



OPEN ACCESS

EDITED BY

Liqun Zhang,
University of Rhode Island, United States

REVIEWED BY

Mariangela Agamennone,
University of Studies G. d'Annunzio Chieti and
Pescara, Italy
Son Tung Ngo,
Ton Duc Thang University, Vietnam

*CORRESPONDENCE

Andrés Fernando González Barrios,
✉ andgonza@uniandes.edu.co
Yasser Hayek-Orduz,
✉ y.hayek10@uniandes.edu.co
Dorian Armando Acevedo-Castro,
✉ da.acevedo10@uniandes.edu.co

†These authors have contributed equally to this work

RECEIVED 12 August 2024

ACCEPTED 06 January 2025

PUBLISHED 04 February 2025

CITATION

Hayek-Orduz Y, Acevedo-Castro DA,
Saldarriaga Escobar JS, Ortiz-Domínguez BE,
Villegas-Torres MF, Caicedo PA,
Barrera-Ocampo Á, Cortes N, Osorio EH and
González Barrios AF (2025) dyphAI dynamic
pharmacophore modeling with AI: a tool for
efficient screening of new
acetylcholinesterase inhibitors.
Front. Chem. 13:1479763.
doi: 10.3389/fchem.2025.1479763

COPYRIGHT

© 2025 Hayek-Orduz, Acevedo-Castro,
Saldarriaga Escobar, Ortiz-Domínguez,
Villegas-Torres, Caicedo, Barrera-Ocampo,
Cortes, Osorio and González Barrios. This is an
open-access article distributed under the terms
of the [Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in
other forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in this
journal is cited, in accordance with accepted
academic practice. No use, distribution or
reproduction is permitted which does not
comply with these terms.

dyphAI dynamic pharmacophore modeling with AI: a tool for efficient screening of new acetylcholinesterase inhibitors

Yasser Hayek-Orduz^{1*†}, Dorian Armando Acevedo-Castro^{1,2*†},
Juan Sebastián Saldarriaga Escobar³,
Brandon Eli Ortiz-Domínguez³, María Francisca Villegas-Torres⁴,
Paola A. Caicedo³, Álvaro Barrera-Ocampo⁵, Natalie Cortes⁶,
Edison H. Osorio⁶ and Andrés Fernando González Barrios^{1*}

¹Grupo de Diseño de Productos y Procesos (GDPP), Department of Chemical and Food Engineering, Universidad de los Andes, Bogotá, Colombia, ²Computational Bio-Organic Chemistry (COBO), Department of Chemistry, Universidad de los Andes, Bogotá, Colombia, ³Grupo Natura, Facultad de Ingeniería, Diseño y Ciencias Aplicadas, Departamento de Ciencias Biológicas, Bioprocesos y Biotecnología, Universidad ICESI, Cali, Colombia, ⁴Centro de Investigaciones Microbiológicas (CIMIC), Department of Biological Sciences, Universidad de los Andes, Bogotá, Colombia, ⁵Grupo Natura, Facultad de Ingeniería, Diseño y Ciencias Aplicadas, Departamento de Ciencias Farmacéuticas y Químicas, Universidad ICESI, Cali, Colombia, ⁶Grupo de Investigación en Química Bioorgánica y Sistemas Moleculares (QBOSMO), Faculty of Natural Sciences and Mathematics, Universidad de Ibagué, Ibagué, Colombia

Therapeutic strategies for Alzheimer's disease (AD) often involve inhibiting acetylcholinesterase (AChE), underscoring the need for novel inhibitors with high selectivity and minimal side effects. A detailed analysis of the protein-ligand pharmacophore dynamics can facilitate this. In this study, we developed and employed **dyphAI**, an innovative approach integrating machine learning models, ligand-based pharmacophore models, and complex-based pharmacophore models into a pharmacophore model ensemble. This ensemble captures key protein-ligand interactions, including π -cation interactions with Trp-86 and several π - π interactions with residues Tyr-341, Tyr-337, Tyr-124, and Tyr-72. The protocol identified 18 novel molecules from the ZINC database with binding energy values ranging from -62 to -115 kJ/mol, suggesting their strong potential as AChE inhibitors. To further validate the predictions, nine molecules were acquired and tested for their inhibitory activity against human AChE. Experimental results revealed that molecules, 4 (P-1894047), with its complex multi-ring structure and numerous hydrogen bond acceptors, and 7 (P-2652815), characterized by a flexible, polar framework with ten hydrogen bond donors and acceptors, exhibited IC_{50} values lower than or equal to that of the control (galantamine), indicating potent inhibitory activity. Similarly, molecules 5 (P-1205609), 6 (P-1206762), 8 (P-2026435), and 9 (P-533735) also demonstrated strong inhibition. In contrast, molecule 3 (P-617769798) showed a higher IC_{50} value, and molecules 1 (P-14421887) and 2 (P-25746649) yielded inconsistent results, likely due to solubility issues in the

experimental setup. These findings underscore the value of integrating computational predictions with experimental validation, enhancing the reliability of virtual screening in the discovery of potent enzyme inhibitors.

KEYWORDS

docking, MD simulations, pharmacophore, acetylcholinesterase, machine learning

1 Introduction

The global average life expectancy reached 73.4 years in 2019, with healthy life expectancy increasing by 8% due to reduced mortality rates (Mortality and global health estimates; Li et al., 2023). However, the length of time spent on health issues remained constant. This raises challenges to global health systems due to the prevalence of age-related diseases like Alzheimer's disease, which is the seventh leading cause of death globally and impacts over 416 million individuals (Prince et al., 2016; Anonymous, 2021a; Anonymous, 2021b; Gustavsson et al., 2023). Factors such as environmental conditions, genetics, age (Silva et al., 2019; Stepkowski et al., 2020), and others contribute to Alzheimer's disease (Breijyeh and Karaman, 2020; Scheltens et al., 2021), necessitating thorough research and efficient management to improve the wellbeing of the elderly population (Crestini et al., 2022; Whitfield et al., 2023; Delacourte, 2006; Jomova et al., 2023; Tolar et al., 2024; Vojtechova et al., 2022; Korczyn and Grinberg, 2024). Researchers have proposed three fundamental hypotheses to explain the origins of Alzheimer's disease (Leng and Edison, 2020; Wang et al., 2023), including the amyloid hypothesis, copper toxicity (Ejaz et al., 2020), and cholinergic dysfunction (Wilkinson et al., 2004; Camps and Munoz-Torrero, 2005; Giacobini et al., 2022). The final hypothesis suggests reduced levels of the neurotransmitter acetylcholine in brain areas linked to cognition contribute to Alzheimer's disease progression (Amenta et al., 2001; Carotenuto et al., 2022). Consequently, pharmacological interventions aimed at elevating acetylcholine (ACh) concentration represent a promising avenue for potential palliative treatment.

Acetylcholinesterase (AChE) is a hydrolase enzyme that breaks down ACh (Walczak-Nowicka and Herbet, 2021) through hydrolysis. Figure 1C illustrates this process (Komersová et al., 2005; Aroniadou-Anderjaska et al., 2023). Inhibiting AChE can maintain elevated levels of ACh by obstructing its degradation. Several compounds have demonstrated the ability to elevate ACh levels through AChE inhibition (Carotenuto et al., 2022; Akhoo et al., 2020; Marucci et al., 2021), indicating a potential avenue for enhancing Alzheimer's disease treatment. However, these compounds often cause side effects (Akhoo et al., 2020). Finding more selective and powerful inhibitors is crucial to reduce these side effects, which are likely caused by their non-specific interactions with other enzymes.

The human acetylcholinesterase (huAChE) protein consists of five domains: S1, S2, S3, S4, and the Ω -loop (Cheng et al., 2017; Cheung et al., 2012; Dvir et al., 2010). The active site of huAChE exhibits a gorge-like structure (Figure 1A), with approximate dimensions of 20 Å in height, 5 Å in width, and 5 Å in length (Moghadam et al., 2021). The active site has two subsites: the

peripheral anionic site (PAS) and the catalytic anionic site (CAS) (Figure 1B). The CAS has four areas: the acyl pocket, the catalytic triad, the oxyanion hole, and the anionic site (Wlodek et al., 1997). The PAS acts as a bridge for ligands to enter and leave the CAS. These areas are crucial for the acetylcholine hydrolysis reaction (Figures 1C, D) (Silman and Sussman, 2017).

Research on the steric and electronic properties of AChE inhibitors has been extensive (Wu et al., 2020; Zhou et al., 2017; Jang et al., 2018), but there is a lack (Sanson et al., 2011) of comprehensive study across all AChE inhibitor families (Akhoo et al., 2020; Gao et al., 2022; Pérez-Sánchez et al., 2021). While some studies have examined the conformational plasticity of protein-ligand interactions using advanced simulation methods (Lazim et al., 2020; Heilmann et al., 2020; Kaynak et al., 2022; Yasuda et al., 2020; Célerse et al., 2024), it is not common to use energetically unfavorable conformations to find protein inhibitors. Pharmacophore models (Langer et al., 2006) are widely used in drug discovery techniques (Qing et al., 2014; van Drie, 2012; Khedkar et al., 2007), but ensemble pharmacophore modeling can further harness their potential by combining multiple complex-based models. Artificial intelligence and machine learning models have gained significance in drug discovery (Carracedo-Reboredo et al., 2021; Dara et al., 2022; Li et al., 2024; Geng et al., 2024; Gupta et al., 2021), potentially playing a crucial role in virtual screening procedures.

Here, we proposed a workflow to find new acetylcholinesterase inhibitors by identifying key molecular features needed to target the huAChE enzyme specifically. This approach involved an extensive *in silico* protocol (illustrated in Figures 2A, B, Supplementary Figures S1–S6), that included database management, ligand clustering, RMSD calculations, induced-fit docking, molecular dynamics simulations, TRAPP physicochemical analyses (Kokh et al., 2013), ensemble docking, pharmacophore modeling, and machine learning techniques. The method involved looking for different structures in new inhibitors, exploring the space of receptor conformations, combining computer and lab results, using machine learning models to find new inhibitors, and choosing which inhibitors to test in experiments.

We identified nine distinct inhibitor families, each associated with a predictive machine learning model, a ligand-based pharmacophore model, and an ensemble pharmacophore model derived from ligand-receptor complexes. The models enabled virtual screening of the ZINC₂₂ database (Tingle et al., 2023), leading to the identification of 18 molecules with significant inhibitory potential against AChE.

The developed methodology and models, along with the insights gained, provide valuable contributions to the field and serve as a robust foundation for future research aimed at designing novel therapeutics for Alzheimer's disease and related conditions.

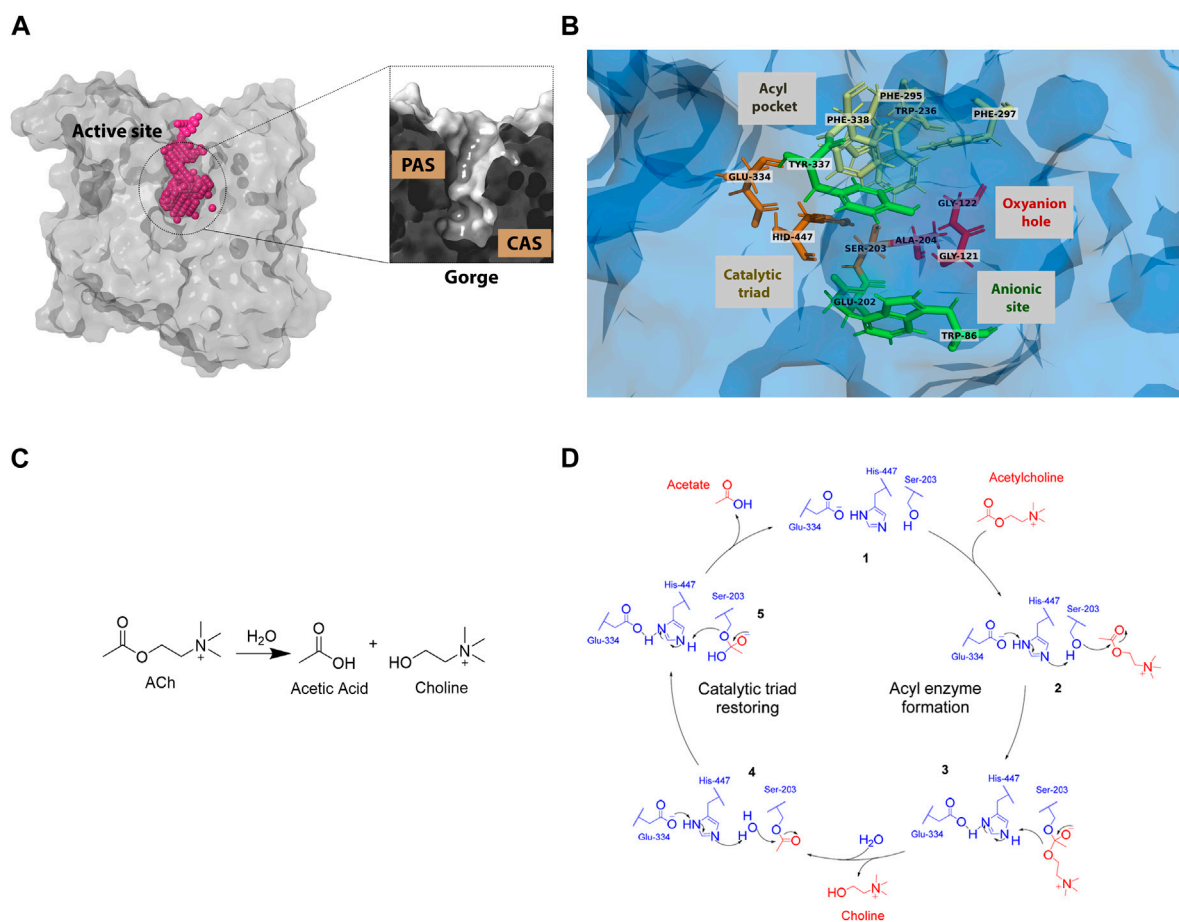


FIGURE 1
Human acetylcholinesterase's active site: **(A)** structure and surface **(B)** active site residues with subsites highlighted. (PDB entry: 4EY6) **(C)** Acetylcholine hydrolysis reaction **(D)** Catalytic cycle of the acetylcholinesterase enzyme.

