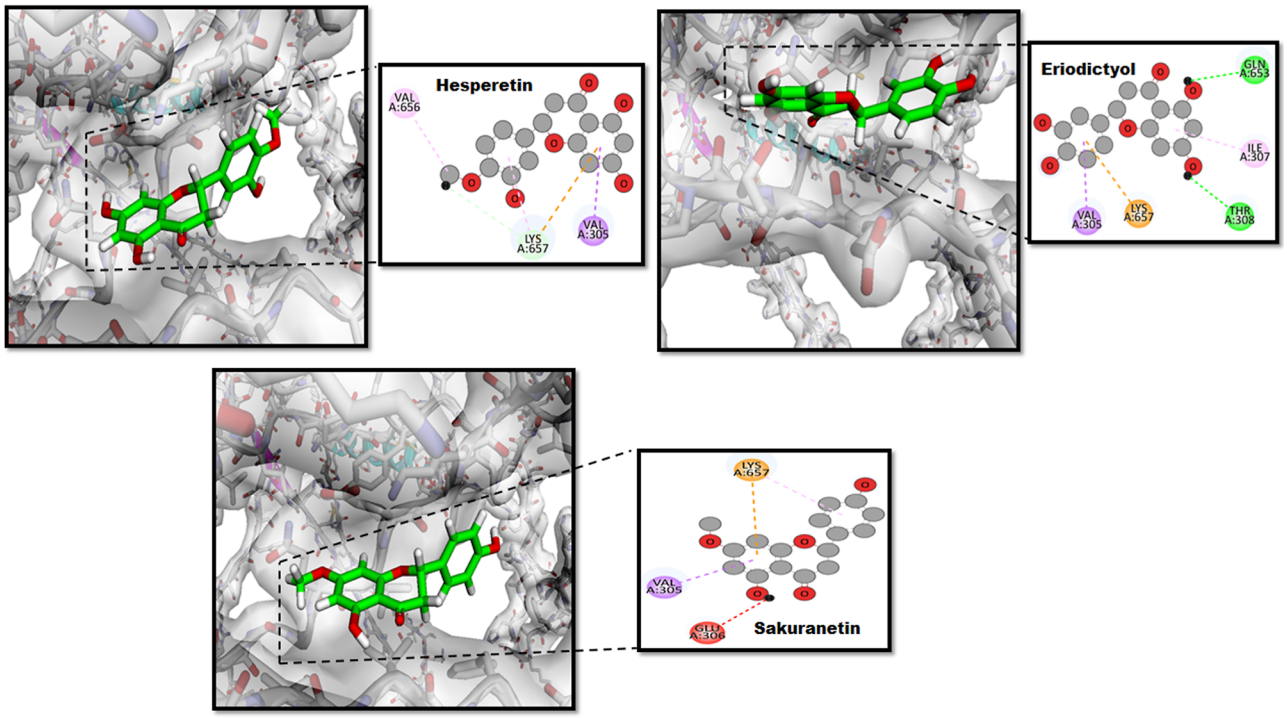


Figure 4. Predicting the top three near-native binding poses and analysis for protein−ligand complexes based on the MM/PB(GB)SA computation. The coloring scheme shows green—conventional H-bonds; light green—van der Waals; dark purple—π−π stacked; light purple—π-alkyl; red—unfavorable; and yellow—π-cation/π-anion.

mutations has important implications for disease pathology and drug discovery, especially in the management of AD.

The use of evolutionary conservation analyses to predict the functional significance of mutations is a novel approach that can help identify SNPs, which may contribute to disease pathogenesis.[44] The identification of three highly conserved transcripts (201, 205, and 212) in our study suggests that these SNPs are more likely to have deleterious effects on the production and processing of tau protein. These findings shared valuable insights regarding the exploration of potentially (evolutionary) conserved transcripts of the MAPT gene observed with specific mutation and its relevance to their possible role in disease progression.

Moreover, we also conducted SDM analysis to investigate the impact of amino acid substitutions on the structural environment of MAPT. This analysis (transcript-sequence-based) involved creating mutant proteins with a single amino acid substitution in specific regions and predicting their structures using a homology modeling method.[49] Our results showed that these substitutions led to variations in the MAPT protein, suggesting that it has the potential to impact the protein's function and stability. The SDM analysis conducted in our study provided valuable information regarding the specific regions of the protein that are affected by amino acid substitutions. This information can serve as a particular contribution to guide future studies aimed at understanding the molecular mechanisms of disease pathology. Besides this, we also utilized the DynaMut computational tool, which predicts changes in protein stability resulting from amino acid substitutions using NMA.[50] The tool outputs predicted changes in free energy and increased vibrational entropy energy between the WT and mutant proteins of MAPT's transcript, which were used to interpret the change in protein stability.

