
Revisiting BACE-1: How Machine Learning and Molecular Dynamics Unveiled Potential Anti-Alzheimer's Activity of a Cysteinylyl Sulfoxide Derivative

[Shadrach C. Eze](#)*, [Stephen C. Nnemolisa](#), Joy C. Onyesoro, [Toluwalope D. Ilori](#), Victor S. Uche, [Augustine C. Madueke](#), [Wande M. Oluyemi](#), [Adeniyi T. Adewumi](#), [Salerwe Mosebi](#), [Innocent U. Okagu](#)*

Posted Date: 23 March 2026

doi: 10.20944/preprints202603.1753.v1

Keywords: Alzheimer; BACE1; machine learning; molecular dynamics; flap dynamics; cysteinylyl sulfoxide



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Revisiting BACE-1: How Machine Learning and Molecular Dynamics Unveiled Potential Anti-Alzheimer's Activity of a Cysteinyll Sulfoxide Derivative

Shadrach C. Eze ^{1,*}, Stephen C. Nnemolisa ², Joy C. Onyesoro ³, Toluwalope D. Ilori ⁴, Victor S. Uche ², Augustine C. Madueke ², Wande M. Oluyemi ¹, Adeniyi T. Adewumi ⁵, Salerwe Mosebi ⁵ and Innocent U. Okagu ^{2,*}

¹ Department of Pharmaceutical and Medicinal Chemistry, College of Pharmacy, Afe Babalola University, Ado-Ekiti, Nigeria

² Drug Discovery Unit, Department of Biochemistry, University of Nigeria, 410001, Nsukka, Nigeria

³ Department of Pharmacy, Federal Medical Centre Makurdi, Benue State Nigeria

⁴ Federal University of Agriculture, Abeokuta Nigeria

⁵ Department of Life and Consumer Sciences, Calabash Building, Room 02-026, University of South Africa, Johannesburg, South Africa

* Correspondence: shadrachchinecheremeze@gmail.com (S.C.E.); innocent.okagu@unn.edu.ng (I.U.O.); Tel: +2349160253290 (S.C.E.); +2348034627630 (I.U.O.)

Abstract

Beta Secretase (BACE-1) is a well-validated target for Alzheimer's therapies, but have seen attrition in drug development. Herein, we leveraged machine learning (ML), virtual screening and molecular dynamics (MD) to identify novel compounds with potential activity against BACE1. We developed ML algorithms to distinguish active and inactive compounds from public databases. Molecular docking and MD (250ns) was used to explore the inhibition mechanism, thermodynamic stability, and the flap dynamics of the BACE1-ligand complexes. Random Forest Classifier (RF) showed excellent metrics (accuracy: 0.9807; F1 score: 0.9804; specificity 0.9977), compared to other models. Molecular docking with predicted actives revealed compounds BA1, BA2, and BA3 with strong affinity for the BACE1. Compound BA2, a cysteinyll sulfoxide derivative, showed good stability (RMSD) during simulations (1.307 ± 0.109 Å) compared to Verubecestat (1.602 ± 0.159 Å). MMGBSA-based binding free energy (ΔG_{bind} ; kcal/mol) showed that BA2 (-33.820 ± 4.254) had comparatively lower energy than Verubecestat (-21.090 ± 6.183). BA2 maintained electrostatic interactions with the catalytic dyad (Asp36 & Asp232) and Thr76 of the flap. BA2 also maintained the flaps in a semi-open conformation ($d_0: 11.807 \pm 0.401$ Å) throughout the simulation. Our study clearly demonstrate the utility of ML in prioritization of compounds before molecular docking and MD in early phases of drug discovery.

Keywords: Alzheimer; BACE1; machine learning; molecular dynamics; flap dynamics; cysteinyll sulfoxide

1. Introduction

The cases of senile dementia are becoming a major worldwide concern as the aging population increases. Alzheimer's disease (AD), a neurodegenerative disease, is one of the most frequently diagnosed cases of senile dementia in the world, with high prevalence in elderly people (>65 years) [1]. It is estimated that over 55 million people worldwide suffer from AD [2]. By 2055, there will be one new AD case every 33 seconds, reaching approximately 152 million cases worldwide [3]. AD is

characterized by the accumulation of extracellular neuritic plaques, which are primarily composed of amyloid- β peptide ($A\beta$) and intracellular neurofibrillary tangles comprising hyperphosphorylated tau [4]. Sequential cleavage of amyloid precursor protein (APP) by β -site APP-cleaving enzyme (BACE1) and γ -secretase in the amyloidogenic pathway produces amyloid- β peptide ($A\beta$). BACE1 is an aspartyl protease of the pepsin family that plays an essential role in the production of all monomeric forms of amyloid- β ($A\beta$), including $A\beta_{42}$, which aggregates into bioactive conformational species and initiates toxicity in Alzheimer's disease [5]. Considering that BACE1 catalyzes the initial cleavage in $A\beta$ generation, and is putatively rate-limiting, this makes it a primary drug target for AD treatment [4]. Intensive research has been conducted over the last few decades to develop potent, selective, safe, orally bioavailable, and brain-penetrant BACE1 inhibitors. However, they have faced significant challenges. Poor bioavailability, BBB permeability, retinal damage, skin rashes, weight loss, and worsening decline in cognitive function are some of the challenges associated with BACE1 inhibitors. Despite these challenges, which led to the termination of a few clinical studies [4]. BACE1 continues to be a well-validated therapeutic target for AD, thus necessitating a shift in approach to overcome these challenges.

BACE1 protein, as in other aspartate proteases, is made of a catalytic dyad of two ASP residues with a flap that encloses the active site to trap the substrate/inhibitor upon binding (**Figure 1A**). The overall conformational flexibility of enzymes is important [6,7]. The opening and closing dynamics of this flap are important in ligand binding and enzymatic action/inhibition [8]. The flaps undergo conformational changes that control the access of incoming substrates or ligands. Various conformations, such as closed and open states, have been observed in the crystal structures of different proteases [9]. The distance between the Ca atoms of the tip residue of the flap and the opposite ASP residue in the catalytic dyad has been used to characterize this flexibility [8,10]. Since the distance alone is limited in the information about the twisting and curling motions of the flaps, additional parameters such as the angle and torsion between the flaps, catalytic dyad, and the opposite loop have been added in the toolkit [6,9]. (**Figure 1B**).

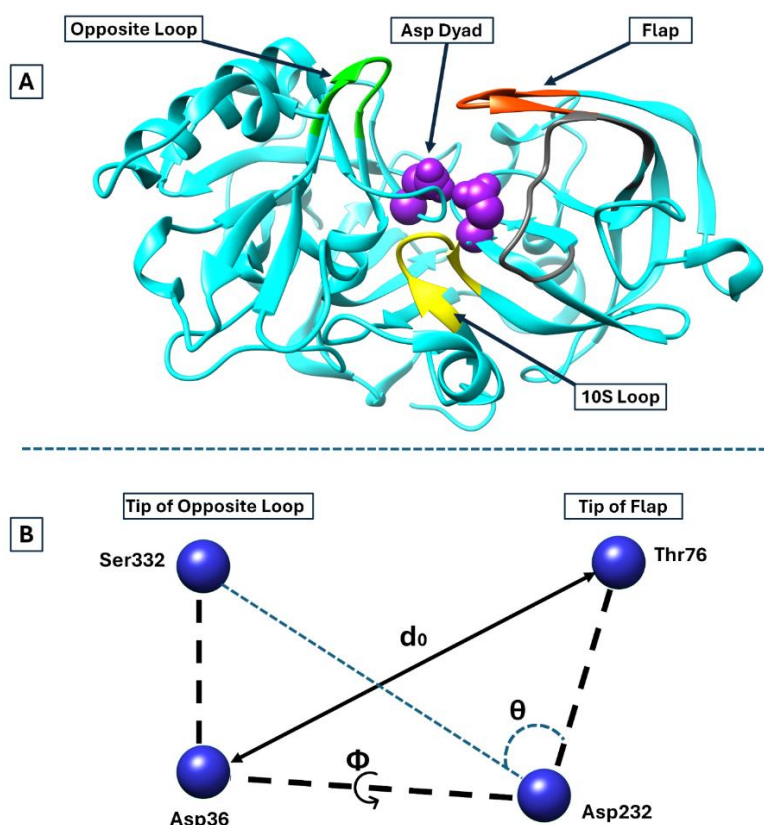


Figure 1. (A) 3D crystal structures of the human brain memapsin 2 (beta-secretase) (BACE1; PDB code: 1XS7) showing the opposite loop, flap, 10S loop and the catalytic dyad; (B) Parameters used to measure the conformational flexibility of the flap over the catalytic dyad. D0: distance between C α of Thr76 and Asp36; θ : angle formed by the C α of Thr76-Asp232-Ser332; Φ : torsion angle formed by C α of Ser332-Asp36-Asp232-Thr76.

Computational biology has revolutionized drug discovery and development, with Machine learning (ML) emerging as a powerful tool in drug discovery [11]. ML, a subset of artificial intelligence, uses algorithms and statistical models to enable computers to carry out tasks without explicit instructions, learn from data patterns, and make decisions. In drug discovery, ML can be applied to various stages, from predicting the biological activity of compounds against a specific target and optimizing the properties of lead compounds to identifying potential toxicity problems early in the drug development process [11]. ML can navigate the huge chemical space, understand intricate biological interactions, predict therapeutic outcomes, and identify potential inhibitors with high accuracy [12]. A range of ML techniques have been employed to predict protein inhibitor activity with high accuracy [13]. Many of these ML techniques have also seen applications in screening for inhibitors against BACE1 protein. Sangeet [12] utilized DrugGPT to generate novel scaffolds and employed molecular docking and dynamics simulations to study the interactions of these ligands with the BACE1 protein [12]. In other studies, Dhamodharan & Mohan employed 2D-QSAR (2D quantitative structure activity relationships), ANN (artificial neural network), SVM (support vector machine) and GFA (genetic function approximation) for modelling dual inhibitors of BACE1 and AchE (Acetylcholinesterase enzyme) while Tung *et al* used GROMACS in the dynamic simulation of promising ligands from an initial machine learning and molecular docking strategy [14,15]. Generative models such as DrugGPT often suffer from poor diversity of generated molecules from input structures. While deep learning models are plausible, studies have shown that a good selection of descriptors with very simple models can outperform heavier architectural models [16,17]. Also, these previous studies lack a robust assessment of the inhibitor-binding by the parameters defined in **Figure 1**. Does the flap close or open upon inhibitor binding? What are the relative motions of the flaps during the simulation? Does the binding free energies correspond to favorable flap dynamics essential for inhibitor binding and activity?

In this study, we aimed to enhance model performance in predicting BACE1 inhibitor activity by utilizing small yet informative molecular descriptors in conjunction with structural fingerprints. We employed an array of machine learning algorithms: Logistic Regression (LR), Support Vector Machine (SVM), Random Forest (RF), Adaptive Boosting (AB), Gradient Boosting (GB), and Extreme Gradient Boosting (XGB) to train and validate our models. These algorithms were fine-tuned using hyper-parameterization and Y-scrambling test to avoid overfitting. After the initial screening with the best performing model, we employed molecular docking to promising binders against BACE1 protein. Three compounds code-named BA1 to BA3 were subjected to 250ns simulation to assess the mechanism of binding and conformational changes in BACE1 upon ligand binding. We evaluated the RMSD, RMSF, and energetics (via MMGBSA) of these ligand-bound complexes. To assess conformational flexibility of the flaps and the whole protein, we measured the distance (d0) between C α of Thr76 and Asp36, angle (θ) formed by the C α of Thr76-Asp232-Ser332, and torsion (Φ) formed by C α of Ser332-Asp36-Asp232-Thr76 as shown in **Figure 1**. We showed that BA2, a derivative of cysteine sulfoxide, possessed great binding properties (binding affinity, binding free energy) and maintained the BACE1 protein in a semi-closed conformation with good binding to the catalytic dyad. Combining molecular dynamics simulation and ML algorithms has shown the potential to improve the accuracy of virtual screening and optimization of compound selection for experimental validation. Additionally, assessing the conformational flexibility of BACE1, a known aspartate protease, through its flap movements provides critical insights into the mechanism and potential of ligand binding. This study, therefore, opens possibilities for the discovery of novel BACE1 therapeutics.

2. Materials and Methods

Machine Learning based Screening

The general computational workflow used in this study is shown in **Figure 2**.

Dataset Curation

A total of 4290 active inhibitors of BACE1 protein target were extracted from the Binding DB database (<https://www.bindingdb.org>). From the database of the Directory of Useful Decoys-Enhanced (<https://dude.docking.org/>), 3,014 decoy molecules were generated. These compounds were considered inactive. In the active inhibitor dataset, duplicate compounds were identified, and the first instance of each was retained. To denote their activity, the active inhibitor dataset (3044 compounds) was labeled as “active,” and the decoy compounds were labeled as “inactive”. The two datasets were then combined into a single CSV file.

Generation of molecular descriptors and preprocessing of data

To predict the activity of compounds, it is necessary to transform the molecular characteristics of a compound into numerical data that the algorithm can understand [18]. The SMILES notation, a feature in the dataset, was used to specify the molecular characteristics of the compound in our dataset. Using the RDKit library (<https://www.rdkit.org/>), the SMILES notations were quantified, computing 6 important features: molecular weight (MW), number of hydrogen donors (nHDs), number of hydrogen acceptors (nHAs), topological polar surface area (TPSA), lipophilicity (cLogP), and the number of rotatable bonds (nrBs). The chemical structural information of the compounds was further encoded into a binary vector using RDKit’s Morgan Fingerprint [19]. A 1024-dimensional Morgan fingerprint was obtained with a radius of 2. To ensure that an efficient and robust predictive machine learning model is built, the target variable was mapped to ‘1’ and ‘0’, respectively, for active and inactive in the dataset.