2 Materials and methods

2.1 Yasser number one (YN1) affinity ranking

The affinity ranking of YN1 relies on the dynamic characteristics of the biological system, requiring the integration of computational and experimental data. A comprehensive workflow diagram illustrating this process is provided in **Supplementary Figure S3**.

2.1.1 Similarity clustering of inhibitor structures

Inhibitors with an $IC_{50} \leq 99,000$ nM against the huAChE were extracted from the Binding Database (BD) (<https://www.bindingdb.org/>) in isomeric SMILES format. These inhibitors were entered into the LigPrep tool of the Schrödinger 2023.1 suite (Schrödinger, LCC) to generate 3D structures using a pH of 7.4 ± 0.2 (Wessler et al., 2015).

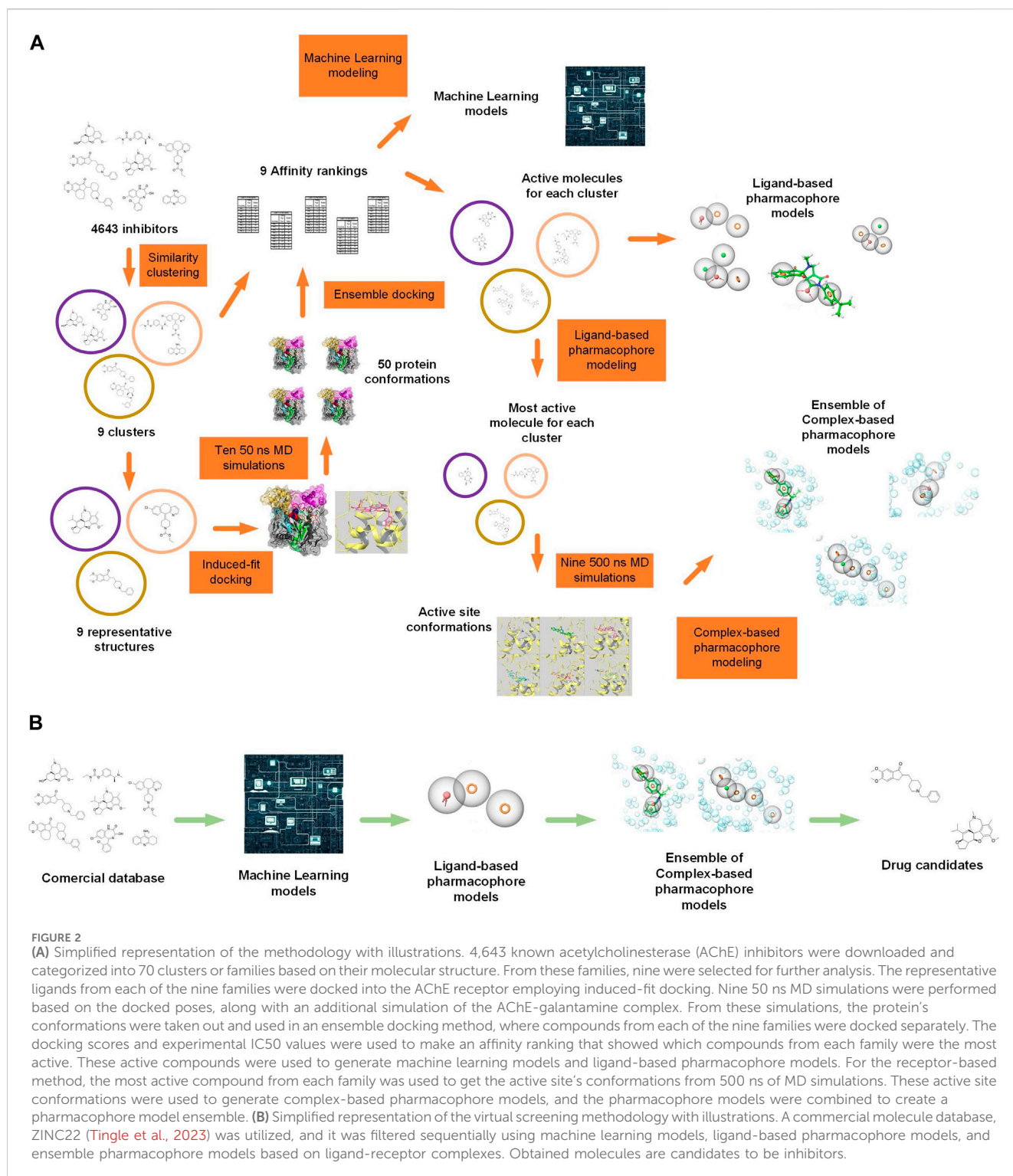
We conducted structural similarity clustering of the molecules from the BD using the Canvas module within the Schrödinger 2023.1 suite (Schrödinger, LCC), employing specific settings including 64-bit precision, radial fingerprint type, Tanimoto similarity criterion, and the average linkage method. Subsequently, utilizing the Canvas results and the Kelley penalty value, we determined the optimal number of clusters, crucially

balancing between over-clustering and under-clustering to ensure the resulting clusters meaningfulness.

A total of 70 ligands clusters were obtained from the BD. For each cluster, we computed statistical metrics including the average IC_{50} , standard deviation of IC_{50} , variation coefficient of IC_{50} , and cluster size. These statistical parameters, alongside other selection criteria, were used to select 9 representative clusters for further analysis.

2.1.2 Induced-fit docking of representative molecules from inhibitor clusters

To generate the Glide grid for induced-fit docking box, RMSD paired calculations were employed (*for more information, refer to the Supplementary Material on RMSD paired calculations*). The PDB files corresponding to the UNIPROT code P22303 of huAChE were downloaded, resulting in a total of 61 structures. Structures that were not wild-type, lacked protein-ligand complexes with covalent ligands, or had a resolution higher than 3.2 \AA were excluded. This filtering process left a total of 18 structures with the following PDB codes: 4BDT, 4EY5, 4EY6, 4EY7, 4M0E, 4M0F, 5FOQ, 6CQU, 6F25, 6NEA, 6O4W, 6O4X, 6O5R, 6O5V, 6O5O, 6O52, 6U3P, and 6U34. Employing RMSD pocket residues, paired calculations facilitated the identification of the key residues



comprising the active site (key pocket), including Tyr-72, Asp-74, Trp-86, Gly-120, Gly-121, Tyr-124, Tyr-133, Glu-202, Ser-203, Trp-286, Phe-295, Phe-297, Tyr-337, Phe-338, Tyr-341, His-447, and Gly-448. The box was manually positioned at the center of the active site using a centroid formed by these residues.

Induced-fit docking validation was carried out using calculations for the huAChE protein-ligand system involving acetylcholine and galantamine molecules. We utilized the protein

crystal structure with PDB code 4EY6 and addressed any missing residues using the Prime module within the Schrödinger 2023.1 suite (Schrödinger, LCC). Protonation states were assigned using PROPKA at a pH of 7.4 (Wessler et al., 2015). Energetic minimization of the crystal structures was carried out, restricting the maximum RMSD of heavy atoms to 0.18 Å. Standard protocols were applied with specific settings: redocking within 30 kcal/mol and 20 structures overall, employing XP precision, utilizing the OPLS3e

force field, setting receptor van der Waals scaling to 0.5, ligand van der Waals scaling to 0.5, refining residues within 5 Å of ligand poses, enabling optimization of side chains, setting the maximum number of poses to 20, and sampling ring conformations within an energy window of 2.5 kcal/mol.

Following selection of representative molecules, based on molecular structure (cluster centroid), from the nine cluster families (refer to [Section 2.1.1](#)) were extracted. Subsequent induced-fit docking calculations were conducted on these molecules, employing the huAChE key pocket residues and adhering to the previously outlined docking settings.

2.1.3 50 ns MD simulation of protein-ligand complexes of representative molecules from clusters of human acetylcholinesterase inhibitors and conformational sampling

The poses corresponding to the 9 representative molecular structures associated with families C31, C4, C19, C50, C35, C20, C36, C42, and C23 were extracted from the induced-fit calculations. In addition, the experimental pose of the galantamine ligand from crystal 4EY6 was also extracted. These 10 poses were used to carry out MD simulations with a production phase duration of 50 ns. The complete molecular dynamics procedure is reported in [Supplementary Material](#).

The 10 trajectories obtained from the 50 ns molecular dynamics simulations were used to extract poses of the active site. The backbone RMSD value was monitored for the key pocket (refer to [Section 2.1.2](#)), and Trp-236, Glu-334, Ala-204, Gly-122 throughout the entire trajectory.

The RMSD values of: average \bar{x}_{RMSD} , average plus, and minus twice the standard deviation $\bar{x}_{RMSD} \pm 2\delta$ were computed. Additionally, both the two smallest l_{RMSD} , $l + 1_{RMSD}$ and the two largest h_{RMSD} , $h - 1_{RMSD}$ values were found. This process yielded a total of 7 conformations per trajectory.

The seven conformations were analyzed using paired RMSD (for more information, refer to the [Supplementary Material](#) on RMSD paired calculations). Based on the RMSD, five conformations were selected from the original seven conformations for each trajectory.

2.1.4 Ensemble docking and affinity ranking

The 50 key conformations were used as input for an ensemble docking procedure, where all compounds belonging to the key clusters 31, 4, 19, 50, 35, 20, 36, 42, and 23 were docked. The box was located manually in the active site center using a centroid composed of key pocket (refer to [Section 2.1.2](#)). Sequential docking of Glide HTVS, Glide SP, and Glide XP was applied. for docking the Epik module to state penalties was utilized.

The XP GScore values obtained from ensemble docking and the experimental inhibition values (IC₅₀) of the compounds in key clusters were used to design an affinity score to rank the compounds in each cluster according to their experimental and computer-predicted inhibition capacities. Hence, the following affinity score, called Yasser's number one (YN1), was employed:

$$LE_i = -\frac{XP GScore_i}{MW_i} \quad i = 1, 2, \dots, N$$

$$YN1_i = \frac{LE_i - LE_{min}}{LE_{max} - LE_{min}} + \left(1 - \frac{\log_{10} IC_{50_i} - \log_{10} IC_{50_{min}}}{\log_{10} IC_{50_{max}} - \log_{10} IC_{50_{min}}} \right)$$

where LE_i is ligand efficiency, MW_i is molecular weight, $\log_{10} IC_{50_i}$ is the log base 10 of IC₅₀, max is the maximum value inside the cluster, min is the minimum value inside the cluster, and N is the number of molecules of the cluster.

2.2 Machine learning modeling to first accurate and fast screening

A schematic of this procedure is presented in [Supplementary Figure S4](#).

2.2.1 Data splitting for training and testing

All molecules with an XP GScore value from ensemble docking were considered active molecules in this stage. Thus, active molecules were extracted from clusters 31, 4, 19, 50, 35, 20, 36, 42, and 23. These active molecules were randomly divided into two groups: 20% for testing and 80% for training. The training and testing sets of active molecules were used separately for decoy calculation, using the DUDE (Database of Useful Decoys: Enhanced) package, generating approximately 50 decoys per active molecule ([Mysinger et al., 2012](#)). Consequently, two libraries were obtained, one for testing and another for training, each containing actives and approximately 50 decoys for every active molecule.

2.2.2 Clustering of molecular descriptors for decoys from the training library

A calculation of molecular descriptors was performed using RDKit for the decoys from the training library (RDKit: Open-source cheminformatics) ([Landrum, 1997](#)). A correlation analysis of the molecular descriptors was conducted, in which the correlation matrix was calculated, and the absolute value of each Pearson correlation coefficient was computed. Next, a feature scaling of the molecular descriptors was carried out, and variances of the scaled variables were calculated. It was considered that two variables are correlated if they have an absolute value of Pearson correlation coefficient greater than or equal to 0.9. To address correlated variables, the variable with the highest variance was chosen, and the rest were discarded. As a result, molecular descriptors that are not correlated were obtained.

The uncorrelated molecular descriptors were used to perform a k-means clustering, where the desired number of clusters was set equal to the number of actives in the training library. Later, the decoy corresponding to the centroid of each cluster was extracted, so that the centroids were used to generate a balanced training dataset.

2.2.3 Training machine learning models

The actives and decoys from the balanced training dataset were taken, molecular descriptors were calculated using RDKit (RDKit: Open-source cheminformatics), and a correlation analysis was conducted. Then, a 5-fold cross-validation was performed for hyperparameter optimization using the following binary classification algorithms: logistic regression, support vector machine, decision trees, and random forests. Optuna was employed for optimization, aiming to find hyperparameter configurations that maximize the ROC-AUC. Once the optimal

hyperparameters were determined, the four algorithms were trained using the complete balanced dataset.

2.2.4 Testing machine learning models

The actives and decoys from the unbalanced test dataset were taken, and molecular descriptors were calculated using RDKit. Correlated variables found during training were eliminated. The four classification algorithms were then tested with the test molecules.

2.3 Ligand-based pharmacophore modeling

In the generation of a ligand-based pharmacophore model, both active molecules and decoys are essential components. Active molecules, identified by a YN1 score equal to or exceeding 1, were selected. Decoys corresponding to the identified active molecules were generated using the DUDE approach, resulting in the production of 50 decoys per active compound (Mysinger et al., 2012).

Ligand-based pharmacophore modeling was configured to identify the best alignment and the most common features. A target of 50 conformers was set, while the number of features in the hypothesis ranged from 3 to 6. The threshold percentage of actives required to match a hypothesis was established at 50%. The scoring function employed was the Phase Hypo Score. All calculations were conducted using the Phase module within the Schrödinger 2023.1 suite.

A schematic of this procedure is presented in [Supplementary Figure S5](#).

