Journal of
Food and Pharmaceutical Sciences

Original Article

A Structure-Based Virtual Screening Protocol Utilizing PyPLIF HIPPOS and Vina for Targeting BACE-1

Mala Hikmawan Primana[1], Mitsue Oka[1], Enade Perdana Istyastono[2], Florentinus Dika Octa Riswanto[1,3]*

[1]Magister of Pharmacy Study Program, Faculty of Pharmacy, Sanata Dharma University, Yogyakarta, Indonesia
[2]Department of Pharmacy, School of Medicine and Health Sciences, Atma Jaya Catholic University of Indonesia, DKI Jakarta, Indonesia
[3]Research Group of Computer-Aided Drug Design and Discovery of Bioactive Natural Products, Sanata Dharma University, Yogyakarta, Indonesia
*Corresponding author: Florentinus Dika Octa Riswanto | Email: dikaocta@usd.ac.id

Received: 22 September 2025; Revised: 17 November 2025; Accepted: 4 December 2025; Published: 20 December 2025

Abstract: Alzheimer's Disease (AD) is a leading cause of dementia, characterized by progressive cognitive decline driven in part by amyloid-β (Aβ) accumulation. The β-site amyloid precursor protein cleaving enzyme 1 (BACE-1) is responsible for initiating Aβ generation, making it a central therapeutic target. Yet, developing effective BACE-1 inhibitors has proven difficult due to structural complexity and pharmacological limitation. This study aimed to construct and validate a structure-based virtual screening (SBVS) workflow combining Vina and PyPLIF HIPPOS to facilitate the identification of promising BACE-1 ligands. The protocol was validated through 100 independent redocking experiments of the native ligand (PDB ID: 3L5F), all reproducing the crystallographic pose with RMSD < 2.0 Å. Large-scale screening of the DUDE dataset (283 active ligands; 18,100 decoys) generated interaction fingerprint, which were subsequently analyzed using Recursive Partitioning and Regression Trees (RPART) under varying prior probabilities. At the optimal prior ratio of 0.82:0.18, the model achieved an enrichment factor (EF) of 10.03, surpassing the DUDE benchmark (EF = 8.1). analysis consistently highlighted ionic interactions with Asp289 and hydrophobic contacts with Trp137 as key determinants of ligand activity. From 283 active ligands, 32 were classified as true positives, narrowing the pool of candidates and interpretable SBVS protocol and proposing a dual anchoring strategy involving Asp289 and Trp137 as a rational design principle for novel BACE-1 inhibitors in AD therapy.

Keywords: azheimer; BACE-1; RPART; SBVS; vina

1. INTRODUCTION

Alzheimer's Disease (AD) is a chronic neurodegenerative disorder that progresses gradually and irreversibly, making it the most prevalent cause of dementia in older adults [1]. Patients with AD typically experience progressive deterioration of memory, reasoning and other cognitive domains, which eventually disrupts daily life [2-3]. Globally, the incidence of dementia continues to increase, with epidemiological reports estimating that a new case arises every three seconds [3]. In 2019, the number of individuals living with AD reached approximately 55 million, and this figure is projected to escalate around 139 million by 2050 [2-4].

A central hallmark of AD is the deposition of amyloid-β (Aβ) plaques in brain tissue [5]. The production of Aβ begins with cleavage of the amyloid precursor protein (APP) by β-site cleaving enzyme 1 (BACE-1) [6]. Because this enzymatic step triggers the amyloidogenic pathway, BACE-1 has been widely investigated as a therapeutic target [5,7]. Despite extensive efforts, however, developing inhibitors that combine potency, selectivity and favorable pharmacokinetics remain a

formidable challenge [8]. Traditional drug discovery approaches are limited by high costs, lengthy development cycles and low success rates, creating demand for more efficient alternatives [9-11].

Structure-based virtual screening (SBVS) [10,12] provides a computational route for exploring large chemical libraries against the three-dimensional (3D) structure of target proteins [12]. By integrating molecular docking with scoring functions, SBVS helps identify molecules likely to bind with high affinity [10]. Recent advances further enhance this process by incorporating artificial intelligence approaches such as machine learning and deep learning [13].

Among widely used docking platform, Vina [14] is notable for its balance of speed, open accessibility and reasonable accuracy [15]. Its scoring function estimates the affinity of candidate ligands, thereby assisting researchers in prioritizing molecules for further investigation. This ease of use, open-source availability and speed make Vina a go-to choice for drug discovery and academic research around the world [14-15].

However, docking scores alone are insufficient for fully understanding ligand-protein recognition. PyPLIF HIPPOS [16] complements docking by generating ligand-protein interaction fingerprints, which encode residue-level contacts into digital representations for comparison or integration into classification models [16-18]. PyPLIF HIPPOS builds upon its predecessor, PyPLIF, but it is noticeably faster, more flexible and better suited to modern docking software like Vina, making it highly useful for SBVS projects [13].