Studies have reported the role of flavonoids like quercetin and EGCG (epigallocatechin gallate) in the management of AD. They have antioxidant and anti-inflammatory properties that reduce oxidative stress and inflammation in the brain.[53] They also inhibit β-amyloid plaque formation and aggregation, a hallmark of AD pathology.[17] Similarly, curcumin and resveratrol have been shown to protect neurons from oxidative damage and improve cognitive function in animal models of AD.[67] Keeping the therapeutic potential of bioactive flavonoids in mind, we conducted an extensive literature search to create a 3D compound library of flavonoids and assessed their drug likeness.[17] We used a virtual screening approach through the

https://doi.org/10.1021/acsomega.3c05769
ACS Omega 2023, 8, 40695−40712

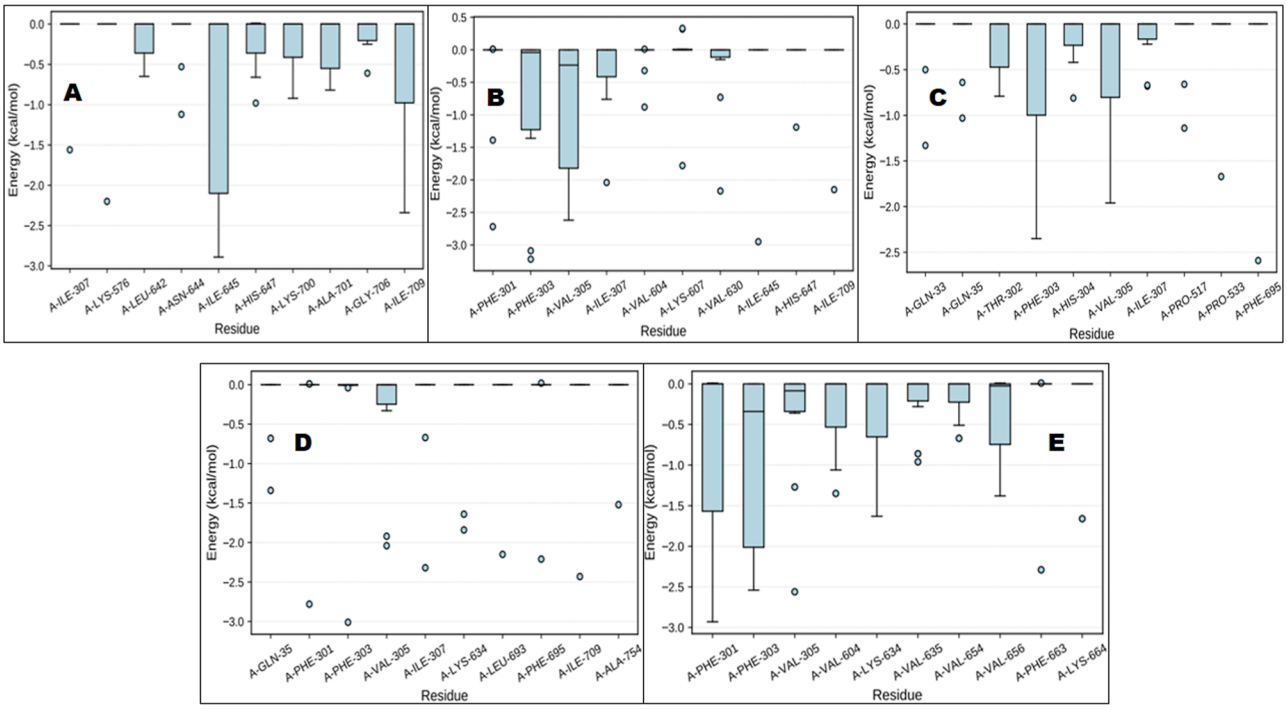


Figure 5. Energy plot of top 10 residues of best-ranked docked compounds with the MAPT proteins. Compounds (A—7-*O*-methyleriodictyol; B—hesperetin; C—alpinetin; D—eriodictyol; and E—sakuranetin). *X*-axis showing the top ten energetically interacting residues of MAPT protein with the respective flavonoids. The *Y*-axis shows estimated energy in kcal/mol.