Machine learning models

To classify the compounds in the dataset as either active or inactive, different machine learning models were developed. The models employed in this study were Logistic Regression (LR), Support Vector Machine (SVM), Random Forest (RF), Adaptive Boosting (AB), Gradient Boosting (GB), and Extreme Gradient Boosting (XGB) [13]. Logistic regression is a popular machine learning model used to model a categorical dependent variable. It assumes a linear relationship with the natural logarithm of the odds outcome [20]. This model, which has been applied in drug discovery [20], enables researchers to hold constant the values of other independent variables while investigating the relationship between each variable and the binary outcome. SVM, one of the well-established supervised machine learning techniques in virtual screening studies for inhibitor identification, is known for its effectiveness in high-dimensional environments and its ability to resist overfitting [21]. It divides the data into two classes using a hyperplane, thereby maximizing the margin between the classes. Random forest is a robust machine learning technique known for its ability to overcome overfitting, handle high-dimensional data, and achieve excellent accuracy [22,23]. It can be used for both classification and regression tasks. For the classification task, numerous decision trees are used during the training phase, with the output representing the mode of the classes [13]. Adaptive Boosting (AdaBoost) is an ensemble technique that creates a robust model by combining multiple weak learners. In AdaBoost, each data point is assigned a weight, and these weights are adjusted during each iteration to focus on previously misclassified instances [13]. Its ability to adapt to various data distributions and handle complex classification tasks makes AdaBoost a valuable technique for enhancing prediction performance. Unlike AdaBoost, Gradient Boosting enhances a model’s performance by fitting weak learners to the residuals or errors of the previous predictions, using a gradually increasing number of iterations. Extreme Gradient Boosting (XGBoost) was built on top of the gradient boosting technique. The ease of use and parallelization, as well as the ability to handle diverse and complex feature spaces of descriptors, while delivering strong predictive performance, make XGBoost vastly popular in drug discovery and among machine learning practitioners.

Model evaluation

Initially, each model was trained with ten-fold cross-validation without fine-tuning the hyperparameters. To guide the selection of the best-performing model, numerous evaluation techniques were employed. The model's efficacy was evaluated using precision, F1 score, accuracy, sensitivity, MCC, and specificity, and area under the receiver operating characteristic curve (AUC-ROC) [24]. The hyperparameters of the best-performing model were fine-tuned using RandomizedSearchCV. RandomizedSearchCV is faster, more efficient, and allows a broad hyperparameter search with fewer iterations [25]. It ensures that the best possible set of hyperparameters is chosen, ultimately leading to optimal model performance.

Y-scrambling test

To rule out any chance correlation and examine the robustness and validity of the best-performing model, we implemented the Y-scrambling technique [26]. In the Y-scrambling test, the target feature was reshuffled with the descriptors left unchanged, and several trials were performed on the model. The new model was expected to have low accuracy compared to the original model.

Screening of a new dataset

To apply our model in the discovery of novel compounds with potential activity against the BACE1 target, we searched and curated a new library of compounds from public databases: PubChem (1152 compounds) [27], ZINC (2657 compounds) [28], and COCONUT (6000 compounds) [29] with the keywords "phytochemical" or "natural products". These databases are open-source chemical databases that provide comprehensive chemical information on chemical compounds, making them suitable for virtual screening. A total of 9809 compounds were retrieved. The strategies used to prepare the initial dataset were applied to this dataset. Using our model, we screened the preprocessed data, categorizing the prepared novel dataset into active or inactive. Our model predicted 3379 compounds to be active. To enhance the drug-like properties of the active compounds, the Lipinski rule of five was employed, focusing on compounds that are both orally active and can also cross the blood-brain barrier. Finally, the model predicted 1,388 active compounds for molecular docking.

Repository for our Models

All the codes, datasets, and algorithms used in training and validation in this study can be found https://github.com/nnemolisastephen/BACE_1_project.

Molecular docking

Protein Preparation and Grid Generation

The 3-dimensional (3D) structure of BACE1 was retrieved from the RSCB PDB (<http://www.rscb.org/>) with a PDB ID: 1XS7 (**Figure 2A**). The protein PDB structure was refined using the protein preparation wizard of the Schrödinger suite before molecular docking [30]. Water molecules were deleted. Charges, bond orders, hydrogen atoms, and missing residues were assigned to the proteins. The heavy atom's Root-Mean-Square Deviation (RMSD) was kept at 0.30 Å using the optimal Potentials for Liquid Simulations (OPLS4) force field. The receptor glide grid was generated using the co-crystal ligand as a reference. The charge cutoff for polarity was set to 0.25, and the van der Waals radii were scaled by a factor of 1.0. The compounds were then docked to the receptor.

Ligand Preparations

The SMILES of the compounds classified as active were extracted and transformed into 2D structure data file using Data Warrior v 5.5 [31]. The compounds were prepared using the Ligprep tool (Schrödinger, 2017) within the Schrödinger suite with OPLS4 force field [30]. Correct bond ordering and geometry optimization were carried out.

Molecular Docking Analysis

The molecular docking method was carried out using the Glide implemented on the Schrödinger suites [30]. The Standard Precision (SP) algorithm was initially used to screen less promising poses and prioritize the best binding poses in the protein's active site. The prepared compounds were made flexible, and the receptor was treated as a rigid body. The top-scoring poses from the SP scoring protocol were rescored using the more rigorous extra precision (XP) algorithm [30]. The compounds were then ranked based on their binding affinities with the target.

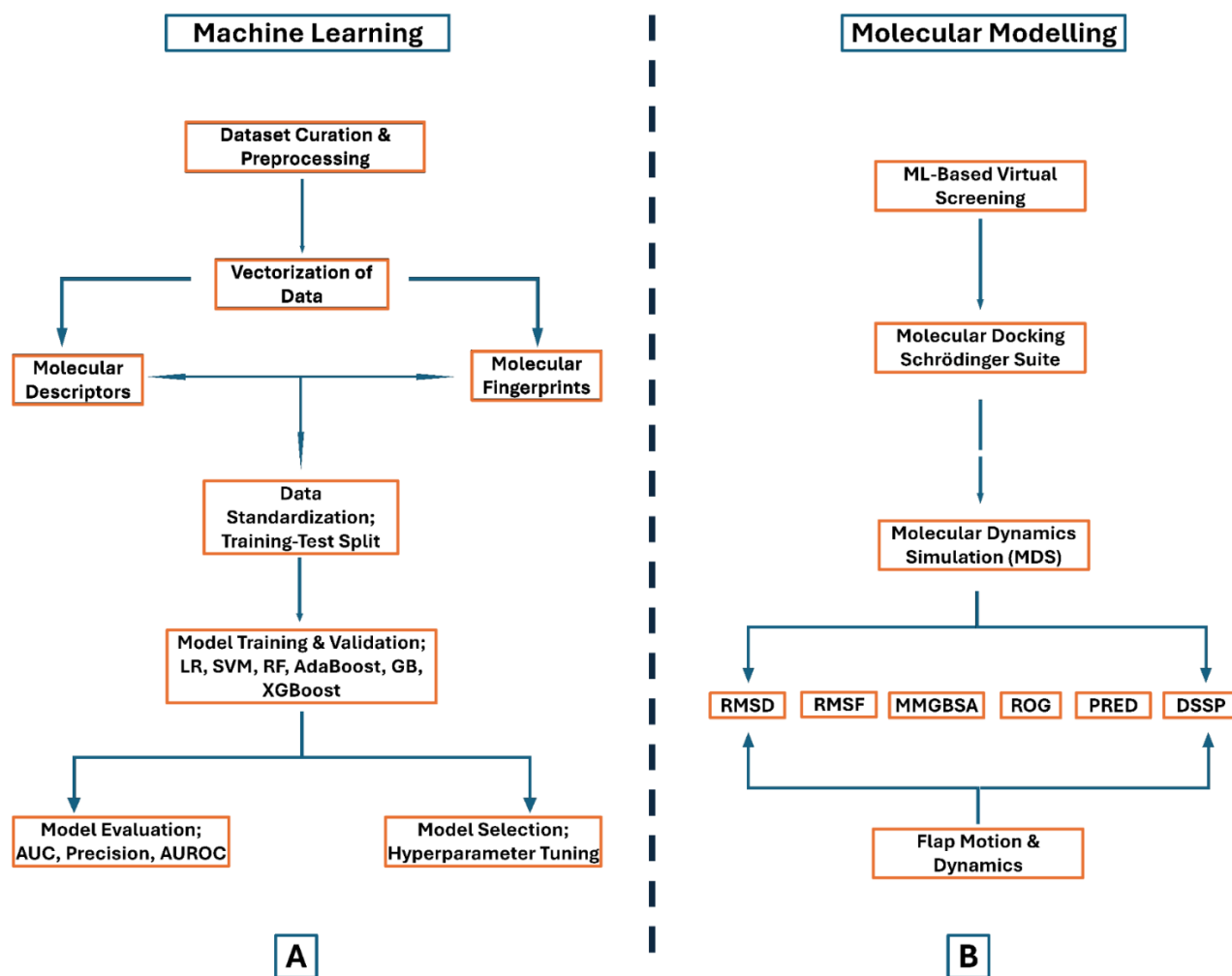


Figure 2. The general Computational workflow used in this study (image generated in Microsoft PowerPoint 2021). (A) Machine learning workflow and (B) computational modeling workflow.

Molecular Dynamics Simulations (MDS)

The UCSF Chimera tool was utilized to prepare the four ligands (**Figure 3**) and the apo protein for the MD simulations [32]. All hydrogen atoms were deleted from the receptor, and then it was exported as rec.pdb. For the ligands, hydrogen atoms and charges were added, and then the resulting structure was exported as lig.mol2.

The MD simulations analyses were performed for the three ligands (and the standard), each complexed with BACE1 protein (**Figure 3**). The graphics processing unit (GPU) version of the PMEMD.CUDA engine equipped with the AMBER Force Field (FF18SB variant) package was used to parameterize the protein [33]. MD simulations helped in examining the stability, compactness, and structural dynamics of the docked complexes [34]. Partial charges were added to the ligands through ANTECHAMBER. The LEAP module of AMBER18 was utilized to introduce Na⁺ counter ions to neutralize and solvate the bound and unbound bio-complexes. In addition, all-atom explicit solvation was performed in an orthorhombic TIP3P box of water molecules of 12 Å size. The initial minimization, consisting of 2500 steps with a restrained potential of 500 kcal/mol, was considered. This is followed by full minimization (5000 steps), which was performed using a conjugate algorithm without applying restraints. Next, the systems were gradually heated from 0 to 300 K for 50 ps to maintain a fixed volume and number of atoms, considering a canonical ensemble (NVT). The shake algorithm was used to reduce the hydrogen bond constraint by applying one-bar pressure, used by the Barendsen-Barostat [35,36]. Production MD was run for 250 ns for all systems, using a time scale of 2 fs and a Langevin thermostat at 300 K and 1 bar constant pressure. The coordinates of the Apo-

receptor and ligand-receptor complexes were saved every 1 ps, and the MD trajectories were analyzed with CJTRAJ and PTRAJ module in AMBER18 [37].

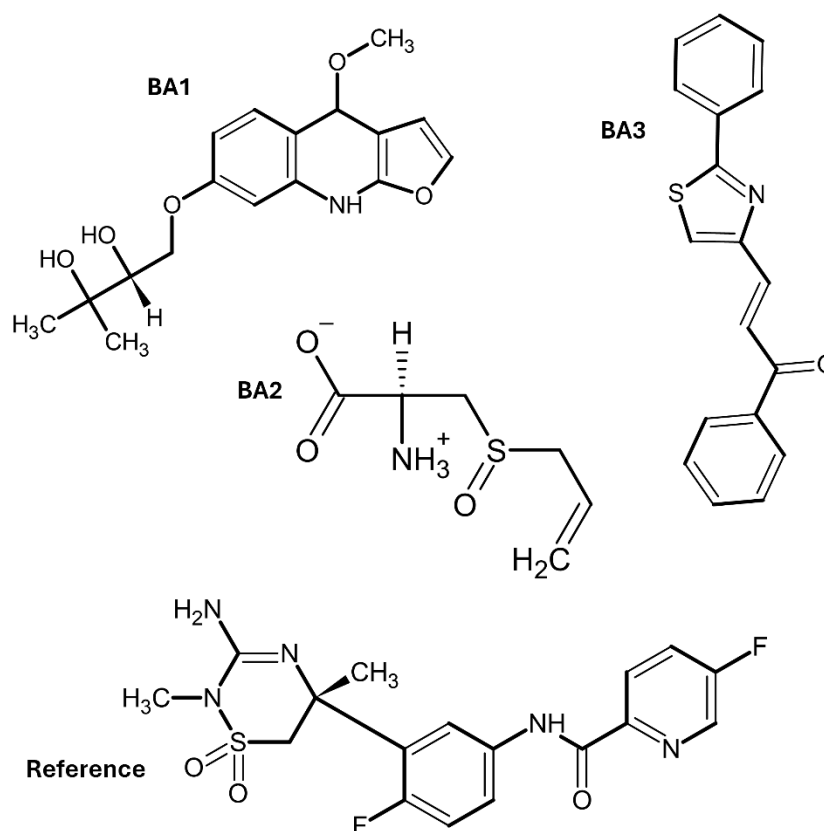


Figure 3. 2D structures of ligands investigated in this study using molecular dynamics simulation. BA1: 1-[(4-methoxy-4,9-dihydrofuro[2,3-b]quinolin-7-yl)oxy]-3-methylbutane-2,3-diol; BA2: S-(2-propenyl)-L-cysteine sulfoxide (Allin); BA3: (2E)-1-phenyl-3-(2-phenyl-1,3-thiazol-4-yl)prop-2-en-1-one; Reference is Verubecestat (MK-8931).

Post MD Analysis

The MD parameters studied include the C- α root mean square of deviation (C- α RMSD), root mean square of fluctuation (C- α RMSF), radius of gyration (RoG), principal component analysis (PCA), and solvent-accessible surface area (SASA) [37]. Also, distances, angles and torsion of the flap dynamics were also measured as described in **Figure 2**. Post-MD structure analysis and visualization of ligand-receptor interactions were done using Chimera and Discovery Studio 2021 client [38]. Plots of RMSD, RMSF, RoG, PRED, and PCA were obtained by Microcal Origin 6.0 software for data analysis [39,40].

Binding Free Energy (BFE) Analysis

Molecular dynamics simulation has proven to be a valuable computational technique for elucidating the mechanisms underlying biological processes. The thermodynamic binding free energy (BFE) provides critical information on the binding interactions of a ligand to a receptor. The Molecular Mechanics Bonn Surface Area (MMGBSA) is a method used to calculate the BFE of a ligand to a protein [41]. The calculated energies utilized over 250000 snapshots from the 250 ns trajectories. In the MMG/PBSA procedure, solvents were completely removed and replaced by a dielectric continuum. The expression below was used to determine the BFE (ΔG) for the ligand-bound receptor systems:

$$\Delta G_{\text{bind}} = \Delta G_{\text{complex}} - \Delta G_{\text{receptor}} - \Delta G_{\text{ligand}}$$

$$\Delta G_{\text{bind}} = E_{\text{gas}} + G_{\text{sol}} - T\Delta S$$

$$E_{\text{gas}} = E_{\text{int}} + E_{\text{vdw}} + E_{\text{ele}}$$

$$G_{sol} = G_{GB} + G_{SA}$$

$$G_{SA} = \gamma SASA$$

ΔG_{bind} = gas-phase summation; E_{gas} = gas-phase energy; G_{sol} = free solvation energy; $T\Delta S$ = entropy; E_{int} = internal energy; E_{ele} = electrostatic energies; E_{vdw} = *van der Waals* energy [35,42].