2.4 Ensemble complex-based pharmacophore modeling

For each of the nine molecular dynamics (MD) trajectories, encompassing 500 nanoseconds each, both stable and unstable conformations of the active site were identified. These conformations were subsequently utilized to construct complex-based pharmacophore models by employing the ‘Develop Pharmacophore Model’ tool within the Schrödinger Phase module. Whilst default settings were initially applied, adjustments were made to the maximum number of features and mandatory features to refine and optimize the receiver operating characteristic-area under the curve (ROC-AUC) values.

A schematic of this procedure is presented in [Supplementary Figure S6](#).

2.4.1 Conformational clustering of ligand poses from ensemble docking

The compounds with the highest YN1 score were selected for each family, totaling 9 compounds. The poses of these selected compounds, acquired through ensemble docking, were employed in conformational clustering using Schrödinger 2023.1’s Conformer Cluster tool. The clustering used the average linkage method, atomic RMSD, and specified comparison regions involving Heavy atoms + OH, SH. We identified the most frequently observed pose among the various poses obtained from ensemble docking for each of the

9 ligands. These 9 poses were used to carry out MD simulations with a production phase duration of 500 ns, using the simulation procedure reported in [Supplementary Material](#).

2.4.2 TRAPP calculations and extraction of active site conformations from 500 ns MD simulations

The RMSF values for heavy atoms (backbone and side chain) were calculated for active site residues, key pocket (refer to [Section 2.1.2](#)). These RMSF values were sorted from highest to lowest. From each simulation, the 6 residues with highest RMSF values were selected for analysis.

The heavy atoms RMSD for the individual 6 active site residues, backbone RMSD for the whole protein, backbone RMSD for the active site residues, and RMSD for ligand heavy atoms were calculated for all MD simulations. An exponential moving average (EMA) with a period of 5,000 was computed for the 9 RMSD data sets, each consisting of 50,000 frames. A 5000-period was chosen for the EMA to maintain a 1:10 ratio with the total number of frames. To normalize the EMA values to a range between 0 and 1, the following process was applied: the smallest value in the dataset was identified and subtracted from each value, shifting the range so that the minimum became zero. Then, each adjusted value was divided by the range of the dataset, calculated as the difference between the maximum and minimum values. The normalized moving average was used to compute the numerical derivative. This was done by estimating the absolute value of the derivative at a given point as the difference between the function values at that point and at a slightly larger point, divided by the distance between the two points. Heatmaps were then generated using both the normalized moving average and the numerical derivative.

Moreover, graphs were generated utilizing the RMSD normalized moving average and numerical derivative to emphasize regions exhibiting a moving average and numerical derivative exceeding 0.6. The concurrent utilization of the moving average and numerical derivative facilitated the identification of time intervals where both stable and unstable conformations were observed. These graphs were crafted using MATLAB software, and the accompanying scripts are detailed in [Supplementary Data 2](#).

The TRAPP-Analysis clustering tool was used to extract frames from time ranges of stable conformations (Yuan et al., 2020). On the other hand, the TRAPP-Pocket tool was used to extract frames from unstable conformations (Kokh et al., 2013). For TRAPP-Analysis the parameter “cluster” was set to 3 Å, “kmeans” was set to 3 Å, “zmax” was set to 8 Å, and “bb” was set of “on”. For TRAPP-Pocket the parameter “radius” was set to 5 Å, the active site residues were entered manually using the 17 residues previously identified, and the other settings were retained the same as the AR-TRAPP example found in the TRAPP example folders. The TRAPP-pocket python scripts were edited to enable the computing of 50,000 frames.

The frames obtained from clustering with TRAPP-Analysis of stable time ranges were called stable conformations. On the other hand, graphs of the descriptors of the physicochemical properties of the active site were made from the results of TRAPP-pocket, and frames showing atypical descriptor values were extracted. These frames were called unstable conformations.

2.4.3 Complex-based pharmacophore modeling

2.4.3.1 TRAPP calculations and extraction of active site conformations from 500 ns MD simulations

For each of the 9 molecular dynamics (MD) trajectories of 500 ns, the frames of stable and unstable conformations from the active site were employed to construct complex-based pharmacophore models using the Develop Pharmacophore Model tool in the Schrodinger 2023.1 suite Phase module. Default settings were utilized, and variations were made to the maximum number of features and mandatory features within the ranges of 6 to 5 and 3 to 2, respectively. The pharmacophoric hypotheses were combined, forming what we refer to as a hypothesis ensemble or pharmacophore model ensemble.

2.4.3.2 Validation of hypothesis ensembles without YN2

A molecule is considered active if it can match at least one pharmacophore model from the ensemble. Conversely, a molecule is deemed inactive if it cannot match any pharmacophore model from the ensemble. We define the term “ensemble phase score” as the highest phase score obtained by a molecule among all the phase scores obtained for each pharmacophore model in the ensemble. Naturally, if a molecule does not match any model, it will have a score of zero. The accompanying scripts are detailed in [Supplementary Data 2](#).

To validate the 9 ensembles, actives and decoys from each cluster were extracted. Compounds with a YN1 score greater than or equal to 1 were considered active for this stage. Decoys were obtained from the DUDE package ([Mysinger et al., 2012](#)), generating 50 decoys per active. ROC curves, BACC, recall, and specificity values were generated for each validation using the ensemble phase score and the enrichment calculator tool of Schrodinger 2023.1 suite. We selected the configuration of maximum number of features and mandatory features that exhibited the highest ROC-AUC value for the ensemble.

2.4.3.3 Validation of hypothesis ensembles with YN2

To further refine the results and behavior of the ensembles, we introduced a metric called Yasser's number two (YN2), which combines the values of ensemble phase score, the number of matching stable conformations, and the number of matching unstable conformations. This way, molecules with few matches to the pharmacophore models are penalized.

$$\hat{C}_{unstable_i} = \frac{C_{unstable_i}}{C_{unstable\ total}} \quad \hat{C}_{stable_i} = \frac{C_{stable_i}}{C_{stable\ total}} \quad i = 1, 2, \dots, N$$

$$\text{if } Ph_{score_i} \geq Ph_{score\ min} \quad \text{then} \quad \widehat{Ph}_{score} = \frac{Ph_{score_i} - Ph_{score\ min}}{Ph_{score\ max} - Ph_{score\ min}} \quad \text{and}$$

$$YN2_i = \frac{(\hat{C}_{unstable}) + (\hat{C}_{stable}) + (\widehat{Ph}_{score})}{3}$$

$$\text{if } Ph_{score_i} < Ph_{score\ min} \quad \text{then} \quad \widehat{Ph}_{score} = 0 \quad \text{and} \quad YN2_i = 0$$

where N is the number of molecules in the validation or virtual screening procedure, $C_{unstable_i}$ is the number of pharmacophore models derived from unstable conformations that fit with molecule i , $C_{unstable\ total}$ is the total number of pharmacophore models from unstable conformations, C_{stable_i} is the number of pharmacophore models derived from stable conformations that fit with molecule i , $C_{stable\ total}$ is the total number of

pharmacophore models from stable conformations, Ph_{score_i} is the ensemble phase score value for molecule i , $Ph_{score\ min}$ is the minimum ensemble phase score value among known active molecules, and $Ph_{score\ max}$ is the maximum ensemble phase score value among known active molecules.

We utilized YN2 for another round of validation. Initially, YN2 calculation was performed for all molecules. For a molecule to be considered active, it must have a value greater than or equal to a YN2 threshold; conversely, for a molecule to be considered inactive, it must have a value lower than a YN2 threshold. To determine the YN2 threshold, we generated an objective function that is the sum of recall and specificity. We optimized the objective function, obtaining the YN2 value that yielded the maximum value for the objective function.

Thus, from each ensemble validation, we extract the YN2 threshold, the value of $Ph_{score\ max}$, and $Ph_{score\ min}$. These three values, along with the number of stable and unstable conformations, serve as parameters for a subsequent virtual screening run. The accompanying scripts are detailed in [Supplementary Data 2](#).

2.5 Database virtual screening

From the ZINC database, molecules with a molecular weight equal to or less than 500 Da and were in stock were downloaded, resulting in the retrieval of 11,012,710 molecules. The molecules underwent a virtual screening protocol consisting of three stages: machine learning models, ligand-based pharmacophore models, and ensemble of complex-based pharmacophore models. Initially, the ZINC molecules underwent screening using 9 machine learning models, each designed for families 31, 4, 19, 50, 35, 20, 36, 42, and 23. Subsequently, the identified actives from each model were combined, and duplicates were removed. The resulting library then underwent screening using ligand-based pharmacophore models designed for key clusters (refer to [Section 2.3](#)). The identified actives then proceeded to filtering with the ensemble models of pharmacophores corresponding to their respective families. Finally, we selected the top 500 ligands with the highest YN2 values from each cluster and subjected them to an ensemble docking procedure, incorporating HTVS, SP, XP, and MMGBSA filters. A schematic of this procedure is presented in [Supplementary Figure S2](#).

2.6 Human acetylcholinesterase activity inhibition assay

The enzymatic activity of human acetylcholinesterase (hAChE) was quantified using a modified version of Ellman's method. The assay was carried out in a 96-well plate format, with all fresh prepared components prior to use. The assay buffers included Buffer A (BA): 50 mM Tris-HCl, pH 8.0; Buffer B (BB): 50 mM Tris-HCl, pH 8.0, with 0.1% bovine serum albumin (BSA); Buffer C (BC): 50 mM Tris-HCl, pH 8.0, 0.1 M NaCl, and 20 mM MgCl₂·6H₂O; Acetylthiocholine (ATCI): 15 mM; Ellman's reagent (DTNB): 3 mM in BC and Acetylcholinesterase (AChE): 0.22 U/mL in BA.

The assay was performed by first preparing a stock solution of each test compound at a concentration of 500 µg/mL. These

TABLE 1 Identification of the molecules used in the inhibition assay against hAChE.

Molecule ID	Concentration (nM) ^a	MW (g/mol)	URL
1: P-14421887/ZINC10274013	1,070,400	420,42	https://mcule.com/P-14421887/
2: P-25746649/ZINC17245175	1,197,000	375,93	https://mcule.com/P-25746649/
3: P-617769798/ZINC16363937	903,700	497,97	https://mcule.com/P-617769798/
4: P-1894047/ZINC2256546	936,500	480,52	https://mcule.com/P-1894047/
5: P-1205609/ZINC18196844	1,295,000	347,41	https://mcule.com/P-1205609/
6: P-1206762/ZINC18197226	1,238,000	363,45	https://mcule.com/P-1206762/
7: P-2652815/ZINC100759000	942,400	477,51	https://mcule.com/P-2652815/
8: P-2026435/ZINC13691092	981,400	458,51	https://mcule.com/P-2026435/
9: P-533735/ZINC852356	1,129,400	398,45	https://mcule.com/P-533735/
Control: Galantamine hydrobromide	829,180	368,3	https://pubchem.ncbi.nlm.nih.gov/compound/Galantamine-Hydrobromide

^aThe concentrations expressed in nM correspond to the initial concentration of 450 µg/mL.

compounds were then serially diluted from 1.4 µg/mL to 450 µg/mL in 50% DMSO. Galantamine hydrobromide, prepared at 306 µg/mL, served as the positive control. A total of nine test molecules derived from each cluster were included in the assay, and their concentrations in nM are summarized in Table 1.

Each assay well contained a final volume of 250 µL, consisting of: 25 µL of ATCI, 125 µL of DTNB and 50 µL of BB for all plates.

For the blank control, 50 µL of 50% DMSO in BA was added. The negative control consisted of 25 µL of AChE and 25 µL of BA. Each test molecule was added as 25 µL of a 1:1 serial dilution in BA, combined with 25 µL of AChE.

The assay plates were prepared using an OpenTron OT-2 liquid handler, ensuring precise and reproducible liquid handling. The reactions were monitored kinetically at 407 nm using a Biotek Synergy H1 Multi-mode reader, set to 37°C. Absorbance was recorded every 45 s over a 15-min period.

The initial velocity of each reaction was determined from the early linear portion of the absorbance curve, and the percentage of inhibition was calculated using the following formula:

$$\%Inhibition = 1 - \frac{V_0(Sample)}{V_0(Control)} * 10$$

where V₀ (Sample) corresponds to the initial velocity of the reaction with the test sample, and V₀ (Control) is the initial velocity in the absence of the test sample (control). IC₅₀ values were determined by plotting inhibitor concentration against percentage inhibition and calculating the point at which 50% inhibition was observed.

3 Results and discussion

3.1 Affinity ranking based on the dynamism of the biological system and experimental data

Here, we retrieved 4,643 huAChE inhibitors from the BD. To scrutinize the chemical composition of these inhibitors and uncover

possible molecular patterns, we conducted a clustering analysis of their molecular structures. This analysis enabled us to categorize the inhibitors into separate clusters based on their molecular structure. The cluster analysis indicated that the optimal number of clusters was 383, accompanied by a Kelley penalty of 1,372 (as illustrated in Supplementary Figure S9). We also computed Kelley penalty values for cluster sizes of 50, 70, and 100, yielding 1,780, 1,570, and 1,540 Kelley penalty, respectively. Striking a balance between computational efficiency and reliability, we chose to proceed with 70 clusters. Consequently, the 4,643 inhibitors were grouped into 70 distinct clusters, synonymous of 70 compound families, based on their molecular structure.

Then, we decreased the number of clusters from 70 to reduce the computational workload in subsequent stages of the research, and to focus on families with a higher concentration of molecules. We leveraged the availability of IC₅₀ values for each molecule in the BD, which reflect their respective inhibitory capacities by evaluating the average, standard deviation and coefficient of variation of IC₅₀ and cluster size for each family. Additionally, we assigned a cluster ID to facilitate identification for each family.

Subsequently, we identify clusters from the initial set of 70 that demonstrated the following characteristics: the lowest average IC₅₀ values (indicating high inhibitory potential), minimal IC₅₀ standard deviation (reflecting low variability within the cluster), a moderate coefficient of variation (indicating consistent and uniform data), and a large population size (providing a broader pool of compounds for analysis). To achieve this goal, we established a threshold based on the mean values of IC₅₀ average, IC₅₀ standard deviation, variation coefficient, and population calculated across all families. The determined cutoff values were IC₅₀ average of 16,268 nM, an average IC₅₀ deviation of 14,901 nM, an average variation coefficient of 1.05, and an average population of 66 compounds.