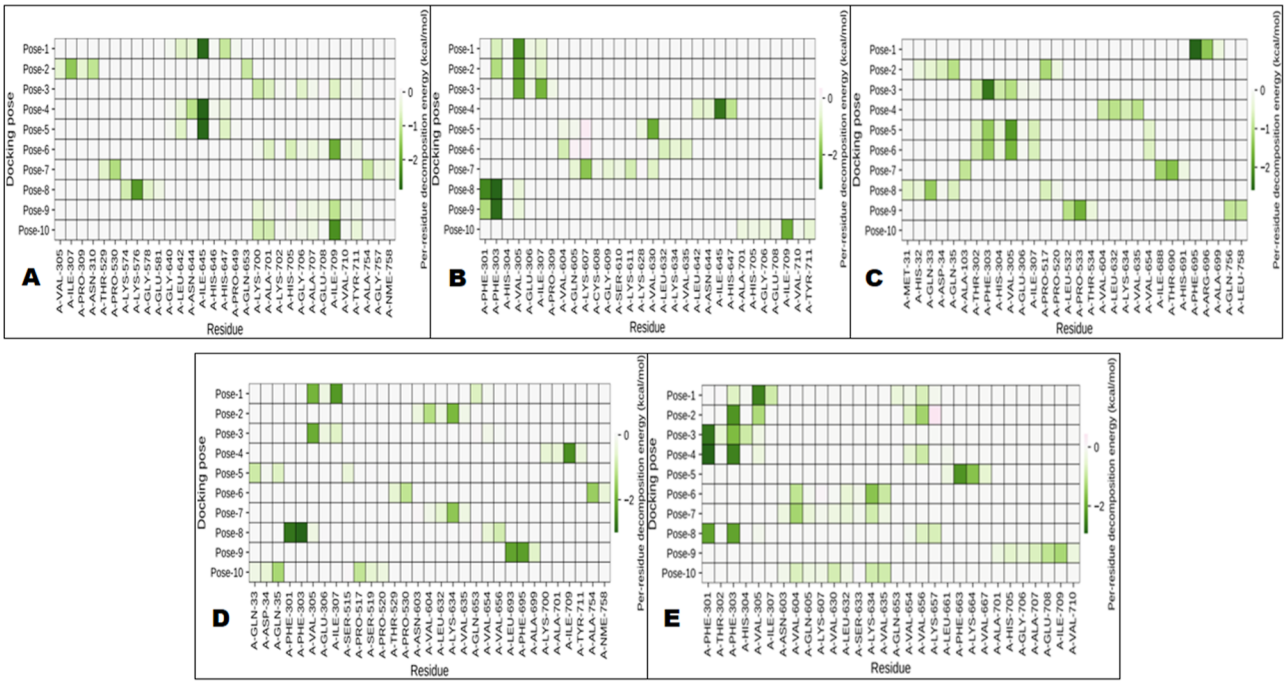


Figure 6. Heat-map of top 30 residues of best-ranked docked compounds with the MAPT proteins. Compounds (A—7-*O*-methyleriodictyol; B—hesperetin; C—alpinetin; D—eriodictyol; and E—sakuranetin). *X*-axis showing the top 30 energetically interacting residues of MAPT protein with the respective flavonoids. The *Y*-axis shows docking poses.

iGEMDOCK tool to identify potential drug candidates that can bind to our target protein.[59] Our results identified several top-ranked compounds with favorable binding interactions that have potential as drug candidates. This computational approach can effectively narrow the list of potential compounds and guide further experimental testing, ultimately leading to the development of effective drugs for the treatment of various diseases.

Our main interest was to investigate the druggable protein cavities in MAPT showing prominent molecular interactions of bioactive flavonoids, for which we utilized the CB-dock methodological approach.[60] This protein−ligand-molecular docking approach identified the protein's binding sites by

https://doi.org/10.1021/acsomega.3c05769
ACS Omega 2023, 8, 40695−40712

computing the center and size and customizing the docking box size according to the query ligands. Through 2D ligand interaction, we identified key consensus residues, including VAL305, GLU655, and LYS657, that showed the best affinity bindings with the top-ranked compounds. Finally, we employed fastDRH, a web server and computational tool for docking, to predict the binding affinity between the protein receptor and a small molecule ligand based on their 3D structure. This tool incorporated features derived from both the protein and the ligand, such as physicochemical properties, geometric descriptors, and intermolecular contacts, to predict the binding affinity between them. fastDRH utilized MM/GBSA analysis to estimate the free energy of binding between molecules.[61] This approach combined molecular mechanics calculations to describe the internal energy of the system with a solvation model based on the generalized Born theory to account for the solvation energy and a surface area term to consider the entropy of solvent molecules. The analysis was a valuable tool for predicting and understanding the thermodynamics of molecular interactions, which is essential in drug discovery and protein engineering. By employing this computational tool, we gained further insight into the binding interactions between the top-ranked flavonoids and the MAPT protein, providing a basis for the development of potential drugs for the treatment of neurodegenerative diseases.

