Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Rational Design and In Silico Evaluation of Benzimidazole-Fused Pyran Derivatives as Acetylcholinesterase Inhibitors for Alzheimer's Disease

Shivank Sharma
Lovely Professional University
Vanktesh Kumar
Lovely Professional University
Stalin Arulsamy
Lovely Professional University
Pankaj Wadhwa
Ummeed Institute of Pharmaceutical Sciences and Research Center,Khatlabana
Shubham Kumar
Shubhamm771@gmail.com
Lovely Professional University

Research Article

Keywords: Acetylcholinesterase inhibitors, Benzimidazole-dihydropyran hybrids, Alzheimer's disease, Structure-based drug design, Machine learning-assisted QSAR

Posted Date: April 17th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8993822/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Rational Design and *In Silico* Evaluation of Benzimidazole-Fused Pyran Derivatives as Acetylcholinesterase Inhibitors for Alzheimer's Disease

Shivank Sharma[1], Vanktesh Kumar[1], Stalin Arulsamy[1], Pankaj Wadhwa[2]*, Shubham Kumar[1]*

[1]*Dept. of Pharmaceutical Chemistry, School of Pharmaceutical Sciences, Lovely Professional University, Jalandhar-Delhi G.T. Road, Phagwara, Punjab 144411, India.*

[2]*Department of Pharmaceutical Chemistry, Ummeed Institute of Pharmaceutical Sciences and Research Center,Khatlabana, Sriganganagar,Rajasthan, 335001 IN*

* Corresponding Author

Dr Pankaj Wadhwa

Dr Shubham Kumar

Department of Pharmaceutical Chemistry,

School of Pharmaceutical Sciences, Lovely Professional University,

Jalandhar-Delhi G.T. Road (NH-1), Phagwara (Punjab) 144411, India;

Phone No. (+) 91-9461754780

E-mail address: pankajwadhwa88@gmail.com

Shubhamm771@gmail.com

Abstract

Alzheimer's disease (AD) is a progressive neurodegenerative disorder marked by cholinergic dysfunction, for which acetylcholinesterase (AChE) inhibition remains a clinically validated symptomatic treatment strategy. In the present study, a series of benzimidazole-dihydropyran hybrid derivatives (**4a-n**) were rationally designed and evaluated using an integrated computational workflow to identify promising AChE inhibitors. Molecular docking demonstrated favorable binding of several derivatives within both the catalytic active site and peripheral anionic site of AChE, stabilized by π-π stacking, hydrogen bonding, and hydrophobic interactions with key residues. ADMET profiling predicted acceptable oral bioavailability, blood-brain barrier permeability, and drug-like properties for the prioritized compounds. Density functional theory analysis revealed that the most active candidates possessed narrow HOMO-LUMO energy gaps, indicating enhanced electronic complementarity and charge-transfer capability within the AChE binding gorge. Machine learning-assisted QSAR modeling further supported the activity trends and highlighted compounds **4b, 4c, 4f,** and **4i** as the most promising leads. These results establish the benzimidazole-dihydropyran scaffold as a viable platform for the development of next-generation AChE inhibitors and provide a robust computational basis for subsequent experimental validation toward Alzheimer's disease therapy.

Keywords: Acetylcholinesterase inhibitors; Benzimidazole-dihydropyran hybrids; Alzheimer's disease; Structure-based drug design; Machine learning-assisted QSAR

1. Introduction

The most prevalent type of dementia, Alzheimer disease (AD), affects more than 50 million people worldwide and is expected to reach 152 million by 2050, and associated healthcare costs expected to reach $2 trillion by 2030.[1] Along with non-cognitive symptoms including depression, hallucinations, and behavioral abnormalities, AD is typified by cognitive deficits like memory loss, language barriers, and thinking issues.[2] It is the leading cause of dementia in older adults (major cases after 65) and the multifactorial nature of AD includes oxidative stress, neuroinflammation, tau hyperphosphorylation, amyloid-beta (Aβ) aggregation, and mitochondrial dysfunction.[3, 4] Along with Aβ plaques and tau tangles, AD is also impacted by neuroinflammation, cardiovascular health, and genetic variables.[5] Enzymes called monoamine oxidase A (MAO-A) and monoamine oxidase B (MAO-B) degrade neurotransmitters such as serotonin and dopamine.[6, 7] Inhibiting MAO-B is a promising treatment approach for AD since it may help prevent oxidative damage and maintain cognitive function. Treatment of depression and mood modulation are more closely linked to MAO-A inhibition.[8] A classical theory in AD pathology proposed in 1976, states that "the neurotransmitter ACh, in the hippocampus and the neocortex regions, plays a key role in learning and memory, and the remarkable decline in the level of ACh is primarily responsible for the dementia of AD patients."[9-11] Under physiological conditions, Acetylcholinesterase (AChE) can hydrolyze ACh to choline and acetate efficiently.[12, 13] Thus, inhibition against AChE is an advisable way to enhance the concentration of ACh in the synaptic cleft, making it a crucial target in the cholinergic system.[14] Enzymes including AChE, GSK-3β, CDK5, secretases (β- and γ-secretase), and NMDA receptors all play crucial roles in the complex pathophysiology of Alzheimer's disease (AD).[15, 16] One important therapeutic approach is AChE inhibition, which elevates synaptic acetylcholine (ACh) levels, thereby improving cholinergic neurotransmission and offering symptomatic relief.[17] Moreover, inhibition of AChE has been shown to reduce formation of NFTs through the modulation of hyperphosphorylation of tau protein and/or the attenuation of Aβ aggregate production (indirect effects) [18]. Other approaches targeting the 2 major kinases needed for the hyperphosphorylation of tau are GSK-3β and CDK5; and secretase inhibitors (aiming to reduce aggregation of Aβ peptides) act by disrupting the cleaving of amyloid precursor protein (APP) [19]. There are 8 pathological processes involved in Alzheimer's disease; the pathways associated with these processes all contribute to the neurodegenerative process in terms of both structure and function, thereby causing cognitive decline, and are summarized in Table 1 below.

Aβ deposition, tau hyperphosphorylation, loss of cholinergic function, activation of neuroinflammation, valvular degeneration, mitochondrial dysregulation, calcium overload due to stimulating glutamate receptors, and disruption of biometal homeostasis directly interfere with and ultimately produce a cascade that destroys the intact neuronal signalling and creates a variety of different signs and symptoms including loss of memory, altered behaviour, and cognitive dysfunction.

Table 1: Mechanism, description & effect on AD progression

Mechanism	Description	Effect on AD Progression
Aβ plaques	Abnormal cleavage of amyloid precursor protein (APP) leads to toxic Aβ1-42 aggregation, forming plaques.	Disrupts neuronal communication, triggers inflammation, and induces synaptic dysfunction.[20]
Tau protein tangles	Hyperphosphorylation of tau causes it to detach from microtubules and form neurofibrillary tangles (NFTs).	Destroys microtubule networks, leading to cell transport failure and neuronal death.[21]
Cholinergic dysfunction	Low ACh levels due to increased AChE activity in AD patients.	Impairs memory, learning, and cognition.[22]
Neuroinflammation	Microglial activation and astrocyte response release pro-inflammatory cytokines (TNF-α, IL-1β, IL-6).	Exacerbates neuronal damage and synaptic loss.[23]
Oxidative stress	Excessive reactive oxygen species (ROS) damage lipids, proteins, and DNA.	Leads to neuronal dysfunction, apoptosis, and brain atrophy.[24]
Excitotoxicity	Glutamate overactivation causes excessive calcium influx through NMDA receptors.	Leads to mitochondrial damage, synaptic loss, and cell death.[25]
Mitochondrial dysfunction	Impaired energy production and calcium dysregulation in brain cells.	Reduces ATP production, increasing cell vulnerability to damage.[26]

Biometal dysregulation	Zinc, copper, and iron imbalance promotes Aβ aggregation.	Enhances plaque formation and oxidative stress.[27]

The mechanism of AChE inhibition **(Figure 1)**, highlighting how blocking AChE prevents the rapid hydrolysis of acetylcholine in the synaptic cleft. By inhibiting the enzyme, ACh levels remain elevated, strengthening cholinergic neurotransmission and improving memory and cognitive function impaired in AD.

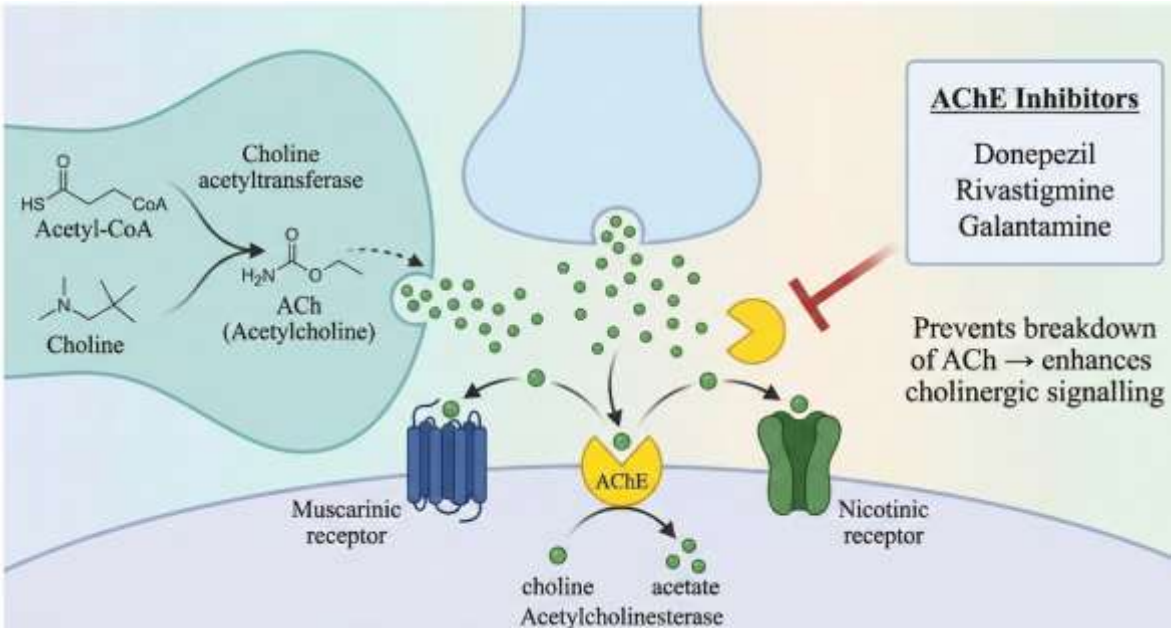


Figure 1: Mechanism for Acetylcholinesterase inhibition

1.1. Marketed Drugs

Figure 2 presents the major marketed drugs currently approved for AD therapy, including cholinesterase inhibitors and NMDA receptor antagonists. These agents provide symptomatic relief by enhancing cholinergic signaling or reducing glutamate-mediated excitotoxicity. Although effective in improving cognition and daily functioning, their benefits remain limited, and they do not halt the underlying neurodegenerative progression, underscoring the need for new therapeutic scaffolds.