Then, we employed these thresholds as a basis for creating cluster filters, refining them manually to maximize the population of clusters. The established criteria were: (i) average IC₅₀ below 10,000 nM, (ii) IC₅₀ standard deviation under 14,900 nM, (iii) population exceeding 58 molecules, and (iv) variation coefficient less than 1.7. A total of 17 clusters (Supplementary Table S2) satisfied at

least 3 conditions (e.g., clusters 41, 1, 25, 51, 22, 56, 60, 46, 36, 31, 4, 19, 50, 35, 20, 42, 23), and among them, 3 clusters fulfilled all conditions (clusters 4, 19, 50). We opted to retain these 3 clusters that met all 4 conditions, along with the 6 clusters that satisfied 3 conditions and had the largest populations. The resultant clusters, namely, 31, 4, 19, 50, 35, 20, 36, 42, and 23, were utilized in subsequent stages. Moreover, we retrieved the compounds associated with the centroid of each cluster ([Supplementary Table S3](#)), which theoretically should represent an average molecular structure for each respective family.

We required the initial poses of the 9 centroid molecules on the AChE active site employing induced-fit docking. To validate the efficacy of the induced-fit docking, we performed induced-fit docking calculations for the huAChE protein-ligand system (PDB 4EY6) using acetylcholine and galantamine molecules. The induced-fit docking successfully reproduced the authentic pose of galantamine ([Supplementary Figure S10](#)), yielding an RMSD value of 0.93 between the predicted and actual poses. In the case of acetylcholine, induced-fit docking accurately replicated the documented pose from the literature ([Supplementary Figure S11](#)) ([Raves et al., 1997](#)). After validation, the centroid structures were successfully subjected to induced-fit docking, and the resulting ligand poses were used to perform 9 MD simulations of 50 ns. In addition, a 50 ns MD simulation was performed for the 4EY6 crystal, which contains the co-crystallized galantamine ligand. Energy, volume, temperature, and pressure were obtained during the MD production phase to confirm proper functioning of the 10 simulated systems. The kinetic and total energy demonstrated fluctuation around a stable value, while the volume graphs exhibited constant values. The temperature and pressure oscillated near 310 K and 1 bar, respectively ([Supplementary Data S1](#)).

The overall backbone RMSD value was monitored for 21 active site residues throughout the entire trajectory ([Supplementary Figure S12](#)). These included 17 residues (key pocket) identified in the RMSD paired calculations and 4 residues chosen for their crucial role in the subsite behavior based on literature (Trp-236, Glu-334, Ala-204, and Gly-122) ([Wlodek et al., 1997](#)) and these values were utilized to extract various conformations of the active site of the AChE ([Supplementary Figure S12](#)).

An statistical analysis of the RMSD data was conducted by calculating \bar{x}_{RMSD} , $\bar{x}_{RMSD} \pm 2\delta$, l_{RMSD} , $l + 1_{RMSD}$, h_{RMSD} , and $h - 1_{RMSD}$. This process yielded a total of 7 conformations extracted from each MD simulation, which was then repeated for all 10 MD, resulting in a total of 70 conformations ([Supplementary Table S6](#)). The 7 conformations obtained were treated as if they were different protein crystals and were subjected to the RMSD paired calculations (for further details, refer to the [Supplementary Material](#) on RMSD paired calculations). RMSD paired calculations were carried out using the all-atoms of 21 residues.

Five out of the 21 residues that showed the highest average RMSD from paired RMSD ([Supplementary Table S4](#)) were selected ([Supplementary Table S5](#)). RMSD heatmaps for these 5 residues were used to extract 5 conformations from the initial 7 conformations to achieve the highest conformational diversity. (All heatmaps can be found in [Supplementary Data S1](#)). Upon comparing the results from [Supplementary Table S4](#), it was observed that the residues Asp-74, Glu-334, Tyr-341, His-447, Phe-295, and Phe-338 (as shown in [Supplementary Figure S13](#)) displayed

significant movement in common among the 7 conformations, revealing changes in both the backbone and the side chain. Finally, 50 conformations were obtained and implemented in the ensemble docking procedure ([Supplementary Data S1](#)).

The ensemble docking process utilized 50 active site conformations obtained from 50 ns MD simulations as input, along with all compounds from families (key clusters). The outcome was an affinity ranking based on Glide XP GScore docking scores, encompassing all conformations that matched each ligand and identifying the conformation that best suited each ligand. Subsequently, a metric named Yasser's Number 1 (YN1) was developed to rank compounds based on both their docking scores and experimental inhibition capacity (IC50). YN1 is comprised of two terms: the first term represents the normalized value of ligand efficiency, while the second term involves the inverted normalized logIC50. The inverted logIC50 value is calculated by subtracting the normalized value from 1 to reverse the scale. YN1 score spans from 0 to 2, where a score nearing 2 signifies greater computational and experimental inhibitory potential.

All ligands from each family were ranked using YN1 as an affinity score, obtaining one ranking for each family. Additionally, we evaluated the correlation of logIC50 and YN1; and XP Gscore and YN1 ([Supplementary Figures S14, 15](#)).

[Table 2](#) shows the Recovery percentage of recovered molecules from ensemble docking (%), which represents the percentage of molecules that were selected from the ensemble docking procedure calculated over the number of molecules that entered, due to the fact that some molecules have had a poor XP Gscore so the ensemble docking was not reported by Glide, as they may or may not even have been able to fit into the active site of the receptors.

The ensemble docking process and subsequent YN1 scoring enabled us to identify the most promising huAChE inhibitors. We set a YN1 score threshold above 1 to ensure a balanced analysis so it allows us to identify compounds that either demonstrated average performance in both experimental and computational assessments, exhibited poor experimental performance but strong computational performance, or showed strong experimental performance but weak computational.

In our research, the use of IC50 values is particularly valuable as it complements molecular docking rankings with empirical evidence. IC50 values are derived from *in vitro* biological assays ([Cer et al., 2009](#)), mimicking the protein's physiological environment, hence directly measuring a compound's ability to inhibit the protein. This direct measure offers a stronger correlation with *in vivo* biological activity. Molecular docking, however, is an *in silico* technique predicting compound-protein binding through computational simulations, providing insights into molecular interactions and potential binding sites but not guaranteeing biological activity.

Finally, the ligand with the highest YN1 was obtained from each ranking, resulting in 9 ligands presenting the best equilibrium between docking score and IC50 values ([Supplementary Table S7](#)). Among the selected compounds, inhibitors with IC50 values lower than 420 nM and XP GScore lower than -9.4 kcal/mol were obtained, demonstrating the interesting inhibitory capacity of the selected ligands.

TABLE 2 Results of ensemble docking and YN1 implementation.

Cluster ID	Cluster initial size	Recovered molecules from ensemble docking	Recovery percentage of recovered molecules from ensemble docking (%)	Molecules with YN1 ≥ 1	Recovery percentage of molecules with YN1 (%)
4	59	58	98	25	43
19	64	59	92	30	51
20	261	156	60	66	42
23	1,455	291	20	95	33
31	48	41	85	11	27
35	68	62	91	44	71
36	47	47	100	20	43
42	1,038	241	23	139	58
50	68	46	68	16	35

Cluster initial size: The number of molecules belonging to the family before ensemble docking. Recovered molecules from ensemble docking: The number of molecules that Glide reports with an XP Gscore value. Recovery percentage of recovered molecules from ensemble docking (%): The percentage of molecules reported by Glide with an XP Gscore value relative to the initial number of molecules. Molecules with YN1 ≥ 1 : The number of molecules with a YN1 value greater than or equal to 1. Recovery percentage of molecules with YN1 (%): The percentage of molecules with a YN1 value greater than or equal to 1 over the number of recovered molecules from ensemble docking.

TABLE 3 Data splitting for training and testing.

Family	Training		Testing	
	Number of actives	Number of decoys	Number of actives	Number of decoys
C4	47	2,450	11	637
C19	48	2,351	11	539
C20	125	6135	31	1,519
C23	233	13,261	58	3,649
C31	33	1,617	8	392
C35	50	3,220	12	931
C36	38	3,234	9	833
C42	193	11,502	48	2,695
C50	37	1813	9	441

3.2 Machine learning modeling

For this stage, we took the molecules obtained from the output of each of the 9 ensemble dockings to generate 9 ML models capable of classifying molecules as inactive or active with respect to each identified family of inhibitors.

The molecules recovered from ensemble docking were divided into libraries with a 20% portion allocated for testing and an 80% portion for training. In addition, approximately 50 decoys were generated for each active compound, providing both positive and negative instances for training and testing (Table 3). Consequently, two unbalanced libraries were formed, comprising positive and negative instances, and allocated for training and testing purposes separately. To ensure balanced training, we implemented a protocol involving clustering based solely on the molecular descriptors of the negative instances. This protocol enabled the extraction of the most diverse decoys, resulting in

negative instances with a wide range of molecular descriptor values, thereby facilitating optimal model training. Then, we proceeded to train four classification algorithms: logistic regression, support vector machine, decision trees, and random forests. Cross-validation and hyperparameter optimization techniques were carried out to maximize the performance of each algorithm using the available data. Subsequently, average values for accuracy, recall, specificity, and ROC AUC were evaluated over multiple iterations of cross-validation.

Performance evaluation included assessing the models' capacity to identify decoys excluded during the generation of the balanced training library as inactive, the models' ability to classify the initial 4,643 molecules from the binding database as active or inactive, the capacity to classify all molecules belonging to the cluster/family, and the ability to classify molecules from the binding database that do not belong to the cluster/family. The outcomes of both training and testing the models for the nine

families of inhibitors are detailed in [Supplementary Tables S8–16](#).

In order to select the appropriate model, we employed an objective function comprised of specificity values, the proportion of true negatives identified within decoys excluded from training, the percentage of active compounds identified from the family, and the percentage of compounds from the binding database not associated with the family, then we established an ML model for each family, as delineated in [Table 4](#). These ML models are to be applied in the subsequent sections to execute a virtual screening procedure. The scripts utilized for training the ML models and the .pkl files of the models are provided in [Supplementary Data S2](#).

3.3 Ligand-based pharmacophore modeling

Pharmacophore models were generated using the Phase module of Schrodinger 2023.1 from the molecules derived from the 9 ensemble dockings. We employed the DUDE package ([Mysinger et al., 2012](#)) to generate decoys at a ratio of approximately 50 decoys per active compound to validate the pharmacophore models. Thus, we generated 9 libraries comprising molecules with a YN1 score greater than or equal to 1 along with their respective decoys. The Phase calculation resulted in the creation of multiple ligand-based pharmacophore models for each of the nine clusters. However, only models with the highest area under the ROC curve (ROC AUC) were utilized.

Performance metrics, particularly the ROC AUC values, indicates that most models exhibit high predictive power ([Figure 3](#), [Supplementary Figure S16](#)). For example, models such as C-31 (AHHR-1) and C-4 (ARR-3) achieve ROC AUC values of 0.90 and 0.92, respectively, which are indicative of strong discrimination capabilities. These high AUC values are complemented by other performance metrics like recall (RC) and specificity (SP), with some models achieving perfect recall (e.g., C-4 with an RC of 1), highlighting their ability to correctly identify all active compounds. However, the specificity values vary widely, suggesting that while some models are excellent at identifying actives, they might struggle with negatives, as seen in models

with lower SP values. Models like C-4 and C-50 (APR-2) demonstrate high BACC values, underscoring their balanced approach in handling both true positives and true negatives effectively.

Overall, the ligand-based pharmacophore models present strength in terms of high recall and ROC AUC values, suggesting their potential utility in identifying novel active compounds. However, the variability in specificity and balanced accuracy across different models also points to areas for further refinement, particularly in reducing false positive rates to improve the models' practical applicability in diverse drug discovery scenarios.

3.4 Ensemble complex-based pharmacophore modeling

The nine ligands with the highest YN1 from the nine families were employed to perform nine 500 ns MD simulations ([Supplementary Table S7](#)). To accomplish this, the poses of the ligands were extracted from the ensemble docking. A ligand geometry clustering was performed, where the poses associated with the 50 receptor conformations were clustered to extract the most frequently occurring pose ([Supplementary Figure S17](#)). The optimal number of clusters was determined by finding the global minimum in the Kelley penalty graphs ([Supplementary Figures S18, 19](#)). The pose with the highest frequency was defined as the centroid of the most populated cluster and was extracted as the initial pose for MD simulations.

The root-mean-square-deviation (RMSD) over backbone atoms of whole protein, active site, and ligand heavy atoms are reported in [Supplementary Figures S20–22](#). Per-residue-root-mean-square-fluctuation (RMSF) over backbone atoms for the MD production stage of all systems and all heavy atoms (backbone and chain) were evaluated ([Supplementary Figures S23–26](#)). MD simulations showed kinetic and total energy fluctuation around a stable value, while the volume remained constant. Temperature and pressure oscillated near 310 K and 1 bar, respectively ([Supplementary Data S1](#)).

Regarding protein flexibility ([Supplementary Figures S27, 28](#)) RMSF backbone values for the 9 MD simulations of the

TABLE 4 Metrics of the best machine learning algorithms for each of the families. Random forest, RF; Support vector, SV.