Dietary polyphenols have health-promoting properties, such as antioxidative, antiaging, and anticancer properties.[68] Their potential to combat AD, a common neuropathological disease, is being debated. Polyphenols, such as curcumin and its analogous, resveratrol, and other plant compounds, have strong antioxidant properties and can modulate the Aβ biogenesis pathway in AD.[70] Therefore, in this study, different methodologies of docking experimentation, especially MM/PB(GB)-SA-based findings, described hesperetin as the best-docked compound with a protein target, which, in general, indicates the possible role of this flavonoid in the management of AD progression. Interestingly, the dietary flavonoid "Hesperetin" (an aglycone of the flavanone glycoside hesperidin) mainly present in oranges carries its antioxidant properties through direct radical scavenging and enhancement of antioxidant defense of the cell.[71] Additionally, the antioxidant, anti-inflammatory, and neuroprotective properties of hesperetin have received considerable attention.[71] Researchers have reported that this flavonoid compound has the ability to cross the blood−brain barrier, which has added value to our findings as a natural treatment for various disorders of the central nervous system.[71,72] Various citrus flavanones, including hesperetin, have been well documented to have neuroprotective effects against H_2O_2-induced cytotoxicity as they protect PC-12 cells, likely due to their radical scavenging property.[71,72] In 2018, the most significant protective mechanism of hesperetin reported by Kheradmand and co-workers is related to its ability to decrease hippocampal lipid peroxidation and increase levels of glutathione, as well as other antioxidant enzyme activities.[71] They reported that the medicinal effects of this flavonoid significantly reduced oxidative stress and increased antioxidant enzyme activity in a rat model of AD. In the future, this outcome and our present findings may serve as a basis to further explore the role of hesperetin in the management of AD.[71,72]

Comparably, we performed multiple docking experiments using different software packages, parameters, and validation methods to probe the in silico molecular interactions of selected flavonoids with protein targets. In this case, software-specific force field, scoring function, processing flexibility of ligand−protein, solvent and/or ion effects, conformational sampling, ligand parametrization, and complexity of the system for analytical data interpretation, are the prominent variables of the applied molecular docking methodologies that produce different orders of reactivity of the studied flavonoids with MAPT protein.[73] Moreover, the differences in experimental conditions, such as temperature, pH, and buffer composition, can affect the accuracy of predictions and contribute to variations.[73] However, the findings from applied molecular docking methodologies showed a lowering of the free energy when the binding of drug-like screened flavonoids was investigated with MAPT. In the future, it would be valuable to validate the rare-coding mutations identified through our in silico analysis of the MAPT gene through local population sequencing. Furthermore, the efficacy of the identified flavonoids for the treatment of AD can be clinically evaluated in this population. This localized approach will enhance our understanding of the genetic factors contributing to AD within the population and help establish the relevance and potential clinical implications of these rare-coding mutations.

5. CONCLUSIONS

Bioinformatics tools can provide valuable insights into the genetic basis of diseases like Alzheimer's and aid in the development of personalized therapies. However, it is important to validate these predictions through experimental methods and not rely solely on bioinformatics for clinical decision-making. By combining evolutionary conservation analyses and protein structure predictions, rare-coding deleterious SNPs can be prioritized for further experimental validation, which can aid in the identification of potential therapeutic targets. While more research is needed to fully understand the potential of flavonoids in managing neurodegeneration and AD, the current evidence suggests that these compounds may have promising neuroprotective and cognitive-enhancing effects. The functional annotation of genomes improves the understanding of genetic mechanisms leading to disease risk, symptoms, or prognosis, and is important for the development of personalized medicine. Continued research in this field can ultimately lead to improved treatments and better outcomes for patients with neurodegenerative diseases. Furthermore, we identified the best interaction pattern of MAPT with hesperetin, which may serve as new targets and therapeutic interventions for the management of AD.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.3c05769.

Rare-coding exonic SNPs of *APP*, *MAPT*, and *NEFL* genes; assessment of drug-likeness properties of bioactive flavonoid compounds; and virtual screening of bioactive flavonoid compounds with MAPT protein through the iGEMDOCK tool (XLSX)

■ AUTHOR INFORMATION

Corresponding Author

Muhammad Bilal Azmi − *Department of Biochemistry, Dow Medical College, Dow University of Health Sciences, Karachi*

https://doi.org/10.1021/acsomega.3c05769
ACS Omega 2023, 8, 40695−40712

74400, Pakistan; orcid.org/0000-0001-8320-4479; Email: bilal.azmi@duhs.edu.pk

Authors

Affan Ahmed − *Dow Medical College, Dow University of Health Sciences, Karachi 74400, Pakistan*

Tehniat Faraz Ahmed − *Department of Biochemistry, Dow International Dental College, Dow University of Health Sciences, Karachi 75460, Pakistan*

Fauzia Imtiaz − *Department of Biochemistry, Dow Medical College, Dow University of Health Sciences, Karachi 74400, Pakistan*

Uzma Asif − *Department of Biochemistry, Medicine Program, Batterjee Medical College, Jeddah 21442, Saudi Arabia*

Uzma Zaman − *Department of Biochemistry, Dow International Medical College, Dow University of Health Sciences, Karachi 74200, Pakistan*

Khalid Ali Khan − *Unit of Bee Research and Honey Production, Research Center for Advanced Materials Science (RCAMS) and Applied College, King Khalid University, Abha 61413, Saudi Arabia*

Asif Khan Sherwani − *Research and Development Unit, Jamjoom Pharmaceuticals Co. Ltd, Jeddah 21442, Saudi Arabia*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.3c05769

Author Contributions

○A.A., T.F.A., and F.I. contributed equally to this work.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors extend their appreciation to the Deanship of Scientific Research at King Khalid University, Saudi Arabia, for supporting this work through a large group project under grant no. RGP2/447/44. The authors also acknowledge the support of the Research Center for Advanced Materials Science (RCAMS) at King Khalid University in Abha, Saudi Arabia.