The reliability of any SBVS workflow requires retrospective validation. The Directory of Useful Decoys Enhanced (DUDE) dataset is frequently used for this purpose, as it provides matched sets of active and inactive molecules with similar physicochemical profiles [19]. In this study, Vina and PyPLIF HIPPOS were combined to construct an SBVS workflow targeting BACE-1. Interaction fingerprints were subsequently analyzed using Recursive Partitioning and Regression Trees (RPART) [20], and model performance was assessed with enrichment factor (EF), F1-score and balanced accuracy (BA) [21]. The goal was to establish a validated, reproducible and interpretable computational pipeline that also offers insights to guide rational BACE-1 inhibitor design for AD treatment.

2. MATERIALS AND METHODS

2.1. Materials

2.1.1. Equipment

The computational experiments were carried out on two main instruments: (i) PC-Client: an Asus Vivo Book X415JA laptop equipped with an Intel® CoreTM i3-1005G1 CPU (1.20 GHz), 12 GB RAM, running both Windows 11 Home Single Language 64-bit and Ubuntu 24.04 LTS operating systems. The main software employed in this machine included PyPLIF HIPPOS version 0.2.0, Vina version 1.2.7, Yasara-Structure version 24.4.10, ADFR Suite version 1.0, R software version 4.3.3 with the RPART package, and (ii) Cloud Protein Simulator (CPS): A Google Cloud Platform (Instance ID: 3272493868356570273) with 96 vCPU and 48 GB RAM with Ubuntu 24.04 LTS operating systems. The main software employed in this machine included PyPLIF HIPPOS version 0.2.0 and Vina version 1.2.7. The CPS was provided by MOLMOD ID (https://molmod.id/; accessed August 12–26, 2025).

2.1.2. Datasets

The ligand dataset was obtained from DUDE repository (accessed on July 2nd, 2025) [19]. For BACE-1 target, DUDE provides a curated collection of 283 active ligands along with 18.100 pre-generated decoys. All active ligands included in DUDE have undergone internal curation involving structure standardization, removal of duplicates, verification of valency and confirmation of physicochemical consistency. Decoys in DUDE are automatically generated to match the physicochemical properties of active ligands (such as molecular weight and number of rotatable

bonds) while maintaining topological dissimilarity. No additional filtering criteria were applied in this study beyond DUDE's curation pipeline, and the ligand sets were used as provided. All ligands were initially available in SMILES (.smi) format and were subsequently converted into 3D structures using Yasara-Structure before preparation into pdbqt format for docking simulations.

2.1.3. Protein Target

The crystallographic structure of human BACE-1 in complex with its native ligand (PDB ID: 3L5F) was retrieved from the RCSB Protein Data Bank (https://www.rcsb.org; accessed on March 15th, 2025). The main native ligand of the 3L5F complex is a compound that identified as (2E,5R)-5-(2-cyclohexylethyl)-5-(cyclohexylmethyl)-2-imino-3-methylimidazolidin-4-one (BDX), which was used as reference for molecular docking validation. The selection of PDB ID 3L5F was driven by methodological consistency with the DUDE dataset and by the need for reliable receptor validation. Although DUDE lists PDB ID 3L5D (native ligand BDV) as the reference structure for BACE-1, preliminary redocking validation using 3L5D failed to reproduce the crystallographic pose, with all 100 redocking attempts yielding RMSD values above 2 Å. Because accurate pose reproduction is essential for ensuring a valid docking protocol, an alternative structure with the same residue composition and binding-site arrangement was required. PDB ID 3L5F fulfills these criteria and contains a closely related native ligand (BDX). Redocking of BDX into 3L5F successfully reproduced the experimental binding pose in all 100 iterations, with RMSD values consistently below 2 Å. Therefore, 3L5F was selected as the protein target, as it aligns with DUDE's structural framework while ensuring valid docking reproducibility.

2.2. Methods

2.2.1. Target Preparation and Validation

Target preparation was carried out using Yasara-Structure, where monomer A from the crystal complex was retained as the working model. Missing amino acid residues were reconstructed, water molecules were eliminated, and the system was adjusted to pH 7.4 before energy minimization. Finally, both the protein and its native ligand BDX were converted into pdbqt files through ADFR Suite macros available in Yasara-Structure.

Validation of the docking setup was performed through 100 independent redocking runs of the native ligand against the BACE-1 structure (PDB ID: 3L5F) using Vina 1.2.7 implemented in Yasara-Structure. The resulting poses were exported in pdbqt format and their RMSD values obtained from "rmsd_bestpose_all.txt". The protocol was considered acceptable when over 95% of the redocked conformations reproduced the crystallographic binding mode with RMSD values below 2.0 Å [19].