Family	ML algorithm	Average metrics cross validation				Test metrics			
		Accuracy	Recall	Specificity	ROC AUC	Accuracy	Recall	Specificity	ROC AUC
C4	RF	0.990	0.980	1.000	0.990	0.995	1.000	0.995	0.998
C19	RF	1.000	1.000	1.000	1.000	0.998	1.000	0.998	0.999
C20	RF	0.988	0.976	1.000	0.988	0.992	0.871	0.994	0.933
C23	RF	0.970	0.961	0.979	0.970	0.988	0.948	0.989	0.969
C31	SV	0.969	0.967	0.967	0.967	0.995	0.875	0.997	0.936
C35	RF	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
C36	RF	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
C42	RF	0.995	0.990	1.000	0.995	0.998	0.979	0.999	0.989
C50	RF	0.946	0.918	0.975	0.946	0.982	1.000	0.982	0.991

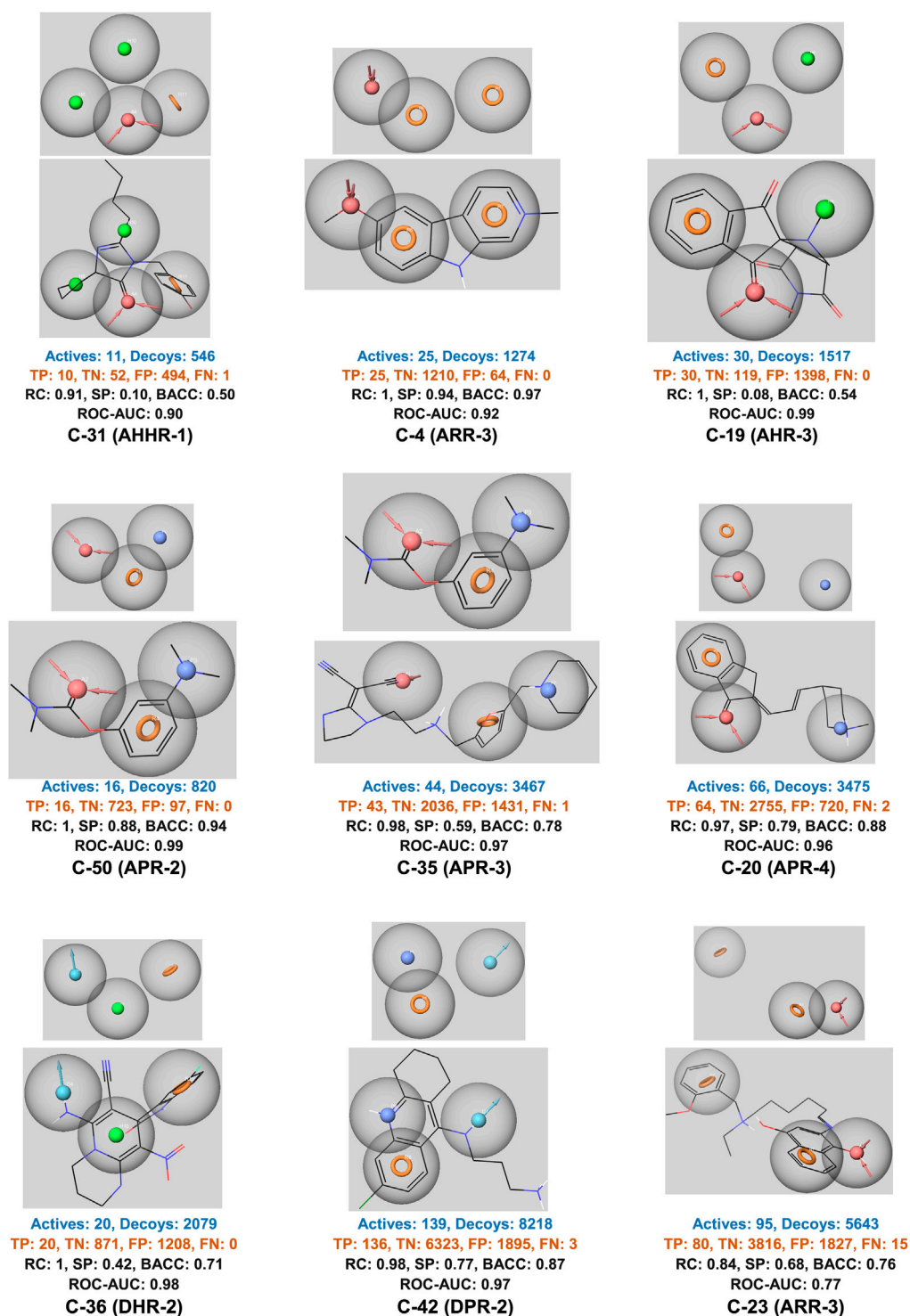


FIGURE 3

The best ligand-based pharmacophore models. H-bond acceptor (A) in red, H-bond donor (D) in light blue, positive point (P) in dark blue, aromatic (R) in orange, and hydrophobic (H) in green. TP, true positive; TN, true negative; FP, false positive; FN, false negative; RC, recall; SP, specificity; BACC, balanced accuracy; ROC-AUC, area under ROC curve.

4EY6 complex showed the S2 domain (228–303) and S3 (331–407) with high fluctuations, suggesting flexibility. On the other hand, the S4 domain (435–457) and S1 domain (117–153) show lower fluctuations, indicating regions with rigidity. The Ω -loop domain

(69–96) also shows moderate fluctuations, which could influence its function or interaction with inhibitors, as it is located directly in the active site. Furthermore, the RMSF for sidechain and backbone showed elevated values due to the higher flexibility of the sidechains.

Regions with high flexibility, such as the S2 domain and Ω -loop, stand out even more when all heavy atoms of the residue are considered. Previous studies suggest that the omega-loop is involved in induced fit phenomena in the active site (Cheng et al., 2017), which is consistent with our findings of flexibility in the omega-loop.

Concerning ligand's behavior in terms of its interaction with the protein we evaluated hydrogen bond occupancy (Supplementary Figures S30, 31), protein-ligand contact for all residues (Supplementary Figures S32–40), and protein-ligand contact for residues that were within 5 Å of the ligand at some point during the MD simulation (Supplementary Figures S41–49) to finally obtain the percentage frequency of interaction occurrences (Supplementary Figures S29).

The following residues obtained a percentage equal to or greater than 80% in terms of hydrophobic contacts for several MD simulations, with the number of total appearances in parentheses: Trp-86 (9), Tyr-341 (9), Tyr-124 (8), Tyr-337 (8), Gly-121 (6), Trp-286 (6), Tyr-72 (6), Asp-74 (4), Gly-448 (4), Hid-447 (4), Met-85 (4), Phe-297 (4), Phe-338 (4), Thr-75 (4), Thr-83 (4), Glu-292 (3), Ser-125 (3), Val-73 (3), Asn-87 (2), Leu-76 (2), Phe-80 (2), Ser-293 (2), Tyr-449 (2), Tyr-77 (2), Val-294 (2), Ala-343 (1), Arg-296 (1), Gln-71 (1), Gly-122 (1), Gly-342 (1), Gly-362 (1), Gly-366 (1), Gly-82 (1), Ile-451 (1), Leu-339 (1), Leu-398 (1), Lys-348 (1), Phe-295 (1), Pro-344 (1), Ser-438 (1), Trp-439 (1), Tyr-133 (1), Val-132 (1), Val-365 (1). Among these, the residues that appeared in 4 or more MD simulations were: Trp-86, Tyr-341, Tyr-124, Tyr-337, Gly-121, Trp-286, Tyr-72, Asp-74, Gly-448, Hid-447, Met-85, Phe-297, Phe-338, Thr-75, and Thr-83.

Nine MD simulations were employed to extract protein conformations for generating complex-based pharmacophore models. To do this, RMSF calculated for heavy atoms (backbone and side chain) were employed. The RMSF for Tyr-72, Asp-74, Trp-86, Gly-120, Gly-121, Tyr-124, Tyr-133, Glu-202, Ser-203, Trp-286, Phe-295, Phe-297, Tyr-337, Phe-338, Tyr-341, His-447, and Gly-448 were selected and ranked, resulting in 6 residues being chosen from each of the 9 simulations (Supplementary Table S17). These residues exhibited the greatest conformational variability throughout the entire simulation.

Individual heavy atoms RMSD for the 6 active site residues, backbone RMSD for the whole protein, backbone RMSD for the 17 active site residues, and RMSD for ligand heavy atoms were calculated for all MD simulations. To smooth the RMSD data, a moving average was constructed for the 9 data sets (Supplementary Figures S50–58). The moving average data was utilized to normalize data. Heatmaps were obtained with the normalized average to compare the behavior of the RMSD for the 9 data sets (Supplementary Figures S59–67).

In addition, the average was utilized to calculate the numerical derivative (Supplementary Figures S59–67). The derivative was used to identify regions of variation in the RMSD. Finally, we highlighted the moments where the moving average and numerical derivative were both greater than 0.6 (Supplementary Figures S68–76).

The parallel implementation of the moving average and numerical derivative enables the identification of stable and unstable conformations. If the RMSD normalized moving average does not show large variations over a time range, it suggests that the active site adopts stable conformations. Conversely, if the RMSD

normalized moving average displays large variations, then the active site adopts unstable conformations. In order for the active site geometry to transition between stable conformations, it has to pass through an unstable conformation that enables the change. Therefore, high values of the numerical derivative can be used to identify time intervals where unstable conformations are more likely to occur, as these represent transition zones between stable conformations.

The zones of stable and unstable conformations can be identified in heatmap (Supplementary Figures S68–76) (A stable zone is identified when the graph remains in a green or blue color over a large time interval while an unstable zone is identified when the graph changes from green to blue and red squares are present in between). For example, the MD trajectory of Cluster 4, displayed three stable zones corresponding to 0–150, 180–430, and 460–500 ns (Supplementary Figures S59, S68) and two unstable zones for 150–180 and 430–460 ns. These regions were used to extract specific frames (stable or unstable conformations) for the generation of complex-based pharmacophore models (Table 5).

The TRAPP-Analysis clustering tool was used to extract frames from stable zones and the TRAPP-Pocket tool was utilized to extract frames from unstable zones (Kokh et al., 2013; Yuan et al., 2020; Kokh et al., 2016). TRAPP-Analysis uses a clustering algorithm that serves to categorize the active site conformations into similar groups (Kokh et al., 2013; Yuan et al., 2020; Kokh et al., 2016), in this manner the conformations obtained correspond to the highest populated (energetically most favorable) conformations in the time range.

In the case of TRAPP-pocket, numerical descriptors related to the physicochemical properties of the active site were calculated for each frame of the trajectory (Figure 4) in order to identify energetically high conformations, so these conformations were assumed to be energetically unfavorable (Supplementary Figures S77–92; Table 5) and viceversa.

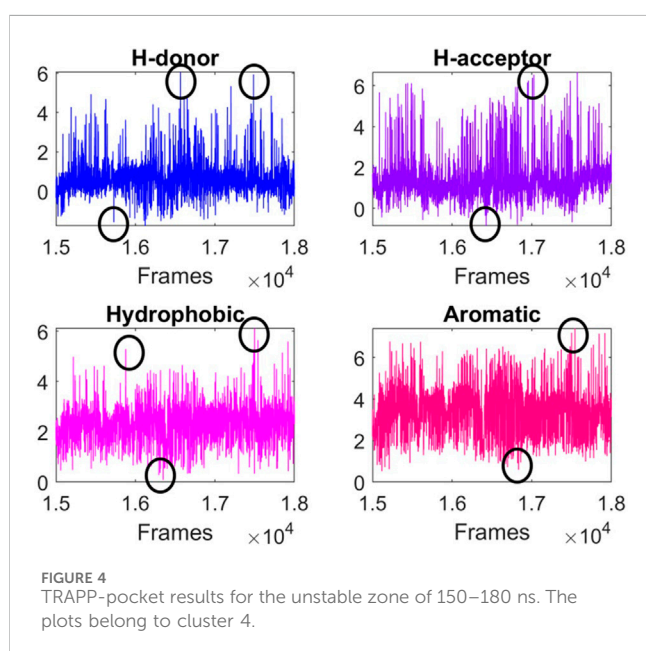
Stable and unstable conformations were utilized to construct complex-based pharmacophore models using the Schrödinger 2023-1 phase module. These models were combined into an ensemble. Within this framework, (Supplementary Figures S93), molecules undergo virtual screening through all pharmacophore models derived from both stable and unstable conformations. Pharmacophore models are linked in the ensemble based on the logical OR operator. This means that for a molecule to be active, it is sufficient for just one individual model to classify it as active, and for a molecule to be inactive, it should not be classified as active by any model. Lastly, we define the ensemble Phase score as the highest Phase score attained by a molecule among all Phase scores obtained for each pharmacophore model in the ensemble.

The pharmacophore model ensembles were subjected to validation using the same actives and decoys employed in the ligand-based model validation. Subsequently, these validation results were introduced into the Schrodinger Enrichment Calculator tool, and ROC curves were generated for each ensemble (Supplementary Figures S94). Additionally, recall, specificity, ACC, and BACC values were obtained (Supplementary Table S18).

To improve the performance of the ensemble, we introduced a metric called Yasser's number two (YN2), which combines the values of ensemble phase score, the number of matching stable

TABLE 5 Time intervals of stable and unstable conformations during 500 ns MD simulations. Number of extracted conformations in parentheses.

Cluster	Stable	Unstable
C4	(0–150(3)), (180–430(2)), (460–500(3))	(150–180(12)), (430–460(13))
C19	(0–60(3)), (100–200(1)), (220–330(3)), (350–500(2))	(60–100(8)), (200–220(12)), (330–350(8))
C20	(0–350(3)), (410–500(1))	(350–410(15))
C23	(0–110(2)), (140–410(3)), (450–500(2))	(110–140(12)), (410–450(12))
C31	(0–90(1)), (130–380(4)), (410–500(2))	(90–130(13)), (380–410(12))
C35	(0–50(3)), (60–210(1)), (250–500(4))	(50–60(15)), (210–250(15))
C36	(0–60(3)), (100–500(4))	(60–100(12))
C42	(0–60(2)), (130–500(2))	(60–130(17))
C50	(0–50(2)), (60–270(5)), (280–500(2))	(50–60(15)), (270–280(9))



conformations, and the number of matching unstable conformations. This metric averages the normalized values of ensemble phase score, the number of matching stable conformations, and the number of matching unstable conformations. A high YN2 value indicates that the molecule in question has an excellent ensemble phase score and fits with the majority of stable and unstable conformations.

For a molecule to be considered active employing YN2, it must have a value greater than or equal to a YN2 threshold and viceversa. To perform virtual screening using pharmacophore model ensembles coupled with YN2, 5 parameters were required: the number of stable and unstable conformations, the minimum ensemble phase score value among known active molecules ($Ph_{score\ min}$), the maximum ensemble phase score value among known active molecules ($Ph_{score\ max}$), and the YN2 threshold. The number of conformations were fixed while the other parameters were obtained in the ensemble validation process.