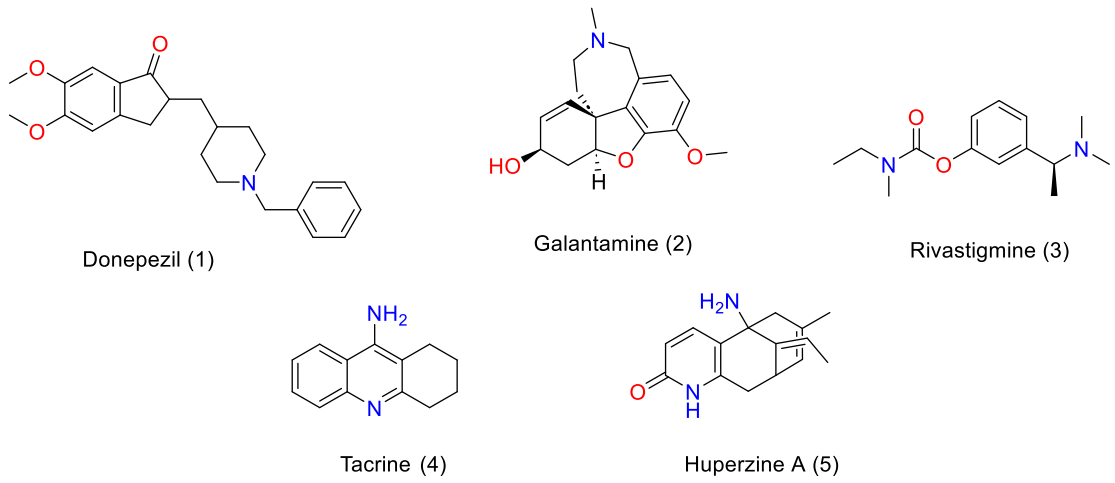


Figure 2: Marketed drugs for Alzheimer Disease

1.2. Rationale

Treatments available currently (including anti-amyloid drugs like Aducanumab and cholinesterase inhibitors such as Donepezil) help temporarily with the symptoms caused by Alzheimer's Disease but cause serious side effects (including hepatotoxicity and gastrointestinal effects) and do not cure the disease [4]. Alzheimer's Disease (AD) is caused by the accumulation of amyloid-beta peptide (Aβ) and the excess phosphorylation of tau proteins, which result in less cognitive ability. AD also has lower levels of acetylcholine (ACh), which makes it difficult for people to think, move and perform daily tasks, and cause these problems. ACh is broken down into acetate and choline by the enzyme acetylcholinesterase (AChE) which prevents nerve impulses. Inhibiting AChE from breaking down ACh allows more ACh to remain in the synaptic cleft, and can potentially improve cognitive function and performance [28].

To address these limitations, I synthesized a number of new benzimidazole derivatives as privileged structures for medicinal chemistry because of their excellent biological activity, metabolic stability, and ability to bind to key biological targets. Benzimidazoles were studied for their ability to inhibit cholinesterase, their antioxidant effects, and their neuroprotective potential as candidates for the treatment of AD. Benzimidazole's unique heterocyclic structure enables the compound to form π-π stacking, hydrogen bonding, and hydrophobic interactions with both the active site and peripheral anionic site of AChE. This dual binding potential can

not only inhibit AChE but also reduce amyloid-beta (Aβ) aggregation, a secondary neurotoxic event linked to AD progression. The main goal is to design and create compounds based on benzimidazoles to investigate novel potential AChE inhibitors. My research attempts to find potent compounds that may be used as therapeutic leads to treat AD. In 2024, Yaseen S. *et al.* created and assessed new benzimidazole compounds conjugated with lipoic acid, coumaric acid, and cinnamic acid as AChE inhibitors for possible treatment of AD. The compounds were prepared by employing thionyl chloride ($SOCl_2$) in anhydrous THF to carboxylate activate p-coumaric acid, cinnamic acid, or lipoic acid, and then coupling with benzimidazole in the presence of triethylamine. And among them, comp-a is referred as most potent compound (IC_{50} = 13.008 ± 0.43 μg/ml) but less potent than the standard drug galantamine (IC_{50} = 9.169 ± 0.48 μg/ml). From SAR it was found that the hydroxyl (-OH) group, conjugated α,β-unsaturated ketone, and benzimidazole core in compound d enhanced the hydrogen bonding, π-π stacking, and hydrophobic interactions making comp-d most potent.[29] In 2023, Adalat B. *et al.* condensated 4-methoxybenzene-1,2-diamine with different substituted benzaldehydes in DMF using sodium meta-bisulfite as a catalyst under reflux conditions to create and assess benzimidazole-based benzaldehyde derivatives. Where comp-b was found to be the most potent AChE inhibitor (IC_{50} = 0.050 μM) but 3 fold less potent than referenced drug donepezil (IC_{50} = 0.016 ± 0.12 μM). SAR findings revealed that the highest AChE inhibition was produced by two chlorine (-Cl) groups at positions three and four of the phenyl ring. According to molecular docking results, comp-b has a higher binding affinity (-56.42 kcal/mol) than the reference medication donepezil (-50.66 kcal/mol), producing strong hydrogen bonds, π-π stacking, and electrostatic interactions.[30] In 2023, Hussain R. *et al*. designed, synthesized and evaluated benzimidazole-based thiazole derivatives as potential inhibitors of AChE for the treatment of AD. The compounds were created using a three-step procedure that included cyclization with substituted phenacyl bromides in ethanol under reflux and basic conditions, condensation with thiosemicarbazide, and S-alkylation of 2-mercaptobenzimidazole. Among all the synthesized compounds, comp-c (binding free energy -11.23 kcal/mol) exhibits the maximum potential with IC_{50} of 0.10 ± 0.05 μM (21 fold) as compared to the standard drug donepezil (IC_{50} 2.16 ± 0.12 μM). SAR study revealed that phenyl rings with EWG (-Cl) enhanced enzyme inhibition, most likely *via* improving interactions inside the active site.[31] In 2022, Khan Y. *et al.* synthesized and evaluated a series of benzimidazole-based triazole derivatives as inhibitors of AChE for potential use in AD treatment. The synthesis of compounds was done by cyclization with an acid catalyst, hydrazide production, reaction with phenylisothiocyanates, and final cyclization with NaOH. Among them, compound d was

highlighted as the most potent AChE inhibitor (IC_{50} = 0.40 ± 0.05 μM) which was 25 fold less potent than the standard drug donepezil (IC_{50} = 0.016 ± 0.12 μM). SAR revealed that enzyme inhibition was greatly increased by EWG (-NO_2) at the *p*-position of the aryl ring.[32] In 2022, Alim Z. *et al.* synthesized and evaluated a series of benzimidazole-based amide derivatives as inhibitors of AChE. In the presence of potassium carbonate in acetonitrile under reflux, molecules were created by nucleophilically substituting benzimidazole derivatives with 2-chloroacetamides, and comp-e was regarded as the most potent compound with IC_{50} of 0.936 μM which was 1.3 fold less potent than standard drug tacrine (IC_{50} = 0.697 μM). According to SAR analysis, 4-chlorophenyl substitution increases enzyme inhibition through halogen bonding and hydrophobic interactions.[33] In 2016, Boulebd *et al.* developed novel imidazopyranotacrines, a new class of tacrine analogues, as potential AD treatments targeting AChE inhibition. These compounds were synthesized using 4H-pyrans derived from 1-methyl-1H-imidazole-2-carbaldehyde and cycloalkanones or naphthol derivatives. Tested *via* Ellman's assay for EeAChE inhibition, the most effective compound (compound f) had an IC_{50} of 6.73 ± 0.52 μM, 74 times less potent than tacrine (IC_{50} = 0.0898 ± 0.0022 μM). SAR studies showed that fusing the 4H-pyran core with a naphthol ring significantly improved inhibition compared to other substituents, likely due to optimized electronic effects. A cyclohexane ring was found to enhance potency over cyclopentane or cycloheptane, suggesting an optimal ring size, with the nitrogen atom likely contributing to enzyme binding through hydrogen bonding or electronic interactions.[34] In 2018, Kumar *et al.* synthesized a series of 2-amino-4-aryl-8-[(E)-arylmethylidene]-5,6,7,8-tetrahydro-4H-pyrano[3,2-c]pyridin-3-carbonitriles and tested their *in vitro* AChE inhibitory activity, using galantamine (IC_{50} = 2.09 ± 0.13 μM) as a reference. The most potent compound, comp-g, showed an IC_{50} of 1.98 ± 0.09 μM. Nitro groups on both aryl rings (C4 and C8) enhanced electron-withdrawing effects, likely improving AChE active site binding. The presence of 2-nitro and para-methyl groups on 2-amino-4H-pyrans boosted inhibition through π-π and hydrophobic interactions, while ortho-halogens and di-substitution reduced potency due to steric hindrance.[35] In 2019, Swetha *et al.* synthesized a series of 3-((4-substituted)sulfonamido)-N-(indole/quinolin-3/6/8-yl)ethan/propanamides as multifunctional hybrid sulfonamides for AD therapy, targeting AChE inhibition. The most potent compound, comp h (IC_{50} = 2.064 ± 0.27 μM), was less potent than donepezil (IC_{50} = 0.023 ± 0.018 μM) but showed non-toxic, multifactorial activity against AD. Biphenyl sulfonamides, like comp h, outperformed monophenyl analogs, with the sulfonamide group stabilizing π-alkyl interactions with Leu289 and Val294. The electron-withdrawing -SO_2- group enhanced π-π and hydrophobic interactions of p-chloro or p-fluoro biphenyls with