■ REFERENCES

(1) Błaszczyk, J. W. Pathogenesis of Dementia. *Int. J. Mol. Sci.* **2022**, *24* (1), 543.

(2) Kodam, P.; Sai Swaroop, R.; Pradhan, S. S.; Sivaramakrishnan, V.; Vadrevu, R. Integrated multi-omics analysis of Alzheimer's disease shows molecular signatures associated with disease progression and potential therapeutic targets. *Sci. Rep.* **2023**, *13* (1), 3695.

(3) 2021 Alzheimer's disease facts and figures. *Alzheimer's Dementia* **2021**, *17* (3) 327−406 doi: .

(4) Mao, Y. M.; Wang, P.; Wang, X. Y.; Ye, D. Q. Global Public Interest and Seasonal Variations in Alzheimer's Disease: Evidence From Google Trends. *Front. Med.* **2021**, *8*, 778930.

(5) Vipin, A.; Satish, V.; Saffari, S. E.; Koh, W.; Lim, L.; Silva, E.; Nyu, M. M.; Choong, T. M.; Chua, E.; Lim, L.; et al. Dementia in Southeast Asia: influence of onset-type, education, and cerebrovascular disease. *Alzheimer's Res. Ther.* **2021**, *13* (1), 195.

(6) World Life Expectancy. Alzheimer's and Dementia, Pakistan. https://www.worldlifeexpectancy.com/pakistan-alzheimers-dementia (accessed April 15, 2023). (b) Adamson, M. M.; Shakil, S.; Sultana, T.; Hasan, M. A.; Mubarak, F.; Enam, S. A.; Parvaz, M. A.; Razi, A. Brain Injury and Dementia in Pakistan: Current Perspectives. *Front. Neurol.* **2020**, *11*, 299.

(7) Knopman, D. S.; Amieva, H.; Petersen, R. C.; Chételat, G.; Holtzman, D. M.; Hyman, B. T.; Nixon, R. A.; Jones, D. T. Alzheimer disease. *Nat. Rev. Dis. Prim.* **2021**, *7* (1), 33.

(8) Tiwari, S.; Atluri, V.; Kaushik, A.; Yndart, A.; Nair, M. Alzheimer's disease: pathogenesis, diagnostics, and therapeutics . *Int. J. Nanomed.* **2019**, *14*, 5541−5554.

(9) Moore, K. B. E.; Hung, T. J.; Fortin, J. S. Hyperphosphorylated tau (p-tau) and drug discovery in the context of Alzheimer's disease and related tauopathies. *Drug discovery today* **2023**, *28* (3), 103487.

(10) Moloney, C. M.; Lowe, V. J.; Murray, M. E. Visualization of neurofibrillary tangle maturity in Alzheimer's disease: A clinicopathologic perspective for biomarker research. *Alzheimer's Dementia* **2021**, *17* (9), 1554−1574.

(11) Sengupta, U.; Kayed, R. Amyloid β, Tau, and α-Synuclein aggregates in the pathogenesis, prognosis, and therapeutics for neurodegenerative diseases. *Prog. Neurobiol.* **2022**, *214*, 102270.

(12) Ahmed, T. F.; Ahmed, A.; Imtiaz, F. History in perspective: How Alzheimer's Disease came to be where it is? *Brain Res.* **2021**, *1758*, 147342.

(13) Klyucherev, T. O.; Olszewski, P.; Shalimova, A. A.; Chubarev, V. N.; Tarasov, V. V.; Attwood, M. M.; Syvänen, S.; Schiöth, H. B. Advances in the development of new biomarkers for Alzheimer's disease. *Transl. Neurodegener.* **2022**, *11* (1), 25.

(14) Cummings, J. The National Institute on Aging-Alzheimer's Association Framework on Alzheimer's disease: Application to clinical trials. *Alzheimer's Dementia* **2019**, *15* (1), 172−178.

(15) Jack, C. R.; Therneau, T. M.; Weigand, S. D.; Wiste, H. J.; Knopman, D. S.; Vemuri, P.; Lowe, V. J.; Mielke, M. M.; Roberts, R. O.; Machulda, M. M.; et al. Prevalence of Biologically vs Clinically Defined Alzheimer Spectrum Entities Using the National Institute on Aging-Alzheimer's Association Research Framework. *JAMA Neurol.* **2019**, *76* (10), 1174−1183.

(16) Faraz Ahmed, T.; Bilal Azmi, M.; Imtiaz, F.; Zaman, U.; Ahmed, A.; Shahbaz, N. Plasma levels of phosphorylated tau and neurofilament light chain as potential biomarkers for Alzheimer's disease: A biochemical analysis in Pakistani population. *Saudi Pharmaceut. J.* **2023**, *31* (7), 1202−1209.