YN2 threshold was defined based on an objective function summing recall and specificity. After evaluating YN2 for all

molecules (based on [Supplementary Table S18](#)) we carried out an optimization of an objective function corresponding to the sum of specificity and recall by varying the value of YN2. This way, we obtained the YN2 value that yielded the maximum value for the objective function ([Supplementary Tables S19–S27](#); [Supplementary Table S28](#)).

The validation of the improved ensembles with the inclusion of YN2 was conducted ([Supplementary Table S29](#)). Furthermore, these findings were integrated into the Schrödinger Enrichment Calculator tool, and ROC curves were generated for each ensemble ([Supplementary Figures S95](#)). [Supplementary Table S30](#) presents the percentage improvements in ensemble performance with and without the evaluation of YN2, indicating a slight decline in recall values and a significant increase in specificity values in some cases, up to approximately 2000%. The ensembles displayed good metrics, with ROC AUC values greater than 0.8 in most cases.

The most common features found among the ensembles correspond to aromatic interactions, including those caused by the residues Trp-86, Tyr-341, Tyr-337, Tyr-124, Tyr-72, and Trp-286 ([Figure 5](#)). Among these, Trp-86 and Trp-286 showed a capacity to generate π - π interactions between aromatic rings and also cation- π interactions with positive groups of the ligands. The role of these residues concerning the ligand entry is reported in other studies ([Branduardi et al., 2005](#)), indicating that these residues interact with the positive charges of the natural substrate acetylcholine and drive it into the active site. The fact of finding this dual capacity of these tryptophan residues in this research confirms their ability to interact with positive charges. On the other hand, the residues Tyr-341, Tyr-337, Tyr-124, and Tyr-72 showed the ability only to generate aromatic π - π interactions. But in some conformations, these tyrosine residues presented hydrogen bond interactions.

Protein dynamics are crucial for understanding how proteins interact with other molecules. Two key models for explaining protein dynamics are the induced-fit model and conformational selection ([Galburt and Tomko, 2017](#)). The induced-fit model suggests that when a ligand binds, it triggers changes in the protein's shape to accommodate the ligand. Conversely, the conformational selection model proposes that the protein exists in various conformations, and ligand binding selects a specific one. Both models underscore the significance of protein dynamics in ligand binding. In our research, we considered the protein's dynamic

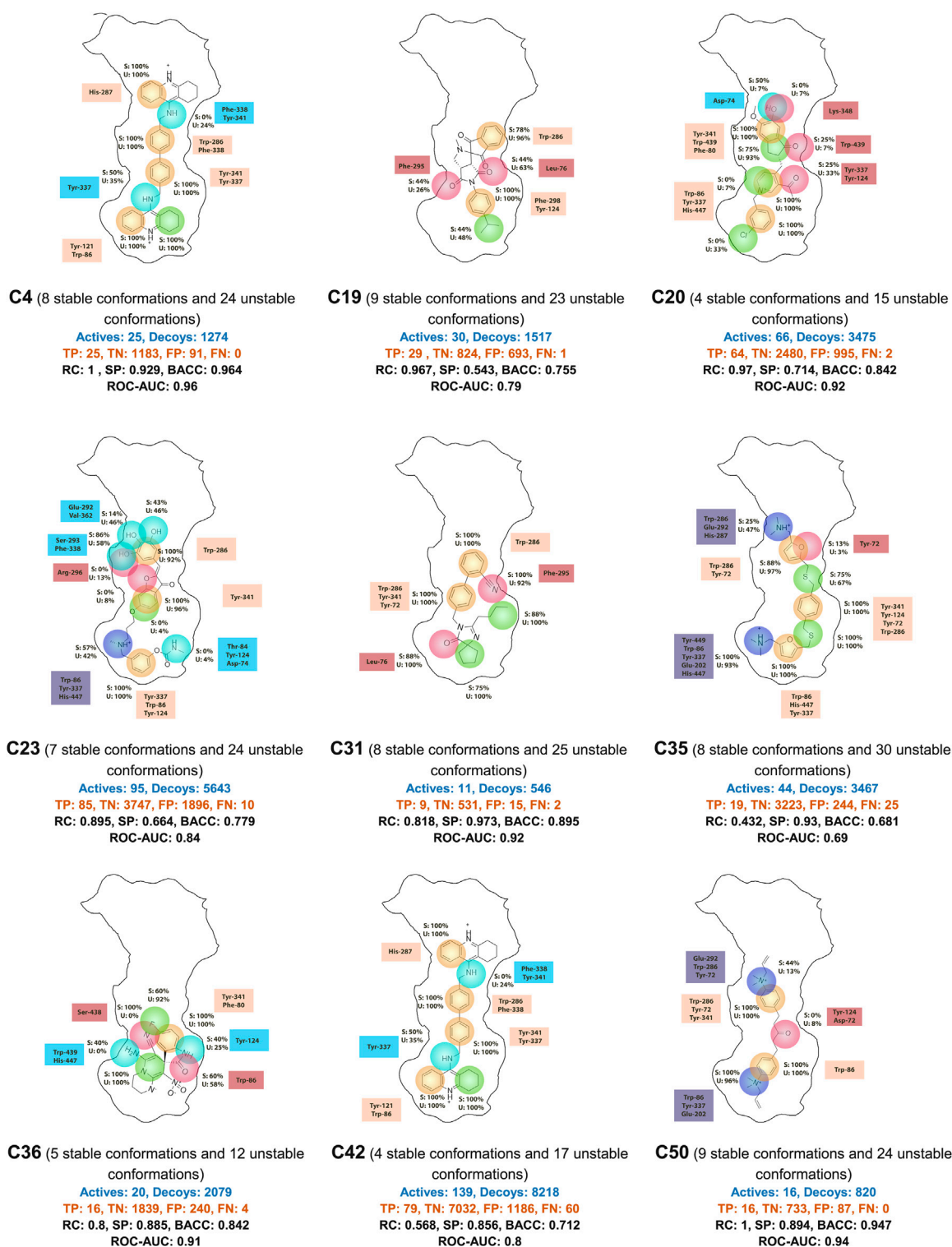


FIGURE 5

Summary of pharmacophore model ensembles for each cluster. Hydrophobic contacts are shown in green, H-bond donors in light blue, H-bond acceptors in red, aromatic interactions in beige, and positively charged groups in dark blue. The crucial residues for each feature are displayed next to the corresponding feature. S indicates stable conformation and U indicates unstable conformation. The percentages shown correspond to the number of times a feature appears in the total stable and unstable conformations. TP, true positive; TN, true negative; FP, false positive; FN, false negative; PC, precision; RC, recall; SP, specificity; BACC, balanced accuracy; ROC-AUC, area under ROC curve.

behavior in various computational calculations. Protein dynamics, including induced-fit and conformational selection, can significantly impact ligand binding. For instance, a molecule might not fit into a specific protein conformation, appearing inactive, yet it could effectively bind to other conformations. By employing ensemble pharmacophore models and other techniques, we gained valuable insights into the diverse protein conformations that a ligand could bind to, enhancing the accuracy of inhibitory activity predictions.

It was pivotal to link pharmacophore models corresponding to various active site conformations using the logical OR and YN2 on identified actives from each model individually. This approach allows for active site flexibility, contrasting with the use of a single pharmacophore model. Besides, we emphasize the importance of exclusion volumes in complex-based models. These volumes mimic the steric effects of the active site, effectively preventing recognition of ligands with unfavorable geometries as active. Exclusion volumes primarily aim to boost specificity and selectivity of pharmacophore models by narrowing the search space, thereby minimizing false positives resulting from improper ligand positioning.

Regarding seeking selective drugs, we consider crucial to individually address energetically unstable conformations. For instance, targeting specific protein conformations can yield more effective drugs with fewer side effects by reducing the likelihood of interference with the normal functioning of other enzymes. In molecular dynamics research, clustering techniques are commonly employed to extract representative frames of protein conformations observed during simulations. Traditional clustering groups frames based on their similarity in conformation, prioritizing the most frequently occurring ones in the simulation. However, this can lead to the loss of important information from less populated conformations, which may be equally relevant for understanding the binding's behavior. This is the reason why the combined use of stable and unstable conformations is important.

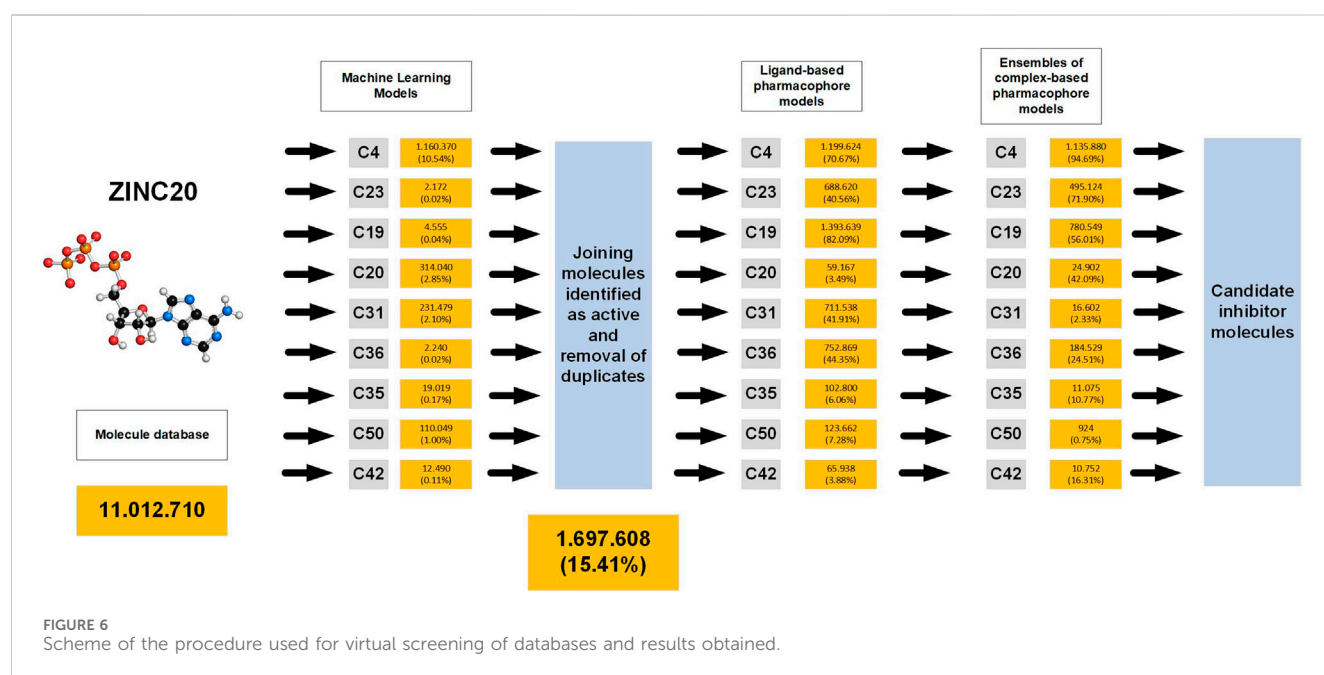
3.5 Database virtual screening

We aimed to implement the predictive models developed in this research to identify molecules that could function as novel inhibitors for AChE. To achieve this goal, we designed the virtual screening scheme depicted in [Supplementary Figures S2](#). The scheme consists of a sequential use of machine learning models, ligand-based pharmacophore models, and complex-based pharmacophore model ensembles ([Figure 6](#)).

In order to find the inhibitors we extracted 11,012,710 molecules from the ZINC database. These molecules underwent virtual screening with the 9 ML models, and the active molecules identified by each model were aggregated into a combined library, resulting in a total of 1,697,608 retrieved molecules, which represents 15.41% of the original ZINC library. Subsequently, these molecules were subjected to screening with 9 ligand-based pharmacophore models and 9 ensembles of complex-based pharmacophore models. 500 molecules with the highest YN2 scores were selected from each final list, and subjected to ensemble docking calculations and MMGBSA free energy calculations to obtain those with the most interesting interactions ([Supplementary Figures S96, 97; Figure 7](#)).

[Supplementary Figures S98–106](#) display the 2D interaction diagrams of the poses in the protein-ligand complex for these compounds. Free energy value varies from ZINC22983357 with -62.77 kJ/mol, which comes from the C31 family to ZINC18206771 from the C35 family, with -115.69 kJ/mol.