2.2.2. Ligand Preparation

Ligand preparation followed a standardized protocol. The DUDE ligands were converted from SMILES into 3D pdb format with Yasara-Structure macros, followed by energy minimization. They were subsequently converted to pdbqt files using ADFR Suite tools to ensure compatibility with Vina. As DUDE ligands had previously undergone internal filtering and structural validation, no further manual screening was performed in this study. All ligands were used exactly as curated by the DUDE repository.

2.2.3. Molecular Docking Simulations

Docking simulations of both active ligands and decoys were performed using Vina 1.2.7 with a protocol adapted from Istyastono (2024) [13] and further optimized specifically for the BACE-1 binding pocket. The grid box was centered on the crystallographic pose of the native ligand BDX in the 3L5F structure, with coordinates set at x = −7.614, y = 6.039, and z = −5.641, and grid dimensions

of 26.194 × 26.194 × 26.194 Å to fully cover the catalytic dyad. In contrast to the 100 redocking repetitions used solely for receptor validation, each ligand in the SBVS dataset (283 active ligands and 18,100 decoys) underwent five independent docking runs to minimize stochastic variation inherent to Vina. For each ligand, the best-scoring pose (lowest binding free energy) among the five runs was selected for subsequent processing. All docking jobs were executed automatically using the Cloud Protein Simulator (CPS) platform to ensure consistent parameterization and full reproducibility across the dataset. The resulting docking poses were then analyzed using PyPLIF HIPPOS to generate interaction fingerprints for retrospective model development and validation.

2.2.4. Interaction Fingerprinting

Top-ranked docking poses obtained from Vina were analyzed with PyPLIF HIPPOS to identify protein-ligand interactions. Each contact was mapped to BACE-1 residues and translated into binary fingerprints vectors. The compiled interaction dataset, combining fingerprints with affinity values, served as input for classification analysis.

2.2.5. Classification and Validation

Interaction fingerprints and binding affinities were used as predictors in a decision tree classification model constructed with the RPART package in R version 4.3.3 [20]. Several prior probability ratios were tested to optimize model performance [13,21]. The classification was evaluated using three metrics: enrichment factor (EF) to measure early enrichment of active ligands [19], F1-Score to balance precision and recall and balanced accuracy (BA) to account for dataset imbalance. The decision tree model that yielded the highest EF was selected as the optimal classification scheme.

3. RESULTS AND DISCUSSION

3.1. Results

The docking protocol against BACE-1 (PDB ID: 3L5F) was first validated through redocking simulations of the native ligand BDX. Out of 100 independent trials, all reproduced the crystallographic binding pose with RMSD values below 2.0 Å, surpassing the comon validation threshold of 95% reproducibility [14,22]. This outcome confirmed the robustness of the docking setup and provided a reliable foundation for large-scale virtual screening.

Following validation, the DUDE dataset consisting of 283 actives and 18,100 decoys was screened. Interaction profiles were subsequently analyzed using PyPLIF HIPPOS. Active ligands displayed consistent molecular fingerprints, most notably ionic interactions with Asp289 and hydrophobic contacts with Trp137 (Figure 1). The identification of Asp289 and Trp137 as key residues was not derived from experimental mutagenesis data but emerged directly from the interaction fingerprint analysis incorporated into the RPART decision-tree model. During classification, RPART evaluates which interaction features represented as binary contact vectors for each residue, most effectively separate active ligands from decoys. Across multiple splits, interactions involving Asp289 (ionic contacts) and Trp137 (hydrophobic contacts) consistently contributed to branches enriched with active ligands, indicating that these residues carry strong discriminative power within the dataset. Therefore, their importance in this study reflects the statistical learning process embedded in RPART rather than prior experimental knowledge. These interactions were rarely observed among decoys, suggesting their importance as structural determinants of activity (Table 1).

Table 1. Key Residues, Interaction Types, and Potential Structural Motifs for Ligand Optimization

Residue	Type of Interaction	Potential Enhancing Moieties of Molecular Design
Asp289	Ionic as the anion	Positively charged groups such as amines or guanidines
Trp137	Hydrophobic	Aromatic fragments, hydrophobic ring systems

Notes: This table summarizes the amino acid residues that play a critical role in ligand binding, the type of molecular interactions involved, and the potential molecular fragments that could enhance inhibitor design. The results were derived from molecular interaction analysis based on docking simulations.