(17) Ayaz, M.; Sadiq, A.; Junaid, M.; Ullah, F.; Ovais, M.; Ullah, I.; Ahmed, J.; Shahid, M. Flavonoids as Prospective Neuroprotectants and Their Therapeutic Propensity in Aging Associated Neurological Disorders. *Front. Aging Neurosci.* **2019**, *11*, 155.

(18) Sayers, E. W.; Beck, J.; Bolton, E. E.; Bourexis, D.; Brister, J. R.; Canese, K.; Comeau, D. C.; Funk, K.; Kim, S.; Klimke, W.; et al. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **2021**, *49* (D1), D10−d17.

(19) Bruford, E. A.; Braschi, B.; Denny, P.; Jones, T. E. M.; Seal, R. L.; Tweedie, S. Guidelines for human gene nomenclature. *Nat. Genet.* **2020**, *52* (8), 754−758.

(20) Howe, K. L.; Achuthan, P.; Allen, J.; Allen, J.; Alvarez-Jarreta, J.; Amode, M. R.; Armean, I. M.; Azov, A. G.; Bennett, R.; Bhai, J.; et al. Ensembl 2021. *Nucleic Acids Res.* **2021**, *49* (D1), D884−d891.

(21) The UniProt Consortium. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* **2019**, *47* (D1), D506−D515.

(22) Safran, M.; Rosen, N.; Twik, M.; BarShir, R.; Stein, T. I.; Dahary, D.; Fishilevich, S.; Lancet, D. The GeneCards Suite. *Practical Guide to Life Science Databases*; Springer, 2021; pp 27−56.

(23) Byrska-Bishop, M.; Evani, U. S.; Zhao, X.; Basile, A. O.; Abel, H. J.; Regier, A. A.; Corvelo, A.; Clarke, W. E.; Musunuri, R.; Nagulapalli, K.; et al. High-coverage whole-genome sequencing of the expanded 1000 Genomes Project cohort including 602 trios. *Cell* **2022**, *185* (18), 3426−3440.e19.

(24) Koch, L. Exploring human genomic diversity with gnomAD. *Nat. Rev. Genet.* **2020**, *21* (8), 448.

(25) Sim, N. L.; Kumar, P.; Hu, J.; Henikoff, S.; Schneider, G.; Ng, P. C. SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res.* **2012**, *40*, W452−W457.

(26) Adzhubei, I.; Jordan, D. M.; Sunyaev, S. R. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr. Protoc. Hum. Genet.* **2013**, *76*, 7.20.

(27) Rentzsch, P.; Witten, D.; Cooper, G. M.; Shendure, J.; Kircher, M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res.* **2019**, *47* (D1), D886−d894.