Trp286 and Tyr337, and a three-carbon linker improved binding over a two-carbon linker. The sulfonamide's polarity and flexibility supported interactions with 8-quinolinyl and 3-indolyl cores, though aryl and heterocyclic components primarily drove potency.[36] The design rationale **(Figure 4)** behind the benzimidazole-dihydropyran-tosyl hybrid scaffold. The strategy integrates three pharmacophoric elements expected to enhance AChE inhibition: the benzimidazole ring for π-π stacking and hydrogen bonding, the dihydropyran core for improved rigidity and spatial orientation within the active site, and the tosyl group for additional hydrophobic interactions. Together, these features are intended to support dual engagement at the catalytic and peripheral anionic sites of AChE, potentially improving both inhibitory potency and anti-amyloid activity.

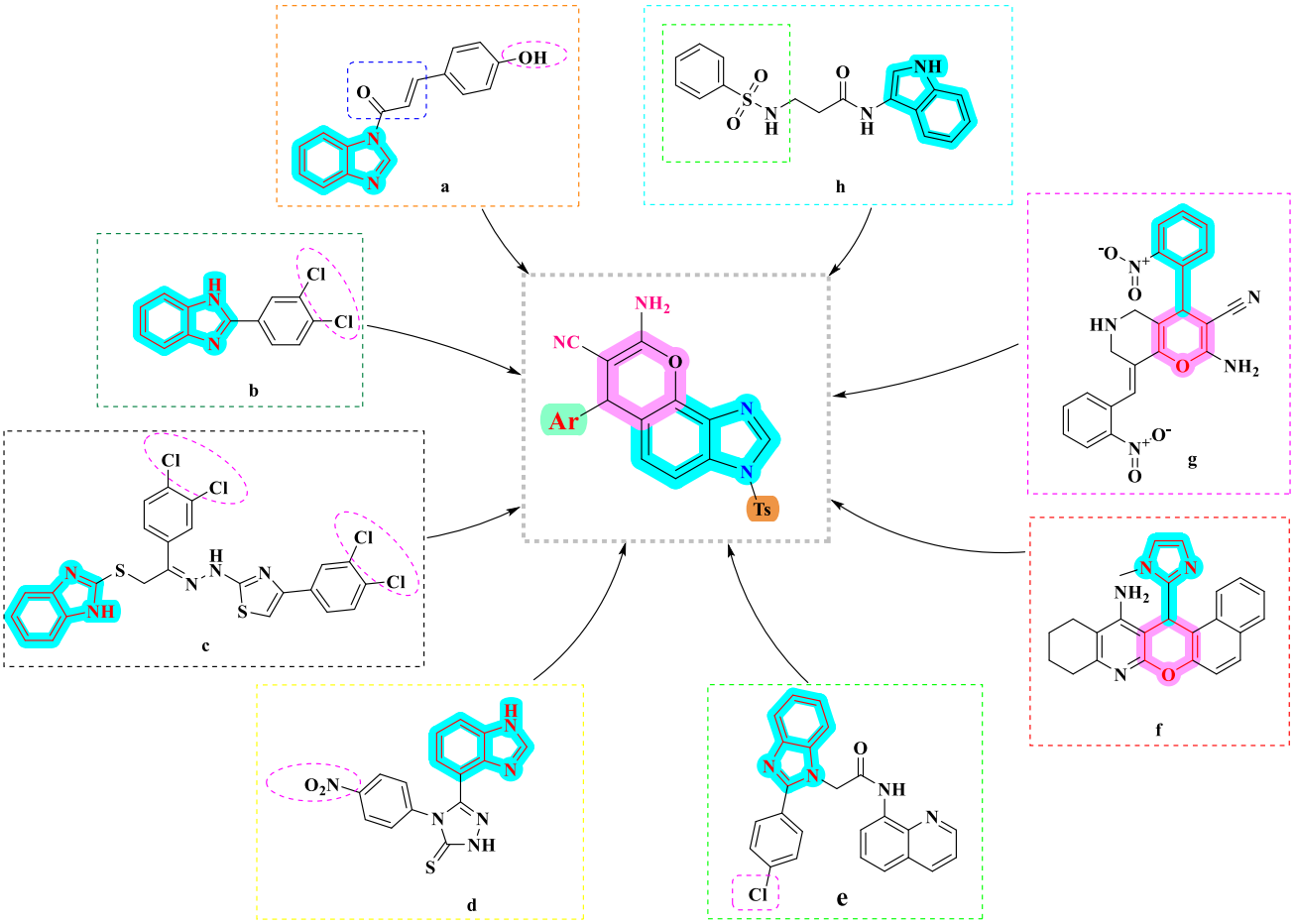


Figure 3: Comparison of reported benzimidazole-based AChE inhibitors (a-h) with the newly designed benzimidazole-fused pyran scaffold