Receptor–Ligand Interactions Systems

The analysis of the receptor's active site residues forming an interaction network with the ligands was done with Discovery Studio Visualizer 2020 Client [43]. Various intermolecular forces that exist between the ligands and the receptor were determined. Snapshots were also taken at different intervals during the MD run to display various interactions and conformations of the protein's secondary structure. Discovery Studio Visualizer was employed to analyze the interacting residues of the receptor and the chemical bonding types within the bound complexes [44].

Per-Residue Energy Decomposition (PRED) Analysis

The PRED estimates the energy contributions of amino acid residues at the protein's binding site to the overall stability of the selected compounds relative to receptor inhibition. PRED analysis enhances and reveals the energetic contributions of individual interacting residues. Overall, it enhances our understanding of protein structure-function relationships. MMGBSA method in AMBER18 GPU was used to analyze the decomposition of the residual energy contribution to BFE in the protein-ligand complexes [34].

Drug-Likeness and Toxicity Prediction

The drug-likeness properties of the ligands were carried out using the SwissADME online server (<http://www.swissadme.ch/index.php>). Molecular weight (MW), Hydrogen bond acceptor (HBA), Hydrogen bond donor (HBD), Clog P, Gastrointestinal Absorption (GIA), Topological polar surface area (TPSA), BBB (Blood-Brain-Barrier), and Water solubility were the parameters investigated. The toxicity profile of the compounds was predicted using ProTox III (<https://tox.charite.de/protox3/>).

3. Results and Discussion

Machine Learning Model Evaluation and Performance

One of the promising lines of inquiry in drug discovery is the identification of compounds that can modulate protein function in pathways related to pathogenesis. The recurrent failures in targeting the amyloid pathway have laid bare the limitations of traditional AD drug development. In this regard, ML has emerged as a transformative force in our approach to therapeutic discovery and optimization. In this study, we leverage the power of ML in screening a large chemical space for the identification of novel BACE1 inhibitors for AD treatment. A dataset containing 4,290 compounds against BACE1 was retrieved from Binding DB, labeled as active, and 3,014 decoy molecules were labeled as inactive (**Figure 4B**). Following preprocessing, which includes the removal of duplicates and missing values, the active compounds data yielded a total of 3044 compounds. The datasets were combined, consisting of a total of 6058 compounds. The SMILES notation of each compound was transformed into numerical descriptors using the RDKit library [45]. Descriptors were selected based on their importance to drug-likeness and physicochemical properties. The dataset was split into training and testing sets, with 70% (4090) of the data allocated to the training set and 30% (1968) to the testing set, respectively (**Figure 4B**). The use of diverse datasets is crucial for developing a robust predictive model. Using a diverse dataset ensures that our model can perform better on unseen data. To explore the diversity of our dataset, we examined MW and CLogP values of the compounds, given their importance in drug discovery and development. **Figure 4A** shows that our training set had compounds with MW ranging between 230 and 2150 Daltons and the CLogP values ranging between -15 and 15. This broad chemical space within our dataset can enhance the predictive power of our model while ensuring that compounds that do not meet the drug-likeness guidelines are screened out. The chemical space analysis of the test data is presented in Supplementary **Fig. S0**.

Our study utilized various machine learning models, including Logistic Regression, Support Vector Machines (SVM), Random Forest Classifier (RF), Adaptive Boosting Classifier (AdaBoost), Gradient Boosting Classifier (XGBoost), and Extreme Gradient Boosting (XGB), to classify active

compounds from inactive compounds. The performance of each model was evaluated based on accuracy, sensitivity, Matthews Correlation Coefficient (MCC), and Area under the curve (**Figures 4C & 4D; Table 1**). Out of all the models used, Adaptive Boosting performed fairly-poorly with accuracy (0.9394), F1 score (0.9397), specificity (0.9366), sensitivity (0.9428), and MCC (0.879). SVM, RF, and XGB had a comparable accuracy (0.9801 vs. 0.9807 vs. 0.9823), F1 score (0.9800 vs. 0.9804 vs. 0.9822), specificity (0.9911 vs. 0.9977 vs. 0.9899), and MCC value. However, XGB showed the highest sensitivity. To further assess the performance of the various models in classifying our data, we employed a confusion matrix. The confusion matrix (**Figure 5A**) showed that the different models had a strong capability in identifying active and inactive compounds. RF exhibited excellent performance in predicting inactive compounds, misclassifying on 2 instances compared to other models. Based on its robust performance in correctly identifying both active and inactive compounds and its high accuracy, we selected RF as the best-performing model. To optimize our model's ability to make accurate predictions, we fine-tuned its hyperparameters (**Figure 5B**).

The accuracy (0.99) and confusion matrix of our fine-tuned RF model demonstrated an improved predictive ability of the model (**Figure 5B**). The fine-tuned model misclassified zero instances of the inactive compound. To validate the fine-tuned model, we employed the Y-scrambling test (**Figure 5C**). The fine-tuned model significantly outperformed the base model in terms of the random forest's accuracy score, affirming the robustness of our model. Additionally, it suggests that our model captured the underlying pattern in the data. To find a promising inhibitor against BACE1, we curated and screened a library of 9280 compounds using the established model. We found 3379 to be active against BACE1. To shortlist only compounds with permeation and absorption potential, we employed the Lipinski rule of five. Interestingly, only 1382 compounds fulfilled the Lipinski Rule of Five criteria. These compounds were further explored as drug candidates through molecular docking studies.

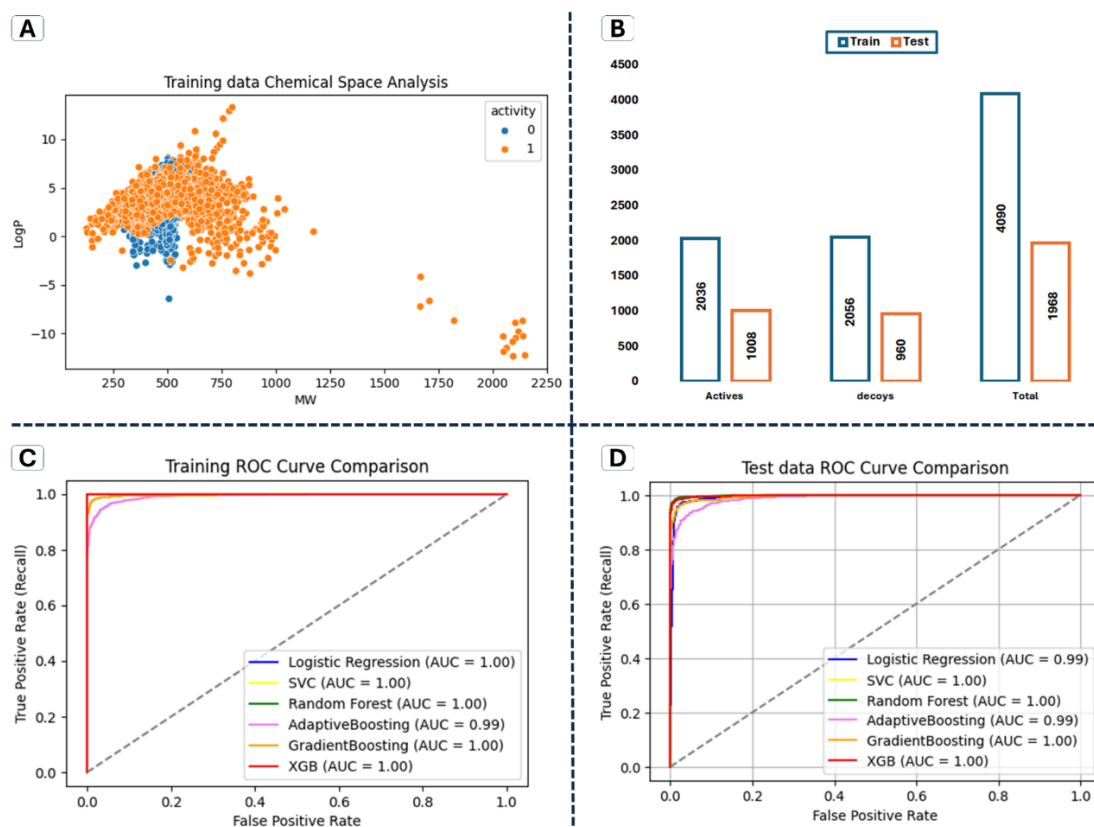


Figure 4. (A) Chemical space occupied by the training data on the molecular weight-cLogP landscape (B) distribution of training and test data used in this study including the decoys employed (C) The ROC curve of the different models on the training dataset (D) The ROC curve of the different models on the test dataset.

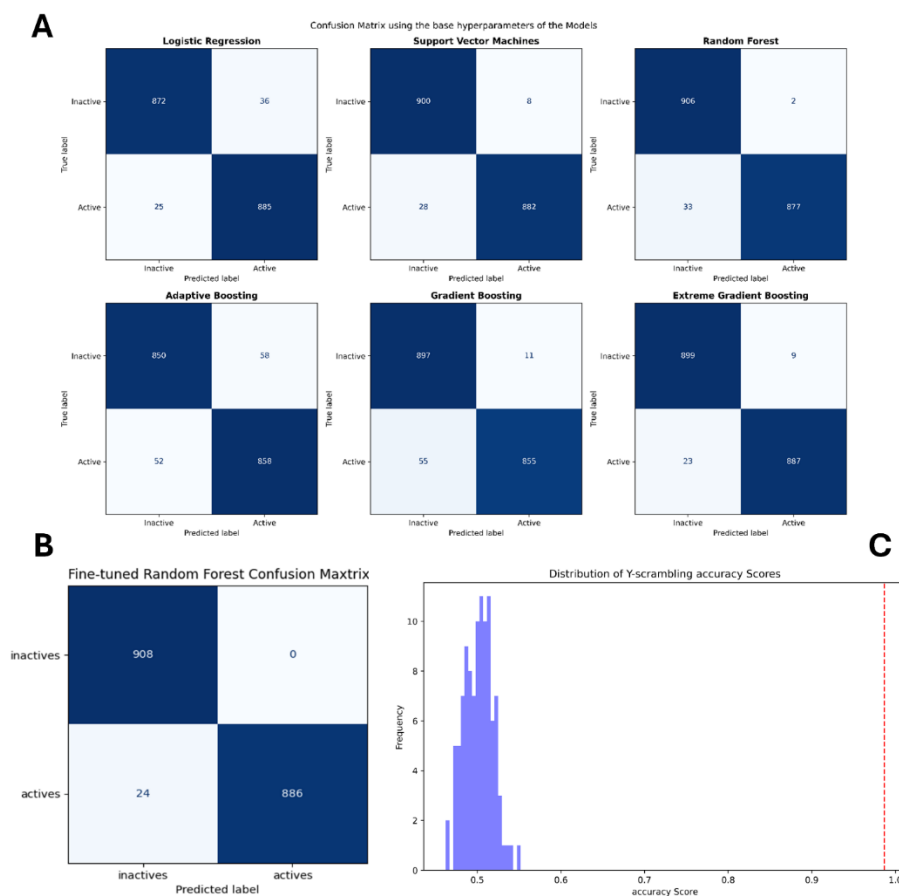


Figure 5. (A) Confusion matrix for different models used in this study (B) Confusion matrix for fine-tuned model (Random forest) (C) Y-scrambling test of the fine-tuned model (RF) to validate model's performance.

Table 1. Metric Evaluation for all the models used in this study.

Model name	Accuracy	F1 score	Precision	Sensitivity	Specificity	MCC
Logistic Regression	0.966	0.967	0.961	0.973	0.960	0.933
SVM	0.980	0.980	0.991	0.969	0.991	0.961
RF	0.981	0.980	0.998	0.964	0.998	0.962
AdaBoost	0.939	0.940	0.937	0.943	0.936	0.879
Gradient Boost	0.964	0.963	0.987	0.939	0.988	0.928
XGB	0.982	0.982	0.990	0.975	0.990	0.965

Binding Affinity and Binding Interactions

Aspartate enzymes, such as BACE1, all possess the aspartic acid dyad units critical for the catalytic function of the enzyme [6]. This catalysis involves electron and proton transfer, and a stable interaction of the aspartate residues through electrostatic interactions is critical for an inhibitor of this enzyme. After energy minimization, the three study ligands (BA1, BA2 & BA3) and the reference

compound were docked into the active site of BACE1 using the Glide algorithm (Schrodinger). The results are shown in **Table 2**. BA1 had the highest binding affinity (-8.33 kcal/mol), approximately twice that of the reference compound (-4.77 kcal/mol). BA2 (-6.8 kcal/mol) and BA3 (-6.5 kcal/mol) had nearly equal binding affinities, as also shown in Table 2. Glide uses a systematic search of the conformational space of the docked ligand, followed by energy optimization on an OPLS-AA nonbonded potential grid and a final refinement using Monte Carlo sampling [30] BA1 made an electrostatic contact with the aspartic dyad (Asp232 & Asp36) through hydrogen bonding while Tyr75 stabilized the rings of the ligand through π - π stacking interactions (**see later sections**).

The aspartic dyad was involved in electrostatic charged interactions with the quaternary amine of BA2, while Thr76 from the flaps also maintained hydrogen bonding interactions with the carboxylic end (**see later sections**). Hydrophobic interaction to BA3 is provided by Tyr202 through pi-alkyl interaction. BA3 enjoyed a robust hydrophobic interaction with aromatic residues such as Tyr75, Trp119, and Phe112, while Trp80 formed a hydrogen bond interaction with the keto group (**Table 2, see later sections**). The reference compound (Verubecestat (MK-8931), which failed at phase III trial as a BACE1 inhibitor, had been characterized to bind to Asp32 through the secondary piperidine nitrogen and to Asp232 through the amidine nitrogen bonded to the piperazine. The amide nitrogen is also in a hydrogen bond with Gly234 [46]. In this study, however, Verubecestat made no contact with the aspartate residues but formed a hydrogen bonding interaction with Phe112 on the amidine nitrogen of the piperazine ring (**see later sections**). Additionally, we observed that the fluoro-piperidine ring of Verubecestat extended into the opposite loop (**see later sections**) and maintained a pi-alkyl interaction with Arg239. This agrees with the study of Kennedy *et al.*, who reported similar interaction [46].