(28) Ioannidis, N. M.; Rothstein, J. H.; Pejaver, V.; Middha, S.; McDonnell, S. K.; Baheti, S.; Musolf, A.; Li, Q.; Holzinger, E.; Karyadi, D.; et al. REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. *Am. J. Hum. Genet.* **2016**, *99* (4), 877−885.
(29) Hassan, M. S.; Shaalan, A. A.; Dessouky, M. I.; Abdelnaiem, A. E.; ElHefnawi, M. A review study: Computational techniques for expecting the impact of non-synonymous single nucleotide variants in human diseases. *Gene* **2019**, *680*, 20−33.
(30) Frousios, K.; Iliopoulos, C. S.; Schlitt, T.; Simpson, M. A. Predicting the functional consequences of non-synonymous DNA sequence variants−evaluation of bioinformatics tools and development of a consensus strategy. *Genomics* **2013**, *102* (4), 223−228.
(31) Capriotti, E.; Fariselli, P. PhD-SNPg: a webserver and lightweight tool for scoring single nucleotide variants. *Nucleic Acids Res.* **2017**, *45* (W1), W247−W252.
(32) Tang, H.; Thomas, P. D. PANTHER-PSEP: predicting disease-causing genetic variants using position-specific evolutionary preservation. *Bioinformatics* **2016**, *32* (14), 2230−2232.
(33) Pejaver, V.; Urresti, J.; Lugo-Martinez, J.; Pagel, K. A.; Lin, G. N.; Nam, H. J.; Mort, M.; Cooper, D. N.; Sebat, J.; Iakoucheva, L. M.; et al. Inferring the molecular and phenotypic impact of amino acid variants with MutPred2. *Nat. Commun.* **2020**, *11* (1), 5918.
(34) Johnson, A. D.; Handsaker, R. E.; Pulit, S. L.; Nizzari, M. M.; O'Donnell, C. J.; de Bakker, P. I. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics* **2008**, *24* (24), 2938−2939.
(35) Capriotti, E.; Altman, R. B.; Bromberg, Y. Collective judgment predicts disease-associated single nucleotide variants. *BMC Genom.* **2013**, *14*, S2.
(36) Hecht, M.; Bromberg, Y.; Rost, B. Better prediction of functional effects for sequence variants. *BMC Genom.* **2015**, *16*, S1.
(37) López-Ferrando, V.; Gazzo, A.; de la Cruz, X.; Orozco, M.; Gelpí, J. L. PMut: a web-based tool for the annotation of pathological variants on proteins, 2017 update. *Nucleic Acids Res.* **2017**, *45* (W1), W222−W228.
(38) Calabrese, R.; Capriotti, E.; Fariselli, P.; Martelli, P. L.; Casadio, R. Functional annotations improve the predictive score of human disease-related mutations in proteins. *Hum. Mutat.* **2009**, *30* (8), 1237−1244.
(39) Bendl, J.; Stourac, J.; Salanda, O.; Pavelka, A.; Wieben, E. D.; Zendulka, J.; Brezovsky, J.; Damborsky, J. PredictSNP: robust and accurate consensus classifier for prediction of disease-related mutations. *PLoS Comput. Biol.* **2014**, *10* (1), No. e1003440.
(40) Chen, C. W.; Lin, M. H.; Liao, C. C.; Chang, H. P.; Chu, Y. W. iStable 2.0: Predicting protein thermal stability changes by integrating various characteristic modules. *Comput. Struct. Biotechnol. J.* **2020**, *18*, 622−630.
(41) Khan, S.; Vihinen, M. Performance of protein stability predictors. *Hum. Mutat.* **2010**, *31* (6), 675−684.
(42) Capriotti, E.; Calabrese, R.; Casadio, R. Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. *Bioinformatics* **2006**, *22* (22), 2729−2734.
(43) Savojardo, C.; Fariselli, P.; Martelli, P. L.; Casadio, R. INPS-MD: a web server to predict stability of protein variants from sequence and structure. *Bioinformatics* **2016**, *32* (16), 2542−2544.
(44) Ashkenazy, H.; Abadi, S.; Martz, E.; Chay, O.; Mayrose, I.; Pupko, T.; Ben-Tal, N. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* **2016**, *44* (W1), W344−W350.
(45) Angamuthu, K.; Piramanayagam, S. Evaluation of in silico protein secondary structure prediction methods by employing statistical techniques. *Biomed. Biotechnol. Res. J.* **2017**, *1* (1), 29−36.
(46) Klausen, M. S.; Jespersen, M. C.; Nielsen, H.; Jensen, K. K.; Jurtz, V. I.; Sønderby, C. K.; Sommer, M. O. A.; Winther, O.; Nielsen, M.; Petersen, B.; et al. NetSurfP-2.0: Improved prediction of protein structural features by integrated deep learning. *Proteins* **2019**, *87* (6), 520−527.
(47) Parthiban, V.; Gromiha, M. M.; Schomburg, D. CUPSAT: prediction of protein stability upon point mutations. *Nucleic Acids Res.* **2006**, *34*, W239−W242.
(48) Kelley, L. A.; Mezulis, S.; Yates, C. M.; Wass, M. N.; Sternberg, M. J. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **2015**, *10* (6), 845−858.
(49) Worth, C. L.; Preissner, R.; Blundell, T. L. SDM−a server for predicting effects of mutations on protein stability and malfunction. *Nucleic Acids Res.* **2011**, *39*, W215−W222.
(50) Rodrigues, C. H.; Pires, D. E.; Ascher, D. B. DynaMut: predicting the impact of mutations on protein conformation, flexibility and stability. *Nucleic Acids Res.* **2018**, *46* (W1), W350−w355.
(51) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596* (7873), 583−589.
(52) Oladejo, D.; Oduselu, G.; Dokunmu, T.; Isewon, I.; Okafor, E.; Iweala, E. E.; Adebiyi, E. F. In silico evaluation of inhibitors of Plasmodium falciparumAP2-I transcription factor. *FASEB J.* **2022**, *36*, L7455.
(53) Uddin, M. S.; Kabir, M. T.; Niaz, K.; Jeandet, P.; Clément, C.; Mathew, B.; Rauf, A.; Rengasamy, K. R. R.; Sobarzo-Sánchez, E.; Ashraf, G. M.; et al. Molecular Insight into the Therapeutic Promise of Flavonoids against Alzheimer's Disease. *Molecules* **2020**, *25* (6), 1267.
(54) Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B. A.; Thiessen, P. A.; Yu, B.; et al. PubChem 2019 update: improved access to chemical data. *Nucleic Acids Res.* **2019**, *47* (D1), D1102−d1109.
(55) Yoshikawa, N.; Hutchison, G. R. Fast, efficient fragment-based coordinate generation for Open Babel. *J. Cheminf.* **2019**, *11* (1), 49.
(56) Benet, L. Z.; Hosey, C. M.; Ursu, O.; Oprea, T. I. BDDCS, the Rule of 5 and drugability. *Adv. Drug Deliv. Rev.* **2016**, *101*, 89−98.
(57) Ursu, O.; Rayan, A.; Goldblum, A.; Oprea, T. I. Understanding drug-likeness. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2011**, *1* (5), 760−781.
(58) Dalafave, D. Design of druglike small molecules for possible inhibition of antiapoptotic BCL-2, BCL-W, and BFL-1 proteins. *Biomed. Eng. Comput. Biol.* **2010**, *2*, BECB.S5575.
(59) Hsu, K. C.; Chen, Y. F.; Lin, S. R.; Yang, J. M. iGEMDOCK: a graphical environment of enhancing GEMDOCK using pharmacological interactions and post-screening analysis. *BMC Bioinf.* **2011**, *12*, S33.
(60) Liu, Y.; Yang, X.; Gan, J.; Chen, S.; Xiao, Z. X.; Cao, Y. CB-Dock2: improved protein-ligand blind docking by integrating cavity detection, docking and homologous template fitting. *Nucleic Acids Res.* **2022**, *50* (W1), W159−W164.
(61) Wang, Z.; Pan, H.; Sun, H.; Kang, Y.; Liu, H.; Cao, D.; Hou, T. fastDRH: a webserver to predict and analyze protein-ligand complexes based on molecular docking and MM/PB(GB)SA computation. *Briefings Bioinf.* **2022**, *23* (5), bbac201.
(62) Dumitrescu, L.; Mahoney, E. R.; Mukherjee, S.; Lee, M. L.; Bush, W. S.; Engelman, C. D.; Lu, Q.; Fardo, D. W.; Trittschuh, E. H.; Mez, J.; et al. Genetic variants and functional pathways associated with resilience to Alzheimer's disease. *Brain* **2020**, *143* (8), 2561−2575.
(63) Andrews, S. J.; Fulton-Howard, B.; Goate, A. Interpretation of risk loci from genome-wide association studies of Alzheimer's disease. *Lancet Neurol.* **2020**, *19* (4), 326−335.
(64) Branco, I.; Choupina, A. Bioinformatics: new tools and applications in life science and personalized medicine. *Appl. Microbiol. Biotechnol.* **2021**, *105* (3), 937−951.
(65) Sánchez-Juan, P.; Moreno, S.; de Rojas, I.; Hernández, I.; Valero, S.; Alegret, M.; Montrreal, L.; García González, P.; Lage, C.; López-García, S.; et al. The MAPT H1 Haplotype Is a Risk Factor for Alzheimer's Disease in APOE ε4 Non-carriers. *Front. Aging Neurosci.* **2019**, *11*, 327.
(66) Parthiban, V.; Gromiha, M. M.; Hoppe, C.; Schomburg, D. Structural analysis and prediction of protein mutant stability using