The virtual screening results underscore the effectiveness of an integrated approach that combines computational and experimental data through the development and application of machine learning and pharmacophore modeling. The order of filter application was specifically designed to progress from less to more precision in molecule searching. Initially, the ML models were trained to identify molecules from each family, unlike the ligand-based and



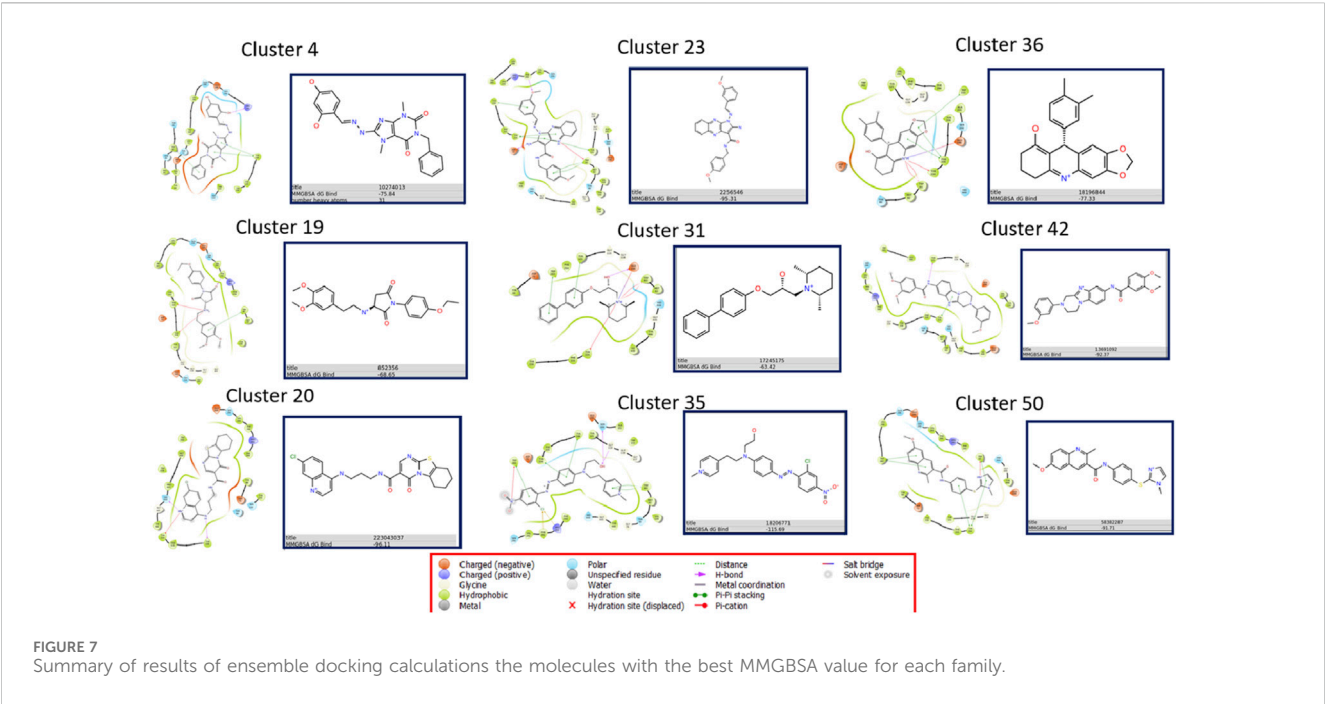


TABLE 6 IC₅₀ values of tested molecules.

Molecule	IC ₅₀ (μM)
1	2.08E-197 ± 0.022
2	3.71E+16 ± 0.021
3	58.41 ± 0.033
4	29.22 ± 0.016
5	40.59 ± 0.015
6	41.60 ± 0.017
7	26.13 ± 0.014
8	24.85 ± 0.019
9	9.98 ± 0.021
Control	41.60 ± 0.011

Data are expressed as the means ± standard deviation (SD) of at least three independent experiments.

complex-based pharmacophore models, which were trained with the molecules exhibiting the highest activity by YN1 from each family. Additionally, since the ML models were located at the first phases, we focused on seeking models with more specificity. This way, we filtered out inactive molecules initially, expecting the ligand-based and complex-based pharmacophore models to identify active molecules. The ML models proved very useful as they required lower computational resources compared to directly using pharmacophore models. This enabled successful management of the initial ZINC library of 11,012,710 molecules, reducing it by up to 15%.

Afterwards, the molecules underwent ligand-based pharmacophore modeling and complex-based pharmacophore ensemble modeling. Pharmacophore modeling is a powerful tool

for drug discovery, but its effectiveness heavily relies on the availability of both active molecules and the biological target (Qing et al., 2014). However, when only the biological target is known, this approach encounters several challenges, making it difficult to identify promising drug candidates. It is assumed that pharmacophoric points represent regions where favorable protein-ligand contacts occur (Ertl et al., 2011). This holds particularly true for complex-based pharmacophores, which utilize the receptor's structure in complex with the ligand for model creation. However, in the case of ligand-based or receptor-based pharmacophore models, the hypothesis might fail to capture locations where favorable binding contacts occur. This is why we positioned ligand-based models in second place, as complex-based models provide more precise information for identifying active molecules. This is evident when comparing the superior metrics of the complex-based pharmacophore ensemble models with those of ligand-based models. Additionally, another reason for prioritizing ligand-based models before complex-based ones is computational cost, as pharmacophore ensembles require more resources. Thus, our aim was to minimize the number of molecules reaching the complex models to optimize computational time and resources.

3.6 Experimental validation

After applying the virtual screening protocol (Supplementary Figures S2), we identified 18 candidate molecules based on their docking scores and MMGBSA free energy calculations. These molecules were prioritized for their predicted binding affinity and potential inhibitory activity against acetylcholinesterase (AChE).

To ensure a diverse representation across molecular families, we aimed to include at least one molecule from each family identified in the virtual screening results. The final selection was also constrained by commercial availability. Following an extensive search among

several suppliers, we acquired nine molecules that were readily available for purchase. The decision to proceed with commercially available compounds allowed us to maintain a balance between molecular diversity and feasibility for experimental validation.

These nine molecules, spanning different families, were subjected to *in vitro* AChE inhibition assays to experimentally validate their predicted activity. This approach ensured that our experimental testing covered a range of chemical scaffolds, providing insights into structure-activity relationships and enhancing the robustness of our validation process.

The IC₅₀ values for each tested molecule are presented in Table 6. Molecules 4, 5, 6, 7, 8, and 9 exhibited IC₅₀ values equal to or lower than the control, indicating potent inhibition of acetylcholinesterase activity. In contrast, molecule 3 displayed a higher IC₅₀ value, suggesting a weaker inhibitory effect. Molecules 1 and 2, however, did not yield consistent IC₅₀ values across replicates, which could be attributed to several experimental factors. Possible reasons for this inconsistency include poor solubility of the compounds in the assay buffer, inadequate interaction between the molecules and the enzyme, or potential interference from DMSO in the assay environment. These experimental variables may have impacted the reproducibility and accuracy of the observed inhibition, leading to unreliable IC₅₀ values for these molecules.

4 Conclusion

The intricate process of identifying potent human acetylcholinesterase enzyme inhibitors relies on the seamless integration of computational and experimental methodologies. A key initial step was the clustering analysis of 4,643 inhibitors from the binding database, which played a crucial role in organizing and managing the extensive dataset. This analysis categorized the inhibitors into 70 distinct clusters based on their molecular structures, significantly reducing computational complexity. Refining these clusters further to focus on those with higher concentrations of molecules and lower IC₅₀ values proved essential. This refinement, which resulted in nine significant clusters, underscored the importance of balancing computational efficiency with result reliability, ensuring that the most promising inhibitors were prioritized for further analysis.

Furthermore, combining the ensemble docking process with the innovative YN1 metric provided a sophisticated means to rank inhibitors by integrating both docking scores and experimental IC₅₀ values. This approach addressed the limitations of relying solely on computational predictions, which can sometimes overlook critical biological nuances. The YN1 score, ranging from 0 to 2, effectively balanced ligand efficiency and IC₅₀ values, offering a comprehensive assessment of a compound's inhibitory potential. The observed correlation between lower log (IC₅₀) values and higher YN1 scores validated this metric's utility. This dual-faceted ranking method successfully identified compounds with both high computational and experimental inhibitory activity, demonstrating the enhanced reliability of integrating diverse data sources.

Moreover, developing and validating ML models further strengthened the screening process. These models were trained on the refined set of active compounds and incorporated extensive decoy generation to ensure high specificity in distinguishing active from inactive compounds. The selection of the best-performing algorithm for each inhibitor family based on a comprehensive objective function ensured that the models were robust and precise. This integration of ML models into the screening pipeline not only enhanced efficiency but also improved the optimization of computational costs for the identification of viable huAChE inhibitors.

Additionally, the use of pharmacophore model ensembles provided a good metrics, showcasing their high capability to identify active molecules. The consideration of stable and unstable conformations to account for the flexibility of the protein's active site resulted in a improved performance compared to ligand-based models. We believe that the application of model ensembles could emerge as a powerful technique for virtual screening processes, especially when dealing with large numbers of molecules.

The experimental validation further strengthened these findings. Seven of the nine molecules tested showed strong inhibitory activity against huAChE, with molecules 4, 5, 6, 7, 8, and 9 demonstrating IC₅₀ values superior to the control, galantamine. It is noteworthy that the observed inhibitory capacities may be underestimated due to partial solubility of these molecules, suggesting that their effective concentrations could be even lower, potentially enhancing their IC₅₀ values. In contrast, molecules 1 and 2 exhibited inconsistent results, likely due to solubility issues in DMSO, which could be resolved by exploring alternative solvents in future studies.

Finally, the successful identification of promising inhibitors through the integrated approach of virtual screening, experimental validation, and advanced ML algorithms represents a significant achievement. This comprehensive methodology not only accelerates the identification of viable huAChE inhibitors but also establishes a reliable and efficient framework for drug discovery. The findings from this research open the door to designing new molecules with high affinity for huAChE, potentially paving the way for novel treatments for Alzheimer's disease. We believe that this study sets a precedent for future virtual screening efforts, combining computational, experimental, and machine learning approaches to discover potent enzyme inhibitors.

Data availability statement

The data that support the findings of this study are openly available in ZENODO: Supplementary Data 1: <https://zenodo.org/doi/10.5281/zenodo.11402870>; Supplementary Data 2: <https://zenodo.org/doi/10.5281/zenodo.11402907>.

Author contributions

YH-O: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing—original draft, Writing—review and editing. DA-C: Conceptualization, Data curation, Formal Analysis, Funding

acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing—original draft, Writing—review and editing. JS: Investigation, Resources, Writing—review and editing. BO-D: Investigation, Resources, Writing—review and editing. MV-T: Funding acquisition, Writing—review and editing. PC: Funding acquisition, Writing—review and editing. AB-O: Funding acquisition, Writing—review and editing. NC: Funding acquisition, Writing—review and editing. EO: Funding acquisition, Writing—review and editing. AG: Funding acquisition, Supervision, Writing—review and editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This work was funded by Ministerio de Ciencia, Tecnología e Innovación, grant No. 86978.

Acknowledgments

The authors gratefully acknowledge financial support from the Ministerio de Ciencia, Tecnología e Innovación, grant No. 86978.

References

- Akhood, B. A., Choudhary, S., Tiwari, H., Kumar, A., Barik, M. R., Rathor, L., et al. (2020). Discovery of a new donepezil-like acetylcholinesterase inhibitor for targeting Alzheimer's disease: computational studies with biological validation. *J. Chem. Inf. Model* 60, 4717–4729. doi:10.1021/acs.jcim.0c00496
- Amenta, F., Parnetti, L., Gallai, V., and Wallin, A. (2001). Treatment of cognitive dysfunction associated with Alzheimer's disease with cholinergic precursors. Ineffective treatments or inappropriate approaches? *Mech. Ageing Dev.* 122, 2025–2040. doi:10.1016/s0047-6374(01)00310-4
- Anonymous (2021a). Alzheimer's disease facts and figures. *Alzheimer's and Dementia* 17, 327–406. doi:10.1002/alz.12328
- Anonymous (2021b). Mortality and global health estimates. Available at: <https://www.who.int/data/gho/data/themes/mortality-and-global-health-estimates>.
- Aroniadou-Anderjaska, V., Figueiredo, T. H., de Araujo Furtado, M., Pidoplichko, V. I., and Braga, M. F. M. (2023). Mechanisms of organophosphate toxicity and the role of acetylcholinesterase inhibition. *Toxics* 11, 866. Preprint at. doi:10.3390/toxics11100866
- Branduardi, D., Gervasio, F. L., Cavalli, A., Recanatini, M., and Parrinello, M. (2005). The role of the peripheral anionic site and cation- π interactions in the ligand penetration of the human AChE gorge. *J. Am. Chem. Soc.* 127, 9147–9155. doi:10.1021/ja0512780
- Breijyeh, Z., and Karaman, R. (2020). Comprehensive review on alzheimer's disease: causes and treatment. *Molecules* 25, 5789. doi:10.3390/molecules25245789
- Camps, P., and Munoz-Torrero, D. (2005). Cholinergic drugs in pharmacotherapy of alzheimer's disease. *Mini-Reviews Med. Chem.* 2, 11–25. doi:10.2174/1389557023406638
- Carotenuto, A., Fasanaro, A. M., Manzo, V., Amenta, F., and Traini, E. (2022). Association between the cholinesterase inhibitor donepezil and the cholinergic precursor choline alphoscerate in the treatment of depression in patients with alzheimer's disease. *J. Alzheimers Dis. Rep.* 6, 235–243. doi:10.3233/adr-200269
- Carracedo-Reboredo, P., Liñares-Blanco, J., Rodríguez-Fernández, N., Cedrón, F., Novoa, F. J., Carballal, A., et al. (2021). A review on machine learning approaches and trends in drug discovery. *Comput. Struct. Biotechnol. J.* 19, 4538–4558. doi:10.1016/j.csbj.2021.08.011
- Célerse, F., Lagardère, L., Bouchibti, Y., Nachon, F., Verdier, L., Piquemal, J. P., et al. (2024). Unveiling the full dynamical and reactivity profiles of acetylcholinesterase: a comprehensive all-atom investigation. *ACS Catal.* 14, 1785–1796. doi:10.1021/ACSCATAL.3C05560
- Cer, R. Z., Mudunuri, U., Stephens, R., and Lebeda, F. J. (2009). IC50-to-Ki: a web-based tool for converting IC50 to Ki values for inhibitors of enzyme activity and ligand binding. *Nucleic Acids Res.* 37, W441–W445. doi:10.1093/nar/gkp253
- Cheng, S., Song, W., Yuan, X., and Xu, Y. (2017). Gorge motions of acetylcholinesterase revealed by microsecond molecular dynamics simulations. *Sci. Rep.* 7 (1), 3219. doi:10.1038/s41598-017-03088-y
- Cheung, J., Rudolph, M. J., Burshteyn, F., Cassidy, M. S., Gary, E. N., Love, J., et al. (2012). Structures of human acetylcholinesterase in complex with pharmacologically important ligands. *J. Med. Chem.* 55, 10282–10286. doi:10.1021/jm300871x
- Crestini, A., Santilli, F., Martellucci, S., Carbone, E., Soric, M., Piscopo, P., et al. (2022). Prions and neurodegenerative diseases: a focus on alzheimer's disease. *J. Alzheimer's Dis.* 85, 503–518. doi:10.3233/jad-215171
- Dara, S., Dhamercherla, S., Jadav, S. S., Babu, C. M., and Ahsan, M. J. (2022). Machine learning in drug discovery: a review. *Artif. Intell. Rev.* 55, 1947–1999. doi:10.1007/s10462-021-10058-4
- Delacourte, A. (2006). The natural and molecular history of Alzheimer's disease. *J. Alzheimer's Dis.* 9, 187–194. doi:10.3233/jad-2006-9s322
- Dvir, H., Silman, I., Harel, M., Rosenberry, T. L., and Sussman, J. L. (2010). Acetylcholinesterase: from 3D structure to function. *Chem. Biol. Interact.* 187, 10–22. doi:10.1016/j.cbi.2010.01.042
- Ejaz, H. W., Wang, W., and Lang, M. (2020). Copper toxicity links to pathogenesis of Alzheimer's disease and therapeutics approaches. *Int. J. Mol. Sci.* 21, 7660. doi:10.3390/ijms21207660
- Ertl, P., Schuffenhauer, A., and Renner, S. The scaffold tree: an efficient navigation in the scaffold universe. *Methods Mol. Biol. Life Sci.* 672, 245, 260. doi:10.1007/978-1-60761-839-3_102011
- Galburt, E. A., and Tomko, E. J. (2017). Conformational selection and induced fit as a useful framework for molecular motor mechanisms. *Biophys. Chem.* 223, 11–16. doi:10.1016/j.bpc.2017.01.004
- Gao, W., Ma, X., Yang, H., Luan, Y., and Ai, H. (2022). Molecular engineering and activity improvement of acetylcholinesterase inhibitors: insights from 3D-QSAR, docking, and molecular dynamics simulation studies. *J. Mol. Graph Model* 116, 108239. doi:10.1016/j.jmgm.2022.108239
- Geng, C., Wang, Z. B., and Tang, Y. (2024). Machine learning in Alzheimer's disease drug discovery and target identification. *Ageing Res. Rev.* 93, 102172. doi:10.1016/j.arr.2023.102172