Table 2. Docking score and interacting residues of study ligands with BACE1.

S/N	Ligands	Binding Affinity (Score)	Hydrogen Bond	Hydrophobic Interactions
Amino acid residues				
1.	BA1	-8.330	Asp232, Asp36	Tyr75
2.	BA2	-6.800	Asp232, Asp36, Thr76	Tyr202
3.	BA3	-6.500	Trp80	Phe112
4.	Reference	-4.700	Phe112	Arg239, Tyr75

Residues shown here are the ones with the most significant interactions.

Structural stability of Ligand-bound BACE1 protein

To assess the structural stability of our protein throughout the course of simulation, the RMSD with respect to the initial optimized structure was calculated as shown in **Table 3 & Figure 6A**. All the systems attained structural stability over the course of the 250ns simulations. In **Table 3**, BA2 (1.30713±0.109Å) and BA1 (1.49245±0.144Å) achieved the lowest deviation from the average structure compared to the reference (1.60151±0.159Å) and the Apo structure (1.51714±0.184Å). This indicates a more stable structure for the ligand-bound systems compared to the reference-bound and apo systems. While RMSD is important, it can differ significantly depending on the average structure used. Structures from NMR, minimized structures, and averages of a simulation trajectory all give slightly different deviations [47]. Kumalo & Soliman reported a stable RMSD of 1.0 - 2.3Å in the apo form of BACE1 for a 300ns simulation [6,48].

While RMSD measures deviations from a global stable state, RMSF measures the fluctuations of α -carbon of the protein backbone from their mean positions. This is important in understanding the structural changes in different domains and motifs of the protein [49]. A high value of RMSF indicates much larger deviations and thus structural flexibility and mobility from the average position of the residues. All the studied ligands had a higher RMSF value than the Apo (0.76017±0.40Å) form of the

protein (**Table 3, Figure 6B**). This suggests that binding of ligands introduced more fluctuations in the mean positions of the residues compared to the unbound protein. Compound BA2 ($0.775\pm 0.461\text{\AA}$) had the lowest RMSF value compared to BA1 ($0.782\pm 0.444\text{\AA}$), BA3 ($0.781\pm 0.578\text{\AA}$), and the reference ($0.831\pm 0.577\text{\AA}$). In **Figure 6B**, significant fluctuations were observed around residues 250-275 and 320-330. We further investigated the fluctuations in the mean positions of the residues in the active site to determine whether these fluctuations are due to the binding mechanisms of these ligands (**Supplementary Figure S1A**). As shown, the fluctuations did not vary much among the ligands, though BA2 maintained somewhat lower fluctuations than others for all the active site residues. These findings suggest that the significant differences in fluctuations seen between the ligands and the apo could be due to movement in other parts of the protein away from the active site residues. Possible causes include movements in the residues of the flaps and the opposite loop [7] as will be discussed in later sections of this paper.

In the 250ns simulation period, we also assessed the compactness of our ligand-bound and apo protein by measuring the ROG. The radius of gyration is important to understand how far apart the atoms of a protein are from its center of mass. This is useful in understanding the folded/unfolded states of native and ligand-bound systems to measure structural stability, solvation, and possible interactions with other biological molecules [50]. The ROG was very stable for all the ligand-bound systems compared to the Apo protein ($20.921\pm 0.077\text{\AA}$) except the BA3-bound systems ($21.118\pm 0.11\text{\AA}$) (**Table 3**). **Figure 6C**, showed that ligand BA3 maintained a relatively higher ROG during simulations. **Figure S1B (supplementary figure)** shows the ROG for the active site residues, revealing similar patterns for all systems, with BA2 and the Apo systems exhibiting significant fluctuations throughout the simulation. Studies have reported that radical conformation changes occur upon the binding of ligands to BACE1 [6]. The relative stability of our ligand-bound systems may be due to the energetics and interactions involved in the binding process.

During MD simulations, the system is forced to explore various conformations representative of the energy states of the system that are favorable. Due to the sheer number of these motions, dimensionality reduction is applied to extract and present the most important motions responsible for the system's behavior in a principal component analysis. PCA helps measure the average structures underlying these significant protein motions [51]. **Figure 6D** shows the two most important principal axes that contain information about the motions of these proteins, and PC1 describes about 90% of these motions. Most of the motions of these bound and unbound systems overlap along both axes. The reference compound explored conformations like those of the BA1 system and the apo proteins. However, there are a few isolated conformations visited by the BA1, BA3, and the apo system. SASA measures the extent to which atoms at the surface of a protein can interact with the solvent around it. Since proteins are soluble in biological systems, folding activities tend to bury the hydrophobic residues deep within the protein's structure to minimize their access to surrounding water molecules [52]. How successful a protein is related to its stability. In implicit solvent models, the energy of solvation is approximated as a linear function of the accessible surface area, treating water molecules as a continuum. In **Figure 7A**, the SASA of the bound and apo proteins fluctuated slightly between 15000\AA and 17500\AA . This suggests that ligand binding provided no significant changes in the residues exposed to the solvent system.

Figure 7B-E shows the correlation matrix for the inhibitor-bound and apo form of BACE1. Correlated protein motions are essential in the study of protein dynamics and function [53]. The correlation matrix is a linear matrix that describes which regions of proteins move together [53]. **Figure 7B-E** is a pairwise correlation that allows us to track how the movement of each residue of our protein influences the motions of the other residues. Due to these motions, changes about a residue can be sensed by a distant residue, thereby inducing a conformational change—the principle of allostery [54]. Compounds BA1, BA2, reference, and apo structures (**Figures 7B-C & 7E-F**) showed similar correlations in their complexed states. All four systems showed correlations between residues 340 to 360 and residues 280 to 320. These are the two regions leading up to the folded opposite loop, as seen in **Figure 2A**. So, these two regions move in sync to change the position of the opposite loop

during ligand binding. However, this correlation is less intense in compound BA2, suggesting less movement and greater stability. There are also regions of correlated motions between residues 60 to 120 for compounds BA1, BA2, the reference, and the apo structures. This region controls the movements of the 10s loop of the BACE1 protein. Compound BA3 showed a slightly different correlation (**Figure 7D**). There is a stronger and wider correlation in the regions of the opposite loop (residues 340 to 360) and the 10s loop (residues 280 to 320), signifying higher instability than in other complexes and the apo forms. Additionally, there is a highly anticorrelated region between residues 240 and 360, and residues 60 and 180. Residues 240 to 360 correspond to the loops hosting the Asp232, while residues 60 to 180 correspond to those of the flaps. This signifies an opposing motion induced by BA3 on the flaps and the catalytic dyad. This could explain the poor binding and interactions of BA3.

Table 3. RMSD, RMSF, and ROG parameters of the ligand-bound and the Apo systems of BACE1.

S/N	Ligands	RMSD (Mean±SD)	(Å)	RMSF (Å) (Mean±SD)	RoG (Å) (Mean±SD)
	BA1	1.492±0.144		0.782±0.444	20.870±0.910
	BA2	1.307±0.109		0.775±0.461	20.931±0.076
	BA3	1.524±0.176		0.781±0.578	21.117±0.110
	Reference	1.602±0.159		0.831±0.577	20.994±0.078
	Apo	1.517±0.184		0.760±0.40	20.921±0.077

RMSD= Root Mean Square Deviation; RMSF= Root Mean Square Fluctuations; ROG= Radius of Gyration.

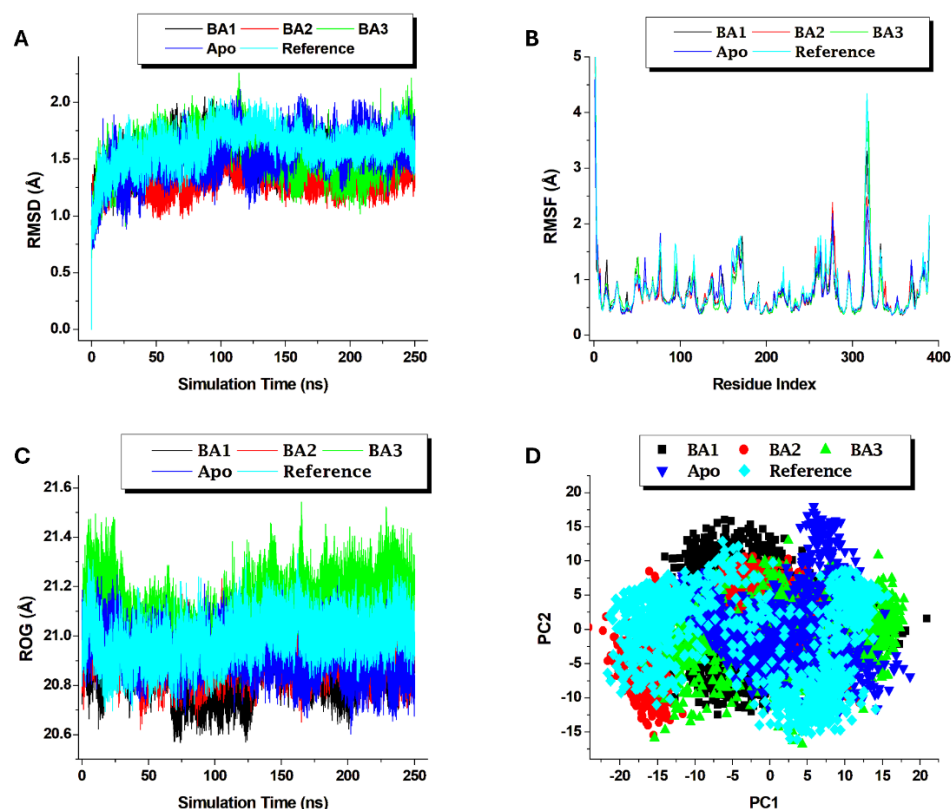


Figure 6. (A) Conformational stability, C-α atoms RMSD of Apo and the BACE1-bound ligands; (B) Flexibility, C-α atoms RMSF of apo and the BACE1-bound ligands; (C) RoG of Apo and BACE1-bound ligands; (D) PCA plot of the apo and the BACE1-bound ligands. All simulations were ran for 250 ns. Apo: unbound receptor.

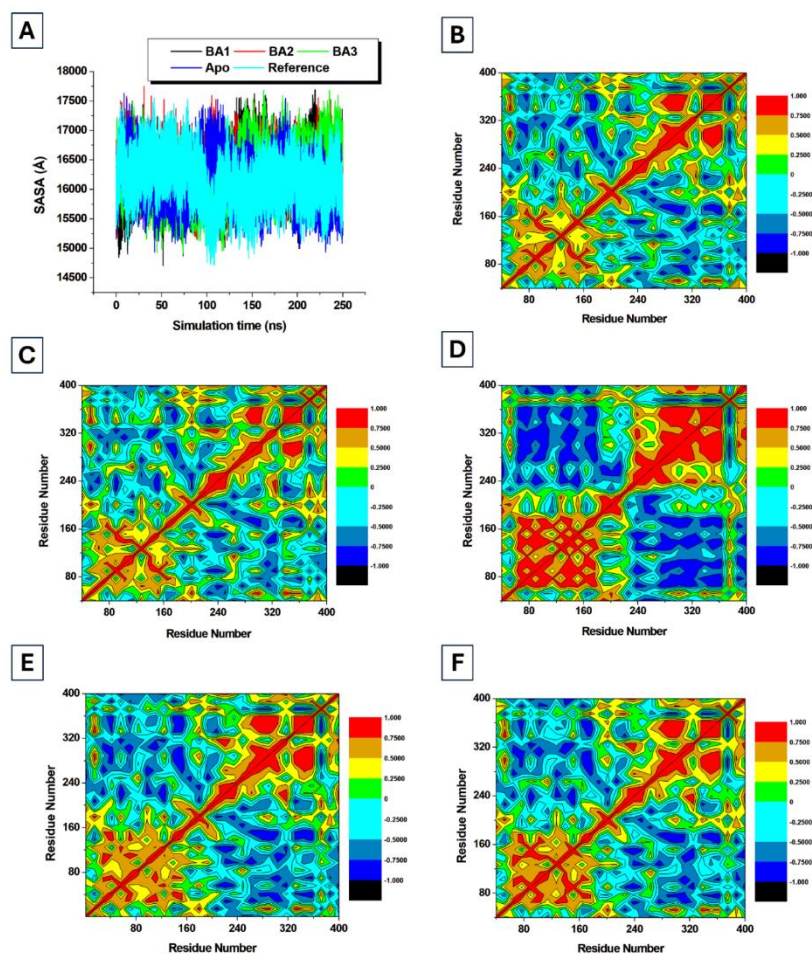


Figure 7. (A) SASA plots for all the ligand-bound BACE1 complexes and the apo protein. Correlation Matrix plots for (B) BA1 (C) BA2 (D) BA3 (E) Reference and (F) apo complexes of BACE1 protein.

Thermodynamic Stability and Energetics of the System

Binding free energy (BFE) is the driver of all molecular processes from chemical reactions to protein folding and interactions. It quantifies the amount of internal energy present in a system that is available to do work, as well as the direction of the thermodynamic process [55]. The value of BFE, and its direction, will determine if a molecular system will remain in its state or transition to a state with lower energy. The MMGBSA has emerged as a valuable tool in estimating this energy of binding/interaction to a certain degree of accuracy. Unlike pathway methods such as FEP (Free energy perturbation) and thermodynamic integration (TI), MMGBSA samples only the initial and final state of the system in an implicit solvent system, thereby cutting computational cost drastically while still maintaining accuracy [56].