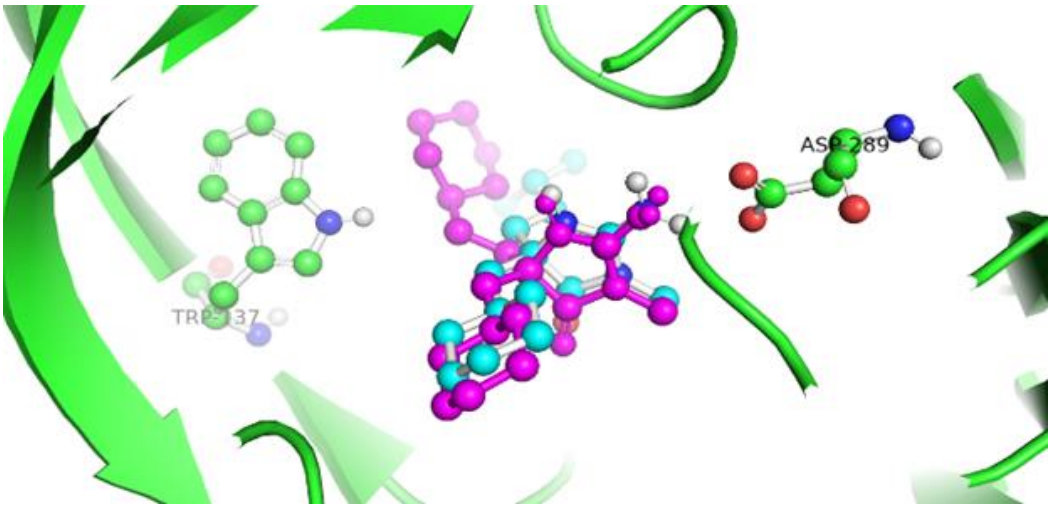

Figure 1. Visualization of the interaction of native ligand BDX in 3D

Notes: The color differences in this figure are used to distinguish the crystallographic pose from the redocked pose and to visualize the surrounding binding-site residues. The native BDX ligand from the crystal structure is shown using the default element-based coloring scheme (carbon = cyan, nitrogen = blue, oxygen = red), whereas the redocked BDX pose is displayed in solid purple to clearly differentiate it from the original crystallographic orientation. All amino-acid residues lining the binding pocket are rendered in solid green for uniformity and to highlight the ligand–protein interface without introducing additional color complexity.

Importantly, from the initial pool of 283 active ligands, the workflow successfully prioritized 32 compounds as active hits. These ligands, detailed in Supplementary Material 1, represent a refined subset with the most promising interaction profiles against BACE-1. By narrowing down the chemical space, this dataset provides a valuable foundation for subsequent in vitro and in vivo investigations. Moreover, these compounds offer strategic starting points for the rational design of next-generation BACE-1 inhibitors.

Classification analysis with the RPART algorithm yielded optimal performance when the prior probability ratio was set at 0.82:0.18. In the context of RPART classification, the prior probability ratio represents the assumed baseline likelihood of a ligand belonging to each class before the model evaluates the interaction features. Because the DUDE dataset for BACE-1 is highly imbalanced (283 active ligands vs. 18.100 decoys), assigning a prior of 0.82 for actives and 0.18 for decoys counteracts the natural dominance of the decoy class. This adjustment forces the decision tree to give greater weight to interaction patterns associated with actives, thereby improving sensitivity toward true actives and preventing the model from being biased toward predicting the majority class [18]. Under this condition, EF reached 10.03, exceeding the DUDE benchmark EF value of 8.1 [19]. At this optimal setting, the model successfully identified 32 true positives (TP), while misclassifying 251 actives as false negatives (FN), 17.896 decoys as true negatives (TN) and 204 as false positives (Table 3).

Table 2. Performance Metrics of the Classification Model under Different Prior Ratios

Prior Ratio	Data (Unit)				F1-Score	EF	BA
	TP	FN	TN	FP			
0.1:0.9	283	0	5.985	12.115	0.044634	1.494016	0.665331
0.2:0.8	279	4	7.631	10.469	0.050585	1.704477	0.703734
0.3:0.7	275	8	7.593	10.507	0.049706	1.673964	0.695617
0.4:0.6	231	52	11.725	6.375	0.067063	2.317522	0.732022
0.5:0.5	227	56	11.843	6.257	0.06709	2.320341	0.728215
0.6:0.4	179	104	14.501	3,599	0.088156	3.180997	0.716835
0.7:0.3	82	201	17.165	935	0.126154	5.609115	0.619048
0.8:0.2	76	207	17.489	611	0.156701	7.955446	0.617397
0.81:0.19	76	207	17.532	568	0.16397	8.557707	0.618585
0.82:0.18	**32**	**251**	**17.896**	**204**	**0.123314**	**10.03256**	**0.550902**

Notes: This table displays the performance of the classification model under different prior ratio settings, indicated by the values of True Positive (TP), False Negative (FN), True Negative (TN), False Positive (FP), F1-score, Enrichment Factor (EF), and Balanced Accuracy (BA). These results were used to determine the most reliable prior ratio for predicting ligand activity toward the target protein. The best performance of the classification model was marked with bold.