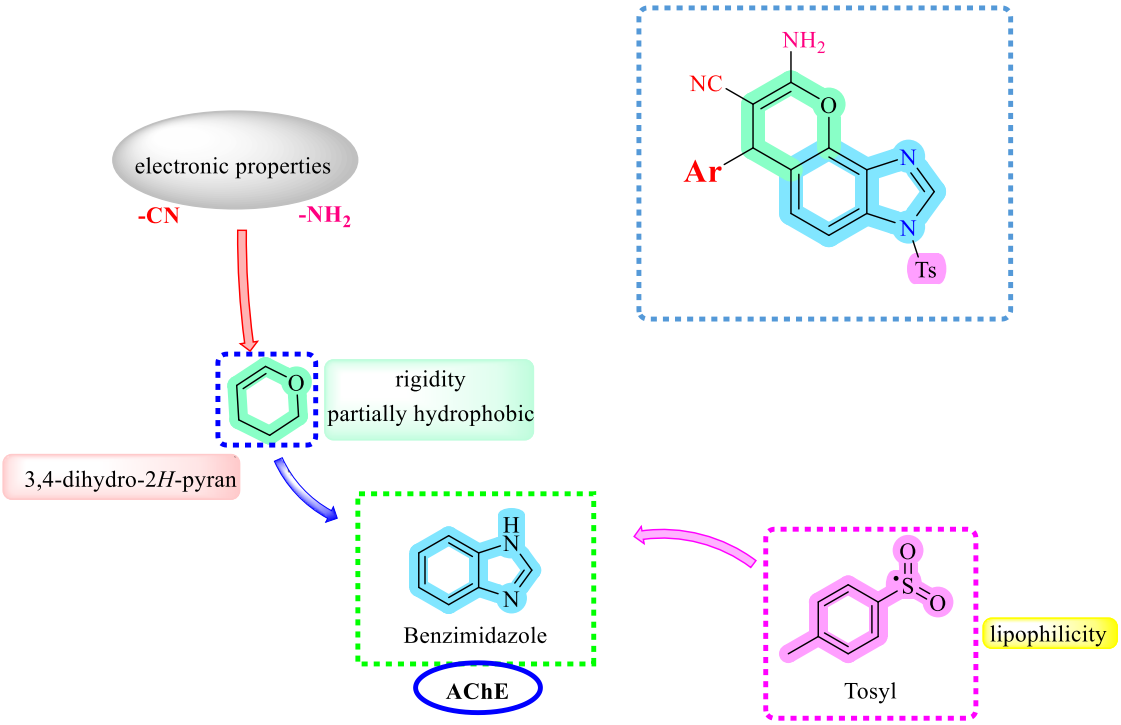


Figure 4: Design rationale of the benzimidazole-dihydropyran-tosyl hybrid

2. Materials and Methods

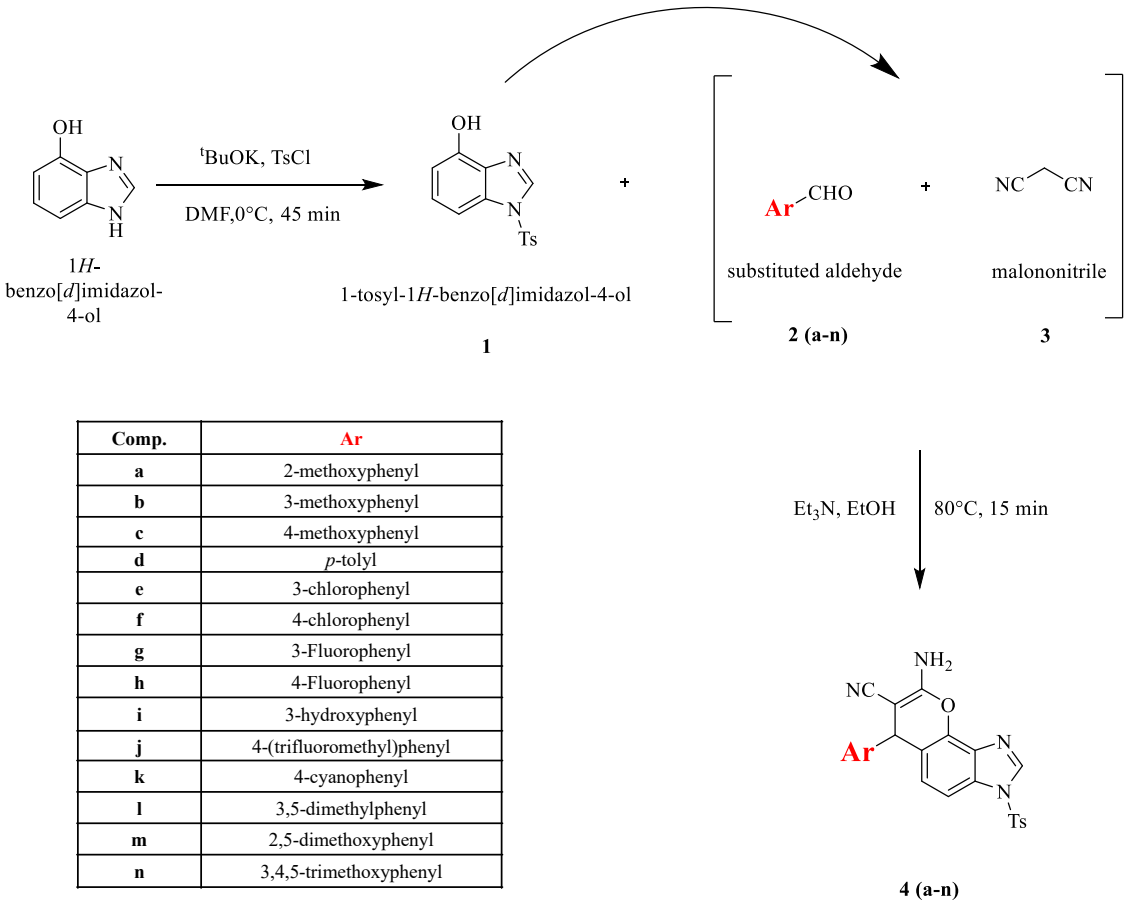
The benzimidazole-dihydropyran derivatives (4a-4m) were sketched in ChemDraw, and their 3D structures were generated and energy-minimized using the MMFF94 force field to obtain stable conformations. The ligands were exported in .mol2 and .pdbqt formats via OpenBabel for compatibility with docking tools. The modeled human acetylcholinesterase (AChE) protein was prepared by removing water molecules, adding polar hydrogens, merging non-polar hydrogens, optimizing protonation states, and assigning Gasteiger charges using AutoDock Tools. The processed structure was saved as a pdbqt file for downstream simulations.

The simulation of the binding of ligands to Acetylcholinesterase (AChE) was investigated using AutoDock Vina, with the ligand being placed in a grid box around two important regions of the active site; the catalytic and the peripheral (anionic) site of AChE. The redocking of donepezil was done as a validity check on the protocol for whether it would reproduce the same type of binding that was reported for donepezil in the literature. Interaction between the ligand and protein during docking was analysed using two programs - PyMOL and Discovery Studio. These programs show the various types of interactions that can occur between ligands and

proteins and the physical orientations of those types of binding (e.g. π-π stacking, hydrogen bonding, and hydrophobic interactions). The ADMET parameters (absorption, distribution, metabolism, excretion, and toxicity) for each ligand were predicted with pkCSM and ProTox-II as a way to assess the pharmacokinetic profiles, the metabolic stability, and the safety profiles of the ligands designed for AChE. The geometries of the molecules were optimised and the electronic properties of the ligands were calculated using DFT at the B3LYP/6-31G(d,p) level. These DFT calculations were used to determine the nature of the HOMO and LUMO in each Atom and the nature of the orbital populated by the electrons (i.e. the shape of the electron cloud around each atom). The relationships between structural, electronic, and steric parameters were studied to understand how ligand behaviour changes based on those factors in the binding pocket of AChE.

Machine learning techniques to predict ligand activity were developed using a QSAR approach. Molecular descriptors were generated using the Mordred program, followed by selection of the most optimal descriptors for training the resulting machine learning models with the RapidMiner and Knime software applications. The trained models were validated based on their predictive ability regarding the inhibition of AChE through the ligand series created for this study. Finally, the synthetic approach involved generating the benzimidazole scaffold, cyclizing it to form the fused dihydropyran core, and introducing structural variations through a Knoevenagel-type condensation, enabling modular synthesis of derivatives 4a-4m.

Figure 5 depicts the complete synthetic pathway for generating the benzimidazole-dihydropyran derivatives (4a-n), which form the core of the proposed AChE inhibitor series. The scheme highlights the stepwise construction of the hybrid scaffold, beginning with formation of the benzimidazole nucleus, followed by incorporation of the dihydropyran ring, and final functionalization with diverse aryl substituents at the C-4 position. This modular strategy enables systematic variation of electronic and steric features to support SAR exploration.



Comp.	Ar
a	2-methoxyphenyl
b	3-methoxyphenyl
c	4-methoxyphenyl
d	*p*-tolyl
e	3-chlorophenyl
f	4-chlorophenyl
g	3-Fluorophenyl
h	4-Fluorophenyl
i	3-hydroxyphenyl
j	4-(trifluoromethyl)phenyl
k	4-cyanophenyl
l	3,5-dimethylphenyl
m	2,5-dimethoxyphenyl
n	3,4,5-trimethoxyphenyl

Figure 5: Synthesis of benzimidazole-dihydropyran derivatives **(4a-n)**

2.1 Dataset Acquisition and Curation

Bioactivity data for acetylcholinesterase (AChE; ChEMBL Target ID: CHEMBL220) were retrieved from the BindingDB public repository (https://www.bindingdb.org), a comprehensive, experimentally validated database of measured binding affinities between proteins of known 3D structure and drug-like small molecules. BindingDB aggregates curated Ki, Kd, IC_{50}, and EC_{50} values from peer-reviewed literature, offering one of the largest publicly accessible repositories for structure-activity relationship (SAR) investigations. All records containing SMILES (Simplified Molecular Input Line Entry System) representations paired with experimentally determined inhibitory concentrations (IC_{50}) against human AChE were extracted. The raw bioactivity values (IC_{50} in nanomolar units) were transformed to the negative logarithmic scale (pIC_{50}) to normalize the distribution and reduce the influence of extreme outliers, following the established QSAR convention: This transformation yields a continuous response variable that more closely follows a Gaussian distribution and is amenable to linear and non-linear regression modeling. Structural data underwent rigorous quality

filtering: records with invalid or unparsable SMILES strings, missing activity values, infinite or non-finite numerical entries, and duplicate SMILES were removed. The final curated dataset was retained for downstream molecular featurization and model development.

2.2 Molecular Featurization

2.2.1 Two-Dimensional Physicochemical Descriptors

Eleven two-dimensional physicochemical descriptors were computed for each compound using the RDKit cheminformatics toolkit (version 2023.09) via Python. These descriptors represent fundamental molecular properties known to govern pharmacokinetic behavior, membrane permeability, and receptor binding affinity, and were selected based on their relevance in Lipinski's Rule of Five and Veber's oral bioavailability criteria: Descriptors included molecular weight (MW), total number of heavy atoms (nAtom), number of heteroatoms (nHET), ring count (nRing), number of rotatable bonds (nRot), number of hydrogen bond donors (HBD), number of hydrogen bond acceptors (HBA), topological polar surface area (TPSA), partition coefficient (logP and SlogP via Wildman-Crippen method), and quantitative estimate of drug-likeness (QED). These descriptors were used for exploratory analysis, correlation profiling, and visualization but not directly as features for the primary ML regression models.

2.2.2 Extended Connectivity Fingerprints (ECFP)

Molecular structural information was encoded as Extended Connectivity Fingerprints (ECFP) at a diameter of 4 (Morgan radius r = 2), commonly referred to as ECFP4. For each compound with a valid SMILES string, the RDKit Morgan algorithm was applied to generate a hashed, bit-vector representation of length 2,048 bits. The ECFP algorithm operates by iteratively updating atom environments over successive bond-radius shells using the following invariant-based hashing approach: where fp_i^(k) is the fingerprint identifier for atom i at iteration k, a_i represents the atomic invariants (atomic number, degree, number of hydrogens, charge, isotope), and N(i) denotes the set of neighbors of atom i. After k = 2 iterations, the resulting atom environment identifiers are hashed into a fixed-length 2,048-bit binary vector, where each bit encodes the presence or absence of a unique circular structural subgraph. This representation captures both local atomic environments and extended molecular topology, providing a rich and interpretable featurization that has demonstrated state-of-the-art performance across diverse QSAR benchmarks.

2.3 Data Preprocessing and Train-Test Partitioning

Prior to model training, the 2,048-dimensional ECFP4 fingerprint matrix X ∈ ℝ^{n×2048} was normalized using zero-mean, unit-variance standardization (StandardScaler, scikit-learn v1.3). For each feature dimension j, the transformation was applied as:

where μ_j and σ_j are the empirical mean and standard deviation of feature j computed exclusively from the training data, and subsequently applied without modification to the held-out test set to prevent data leakage. The dataset was partitioned into training (85%) and test (15%) subsets using stratified random splitting (random_state = 42), ensuring reproducibility across experimental runs.