When a ligand binds to a protein, negative values of Gibbs free energy (ΔG_{bind}) suggest favourable binding, which is more stable than the native or the unbound state. Additionally, the value of ΔG_{bind} indicates the stability of the conformation, whether an inhibitor, agonist, or inverse agonist is bound. To understand the thermodynamic stability of our bound complexes in this study, we implemented the MMGBSA protocol for the first 40ns (Table 4). Ligand-bound systems BA1 and BA2 had the highest BFE of -33.885 ± 5.423 kcal/mol and -33.820 ± 4.254 kcal/mol, respectively. BA3 had a BFE of -28.792 ± 3.813 kcal/mol, which is also higher than that of the reference compound (-21.090 ± 6.183 kcal/mol). The critical role of hydrophobic interactions in designing inhibitors for proteins and nucleic acids cannot be overstated. Every methyl group added to a compound promises a profound effect on the interaction capabilities [57]. Van der Waals forces, through pi-pi stacking and pi-alkyl interactions, are major contributors to the hydrophobic effect. Though weaker than

conventional electrostatic interactions, hydrophobic interactions can accumulate to a high force responsible for improved affinity [58]. This also considers that deeper, active pockets in proteins are shielded from solvents due to important hydrophobic residues hidden within these pockets. In **Table 4**, the reference compounds (-39.226 ± 3.487), BA1 (-38.196 ± 3.204), and BA3 (-37.644 ± 3.245) lead in van der Waals interactions, while BA2 had very low energy due to van der Waals effects. These observations can easily be explained by the structure of these ligands. In **Figure 8A**, BA1 was involved in pi-pi stacking with Tyr75 and many other non-significant pi-alkyl interactions. Phe112, Trp119, and Trp75 were involved in pi stacking with compound BA3; in addition to Tyr75 (**Figures 9A & 9B**), the reference compound interacted with Arg239 from the opposite loop with its fluorophenyl ring (**Figures 9C & 9D**). It is important to note the role of Tyr75 in these interactions. In the paper discussing the co-crystallization of the reference compound with the BACE1 protein, these interactions were not found [48]. However, Tyr75 is reported as a major contributor to hydrophobic interactions in many other studies involving BACE1 [59,60]. Exploiting this residue could be a major link to optimizing hydrophobic interactions in the active site of the BACE1 protein.

The second term in **Table 4** is the electrostatic interaction term. Coulomb's law is employed to calculate the electrostatic interactions of ligands using atomic charges from the molecular mechanics' force field [42]. BA1 (-206.055 ± 12.955) had the highest electrostatic interactions, followed by BA2 (-114.864 ± 10.313). The value of the electrostatic term depends on the charges or protonation state of the ligands and residues in the active site. The active site of BACE1 is made of two acidic, negatively charged critical residues, ASP36 & Asp232 (**Figure 2A**). This makes the active site very attractive to electrophiles and charged functional groups. Popular Amber forcefields parametrize ASP residues as negatively charged, which reflects their normal ionization states in proteins under physiologic conditions [61]. BA2 is a derivative of L-cysteine introduced as a zwitterion in the MD simulation step. In the active site, it caused a transfer of charge from its quaternary amino group to the ASP dyad (**Figures 8C & 8D**). This likely paved the way for the robust electrostatic interactions seen. The mechanism of action could be via charge-induced migration, followed by interactions that stabilize the system. The importance of these charged interactions has also been highlighted by Gueto-Tettay *et al.* in the study of inhibitory pathways of hydroxyl ethanolamine. They reported that the ligand became protonated by the charged Asp232 residue before flap closure, even when the ligand is added in a neutral form [48]. Furthermore, the carboxylate ion of BA2 formed a hydrogen-bonded interaction with Thr76, a unionizing residue. System BA1 recorded the highest electrostatic term, likely due to the polarizability of the atoms. Polarizability of atoms and residues contributes about 30% to the electrostatic term in the MMGBSA protocol [42]. The presence of oxo groups, such as hydroxyl, ether, and cyclo-ethers, may have influenced the polarization of the compound, accounting for the large electrostatic term in **Table 4**. Also, BA1 made a hydrogen bond interaction with Asp36 using its quinoline tertiary amine. Systems BA3 and the reference compound shared similar electrostatic terms in **Table 4**. BA3 consists of two phenyl rings joined by a thiazolyl propenone link. As shown in **Figures 9A & 9B**, hydrophobic interactions predominate the system as Trp80 is the only residue having hydrogen bonding interactions with BA3. The orientation of the reference compound, as already pointed out, excluded it from contact with the ASP dyad, as it made electrostatic interactions with Phe112 under the flap.

Solvation effects of ligands are an important consideration in the design of high-affinity binders [58]. In **Table 4**, E_{GB} & E_{SA} represent the polar and non-polar solvation energies of the ligands. BA1 and BA2 typically have higher polar solvation energies than compounds BA3 and the reference. The polar solvation term represents the electrostatic component of interactions between the ligand and the solvent [42]. Higher polarity also means higher E_{GB} value. BA1 and BA2 are typically highly polar compounds. Highly solvated ligands also pay a penalty of desolvation from the solvent system into the active pocket. This penalty can be offset if the ligand enters the active site, while at the same time exposing a hydrophilic chain into the solvent continuum [58]. In understanding the high binding affinity of biotin to avidin and streptavidin, McConnell *et al* reported that biotin placed its highly solvated carboxylate group at the periphery of the active site of streptavidin to bind two extra water

molecules, thereby reducing the desolvation penalty [58]. In **Figure 10**, while BA2 locked charged interactions with its carboxylate and amino charged groups, it extended the sulfoxide moiety outside of the active site, giving access to water molecules. This is an important strategy that BA2 could have applied to maintain interactions while also becoming solvent. While the reference compound also contains numerous functional groups that could potentially be solvated, Figure 10 shows that its thia-diazinyl group is hidden under the flaps, thereby limiting solvent access. The non-polar solvation term is a derivative of the solvent accessible surface area of the ligand. It represents the attractive and repulsive components of the van der Waals interactions that exist between the solvent and the ligand [42]. As shown in **Table 4**, and **Figure 7A**, this term did not vary significantly among the ligands.

Table 4. Summary of MMGBSA-based computed thermodynamics binding free energy (BFE) obtained for BACE1.

Energy (kcal/mol)	BA1	BA2	BA3	Reference
$\Delta E_{vdw}^{[a]}$	-38.196±3.204	-14.461±3.484	-37.644±3.245	-39.226±3.487
ΔE_{elec}	-206.055±12.955	-114.864±10.313	-17.421±3.311	-17.403±18.145
E_{GB}	215.176±10.628	98.849±8.792	31.784±2.716	40.792±15.303
E_{SA}	-4.810±0.369	-3.344±0.216	-5.511±0.454	-5.253±0.421
ΔG_{gas}	-244.251±14.058	-129.325±9.10	-55.065±5.134	-56.629±20.361
ΔG_{solv}	210.366±10.462	95.505±8.740	26.273±2.547	35.53±15.032
ΔG_{bi}	-33.885±5.423	-33.820±4.254	-28.792±3.813	-21.090±6.183

ΔE_{vdw} = van der Waals, ΔE_{elec} = electrostatic, ΔG_{bind} = calculated total free binding energy, ΔG_{gas} = gas-phase energy, ΔG_{solv} = solvation free energy, E_{SA} = solvation energy, E_{GB} = Gibb's energy.

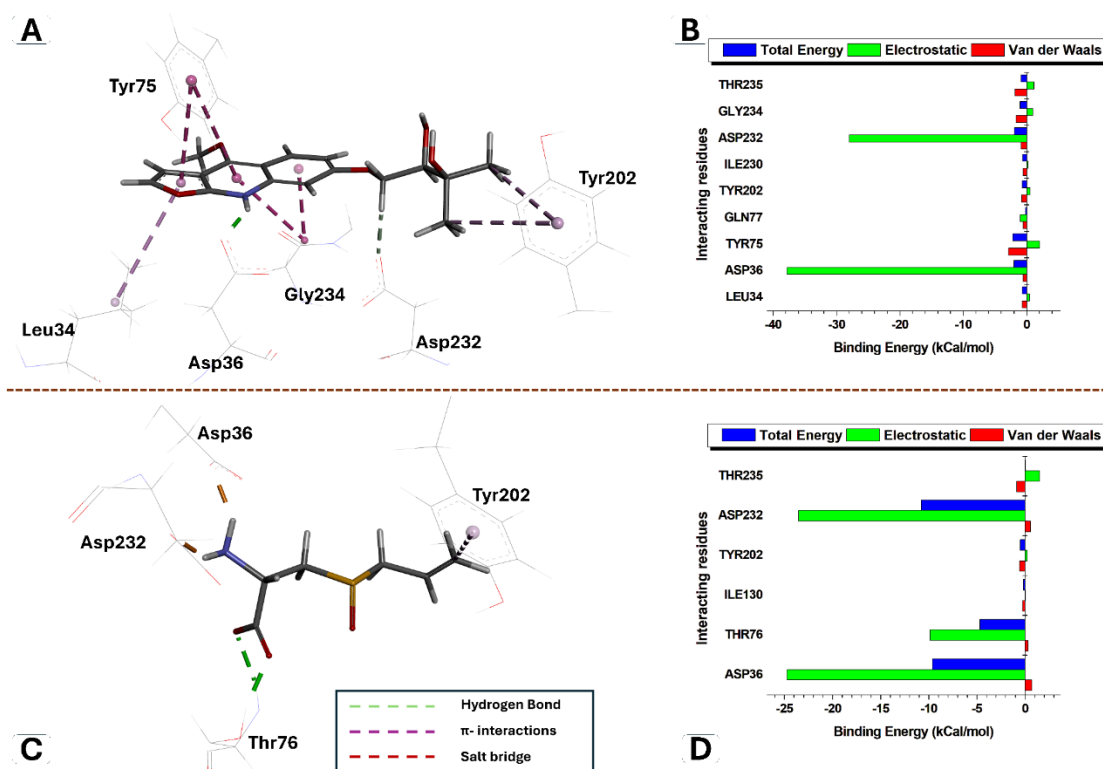


Figure 8. A 3D visualization of the binding interactions and per residue energy decomposition (PRED) of study ligands with BACE1 receptor. (A) Interactions of BA1 at the active site and (B) the contributions of the active site

residues to binding energy via hydrogen bonding to ASP232 and ASP36; (C) interactions of BA2 at the active site and (B) the energy contributions of ASP232 and ASP36 with charged quaternary amine via salt bridge. (Residues are shown as line representations while ligands have stick representations; only residues with significant contributions in the PRED are shown in the interactions diagram).

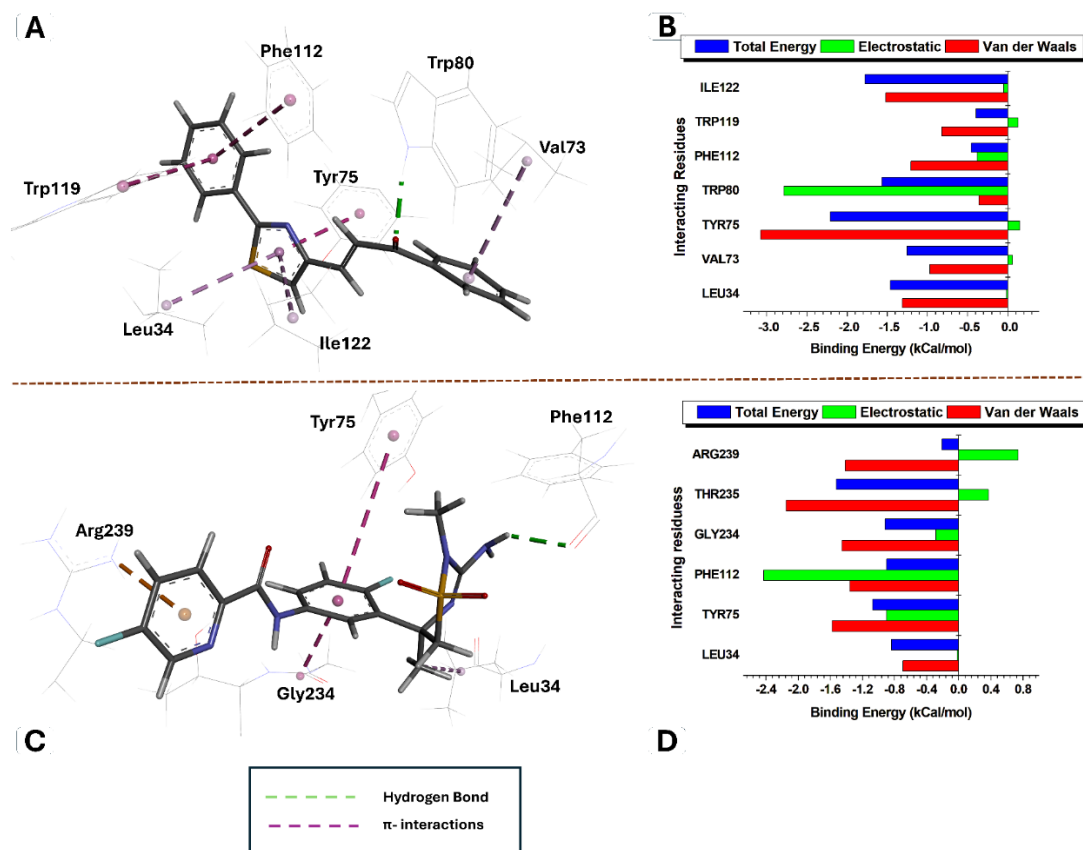


Figure 9. A 3D visualization of the binding interactions and per residue energy decomposition (PRED) of study ligands with BACE1 receptor. (A) Interactions of BA3 at the active site and (B) the contributions of the active site residues to binding energy via hydrogen bonding contact to TRP80; (C) interactions of the reference at the active site and (D) the energy contributions of PHE112 via hydrogen bonding. (Residues are shown as line representations while ligands have stick representations; only residues with significant contributions in the PRED are shown in the interactions diagram).

Binding Dynamics and Flap Motions

The motions undergone by BACE1 upon inhibitor or substrate binding have been well characterized by many other studies [9,10,62]. More importantly, the flaps over the active site play a critical role in substrate recognition and binding. This flap maintains three distinct conformations: open, semi-closed, and closed. The flap should be open when unbound and closed when bound to trap the inhibitor [48]. The most widely used parameter to measure the conformation of this flap is the distance between the backbone carbon of the tip of the flap and the opposite Asp residue in the catalytic dyad. Kumalo *et al*, however, introduced more parameters to account, not only the closing/opening motions but also the twisting motions of the flaps (Figure 2) [6]. In this study, after the 250ns simulation, we employed the distance and angle tools of cpptraj to measure the parameters shown in Figure 2 [37].