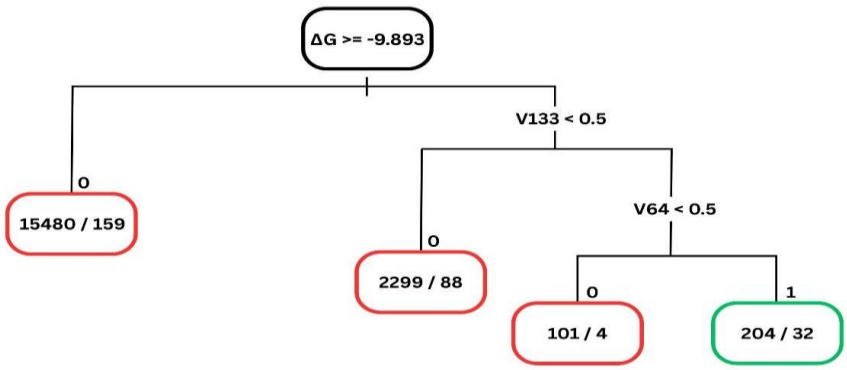

Figure 2. The DUDE optimized decision tree for the SBVS protocol targeting BACE-1

The decision tree (Figure 2) placed the affinity cutoff at ΔG = -9.893 kcal/mol as the first split. Node distributions were represented as active/decoy counts, such as 15,480/159 (decoy dominated), 2,299/88 (higher active proportion), 101/4 (almost exclusively decoys) and 204/32 (an enriched branch of actives). The secondary splits of the decision tree involved interaction fingerprints such as V133 < 0.5 and V64 < 0.5, which indicated the absence of ligand contacts with critical residues. This observation is consistent with the interaction analysis summarized in Table 1, where Asp289 and Trp137 were identified as essential anchoring points for ligand recognition. In other words, when these key contacts were missing, the classification model strongly favored a decoy outcome, reinforcing the central role of ionic interactions with Asp289 and hydrophobic stabilization via Trp137 in driving ligand activity.

3.2. Discussion

The validation step confirmed that the proposed docking protocol was technically sound, as demonstrated by its consistent ability to reproduce the crystallographic pose of BDX with RMSD values below 2.0 Å [14,22]. Beyond validation, the improved EF compared to DUDE benchmark indicates the superior performance of this protocol in prioritizing active compounds at early stages of screening (Table 2) [19]. Such improvement is critical in practical applications, where reducing the number of false leads can significantly lower experimental costs and accelerate drug discovery.

Analysis of the interaction fingerprints revealed key residues responsible for ligand recognition. Among them, Asp289 consistently engaged in ionic contacts with basic groups such as amines,

amidines and guanidinium, functioning as a principal electrostatic anchor [23]. Simultaneously, Trp137 contributed hydrophobic stabilization, often through interactions with aromatic or halogenated substituents (-F, -Cl, -CF_3, -OCF_3) (Table 1). These noncovalent forces are well established as central to protein-ligand recognition [24]. Together, the Asp289-Trp137 dual anchoring motif explained the preferential binding of active ligands compared to decoys. In this study, the prominence of Asp289 and Trp137 arose from the retrospective learning behavior of the RPART classifier. Each residue-ligand interaction detected by PyPLIF HIPPOS serves as an input feature to the decision tree, which then selects the most informative features to split active ligands and decoy populations. Asp289 and Trp137 appeared repeatedly as high-information-gain features, meaning that the presence of interactions with these residues statistically increased the probability of a ligand belonging to the active class. This data-driven emergence highlights that the dual anchoring role of Asp289 and Trp137 is an outcome of model-based interpretation, rather than assumptions derived from experimental mutagenesis studies [18].

From a screening perspective, the selection of 32 active hits out of 283 initial actives effectively narrowed the chemical search space to a more manageable subset (Supplementary Material 1). These compounds constitute promising candidates for further in vitro and in vivo studies, thereby enhancing the efficiency of the drug discovery process. The RPART-based classification model, particularly under the 0.82:0.18 prior probability, provided a clear and interpretable decision tree capable of distinguishing actives from decoys (Table 2, Figure 2). Such interpretability is advantageous for prospective applications, as it not only supports retrospective validation but also highlights practical thresholds for ligand prioritization.

Supplementary Material 1. Active Hits of BACE-1 Obtained from the DUDE Dataset after Virtual Screening