2.4 Machine Learning Regression Algorithms

2.4.1 Gradient Boosting Ensemble Models

Three state-of-the-art gradient-boosted decision tree (GBDT) frameworks were employed: XGBoost (Extreme Gradient Boosting), LightGBM (Light Gradient Boosting Machine), and CatBoost (Categorical Boosting). All GBDT algorithms operate on the principle of iterative residual minimization: at each stage m, a weak learner h_m(x) (a decision tree) is trained to predict the negative gradient of the loss function with respect to the current prediction F_{m-1}(x):

where ν is the learning rate (shrinkage factor). For regression tasks, the squared error loss L(y, F) = ½(y - F)^2 is minimized, and the pseudo-residuals guiding tree construction at step m are:

XGBoost employed histogram-based tree construction (tree_method = 'hist') with n_estimators = 1,500, learning_rate = 0.01, max_depth = 8, and subsample = 0.9. LightGBM utilized leaf-wise (best-first) tree growth with the same n_estimators and learning rate, significantly reducing memory consumption while preserving predictive accuracy. CatBoost applied ordered boosting with symmetric trees and an intrinsic handling of categorical overfitting, with 1,500 iterations, learning_rate = 0.03, and depth = 8. Ridge Regression with L2 regularization (λ = 1) was included as a linear baseline:

2.4.2 Deep Neural Network (DNN)

A fully connected deep neural network (DNN) was implemented using PyTorch (v2.1) to capture complex, non-linear structure-activity relationships latent within the ECFP4 fingerprint space. The network architecture comprised four fully connected (dense) hidden layers with the

following configuration: Input(2048) → Linear(2048, 1024) → BatchNorm1d → ReLU → Dropout(0.3) → Linear(1024, 512) → ReLU → Dropout(0.2) → Linear(512, 128) → ReLU → Linear(128, 1). Batch normalization was applied after the first linear transformation to stabilize training dynamics and mitigate internal covariate shift. Dropout regularization with rates of 0.3 and 0.2 was applied to the first and second hidden layers, respectively, to reduce overfitting.

The network was optimized using the Adam optimizer (learning_rate = 0.001, β_1 = 0.9, β_2 = 0.999, ε = 10^{-8}) and the mean squared error (MSE) loss function:

Training was conducted for 300 epochs over the entire training set in a full-batch mode. Model weights were updated via backpropagation through the computational graph using automatic differentiation. GPU acceleration (CUDA) was employed where available, with automatic fallback to CPU execution.

2.5 Weighted Ensemble Prediction Strategy

To mitigate individual model biases and exploit complementary predictive strengths, a weighted ensemble strategy was adopted. Each base learner k was assigned a normalized weight w_k proportional to its validation-set coefficient of determination (R^2): The final ensemble prediction for a compound x is the weighted linear combination of individual model predictions: where K = 5 (XGBoost, LightGBM, CatBoost, Ridge, DNN). This aggregation approach systematically reduces prediction variance without introducing additional model parameters and has been demonstrated to improve generalization across structurally diverse chemical spaces.

2.6 Model Evaluation Metrics

Model performance was assessed on the held-out test set using three standard regression metrics: the coefficient of determination (R^2), root mean squared error (RMSE), and mean absolute error (MAE): R^2 measures the proportion of variance in pIC_{50} explained by the model (range: ∞ to 1, with 1 indicating perfect prediction). RMSE penalizes large prediction errors quadratically, while MAE provides a scale-interpretable, outlier-robust measure of average absolute deviation. Together, these metrics provide a comprehensive picture of both predictive accuracy and error distribution.

2.7 Chemical Space Visualization and Scaffold Analysis

The diversity and structural landscape of the AChE dataset were characterized through orthogonal dimensionality reduction approaches. Principal Component Analysis (PCA) was applied to the standardized ECFP4 fingerprint matrix to identify the two principal directions of maximum variance, projecting the high-dimensional molecular space into an interpretable 2D plane. The principal components PC1 and PC2 were computed via eigendecomposition of the covariance matrix C = X^TX / (n-1). Complementarily, t-distributed Stochastic Neighbor Embedding (t-SNE) was employed to capture non-linear local structural relationships. t-SNE models pairwise affinities in high-dimensional space using a Gaussian kernel and minimizes the Kullback-Leibler (KL) divergence between the high-dimensional distribution P and the low-dimensional Student-t distribution Q: t-SNE was executed with perplexity = 30 (or n/2 for smaller datasets), n_iter = 1,000, and random_state = 42. Murcko scaffold decomposition was applied to all valid molecular structures to quantify chemical scaffold diversity, providing insight into the structural variety of the compound library.

3. Result and Discussion

3.1 Ramachandran Analysis

The stereochemical quality of the predicted protein model (Figure 6) was examined using PROCHECK to generate a Ramachandran Plot. Of the 456 non-Glycine and non-Proline residues: 87.3% of all residue types resided in the most favoured regions; 11.8% of each residue type was present in the allowed regions; and only 0.7% of total residues were found in the generously allowed region. Of those, 0.2% of total residues were found in the disallowed region. Therefore, from these results, the final predicted model is structurally sound and well-refined. The results also indicate that the predicted protein will have success in being used for downstream molecular simulations and functional studies.

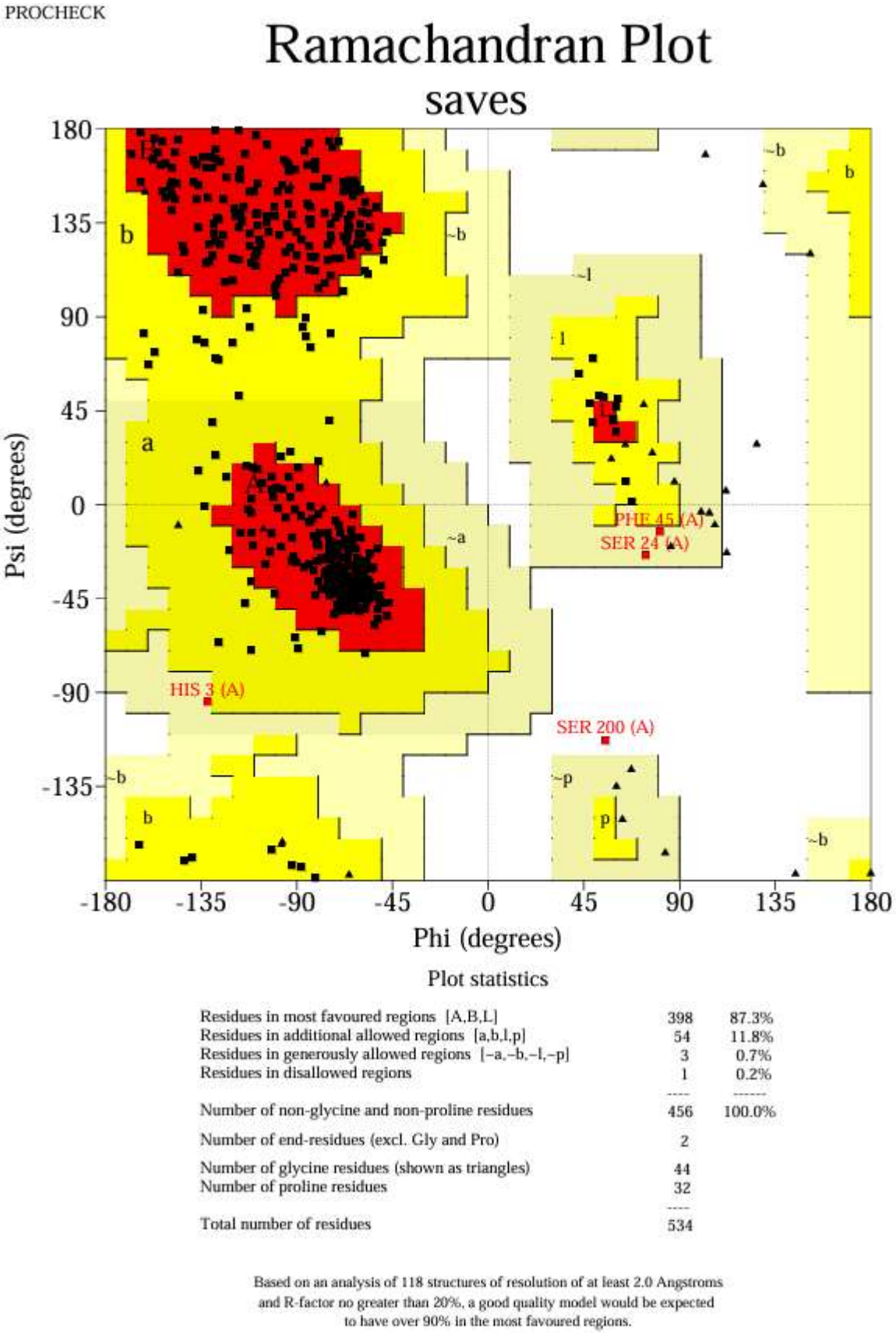


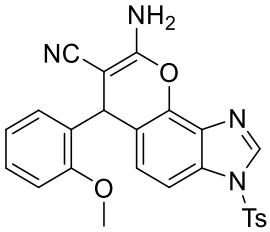
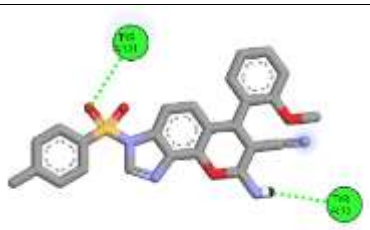
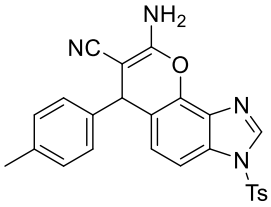
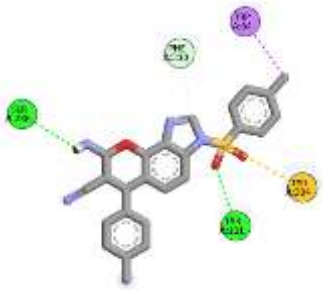
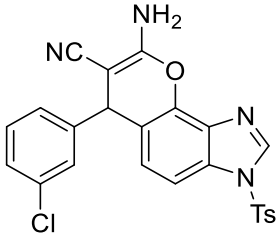
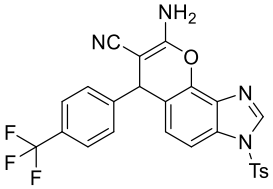
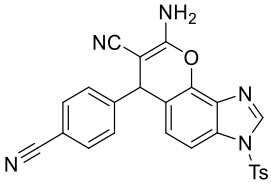
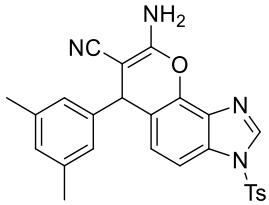
Figure 6: Ramachandran plot validation of the modelled protein structure.