https://doi.org/10.1021/acsomega.3c05769
ACS Omega 2023, 8, 40695−40712

distance and torsion potentials: role of secondary structure and solvent accessibility. *Proteins* **2006**, *66* (1), 41−52.
(67) Cassidy, L.; Fernandez, F.; Johnson, J. B.; Naiker, M.; Owoola, A. G.; Broszczak, D. A. Oxidative stress in alzheimer's disease: A review on emergent natural polyphenolic therapeutics. *Compl. Ther. Med.* **2020**, *49*, 102294.
(68) Brglez Mojzer, E.; Knez Hrnčič, M.; Škerget, M.; Knez, Ž.; Bren, U. Polyphenols: Extraction Methods, Antioxidative Action, Bioavailability and Anticarcinogenic Effects. *Molecules* **2016**, *21*, 901.
(69) Calderaro, A.; Patanè, G. T.; Tellone, E.; Barreca, D.; Ficarra, S.; Misiti, F.; Laganà, G. The Neuroprotective Potentiality of Flavonoids on Alzheimer's Disease. *Int. J. Mol. Sci.* **2022**, *23*, 14835.
(70) Bukhari, S. N. A. Dietary Polyphenols as Therapeutic Intervention for Alzheimer's Disease: A Mechanistic Insight. *Antioxidants* **2022**, *11*, 554.
(71) Kheradmand, E.; Hajizadeh Moghaddam, A.; Zare, M. Neuroprotective effect of hesperetin and nano-hesperetin on recognition memory impairment and the elevated oxygen stress in rat model of Alzheimer's disease. *Biomed. Pharmacother.* **2018**, *97*, 1096−1101.
(72) Babylon, L.; Grewal, R.; Stahr, P. L.; Eckert, R. W.; Keck, C. M.; Eckert, G. P. Hesperetin Nanocrystals Improve Mitochondrial Function in a Cell Model of Early Alzheimer Disease. *Antioxidants* **2021**, *10*, 1003.
(73) Stanzione, F.; Giangreco, I.; Cole, J. C. Use of molecular docking computational tools in drug discovery. *Prog. Med. Chem.* **2021**, *60*, 273−343.

https://doi.org/10.1021/acsomega.3c05769
ACS Omega 2023, 8, 40695−40712