No	Active Ligand ID	Affinity (kcal/mol)	Active Ligand in SMILES Format
1	CHEMBL1082548	-9.895	C[C@@H](NC(=O)c2cc(COC(=O)C(C)(N)Cc1ccccc1)cc(N(C)S(C)(=O)=O)c2)c3ccc(F)cc3
2	CHEMBL1092788	-10.468	CN4C(N)=N[C@](c1ccncc1)(c3cccc(c2cc(F)ccc2F)c3)C4=O
3	CHEMBL1096683	-10.354	C[C@@H](NC(=O)c3cc(N(C)S(C)(=O)=O)cc(c2ncc([C@](C)(N)Cc1ccccc1)o2)c3)c4ccc(F)cc4
4	CHEMBL1209222	-9.979	CC(=O)N[C@@H](Cc1cc(F)cc(F)c1)[C@H](O)CNC4(c3cccc(N2CCCCC2=O)c3)CCCCC4
5	CHEMBL219601	-10.434	C[C@@H](NC(=O)c3cc(N(C)S(C)(=O)=O)cc(c2nnc([C@](C)(N)Cc1ccccc1)o2)c3)c4ccc(F)cc4
6	CHEMBL236851	-10.199	CC(=O)c4ccc(Oc3ccc(c2scc(c1ccccc1Cl)c2CC(=O)N=C(N)N)cc3)cc4
7	CHEMBL244346	-10.032	CN(C1CCCCC1)C(=O)c5cccc(CN4Cc3cc(Oc2ccccc2)ccc3N=C4N)c5
8	CHEMBL252189	-10.427	NC(=N)NC(=O)Cn2c(c1ccccc1)ccc2c4ccc(NC(=O)c3cccc(Br)c3)cc4
9	CHEMBL253237	-10.341	CC(=O)c4ccc(Oc3ccc(c2ccc(c1ccccc1Cl)n2CC(=O)NC(N)=N)cc3)cc4
10	CHEMBL272766	-10.672	Cc6nnc(CNC(\N)=N/C(=O)Cn2c(c1ccccc1)ccc2C45CC3CC(CC(C3)C4)C5)o6
11	CHEMBL375217	-10.122	C[C@@H](NC(=O)c2cc(N(C)S(C)(=O)=O)cc(C(=O)OCC(N)(CO)Cc1ccccc1)c2)c3ccc(F)cc3
12	CHEMBL392662	-9.963	COc3cccc(c2cccc(C1(C)CC(=O)N(C)C(N)=N1)c2)c3

13	CHEMBL404682	-10.001	O[C@H]([C@H]2C[C@@H](OCc1ccccc1)CN2)[C@H](Cc3cc(F)cc(F)c3)NC(=O)[C@@H]5CN(Cc4ccco4)C(=O)C5
14	CHEMBL491659	-10.026	O[C@@H]([C@H](Cc1cc(F)cc(F)c1)NC(=O)C3CN(CC2CC2)C(=O)C3)[C@@H]5NCCN(Cc4ccccc4)C5=O
15	CHEMBL562555	-10.015	CCN3c4cc(C(=O)N[C@H](Cc1ccccc1)[C@@H](O)CNC2CCOCC2)cc5c(CC)cn(CCS3(=O)=O)c45
16	CHEMBL565790	-11.391	COc4ccc(C3(c2cccc(c1cccnc1)c2)N=C(N)N(C)C3=O)cc4C5CCCC5
17	CHEMBL565914	-10.576	NC4=NC(c1ccc(OC(F)(F)F)cc1)(c3cccc(c2cccnc2F)c3)C5=NCCCCN45
18	CHEMBL566112	-10.596	NC4=NC(c1ccc(OC(F)(F)F)cc1)(c3cccc(c2cccnc2F)c3)C5=NCCN45
19	CHEMBL566414	-10.37	CN5C(N)=NC(c2cccc(c1cccnc1)c2)(c4ccc3OCCc3c4)C5=O
20	CHEMBL566603	-10.447	CC(=O)N[C@@H](Cc1cc(F)cc(F)c1)[C@H](O)CN[C@@]3(c2cccc(C(C)(C)C)c2)CC\C(=N\O)NC3
21	CHEMBL566638	-11.132	COc4ccc(C3(c2cccc(c1cccnc1)c2)N=C(N)N(C)C3=O)cc4OC5CCCC5
22	CHEMBL566968	-10.317	NC5=NC(c2cccc(c1cccnc1F)c2)(c4ccc3OCCOc3c4)C6=NCCCN56
23	CHEMBL567430	-10.74	NC4=NC(c1ccc(OC(F)(F)F)cc1)(c3cccc(c2cccnc2F)c3)C5=NOCCN45
24	CHEMBL567477	-10.264	COc4ccc([C@@]3(c2ccc(F)c(c1cccnc1F)c2)N=C(N)N(C)C3=O)cc4C
25	CHEMBL568960	-10.988	NC4=NC(c1ccc(OC(F)(F)F)cc1)(c3cccc(c2cccnc2F)c3)C5=NCCCCCN45
26	CHEMBL582045	-9.966	NC4=N[C@](c1ccc(OC(F)(F)F)cc1)(c3cccc(c2cncnc2)c3)C5=NCCCN45
27	CHEMBL582982	-10.247	COC4=C(F)C3CCC(Cc1ccc(C(F)(F)F)cc1)(c2cn(C)c(N)n2)C3C=C4
28	CHEMBL583034	-10.155	CC(=O)N[C@@H](Cc1cc(F)cc(F)c1)[C@H](O)CN[C@@]3(c2cccc(C(C)(C)C)c2)CC[C@H](NO)NC3
29	CHEMBL583608	-10.16	NC5=NC(c2cccc(c1cccnc1F)c2)(c4ccc3OCOc3c4)C6=NCCCN56
30	CHEMBL589891	-10.085	NC3=NC(C1CCCCC1)(c2ccccc2)C(=O)N3CC4CCC(C(O)=O)CC4
31	CHEMBL595066	-10.746	COCc3cc4NCCCCOc5cccc(C[C@@H]([C@H](O)CNC2(c1cccc(C(C)C)c1)CC2)NC(=O)c(c3)c4)c5
32	CHEMBL598290	-10.159	CCc4cc([C@]3(c2cccc(c1cccnc1F)c2)N=C(N)N(C)C3=O)cc(CC)n4