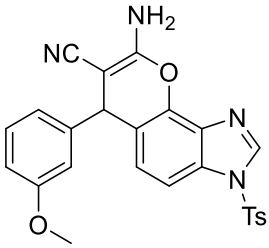
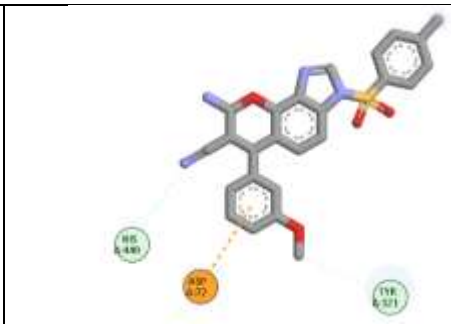
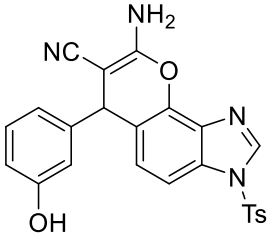
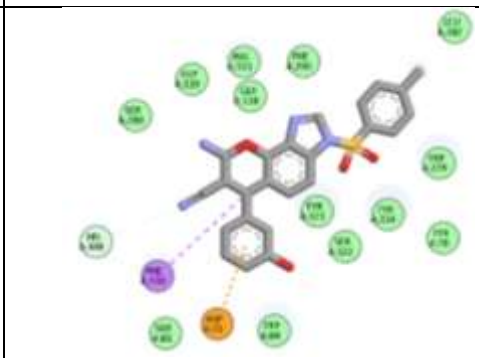
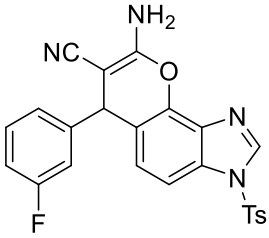
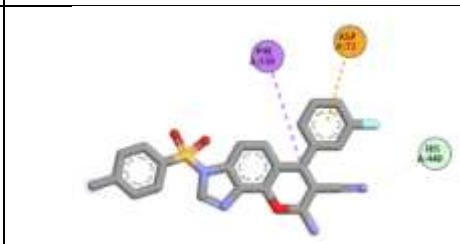
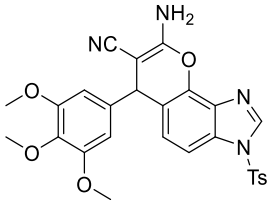
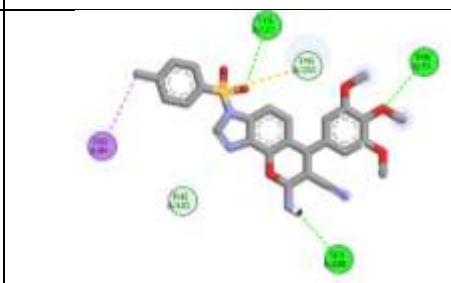
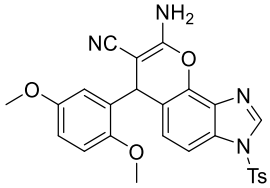
3.2 Molecular Docking Results

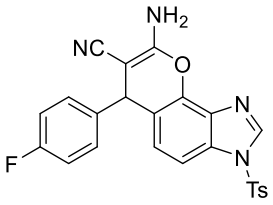
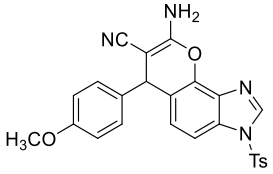
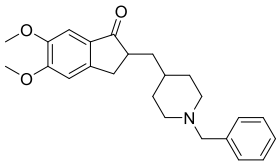
The binding affinities of the benzimidazole-dihydropyran derivatives (4a-4m) are presented in Table 2, and they all exhibit a strong affinity to bind to AChE with values between -10.3 and -12.1 kcal/mol indicating that the designed framework possesses a good alignment with the essential catalytic gorge of the enzyme. The derivatives interact with important catalytic elements of the enzyme, including serine, histidine as well as aromatic residues (His440, Ser122, Tyr121, Phe331 and Tyr70), which are responsible for maintaining the stability of the ligand in both the CAS and PAS of the enzyme. Compound 4b possesses the strongest binding affinity (-12.1 kcal/mol) of the derivatives and forms several hydrogen bonds with Ser286, Tyr121 and Phe330, confirming a very stable binding configuration. The results suggest a two-pronged approach to AChE inhibition by supporting both catalytic stoppage and potentially blocking Aβ clustering at the same time. A number of compounds, including 4c, 4f, 4g, 4l, and 4i, maintain strong interactions with His440, implying a potential to hinder catalytic activity due to disruption of turnover. The docking results also show a high degree of similarity with the reference therapeutic donepezil (-11.0 kcal/mol), therefore supporting the hypothesis that the designed hybrids may possess structural characteristics that will allow them to achieve clinically significant inhibition levels.

Table 2: Docking results of compounds 4a-4m showing binding affinity values, key hydrogen-bond interactions, and corresponding 2D interaction profiles within the AChE active site.

Comp Code	Structure	Binding Affinity (kcal/mol)	H-Bond Interactions	2D Interactions

4a		-10.9	Tyr121, Tyr70	
4b		-12.1	Ser286, Tyr121, Phe330	
4c		-11.6	His440	
4d		-10.6	Ser122, Gly123, Pro86	
4e		-10.6	Tyr121, Ser286, Phe331	
4f		-11.6	His440	

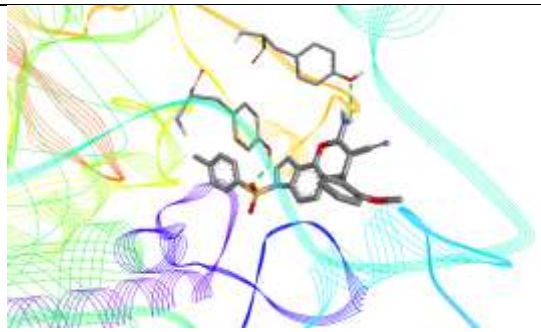
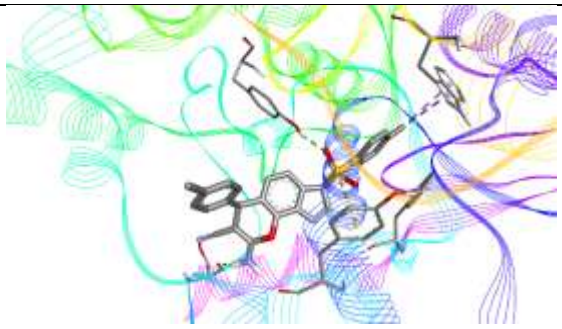
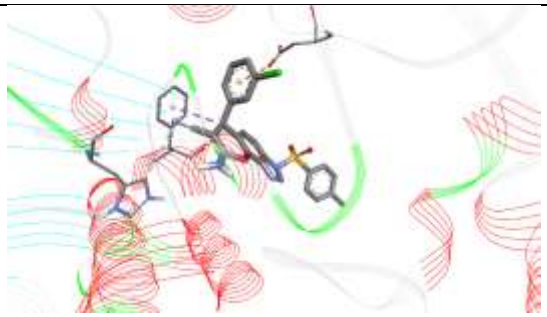
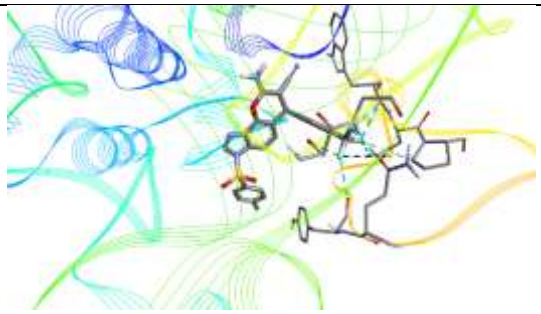
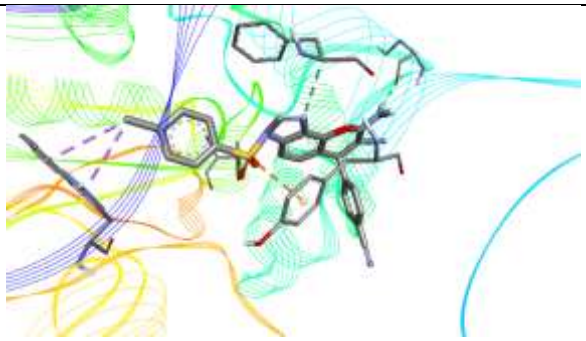
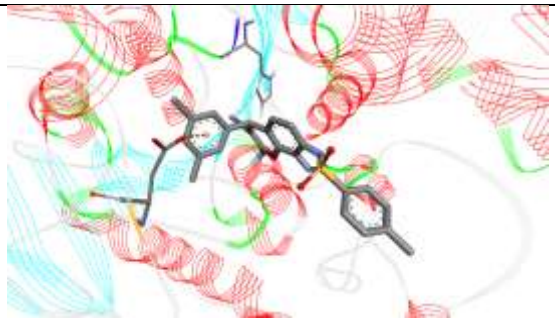
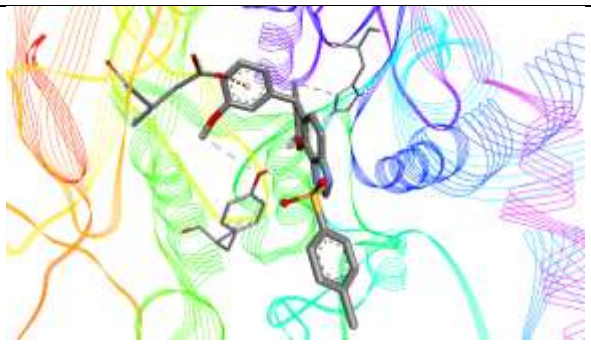
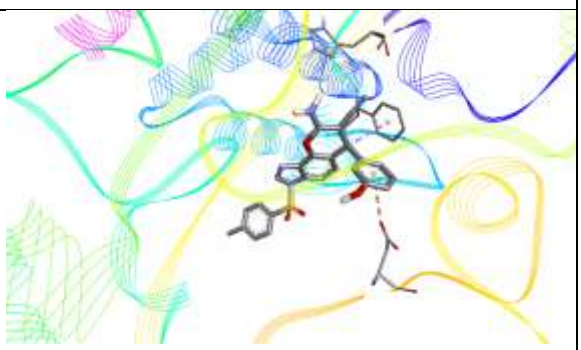
4g		-11.4	His440, Tyr121	
4h		-11.3	His440	
4i		-11.5	His440	
4j		-10.5	Tyr121, Tyr334, Phe331, Ser286, Tyr70	
4k		-10.3	Ser81, Asp72, Trp84, Ser122, Tyr70, Gln69, Asn85	