Figure 11 shows BA1, BA2, BA3, and the reference compound aligned in the active site of BACE1 at 40ns; with the position of their respective flaps, opposite loops, and key interacting residues shown. The most remarkable observation is that the opposite loops have little or no fluctuations, irrespective of the ligand bound. This has been well studied and documented by many other reports [7,8]. On the other hand, the type of bound ligand and its key interacting residues influenced the inward closing

of the flaps. The reference compound, by extending its fluoro-pyridinecarboxamide moiety to the opposite loop, created clashes with the residues of the flap, thereby leaving it in a more open conformation. BA2 was involved in a hydrogen bonding with Thr76 at the tip of the flap, thereby pulling the flap inwards (**Figure 11**). As already discussed, BA3 is well hidden under the flaps, interacting with Phe112, thereby leaving the flaps more flexible than seen in the reference compound. The configuration of the apo structure (brown in **Figure 11**) does not actually reflect the open conformation of the unbound form. This is because the apo used in this study is an inhibitor-bound protein, in which the inhibitor has been removed from the active site.

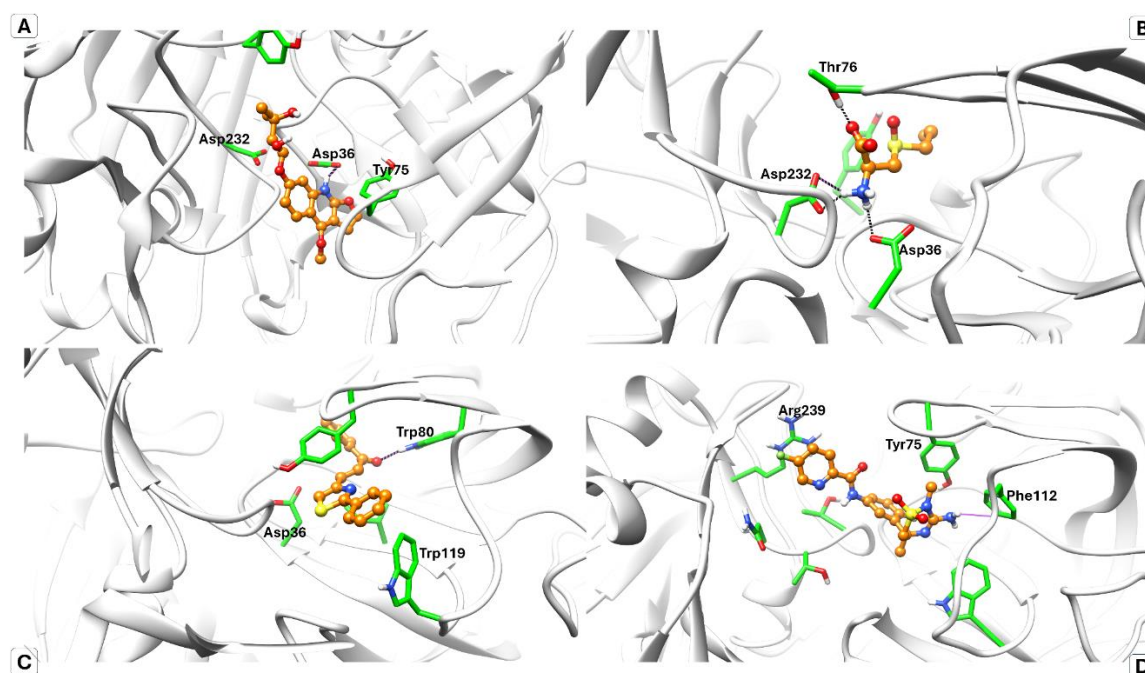


Figure 10. A 3D visualization of the ligands in the active site of BACE1 receptor (protein: white; ligands: yellow; critical interacting residues: green). (A) Interactions of BA1 at the active site (B) Interactions of BA2 at the active site (C) Interactions of BA3 at the active site (D) Interactions of the reference compound at the active site. Dashed lines between residues and ligands denote hydrogen bonding.

To further understand these opening and closing motions as well as the twisting motions of the flaps, we have shown, in **Figure 12**, the distance between Thr76 at the tip of the flap and Asp36 in the catalytic dyad, for the first 80ns. The timescale of 80-100ns is usually sufficient to understand these motions, as they are the initial events that happen upon entry and binding of the substrate to the active site of BACE1 [6,48]. We also provided the distance for the full 250ns in Supplementary **Figure S2**. The average distances explored by the protein-bound complexes, for the full 250ns, are also shown in **Table 5**. To help in the discussion, we adopted the reference distances defined by Kumalo and Soliman in our study [6]. Distances less than 9.00 Å correspond to a closed conformation, while distances greater than 13.00 Å indicate an open conformation [6,9]. So, the semi-closed conformation lies at flap distances between 9.00 Å and 13.00 Å. As shown in **Table 5**, the average distances throughout the simulation are semi-open conformations (Apo:11.724 ± 0.703Å; BA2:11.807 ± 0.401Å), and open conformations (BA1:14.724 ± 0.979Å; BA3:13.785 ± 0.612Å; Reference:13.784 ± 1.246Å). The standard deviations are very low, indicating fewer fluctuations around the average values. Among the ligands, only BA2 adopted an average semi-closed conformation (distance: 11.807 ± 0.401Å). As pointed out earlier, the hydrogen bonding interactions between Thr76 and the carboxylate group of BA2 strongly influenced the semi-closed conformation it adopted. Gueto-Tettay *et al* have also mentioned the importance of such hydrogen bonding on flap closure [48]. Future ligand design should prioritize such contact, in addition to the catalytic dyad, for an early flap closure.

In the detailed flap movements, as shown in **Figures 12B & 13**, in compound BA2, exhibited little fluctuation in the first 40 ns of the simulation, with distances cycling between 10 Å and 13 Å before stabilizing at approximately 11.5 Å for the remainder of the simulation time. The initial fluctuations may occur during the entrance of the ligand and initial binding activities, which stabilize once the inhibitor is bound. Compound BA3 did not significantly influence the flap motions. As shown in **Figures 12C & 13** caused the flaps to fluctuate between 12 Å and 18 Å in the first 5 ns, before settling at around 14 Å for the remainder of the time. Inhibitor design should ensure that the ligand is not trapped under the flap, as this would prevent the flaps from closing and potentially impair key interactions with the catalytic dyad. Such high distances have been reported in the literature [6]. The maximum fluctuations seen in the first 80ns are for compound BA1 (**Figure 12A**). The distance hovered between 13 Å and 16.5 Å, with distinct highs and lows repeating approximately every 10-20ns. An explanation for this is not very clear, but it might be due to the non-conventional C-H-O hydrogen bonding with Asp232 by BA1. Such contacts must have been made/broken, as the ligand moved/relaxed in the binding pocket, thereby influencing the distance. The reference compound started from a semi-open conformation, steadily opened after 10ns, and finally peaked around 15.5 Å for the next 30ns (**Figures 12D & 13**). It then fluctuated in a peak and trough pattern for the last part of 80ns.

The angles, theta (θ) and the dihedral phi (Φ), are two important parameters that measure, not only the opening/closing of the flaps, but also the twisting and curling movements undergone by the flaps [6,48]. While theta is still a measure of distance between the flaps and the opposite loop, phi measures how the flap moves towards or away from the 10S loop (**Figure 2**). The average angle (θ), during the 250ns simulation, is lowest for BA1 ($39.269 \pm 4.304^\circ$) followed by Apo ($40.491 \pm 6.719^\circ$), when compared to other systems (**Table 5**). The Apo in this study (θ : $50.493 \pm 4.699^\circ$) has shown a more closed conformation than the Apo (θ : $57.41 \pm 6.54^\circ$) reported by Kumalo and Soliman, due to reasons previously explained [6]. System BA3 maintained the highest average distance between the flaps and the opposite loop (θ : $40.491 \pm 6.719^\circ$). The variations in the angle throughout the simulation are shown in **Figures S4 & S5**. The trend shows an initial stable angle, followed by a rapid oscillation just after 100 ns (Supplementary **Figure S4**). These oscillations are also supported by the large standard deviations (SD) as recorded in **Table 5**, and the variations in the relative position of the snapshots of the flaps as shown in **Figure 10**. The value of theta for the Apo generally decreased till after 100ns, when it entered an oscillating motion (**supplementary Figure S5**). The average phi (Φ) reported in this study (**Table 5**) is two times lower than in the study by Gueto-Tettay *et al* for BACE1 bound to a known inhibitor (-37.828°) [48]. All the systems exhibited similar trends of bending/torsion of the flap along its vertical axis, with phi ranging from $+5,000^\circ$ to $35,000^\circ$ (for the ligand-bound systems) and the apo system reaching as high as $+10,000^\circ$ (**Figures S3 & S5**). Systems BA2 and BA3 exhibited significant variations in torsion during the first 80ns of simulation, as shown in **Figure 13**. These variations correspond to the opening and closing movements associated with inhibitor binding and interactions [6,48].

Table 5. Average distance, angle and torsion of the ligands and the Apo protein of BACE1.

Systems	Parameters of Flap Motions (Mean \pm SD)		
	Distance, D0 (Å)	Angle, θ ($^\circ$)	Torsion, Φ ($^\circ$)
Apo	11.724 ± 0.703	40.491 ± 6.719	-15.672 ± 6.679
BA1	14.724 ± 0.979	39.269 ± 4.304	-12.907 ± 4.629
BA2	11.807 ± 0.401	45.471 ± 4.357	-18.862 ± 4.276
BA3	13.785 ± 0.612	50.493 ± 4.699	-18.101 ± 4.355
Reference	13.784 ± 1.246	44.376 ± 4.111	-16.092 ± 4.372

All parameters are as defined in Figure 2.

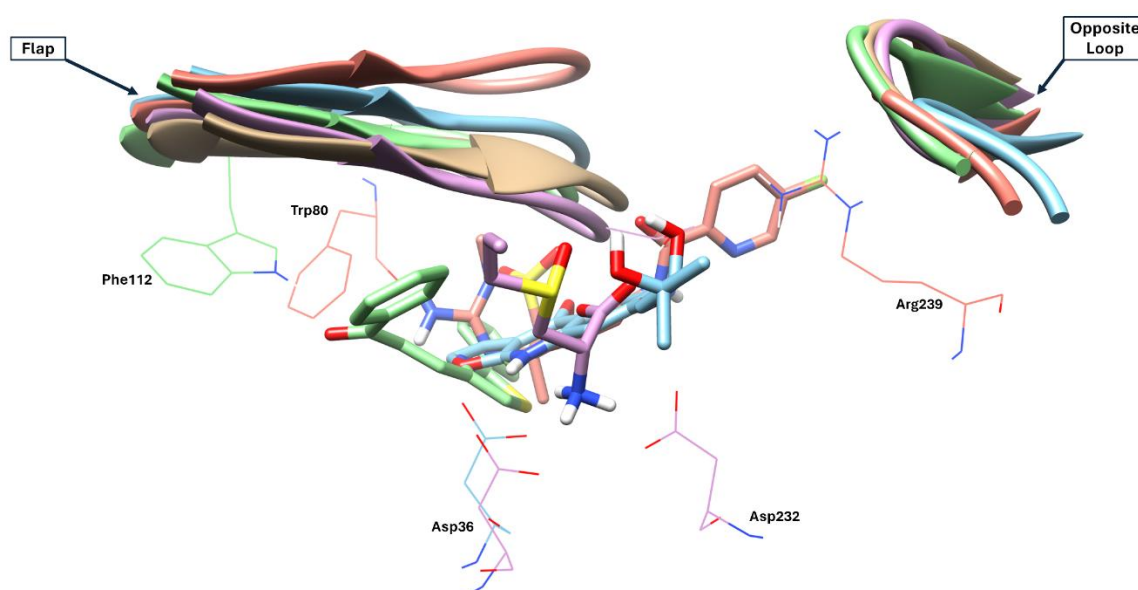


Figure 11. Ligands (stick representations) aligned at the active site of BACE1 protein at 40ns, with the interacting residues (wire representations), flaps and opposite loops shown. BA1 (cyan); BA2 (pink); BA3 (green); Reference (orange); apo (brown).

Pharmacokinetic Properties of the Ligands

Lipinski, in his classical work, described the properties of compounds with the potential to make it into the market as drugs [63]. These properties, called drug likeliness (or the rule of 5), include $MW \leq 500$, $nHBA \leq 10$, $nHBD \leq 5$, and $cLogP \leq 5$. While many drugs have made it to the market with one or more deviations, the rule of 5 remains arguably the most important starting point for the design of small-molecule inhibitors/activators [64]. Since the hit compounds in this study have not been reported as potential inhibitors of BACE1, we assessed these conditions in addition to many other predicted pharmacokinetic properties (**Table 6**). The ligands in this study all passed the Lipinski test, with very good properties, which makes them suitable as potential candidates. Also, leads targeting Alzheimer's disease must be able to cross the blood-brain barrier [65]. The blood-brain barrier (BBB) is composed of tight junctions, receptors, transporters, and metabolizing enzymes, which restrict the passage of drugs/xenobiotics into the brain [66]. Only compound BA2 had a TPSA greater than this threshold, but it is also very water soluble and thus predicted not to cross the BBB. Only compounds BA1 and BA3 may not be able to cross the BBB to target Alzheimer's disease. BACE1 binding sites have additional pockets that could be leveraged to optimize these potential inhibitors without compromising their activity. Many other options to access the BBB include redesigning these inhibitors to be recognized by natural transporters in the brain. Compound BA2, a derivative of cysteine, stands a chance here in utilizing transporters and co-transporters of amino acids in the brain. Apart from this, the transporter-prodrug approach, by derivatizing the drug with an amino acid natural substrate of the transporter, can also be used [67]. In particular, the Alanine, serine, and cysteine transporters (ASCTs) can be exploited in this design [68]. Nano-delivery systems are another ready solution for improving the bioavailability of these potential candidates in the brain. Polymeric nanoparticles, dendrimers, chitosan nanoparticles, lipid-based nanocarriers, liposomes, solid lipid nanoparticles (SLN), and nano-emulsions have been applied in the delivery of drugs targeting Alzheimer's disease [68].