To ensure the stability of docking outcomes within the large-scale SBVS workflow, each ligand was docked five times, and the best-scoring pose was selected for analysis. This multi-run strategy reduces the stochastic variability inherent to Vina while maintaining computational feasibility when processing thousands of ligands. Such an approach is commonly employed in high-throughput screening studies and contributes to the consistency and robustness of the retrospective enrichment results obtained in this work.

Collectively, the findings suggest a rational design strategy for future BACE-1 inhibitors. Specifically, compounds may benefit from a dual anchoring approach: (i) incorporating positively

charged groups to strengthen electrostatic interactions with Asp289, and (ii) embedding aromatic or hydrophobic scaffolds, potentially halogenated to enhance binding through Trp137. This strategy aligns with established pharmacophoric principles, emphasizing the synergy between electrostatic and hydrophobic interactions. Accordingly, the present protocol not only validates a computational screening workflow but also provides a structural framework for the rational design of next-generation BACE-1 inhibitors with potential therapeutic relevance to AD.

4. CONCLUSION

This study successfully established and validated a SBVS protocol that integrates Vina with PyPLIF HIPPOS. The workflow consistently reproduced the crystallographic pose of the native ligand, confirming its reliability for large-scale docking studies. Interaction fingerprinting revealed Asp289 and Trp137 as key residues that differentiate actives from decoys, thereby providing a clear mechanistic explanation for ligand recognition. By adjusting prior probabilities in the RPART model, the protocol achieved an EF superior to the DUDE benchmark, demonstrating improved efficiency in early identification of active compounds. Taken together, these findings highlight the dual anchoring role of Asp289 and Trp137 as a rational framework for the design of novel BACE-1 inhibitors. Beyond offering a validated computational pipeline, the study provides structural insights that may accelerate the discovery of next-generation therapeutic candidates for AD.

Funding: DPPM program of the Government of the Republic of Indonesia
Acknowledgments: This research was financially supported by the DPPM program of the Government of the Republic of Indonesia with contract number 126/C3/DT.05.00/PL/2025 awarded to Dr. Florentinus Dika Octa Riswanto.
Conflicts of interest: The authors declare the following competing financial interest(s): E.P.I. is a co-founder of Molmod Jaya Sejahtera, Ltd. (MOLMOD ID), a contract research organization that provides fee-for-service cloud protein simulators and molecular modeling simulations. The authors declare no other conflicts.

References

[1] Q. Zheng and X. Wang, "Alzheimer's disease: insights into pathology, molecular mechanisms, and therapy," no. March 2024, pp. 83–120, 2025.
[2] A. Arfina, "Pengaruh Edukasi Terhadap Pengetahuan Masyarakat Tentang Deteksi Dini Alzheimer Di Kelurahan Labuh Baru Pekanbaru," *Heal. Care J. Kesehat.*, vol. 10, no. 01, pp. 256–261, 2021.
[3] S. Long, C. Benoist, and W. Weidner, "World Alzheimer Report 2023," p. 94, 2023.
[4] Z. Zhang, X. Liu, S. Zhang, Z. Song, K. Lu, and W. Yang, "A review and analysis of key biomarkers in Alzheimer's disease," *Front. Neurosci.*, vol. Volume 18, 2024, [Online]. Available: https://www.frontiersin.org/journals/neuroscience/articles/10.3389/fnins.2024.1358998
[5] H. Hampel *et al.*, "The Amyloid-β Pathway in Alzheimer's Disease," Oct. 01, 2021, *Springer Nature*. doi: 10.1038/s41380-021-01249-0.
[6] B. Das and R. Yan, "A Close Look at BACE1 Inhibitors for Alzheimer's Disease Treatment," *CNS Drugs*, vol. 33, no. 3, pp. 251–263, 2019, doi: 10.1007/s40263-019-00613-7.
[7] H. Hampel *et al.*, "The β-Secretase BACE1 in Alzheimer's Disease.," *Biol. Psychiatry*, vol. 89, no. 8, pp. 745–756, Apr. 2021, doi: 10.1016/j.biopsych.2020.02.001.
[8] K. Youn and M. Jun, "Biological evaluation and docking analysis of potent BACE1 inhibitors from boesenbergia rotunda," *Nutrients*, vol. 11, no. 3, pp. 1–13, 2019, doi: 10.3390/nu11030662.
[9] G. O. Mendes *et al.*, "Identification of Potential Multitarget Compounds against Alzheimer's Disease through Pharmacophore-Based Virtual Screening," *Pharmaceuticals*, vol. 16, 2023, doi: 10.3390/ph16121645.
[10] C. Gorgulla, "Recent Developments in Ultralarge and Structure-Based Virtual Screening Approaches," *Annu. Rev. Biomed. Data Sci.*, vol. 6, pp. 229–258, 2023, doi: 10.1146/annurev-

biodatasci-020222-025013.