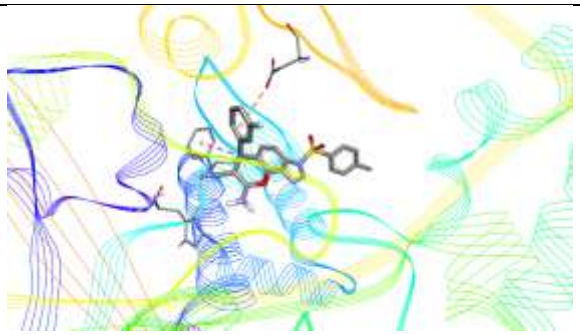
4l		-11.4	His440	
4m		-10.9	Tyr121, Tyr70, Phe331, Ser286	
Donepezil		-11.0	Tyr121	

3.3. Binding Interaction Analysis (3D Profile)

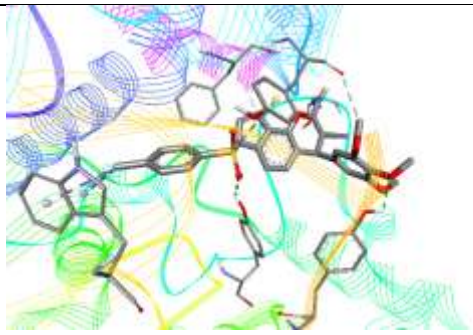
Based on the 3D interaction profiles listed in Table 3, the docking study results were confirmed by showing how the derivatives are located in the binding site of the enzyme. The majority of the compounds are located in the AChE gorge region; therefore, their ability to form extensive π-π interactions with Trp84, Tyr121, Tyr334, and Phe331 (the aromatic residues), are crucial for the compounds to bind to AChE with high affinity. The compounds that were found to form hydrogen bonds with His440, Ser122, Gly123, and Ser286 provided increased stability and increased inhibitory potency. Additionally, some of the derivatives also showed a Donepezil-like orientation, meaning they were found to overlap with both the catalytic and peripheral site. The results provide evidence of the feasibility of combining benzimidazole with the dihydropyran system. Our findings support a greater ability of these scaffolds to target multiple receptors based on the overall interaction capabilities of hydrogen bonding, π-π stacking, and hydrophobic interactions. Therefore, the inclusion of different aryl substituents at the C-4 position of the derivatives greatly influences the binding affinity, binding orientation and retains the overall functionality of the compounds throughout the series.

Table 3: 3D binding interactions of selected derivatives (4a-4m) within the AChE active site

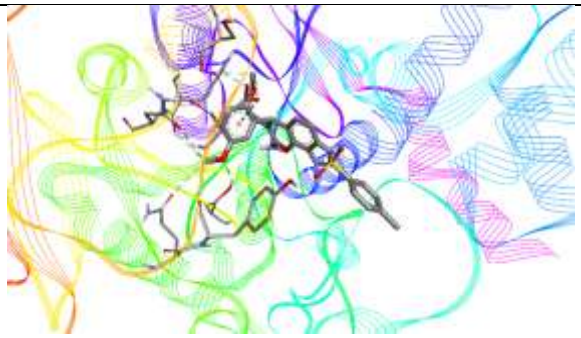
 **4a**	 **4b**
 **4c**	 **4d**
 **4e**	 **4f**
 **4g**	 **4h**



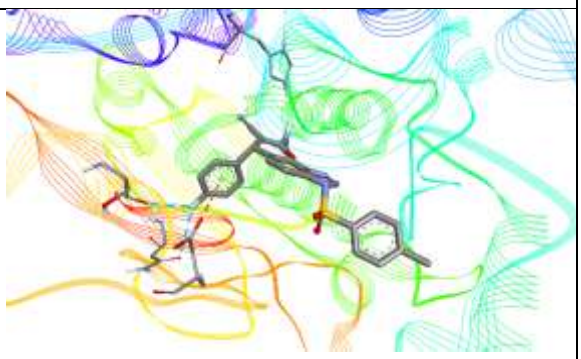
4i



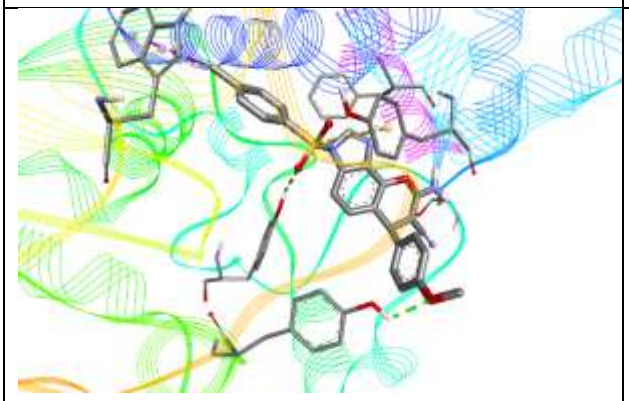
4j



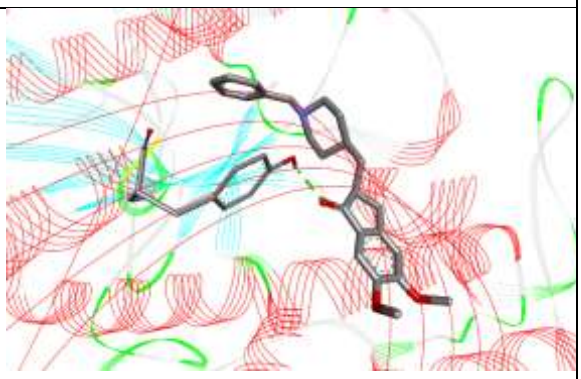
4k



4l



4m



Donepezil

3.4. ADMET Evaluation

Comprehensive predictions related to ADMET properties (Tables 4 to 6) suggest that all derivatives are good candidates for CNS-active AChE inhibition. The intestinal absorption prediction for each compound was > 90 per cent and the predicted values for Caco-2 permeability are at an acceptable level, supporting oral administration of the compounds. All compounds are predicted to have a moderate absorption profile through P-glycoprotein substrate/inhibitor interaction, making them suitable for BBB penetration. P-glycoprotein is one of the principal methods by which drugs cross the BBB. The predicted VDss and fraction unbound (Fu) indicate a reasonable tissue distribution for these compounds and both logBB

and logPS fall within the accepted range for small molecules that penetrate the CNS. These predictions confirm the adequacy of solubility and permeability for the designed scaffold to penetrate through to the central targets.

The metabolic predictions (Table 5) indicate that all derivatives will be substrates for CYP3A4, but will not be substrates for CYP2D6 (which reduces the likelihood of adverse effects due to interindividual variability in CYP2D6). However, the CYP1A2, CYP2C19, and CYP2C9 appear to indicate some variability in the inhibition of these enzymes, which means that moderate metabolic liability still exists, but it should be manageable during optimization. As a result, the total clearance values for this series of compounds are generally low to moderate, indicating that these compounds have good metabolic stability and provide sufficient systemic exposure.

The predicted toxicity data in **Table 6** show that none of the derivatives exhibit hepatotoxicity or skin-sensitizing effects, which is an advantage over donepezil, which is flagged as hepatotoxic. Although occasional **AMES positivity** and hERG-I inhibition signals appear across the series, these findings are typical for early-stage CNS ligands and remain within a manageable range for future lead refinement. Acute and chronic oral toxicity values are consistent across the derivatives and comparable to or safer than the reference drug. Overall, the ADMET assessment supports that the designed derivatives possess drug-like characteristics with a balanced safety, metabolic, and distribution profile suitable for advancing as CNS-directed AChE inhibitors.