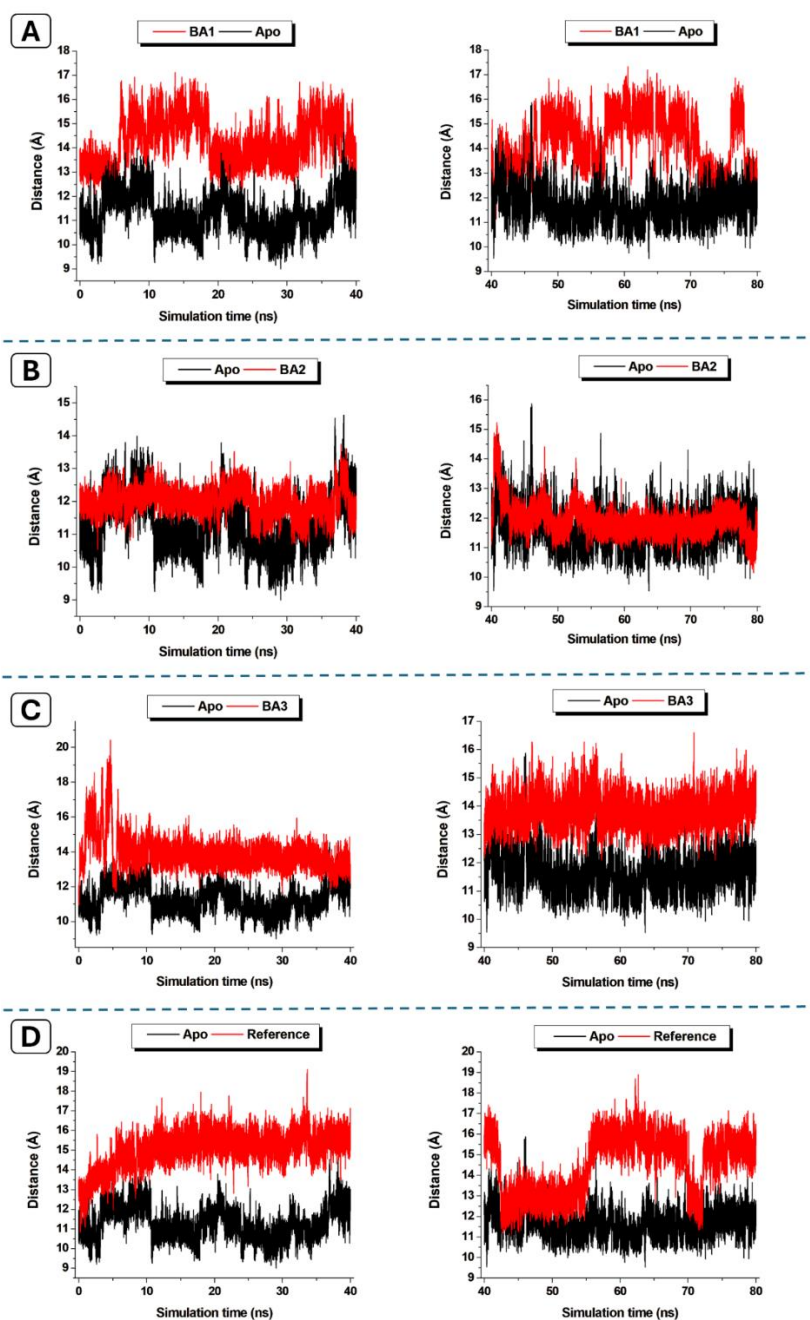


Figure 12. The distance D_0 , between C_α of Thr76 at the tip of the flap and of C_α of Asp36 in the catalytic dyad, for the first 80ns of the simulation timescale. (A)BA1 vs Apo (B) BA2 vs Apo (C) BA3 vs Apo (D) Reference vs Apo.

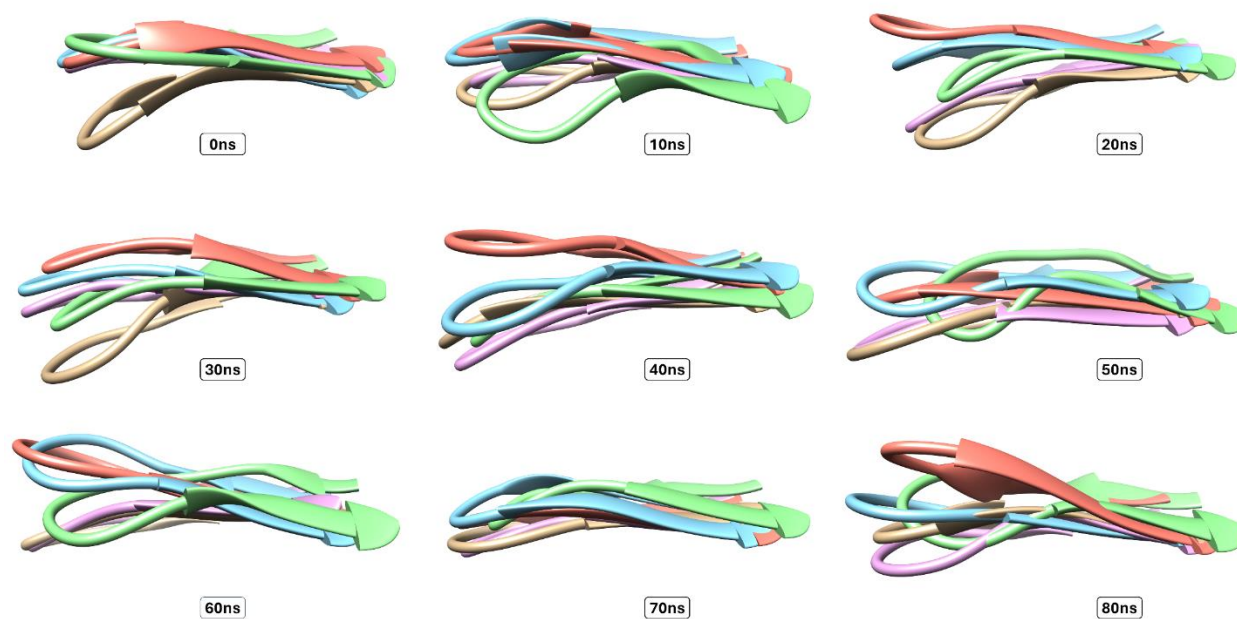


Figure 13. Snapshots of flap motions for every 10ns of the first 80ns of the simulation. Apo (Brown), BA1 (cyan), BA2 (pink), BA3 (green), Reference (Orange).

Table 6. Predicted Pharmacokinetic (ADME) and Toxicity Properties of the study ligands.

Properties	Ligands		
	BA1	BA2	BA3
MW (g/mol)	317.34	177.22	291.37
nHBA	6	4	2
nHBD	2	2	0
TPSA (Å ²)	84.95	99.60	58.20
cLogP	2.17	-1.26	4.13
GIA	High	High	High
BBB	No	No	Yes
Water Solubility	Soluble	Soluble	Poorly Soluble
Number of Violations	No	No	No
Predicted LD ₅₀ value (mg/kg)	1000	8000	1000
Prediction accuracy (%)	54.26%	68.07%	67.938%
Neurotoxicity	Inactive	Inactive	Inactive
Hepatotoxicity	Inactive	Inactive	Active
Cytotoxicity	Inactive	Inactive	Inactive
Carcinogenicity	Inactive	Inactive	Inactive
Mutagenicity	Inactive	Inactive	Inactive

MW: Molecular Weight; nHBA: Number of Hydrogen Bond Acceptor; HBD: Number of Hydrogen Bond Donor; TPSA: Topological Surface Area; cLogP: Octanol-water Partition Coefficient; GIA: Gastrointestinal Absorption; BBB: Blood-brain barrier permeability; Number of violations: Number of violations of Lipinski Rule of 5; LD₅₀: Median lethal dose.

4. Conclusions

The present study employed machine learning tools, molecular docking, and molecular dynamics simulations to identify novel compounds with promising inhibitory activity against the BACE1 enzyme. The conformational flexibility of the BACE1 protein was also assessed by analyzing the dynamics of the flap throughout the simulation. The random forest algorithm proved to be very robust in the initial virtual screening, enabling the identification of hits from public databases. We identified BA1 and BA2 through molecular docking and dynamics, which showed very good binding affinities and binding free energies. Compound BA2 maintained very strong charged interactions and hydrogen bonding with the catalytic dyad (Asp36 & Asp232) while keeping the flaps in a semi-closed conformation necessary for activity. While the methodology and findings presented in this study provide a good starting point for designing novel inhibitors of the BACE1 protein, we propose that future work should extend the simulation window to span several microseconds, allowing for the observation of the full conformational changes of the ligand-bound protein. Additionally, the approximations inherent in the end-state models and implicit solvents of the MMGBSA method introduce errors, making it difficult to distinguish the binding affinities of compounds in a series. We believe the binding affinities in this work serve as a guide, and that more precise methods, such as Free Energy Perturbation (FEP) and Thermodynamic Integration (TI), are needed for validation. Also, biophysical, *in vivo*, and *in vitro* experiments are necessary to validate the findings of this study.

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on Preprints.org.

Author Contributions: **Shadrach C. Eze** ----- Conceptualization, Methodology, Data curation, Formal analysis, Investigation, Visualization, Writing – original draft, Writing – review and editing; **Stephen C. Nnemolisa** ----- Conceptualization, Methodology, Data curation, Formal analysis, Investigation, Visualization, Writing – original draft, Writing – review and editing; **Joy C. Onyesoro** ----- Writing – original draft; **Toluwalope D. Ilori** ----- Writing – review and editing; **Victor S. Uche** ----- Writing – review and editing; **Augustine C. Madueke** ----- Writing – review and editing; **Wande M. Oluyemi** ----- Methodology, Resources, Software, Writing – review and editing; **Adeniyi T. Adewumi** ----- Methodology, Resources, Software, Writing – review and editing; **Salerwe Mosebi** ----- Resources, Software, Supervision, Writing – review and editing; **Innocent, U. Okagu** ----- Supervision, Writing – review and editing.

Funding Sources: The authors did not receive any funding for the conduct of this study.

Data Availability: All data associated with this study have been made available in the Appendix (Supplementary Data). The code for the machine learning is also on GitHub.

Acknowledgments: We are grateful to the Centre for High-Performance Computing, Cape Town, for providing computational resources used in this study. The authors also acknowledge the University of South Africa for their robust support.

Competing Interests: The authors declare no conflict of interest.

References

1. Twarowski B, Herbet M. Inflammatory Processes in Alzheimer's Disease-Pathomechanism, Diagnosis and Treatment: A Review. *Int J Mol Sci* 2023;24. <https://doi.org/10.3390/IJMS24076518>.

2. Liu S, Geng D. A systematic analysis for disease burden, risk factors, and trend projection of Alzheimer's disease and other dementias in China and globally. *PLoS One* 2025;20. <https://doi.org/10.1371/JOURNAL.PONE.0322574>.
3. Huang SY. Comprehensive assessment of flexible-ligand docking algorithms: current effectiveness and challenges. *Brief Bioinform* 2018;19:982–94. <https://doi.org/10.1093/BIB/BBX030>.
4. Bazzari FH, Bazzari AH, Kopustinskiene DM, Bazzari FH, Bazzari AH. BACE1 Inhibitors for Alzheimer's Disease: The Past, Present and Any Future? *Molecules* 2022;27:8823. <https://doi.org/10.3390/MOLECULES27248823>.
5. Hampel H, Vassar R, De Strooper B, Hardy J, Willem M, Singh N, et al. The β -Secretase BACE1 in Alzheimer's Disease. *Biol Psychiatry* 2021;89:745–56. <https://doi.org/10.1016/J.BIOPSYCH.2020.02.001>.
6. Kumalo HM, Soliman ME. A comparative molecular dynamics study on BACE1 and BACE2 flap flexibility. *Journal of Receptors and Signal Transduction* 2016;36:505–14. <https://doi.org/10.3109/10799893.2015.1130058;WGROU:STRING:PUBLICATION>.
7. Mahanti M, Bhakat S, Nilsson UJ, Söderhjelm P. Flap Dynamics in Aspartic Proteases: A Computational Perspective. *Chem Biol Drug Des* 2016;88:159–77. <https://doi.org/10.1111/CBDD.12745>.
8. Cai Y, Kurt Yilmaz N, Myint W, Ishima R, Schiffer CA. Differential Flap Dynamics in Wild-Type and a Drug Resistant Variant of HIV-1 Protease Revealed by Molecular Dynamics and NMR Relaxation. *J Chem Theory Comput* 2012;8:3452–62. <https://doi.org/10.1021/CT300076Y>.
9. Xu Y, Li MJ, Greenblatt H, Chen W, Paz A, Dym O, et al. Flexibility of the flap in the active site of BACE1 as revealed by crystal structures and molecular dynamics simulations. *Acta Crystallogr D Biol Crystallogr* 2012;68:13–25. <https://doi.org/10.1107/S0907444911047251>.
10. Bhakat S, Söderhjelm P. Flap Dynamics in Pepsin-Like Aspartic Proteases: A Computational Perspective Using Plasmepsin-II and BACE-1 as Model Systems. *J Chem Inf Model* 2022;62:914–26. <https://doi.org/10.1021/ACS.JCIM.1C00840>.
11. Udegbe FC, Ebulue OR, Ebulue CC, Chukwunonso &, Ekesiobi S. MACHINE LEARNING IN DRUG DISCOVERY: A CRITICAL REVIEW OF APPLICATIONS AND CHALLENGES. *Computer Science & IT Research Journal* 2024;5:892–902. <https://doi.org/10.51594/CSITRJ.V5I4.1048>.
12. Sangeet S. Machine Learning-Enhanced Drug Discovery for BACE1: A Novel Approach to Alzheimer's Therapeutics. *BioRxiv* 2024. <https://doi.org/10.1101/2024.09.24.614844>.
13. Noviandy TR, Maulana A, Idroes GM, Emran T Bin, Tallei TE, Helwani Z, et al. Ensemble Machine Learning Approach for Quantitative Structure Activity Relationship Based Drug Discovery: A Review. *Infolitika Journal of Data Science* 2023;1:32–41. <https://doi.org/10.60084/IJDS.V1I1.91>.
14. Dhamodharan G, Mohan CG. Machine learning models for predicting the activity of AChE and BACE1 dual inhibitors for the treatment of Alzheimer's disease. *Molecular Diversity* 2021 26:3 2021;26:1501–17. <https://doi.org/10.1007/S11030-021-10282-8>.
15. Dao Quang T, Do D, Mai T, Thai QM, Tran P-T, Ngo Cd ST, et al. Characterizing Potential BACE1 Inhibitors from ChEMBL Database using Knowledge- and Physics-Based Approaches. *Theoretical and Computational Chemistry* 2025. <https://doi.org/10.26434/CHEMRXIV-2025-XCCMH>.
16. Sangeet S. Machine Learning-Enhanced Drug Discovery for BACE1: A Novel Approach to Alzheimer's Therapeutics 2024. <https://doi.org/10.1101/2024.09.24.614844>.
17. Noviandy TR, Maulana A, Emran T Bin, Idroes GM, Idroes R. QSAR Classification of Beta-Secretase 1 Inhibitor Activity in Alzheimer's Disease Using Ensemble Machine Learning Algorithms. *Heca Journal of Applied Sciences* 2023;1:1–7. <https://doi.org/10.60084/hjas.v1i1.12>.
18. Leon M, Perezhohin Y, Peres F, Popovič A, Castelli M. Comparing SMILES and SELFIES tokenization for enhanced chemical language modeling. *Scientific Reports* 2024 14:1 2024;14:1–13. <https://doi.org/10.1038/s41598-024-76440-8>.
19. Irannejad H, Valipour M. Cheminformatics analysis of indoleamine and tryptophan 2,3-dioxygenase inhibitors: A descriptor and fingerprint based machine learning approach to disclose selectivity measures. *Comput Biol Med* 2024;180:108954. <https://doi.org/10.1016/J.COMPBIOMED.2024.108954>.
20. Billones LT, Morales NB, Billones JB. Logistic regression and random forest unveil key molecular descriptors of druglikeness. *Chem-Bio Informatics Journal* 2021;21:39–58. <https://doi.org/10.1273/cbij.21.39>.