[11] K. Fujimoto *et al.*, "Structure-Based Design of Selective β-Site Amyloid Precursor Protein Cleaving Enzyme 1 (BACE1) Inhibitors: Targeting the Flap to Gain Selectivity over BACE2," *J. Med. Chem.*, vol. 62, no. 10, pp. 5080–5095, 2019, doi: 10.1021/acs.jmedchem.9b00309.

[12] E. H. B. Maia, L. C. Assis, T. A. de Oliveira, A. M. da Silva, and A. G. Taranto, "Structure-Based Virtual Screening: From Classical to Artificial Intelligence," *Front. Chem.*, vol. 8, no. April, 2020, doi: 10.3389/fchem.2020.00343.

[13] E. P. Istyastono *et al.*, "PyPLIF HIPPOS-aided construction and retrospective validation of structure-based virtual screening protocol targeting VEGFR2," *Indones. J. Pharm.*, vol. 35, no. 3, pp. 451–458, 2024, doi: 10.22146/ijp.9820.

[14] J. Eberhardt, D. Santos-Martins, A. F. Tillack, and S. Forli, "AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings.," *J. Chem. Inf. Model.*, vol. 61, no. 8, pp. 3891–3898, Aug. 2021, doi: 10.1021/acs.jcim.1c00203.

[15] A. Sarkar, S. Concilio, L. Sessa, F. Marrafino, and S. Piotto, "Advancements and novel approaches in modified AutoDock Vina algorithms for enhanced molecular docking," *Results Chem.*, vol. 7, p. 101319, 2024, doi: https://doi.org/10.1016/j.rechem.2024.101319.

[16] E. P. Istyastono, M. Radifar, N. Yuniarti, V. D. Prasasty, and S. Mungkasi, "PyPLIF HIPPOS: A Molecular Interaction Fingerprinting Tool for Docking Results of AutoDock Vina and PLANTS," *J. Chem. Inf. Model.*, vol. 60, no. 8, pp. 3697–3702, 2020, doi: 10.1021/acs.jcim.0c00305.

[17] E. P. Istyastono, F. D. O. Riswanto, N. Yuniarti, V. D. Prasasty, and S. Mungkasi, "PyPLIF HIPPOS and Receptor Ensemble Docking Increase the Prediction Accuracy of the Structure-Based Virtual Screening Protocol Targeting Acetylcholinesterase," *Molecules*, vol. 27, no. 17, 2022, doi: 10.3390/molecules27175661.

[18] E. P. Istyastono, N. Yuniarti, V. D. Prasasty, and S. Mungkasi, "PyPLIF HIPPOS-assisted prediction of molecular determinants of ligand binding to receptors," *Molecules*, vol. 26, no. 9, pp. 1–12, 2021, doi: 10.3390/molecules26092452.

[19] M. M. Mysinger, M. Carchia, J. J. Irwin, and B. K. Shoichet, "Directory of useful decoys, enhanced (DUD-E): Better ligands and decoys for better benchmarking," *J. Med. Chem.*, vol. 55, no. 14, pp. 6582–6594, 2012, doi: 10.1021/jm300687e.

[20] B. R. Terry Therneau, Beth Atkinson, "Recursive Partitioning for Classification, Regretion and Survival Trees," 2025.

[21] E. O. Cannon *et al.*, "Support vector inductive logic programming outperforms the Naive Bayes Classifier and inductive logic programming for the classification of bioactive chemical compounds," *J. Comput. Aided. Mol. Des.*, vol. 21, no. 5, pp. 269–280, 2007, doi: 10.1007/s10822-007-9113-3.

[22] E. P. Istyastono, *Racikan Autodock Vina dan Yasara-Model dalam Secangkir Kopi: Langkah Praktis dan Gratis Penambatan Molekul pada Sistem Operasi Windows 11*. Yogyakarta: Sanata Dharma University Press, 2025.

[23] W.-A. RS and M. J. S. , Glen F. Rall, "乳鼠心肌提取 HHS Public Access," *J oncol transl res*, vol. 176, no. 1, pp. 139–148, 2018, doi: 10.1021/acs.chemrev.7b00305.Electrostatic.

[24] Y. Xiao and R. J. Woods, "Protein-Ligand CH−π Interactions: Structural Informatics, Energy Function Development, and Docking Implementation," *J. Chem. Theory Comput.*, vol. 19, no. 16, pp. 5503–5515, 2023, doi: 10.1021/acs.jctc.3c00300.