We thank Dirección de Servicios de Información y Tecnología (DSIT) staff at Universidad de los Andes, for the HPC supporting.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fchem.2025.1479763/full#supplementary-material>

- Giacobini, E., Cuello, A. C., and Fisher, A. (2022). Reimagining cholinergic therapy for Alzheimer's disease. *Brain* 145, 2250–2275. doi:10.1093/brain/awac096
- Gupta, R., Srivastava, D., Sahu, M., Tiwari, S., Ambasta, R. K., and Kumar, P. (2021). Artificial intelligence to deep learning: machine intelligence approach for drug discovery. *Mol. Divers* 25, 1315–1360. doi:10.1007/s11030-021-10217-3
- Gustavsson, A., Norton, N., Fast, T., Frölich, L., Georges, J., Holzapfel, D., et al. (2023). Global estimates on the number of persons across the Alzheimer's disease continuum. *Alzheimers Dement.* 19, 658–670. doi:10.1002/alz.12694
- Heilmann, N., Wolf, M., Kozłowska, M., Sedghamiz, E., Setzler, J., Brieg, M., et al. (2020). Sampling of the conformational landscape of small proteins with Monte Carlo methods. *Sci. Rep.* 10 (1), 18211–18213. doi:10.1038/s41598-020-75239-7
- Jang, C., Yadav, D. K., Subedi, L., Venkatesan, R., Venkanna, A., Afzal, S., et al. (2018). Identification of novel acetylcholinesterase inhibitors designed by pharmacophore-based virtual screening, molecular docking and bioassay. *Sci. Rep.* 8, 14921. doi:10.1038/s41598-018-33354-6
- Jomova, K., Raptova, R., Alomar, S. Y., Alwasel, S. H., Nepovimova, E., Kuca, K., et al. (2023). Reactive oxygen species, toxicity, oxidative stress, and antioxidants: chronic diseases and aging. *Arch. Toxicol.* 97, 2499–2574. doi:10.1007/s00204-023-03562-9
- Kaynak, B. T., Krieger, J. M., Dudas, B., Dahmani, Z. L., Costa, M. G. S., Balog, E., et al. (2022). Sampling of protein conformational space using hybrid simulations: a critical assessment of recent methods. *Front. Mol. Biosci.* 9, 832847. doi:10.3389/fmolb.2022.832847
- Khedkar, S., Malde, A., Coutinho, E., and Srivastava, S. (2007). Pharmacophore modeling in drug discovery and development: an overview. *Med. Chem.* 3, 187–197. doi:10.2174/157340607780059521
- Kokh, D. B., Czodrowski, P., Rippmann, F., and Wade, R. C. (2016). Perturbation approaches for exploring protein binding site flexibility to predict transient binding pockets. *J. Chem. Theory Comput.* 12, 4100–4113. doi:10.1021/acs.jctc.6b00101
- Kokh, D. B., Richter, S., Henrich, S., Czodrowski, P., Rippmann, F., and Wade, R. C. (2013). TRAPP: a tool for analysis of Transient binding Pockets in Proteins. *J. Chem. Inf. Model* 53, 1235–1252. doi:10.1021/ci4000294
- Komersová, A., Komers, K., and Zdražilová, P. (2005). (23) Kinetics of hydrolysis of acetylthiocholine and acetylcholine by cholinesterases. *Chem. Biol. Interact.* 157–158, 387–388. doi:10.1016/j.cbi.2005.10.068
- Korczyn, A. D., and Grinberg, L. T. (2024). Is Alzheimer disease a disease? *Nat. Rev. Neurol.* 20 (4), 245–251. doi:10.1038/s41582-024-00940-4
- Langer, T., Pharmacophores, R. D., and Searches, P. (2006) *Pharmacophores and pharmacophore searches*, 32. Wiley.
- Lazim, R., Suh, D., and Choi, S. (2020). Advances in molecular dynamics simulations and enhanced sampling methods for the study of protein systems. *Int. J. Mol. Sci.* 21, 6339. doi:10.3390/ijms21176339
- Leng, F., and Edison, P. (2020). Neuroinflammation and microglial activation in Alzheimer disease: where do we go from here? *Nat. Rev. Neurol.* 2020 17 (3), 157–172. doi:10.1038/s41582-020-00435-y
- Li, Y., Li, X., Wang, W., Guo, R., and Huang, X. (2023). Spatiotemporal evolution and characteristics of worldwide life expectancy. *Environ. Sci. Pollut. Res.* 30, 87145–87157. doi:10.1007/s11356-023-28330-1
- Li, Z., Huang, R., Xia, M., Patterson, T. A., and Hong, H. (2024). Fingerprinting interactions between proteins and ligands for facilitating machine learning in drug discovery. *Biomolecules* 14, 72. doi:10.3390/biom14010072
- Landrum, G. (1997). *RDKit: Open-source cheminformatics*. Available at: <https://www.rdkit.org> (Accessed June, 2024)
- Marucci, G., Buccioni, M., Ben, D. D., Lambertucci, C., Volpini, R., and Amenta, F. (2021). Efficacy of acetylcholinesterase inhibitors in Alzheimer's disease. *Neuropharmacology* 190, 108352. Preprint at. doi:10.1016/j.neuropharm.2020.108352
- Moghadam, B., Ashouri, M., Roohi, H., and Karimi-jafari, M. H. (2021). Computational evidence of new putative allosteric sites in the acetylcholinesterase receptor. *J. Mol. Graph Model* 107, 107981. doi:10.1016/j.jmgm.2021.107981
- Mysinger, M. M., Carchia, M., Irwin, J. J., and Shoichet, B. K. (2012). Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking. *J. Med. Chem.* 55, 6582–6594. doi:10.1021/jm300687e
- Pérez-Sánchez, H., den Haan, H., Pérez-Garrido, A., Peña-García, J., Chakraborty, S., Erdogan Orhan, I., et al. (2021). Combined structure and ligand-based design of selective acetylcholinesterase inhibitors. *J. Chem. Inf. Model* 61, 467–480. doi:10.1021/acs.jcim.0c00463
- Prince, M., Ali, G. C., Guerchet, M., Prina, A. M., Albanese, E., and Wu, Y. T. (2016). Recent global trends in the prevalence and incidence of dementia, and survival with dementia. *Alzheimers Res. Ther.* 8, 23–13. doi:10.1186/s13195-016-0188-8
- Qing, X., Lee, X. Y., De Raeymaecker, J., Tame, J., Zhang, K., De Maeyer, M., et al. (2014). Pharmacophore modeling: advances, Limitations, and current utility in drug discovery. *J. Recept. Ligand Channel Res.* 7, 81–92. doi:10.2147/jrlcr.s46843
- Raves, M. L., Harel, M., Pang, Y. P., Silman, I., Kozikowski, A. P., and Sussman, J. L. (1997). Structure of acetylcholinesterase complexed with the nootropic alkaloid, (–)-huperzine A. *Nat. Struct. Biol.* 4 (1), 57–63. doi:10.1038/nsb0197-57
- Sanson, B., Colletier, J., Xu, Y., Lang, P. T., Jiang, H., Silman, I., et al. (2011). Backdoor opening mechanism in acetylcholinesterase based on X-ray crystallography and molecular dynamics simulations. *Protein Sci.* 20, 1114–1118. doi:10.1002/pro.661
- Scheltens, P., De Strooper, B., Kivipelto, M., Holstege, H., Chêtelat, G., Teunissen, C. E., et al. (2021). Alzheimer's disease. *Lancet* 397, 1577–1590. doi:10.1016/s0140-6736(20)32205-4
- Silman, I., and Sussman, J. L. (2017). Recent developments in structural studies on acetylcholinesterase. *J. Neurochem.* 142, 19–25. Preprint at. doi:10.1111/jnc.13992
- Silva, M. V. F., Loures, C. d. M. G., Alves, L. C. V., de Souza, L. C., Borges, K. B. G., and Carvalho, M. d. G. (2019). Alzheimer's disease: risk factors and potentially protective measures. *J. Biomed. Sci.* 26, 33. doi:10.1186/s12929-019-0524-y
- Stepkowski, D., Studnicki, M., Woźniak, G., and Dębski, K. J. (2020). Diet as a risk factor for Alzheimer's disease. *Postepy Biochem.* 66, 19–22. doi:10.18388/pb.2020.311
- Tingle, B. I., Tang, K. G., Castanon, M., Gutierrez, J. J., Khurelbaatar, M., Dandarchuluun, C., et al. (2023). ZINC-22-A free multi-billion-scale database of tangible compounds for ligand discovery. *J. Chem. Inf. Model* 63, 1166–1176. doi:10.1021/acs.jcim.2c01253
- Tolar, M., Hey, J. A., Power, A., and Abushakra, S. (2024). The single toxin origin of Alzheimer's disease and other neurodegenerative disorders enables targeted approach to treatment and prevention. *Int. J. Mol. Sci.* 25, 2727. doi:10.3390/ijms25052727
- van Drie, J. H. (2012). Pharmacophore discovery - lessons learned. *Front. Med. Chem.* - 2, 511–532. doi:10.2174/978160805205910502010511
- Vojtechova, I., Machacek, T., Kristofikova, Z., Stuchlik, A., and Petrasek, T. (2022). Infectious origin of Alzheimer's disease: amyloid beta as a component of brain antimicrobial immunity. *PLoS Pathog.* 18, e1010929. doi:10.1371/journal.ppat.1010929
- Walczak-Nowicka, Ł. J., and Herbet, M. (2021). Acetylcholinesterase inhibitors in the treatment of neurodegenerative diseases and the role of acetylcholinesterase in their pathogenesis. *Int. J. Mol. Sci.* 22, 9290. Preprint at. doi:10.3390/ijms22179290
- Wang, C., Zong, S., Cui, X., Wang, X., Wu, S., Wang, L., et al. (2023). The effects of microglia-associated neuroinflammation on Alzheimer's disease. *Front. Immunol.* 14, 1117172. doi:10.3389/fimmu.2023.1117172
- Wessler, I., Michel-Schmidt, R., and Kirkpatrick, C. J. (2015). pH-dependent hydrolysis of acetylcholine: consequences for non-neuronal acetylcholine. *Int. Immunopharmacol.* 29, 27–30. doi:10.1016/j.intimp.2015.04.039
- Whitfield, J. F., Rennie, K., and Chakravarthy, B. (2023). Alzheimer's disease and its possible evolutionary origin: hypothesis. *Cells* 12, 1618. doi:10.3390/cells12121618
- Wilkinson, D. G., Francis, P. T., Schwam, E., and Payne-Parrish, J. (2004). Cholinesterase inhibitors used in the treatment of Alzheimer's disease: the relationship between pharmacological effects and clinical efficacy. *Drugs Aging* 21, 453–478. doi:10.2165/00002512-200421070-00004
- Wlodek, S. T., Clark, T. W., Scott, L. R., and McCammon, J. A. (1997). Molecular dynamics of acetylcholinesterase dimer complexed with tacrine. *J. Am. Chem. Soc.* 119, 9513–9522. doi:10.1021/ja971226d
- Wu, J., Pistolozzi, M., Liu, S., and Tan, W. (2020). Design, synthesis and biological evaluation of novel carbamates as potential inhibitors of acetylcholinesterase and butyrylcholinesterase. *Bioorg Med. Chem.* 28, 115324. doi:10.1016/j.bmc.2020.115324
- Yasuda, T., Shigeta, Y., and Harada, R. (2020). Efficient conformational sampling of collective motions of proteins with principal component analysis-based parallel cascade selection molecular dynamics. *J. Chem. Inf. Model* 60, 4021–4029. doi:10.1021/acs.jcim.0c00580
- Yuan, J. H., Han, S. B., Richter, S., Wade, R. C., and Kokh, D. B. (2020). Druggability assessment in TRAPP using machine learning approaches. *J. Chem. Inf. Model* 60, 1685–1699. doi:10.1021/acs.jcim.9b01185
- Zhou, L. y., Zhu, Y., Jiang, Y., ren, Z., jie, X., and Guo, D. (2017). Design, synthesis and biological evaluation of dual acetylcholinesterase and phosphodiesterase 5A inhibitors in treatment for Alzheimer's disease. *Bioorg Med. Chem. Lett.* 27, 4180–4184. doi:10.1016/j.bmcl.2017.07.013