21. Aldakheel FM, Alduraywish SA, Dabwan KH. Integrating machine learning driven virtual screening and molecular dynamics simulations to identify potential inhibitors targeting PARP1 against prostate cancer. *Scientific Reports* 2025 15:1 2025;15:1–19. <https://doi.org/10.1038/s41598-025-97208-8>.
22. Almatroudi A. Integrative Machine Learning, Virtual Screening, and Molecular Modeling for BacA-Targeted Anti-Biofilm Drug Discovery Against Staphylococcal Infections. *Crystals (Basel)* 2024;14:1057. <https://doi.org/10.3390/CRYST14121057/S1>.
23. Khamis MA, Gomaa W, Ahmed WF. Machine learning in computational docking. *Artif Intell Med* 2015;63:135–52. <https://doi.org/10.1016/J.ARTMED.2015.02.002>.
24. Tiwari A, Chugh A, Sharma A. Ensemble framework for cardiovascular disease prediction. *Comput Biol Med* 2022;146:105624. <https://doi.org/10.1016/J.COMPBIOMED.2022.105624>.
25. Ilemobayo JA, Durodola O, Alade O, Awotunde OJ, Olanrewaju AT, Falana O, et al. Hyperparameter Tuning in Machine Learning: A Comprehensive Review. *Journal of Engineering Research and Reports* 2024;26:388–95. <https://doi.org/10.9734/JERR/2024/V26I61188>.
26. Noviandy TR, Maulana A, Irvanizam I, Idroes GM, Mauludya NB, Tallei TE, et al. Interpretable machine learning approach to predict Hepatitis C virus NS5B inhibitor activity using voting-based LightGBM and SHAP. *Intelligent Systems with Applications* 2025;25:200481. <https://doi.org/10.1016/J.ISWA.2025.200481>.
27. Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, et al. PubChem 2019 update: improved access to chemical data. *Nucleic Acids Res* 2019;47:D1102–9. <https://doi.org/10.1093/NAR/GKY1033>.
28. Irwin JJ, Shoichet BK. ZINC - A free database of commercially available compounds for virtual screening. *J Chem Inf Model* 2005;45:177–82. <https://doi.org/10.1021/ci049714+>.
29. Chandrasekhar V, Rajan K, Kanakam SRS, Sharma N, Weißenborn V, Schaub J, et al. COCONUT 2.0: a comprehensive overhaul and curation of the collection of open natural products database. *Nucleic Acids Res* 2025;53:D634–43. <https://doi.org/10.1093/NAR/GKAE1063>.
30. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, et al. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J Med Chem* 2004;47:1739–49. <https://doi.org/10.1021/jm0306430>.
31. Sander T, Freyss J, Von Korff M, Rufener C. DataWarrior: An Open-Source Program For Chemistry Aware Data Visualization And Analysis. *J Chem Inf Model* 2015;55:460–73. <https://doi.org/10.1021/CI500588J>.
32. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera - A visualization system for exploratory research and analysis. *J Comput Chem* 2004;25:1605–12. <https://doi.org/10.1002/jcc.20084>.
33. Lee TS, Cerutti DS, Mermelstein D, Lin C, Legrand S, Giese TJ, et al. GPU-Accelerated Molecular Dynamics and Free Energy Methods in Amber18: Performance Enhancements and New Features. *J Chem Inf Model* 2018;58:2043–50. <https://doi.org/10.1021/acs.jcim.8b00462>.
34. Oluyemi WM, Samuel BB, Adewumi AT, Adekunle YA, Soliman MES, Krenn L. An Allosteric Inhibitory Potential of Triterpenes from *Combretum racemosum* on the Structural and Functional Dynamics of *Plasmodium falciparum* Lactate Dehydrogenase Binding Landscape. *Chem Biodivers* 2022;19. <https://doi.org/10.1002/cbdv.202100646>.
35. Adewumi AT, Elrashedy A, Soremekun OS, Ajadi MB, Soliman MES. Weak spots inhibition in the *Mycobacterium tuberculosis* antigen 85C target for antitubercular drug design through selective irreversible covalent inhibitor-SER124. *J Biomol Struct Dyn* 2022;40:2934–54. <https://doi.org/10.1080/07391102.2020.1844061;ISSUE:ISSUE:DOI>.
36. Ryckaert JP, Ciccotti G, Berendsen HJC. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J Comput Phys* 1977;23:327–41. [https://doi.org/10.1016/0021-9991\(77\)90098-5](https://doi.org/10.1016/0021-9991(77)90098-5).
37. Roe DR, Cheatham TE. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J Chem Theory Comput* 2013;9:3084–95. <https://doi.org/10.1021/CT400341P>.
38. Kalathiya U, Padariya M, Baginski M. Structural, functional, and stability change predictions in human telomerase upon specific point mutations. *Scientific Reports* 2019 9:1 2019;9:1–13. <https://doi.org/10.1038/s41598-019-45206-y>.

39. Seifert E. OriginPro 9.1: Scientific Data Analysis and Graphing Software—Software Review. *J Chem Inf Model* 2014;54:1552. <https://doi.org/10.1021/CI500161D>.
40. Adewumi AT, Soremekun OS, Ajadi MB, Soliman MES. Thompson loop: opportunities for antitubercular drug design by targeting the weak spot in demethylmenaquinone methyltransferase protein. *RSC Adv* 2020;10:23466–83. <https://doi.org/10.1039/D0RA03206A>.
41. Martinelli A, Ortore G. Molecular Modeling of Adenosine Receptors. In: Conn PM, editor. *Methods in Enzymology - G Protein Coupled Receptors Modeling, Activation, Interactions and Virtual Screening*, vol. 522, San Diego, CA: Academic Press; 2013, p. 37–59. <https://doi.org/10.1016/B978-0-12-407865-9.00003-0>.
42. Genheden S, Ryde U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin Drug Discov* 2015;10:449–61. <https://doi.org/10.1517/17460441.2015.1032936>.
43. Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. *J Mol Graph* 1996;14:33–8. [https://doi.org/10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5).
44. Adewumi AT, Oluyemi WM, Adekunle YA, Adewumi N, Alahmdi MI, Soliman MES, et al. Propitious Indazole Compounds as β -ketoacyl-ACP Synthase Inhibitors and Mechanisms Unfolded for TB Cure: Integrated Rational Design and MD Simulations. *ChemistrySelect* 2023;8. <https://doi.org/10.1002/slct.202203877>.
45. Kramer O. Scikit-Learn, 2016, p. 45–53. https://doi.org/10.1007/978-3-319-33383-0_5.
46. Kennedy ME, Stamford AW, Chen X, Cox K, Cumming JN, Dockendorf MF, et al. The BACE1 inhibitor verubecestat (MK-8931) reduces CNS β -Amyloid in animal models and in Alzheimer's disease patients. *Sci Transl Med* 2016;8. <https://doi.org/10.1126/SCITRANSLMED.AAD9704;CTYPE:STRING:JOURNAL>.
47. Kufareva I, Abagyan R. Methods of protein structure comparison. *Methods in Molecular Biology* 2012;857:231–57. https://doi.org/10.1007/978-1-61779-588-6_10.
48. Gueto-Tettay C, Zuchniarz J, Fortich-Seca Y, Gueto-Tettay LR, Drosos-Ramirez JC. A molecular dynamics study of the BACE1 conformational change from Apo to closed form induced by hydroxyethylamine derived compounds. *J Mol Graph Model* 2016;70:181–95. <https://doi.org/10.1016/J.JMGM.2016.10.006>.
49. Alruwaili M, Alhassan HH, Almutary H, Tahir ul Qamar M. Computational identification of aspartic protease inhibitors for antimalarial drug development against Plasmodium Vivax. *Scientific Reports* 2025 15:1 2025;15:1–19. <https://doi.org/10.1038/s41598-025-98516-9>.
50. Lobanov MY, Bogatyreva NS, Galzitskaya O V. Radius of gyration as an indicator of protein structure compactness. *Mol Biol* 2008;42:623–8. <https://doi.org/10.1134/S0026893308040195>.
51. Cossio-Pérez R, Palma J, Pierdominici-Sottile G. Consistent Principal Component Modes from Molecular Dynamics Simulations of Proteins. *J Chem Inf Model* 2017;57:826–34. <https://doi.org/10.1021/acs.jcim.6b00646>.
52. Huang H, Simmerling C. Fast Pairwise Approximation of Solvent Accessible Surface Area for Implicit Solvent Simulations of Proteins on CPUs and GPUs. *J Chem Theory Comput* 2018;14:5797–814. <https://doi.org/10.1021/ACS.JCTC.8B00413>.
53. Tang QY, Kaneko K. Long-range correlation in protein dynamics: Confirmation by structural data and normal mode analysis. *PLoS Comput Biol* 2020;16:e1007670. <https://doi.org/10.1371/JOURNAL.PCBI.1007670>.
54. Collier G, Ortiz V. Emerging computational approaches for the study of protein allostery. *Arch Biochem Biophys* 2013;538:6–15. <https://doi.org/10.1016/J.ABB.2013.07.025>.
55. Homeyer N, Gohlke H. Free energy calculations by the Molecular Mechanics Poisson-Boltzmann Surface Area method. *Mol Inform* 2012;31:114–22. <https://doi.org/10.1002/MINF.201100135;JOURNAL:JOURNAL:15213838>.
56. Wang E, Sun H, Wang J, Wang Z, Liu H, Zhang JZH, et al. End-Point Binding Free Energy Calculation with MM/PBSA and MM/GBSA: Strategies and Applications in Drug Design. *Chem Rev* 2019;119:9478–508. <https://doi.org/10.1021/ACS.CHEMREV.9B00055>.
57. Schönherr H, Cernak T. Profound methyl effects in drug discovery and a call for new C-H methylation reactions. *Angew Chem Int Ed* 2013;52:12256–67. <https://doi.org/10.1002/ANIE.201303207;PAGEGROUP:STRING:PUBLICATION>.

58. McConnell DB. Biotin's Lessons in Drug Design. *J Med Chem* 2021;64:16319–27. <https://doi.org/10.1021/ACS.JMEDCHEM.1C00975>.
59. Kaur N, Gupta S, Pal J, Bansal Y, Bansal G. Design of BBB permeable BACE-1 inhibitor as potential drug candidate for Alzheimer disease: 2D-QSAR, molecular docking, ADMET, molecular dynamics, MMGBSA. *Comput Biol Chem* 2025;116:108371. <https://doi.org/10.1016/J.COMPBIOLCHEM.2025.108371>.
60. Manoharan P, Ghoshal N. Fragment-based virtual screening approach and molecular dynamics simulation studies for identification of BACE1 inhibitor leads. *J Biomol Struct Dyn* 2018;36:1878–92. <https://doi.org/10.1080/07391102.2017.1337590>;REQUESTEDJOURNAL:JOURNAL:TBSD20;WGROU:STRING:PUBLICATION.
61. Cournia Z, Allen B, Sherman W. Relative Binding Free Energy Calculations in Drug Discovery: Recent Advances and Practical Considerations. *J Chem Inf Model* 2017;57:2911–37. <https://doi.org/10.1021/ACS.JCIM.7B00564>.
62. Hong L, Tang J. Flap Position of Free Memapsin 2 (β -Secretase), a Model for Flap Opening in Aspartic Protease Catalysis†. *Biochemistry* 2004;43:4689–95. <https://doi.org/10.1021/BI0498252>.
63. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* 1997;23:3–25. [https://doi.org/10.1016/S0169-409X\(96\)00423-1](https://doi.org/10.1016/S0169-409X(96)00423-1).
64. Degoey DA, Chen HJ, Cox PB, Wendt MD. Beyond the Rule of 5: Lessons Learned from AbbVie's Drugs and Compound Collection. *J Med Chem* 2017;61:2636–51. <https://doi.org/10.1021/ACS.JMEDCHEM.7B00717>.
65. Cummings J, Lee G, Ritter A, Sabbagh M, Zhong K. Alzheimer's disease drug development pipeline: 2019. *Alzheimer's and Dementia: Translational Research and Clinical Interventions* 2019;5:272–93. <https://doi.org/10.1016/J.TRCL.2019.05.008>;ISSUE:ISSUE:DOI.
66. Puris E, Fricker G, Gynther M. Targeting Transporters for Drug Delivery to the Brain: Can We Do Better? *Pharmaceutical Research* 2022 39:7 2022;39:1415–55. <https://doi.org/10.1007/S11095-022-03241-X>.
67. Kratz F, Müller IA, Ryppa C, Warnecke A. Prodrug strategies in anticancer chemotherapy. *ChemMedChem* 2008;3:20–53. <https://doi.org/10.1002/CMDC.200700159>;JOURNAL:JOURNAL:18607187;WEBSITE:WEBSITE:CHEMISTRY-EUROPE;WGROU:STRING:PUBLICATION.
68. Nady DS, Bakowsky U, Fahmy SA. Recent advances in brain delivery of synthetic and natural nano therapeutics: Reviving hope for Alzheimer's disease patients. *J Drug Deliv Sci Technol* 2023;89:105047. <https://doi.org/10.1016/J.JDDST.2023.105047>.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.