Table 4: Predicted absorption and distribution parameters of synthesized derivatives (4a-4m) compared with donepezil

S. No.	ABSORPTION							DISTRIBUTION			
	Water solubility	**Caco 2 permeability**	**Intestinal absorption (human)**	**Skin Permeability**	**P-glycoprotein substrate**	**P-glycoprotein I inhibitor**	**P-glycoprotein II inhibitor**	**VDss (human)**	**Fraction unbound (human)**	**BBB permeability**	**CNS permeability**

	Numeric (log mol/L)	Numeric (log Papp in 10^{-6} cm/s)	Numeric (% Absorbed)	Numeric (log Kp)	Categorical (Yes/No)	Categorical (Yes/No)	Categorical (Yes/No)	Numeric (log L/kg)	Numeric (Fu)	Numeric (log BB)	Numeric (log PS)
4a	-2.895	0.433	93.186	-2.735	Yes	Yes	Yes	-0.073	0.273	-0.823	-2.276
4b	-2.895	0.307	96.55	-2.735	Yes	Yes	Yes	0.087	0.272	-0.614	-1.978
4c	-2.895	0.296	97.316	-2.735	Yes	Yes	Yes	0.081	0.272	-0.782	-1.93
4d	-2.895	0.308	98.394	-2.735	Yes	Yes	Yes	0.076	0.274	-1.048	-1.898
4e	-2.895	0.219	93.731	-2.735	Yes	Yes	Yes	0.081	0.271	-0.775	-2.116
4f	-2.896	0.212	97.127	-2.735	Yes	Yes	Yes	0.08	0.262	-0.609	-1.885
4g	-2.895	0.324	97.878	-2.735	Yes	Yes	Yes	0.005	0.272	-0.835	-2.264
4h	-2.895	-0.043	93.267	-2.735	Yes	Yes	Yes	0.06	0.273	-0.325	-2.262

4i	-2.895	0.357	99.793	-2.735	Yes	Yes	Yes	0.01	0.275	-0.814	-2.141
4j	-2.894	-0.209	94.99	-2.735	Yes	Yes	Yes	-0.154	0.286	-1.28	-3.4
4k	-2.896	0.313	92.67	-2.735	Yes	Yes	Yes	-0.107	0.272	-1.061	-3.191
4l	-2.894	0.37	99.667	-2.735	Yes	Yes	Yes	-0.002	0.276	-0.822	-2.141
4m	-2.894	0.337	97.763	-2.735	Yes	Yes	Yes	-0.006	0.273	-0.839	-2.268
Done	-4.87	1.001	93.074	-2.838	Yes	Yes	Yes	1.262	0.022	0.349	-1.419

Table 5: Predicted metabolism and excretion parameters of synthesized derivatives (4a-4m) compared with donepezil

S. No.	**METABOLISM**							**EXERETION**	
	CYP2D6 substrate	**CYP3A4 substrate**	**CYP1A2 inhibitior**	**CYP2C19 inhibitior**	**CYP2C9 inhibitior**	**CYP2D6 inhibitior**	**CYP3A4 inhibitior**	**Total Clearance**	**Renal OCT2 substrate**

	Categorical (Yes/No)	Categorical (Yes/No)	Categorical (Yes/No)	Categorical (Yes/No)	Categorical (Yes/No)	Categorical (Yes/No)	Categorical (Yes/No)	Numeric (log ml/min/kg)	Categorical (Yes/No)
4a	No	Yes	No	Yes	Yes	No	Yes	0.07	Yes
4b	No	Yes	No	Yes	Yes	No	Yes	0.092	Yes
4c	No	Yes	Yes	Yes	Yes	No	Yes	0.034	Yes
4d	No	Yes	Yes	Yes	Yes	No	Yes	-0.068	Yes
4e	No	Yes	Yes	Yes	Yes	No	Yes	0.11	Yes
4f	No	Yes	Yes	Yes	Yes	No	Yes	-0.017	Yes
4g	No	Yes	Yes	Yes	Yes	No	Yes	0.096	Yes
4h	No	Yes	Yes	Yes	Yes	No	Yes	0.047	Yes
4i	No	Yes	Yes	Yes	Yes	No	Yes	0.02	Yes
4j	No	Yes	No	Yes	Yes	No	Yes	0.118	Yes
4k	No	Yes	No	Yes	Yes	No	Yes	0.076	Yes
4l	No	Yes	Yes	Yes	Yes	No	Yes	0.039	Yes
4m	No	Yes	No	Yes	Yes	No	Yes	0.111	Yes
Done	No	Yes	No	No	No	Yes	Yes	0.998	Yes

Table 6: Predicted toxicity parameters of synthesized derivatives (4a-4m) compared with donepezil

S. No.	TOXICITY									
	AMES toxicity	Max. tolerated dose	hERG I inhibitor	hERG II inhibitor	Oral Rat Acute	Oral Rat Chronic Toxicit	Hepatotoxicity	Skin Sensitisation	*T.Pyriformis* toxicity	Minnow toxicity

		(human)			Toxicity (LD50)	y (LOAEL)				
	Categorical (Yes/No)	Numeric (log mg/kg/day)	Categorical (Yes/No)	Categorical (Yes/No)	Numeric (mol/kg)	Numeric (log mg/kg_bw/day)	Categorical (Yes/No)	Categorical (Yes/No)	Numeric (log ug/L)	Numeric (log mM)
4a	No	-0.018	Yes	Yes	2.606	0.438	No	No	0.285	-0.258
4b	Yes	-0.077	Yes	Yes	2.564	1.436	No	No	0.285	0.206
4c	Yes	-0.082	Yes	Yes	2.562	1.312	No	No	0.285	-0.126
4d	Yes	-0.073	No	Yes	2.567	1.219	No	No	0.285	0.087
4e	Yes	-0.075	Yes	Yes	2.568	1.511	No	No	0.285	0.259
4f	Yes	-0.132	Yes	Yes	2.569	1.387	No	No	0.285	-0.163
4g	Yes	-0.065	No	Yes	2.597	0.383	No	No	0.285	-0.262
4h	Yes	-0.027	Yes	Yes	2.554	0.4	No	No	0.285	1.157
4i	Yes	-0.046	No	Yes	2.594	0.332	No	No	0.285	-0.204

4j	No	0.047	No	Yes	2.559	0.63	No	No	0.285	-1.427
4k	No	-0.001	No	Yes	2.595	0.445	No	No	0.285	-1.16
4l	Yes	-0.051	Yes	Yes	2.592	0.362	No	No	0.285	-0.099
4m	Yes	-0.063	No	Yes	2.596	0.407	No	No	0.285	-0.268
Done	No	0.095	No	Yes	2.886	1.453	Yes	No	0.931	-1.368

3.5 Density Function Theory Analysis

To gain deeper insight into the electronic behaviour and reactivity of the most promising AChE inhibitors, DFT calculations were performed for the top four docked benzimidazole-dihydropyran derivatives, namely 4b, 4c, 4f, and 4i, at the B3LYP/6-31G(d,p) level of theory. For each compound, the frontier molecular orbital (FMO) energies were calculated and expressed in electron volts (eV). The computed HOMO energies lie in the range of approximately -5.58 to -5.88 eV, while the LUMO energies fall around -2.43 to -2.51 eV, giving HOMO-LUMO gaps between 3.15 and 3.45 eV **(Figure 7)**. Among the series, compounds 4b and 4f display the narrowest energy gaps (≈3.16 and 3.15 eV, respectively), followed closely by 4i (≈3.24 eV), whereas 4c exhibits a comparatively wider gap (≈3.45 eV). The smaller energy gaps for 4b, 4f, and 4i indicate higher chemical softness and greater charge-transfer capability, which favour stronger electronic interactions with key residues within the AChE active gorge. The trend seen with the highest scoring compounds corresponds to the theoretical predictions supporting that between compounds with a lower docking score, those that exhibit the greatest degree of electron deflection in combination with the π-π interaction. In contrast to those showing significant differences in the ID of the active site, compounds that are in Higher Areas of Docking Scores (HAD-4c) have significantly lower activities than those showing the same ID and have a more stable electrónico configuration. However, they also have higher

levels of binding by favourable sterics and/or nonbonded interactions to help them compensate for their effectiveness due to a more favourable electronic energy level.

FMO contour plots provide further evidence of the superiority of the pharmacophore designs created from these four derivatives. The majority of both the highest occupied molecular orbital and the lowest occupied molecular orbital densities of all four derivatives are located on the benzimidazole core and associated aryl/group. This portion is the primary electron-sharing location of AChE that will be engaged in catalysis and any associated anionic catalytic or peripheral reactions. Opposite of these locations, the LUMO derived from the dihydropyran-tosyl component is located on the opposite end of the chemical system (in the gorge) and therefore is involved in hydrogen bonding and/or being an electron-accepting site to form a stable hydrolytic interaction on the active site of AChE. The clear spatial distinction between these binding locations and the resultant intramolecular charge transfer characteristics of these compounds supported an explanation for stabilising ligand-enzyme complex formation and therefore the combined factor of polar, π- stacking and hydrophobic contacts between ligands and enzyme complexes is likely to lead to highly stable AChE catalysis.

Overall, the DFT analysis aligns closely with the docking, ADMET, and ML-QSAR findings, reinforcing that the benzimidazole ring primarily governs electron donation and interaction with aromatic residues, while the dihydropyran-tosyl moiety fine-tunes electron acceptance and hydrophobic fit. The relatively low HOMO-LUMO gaps of 4b, 4f, and 4i provide a

quantum-chemical rationale for their enhanced binding affinities and support their prioritization as lead candidates for further experimental validation.

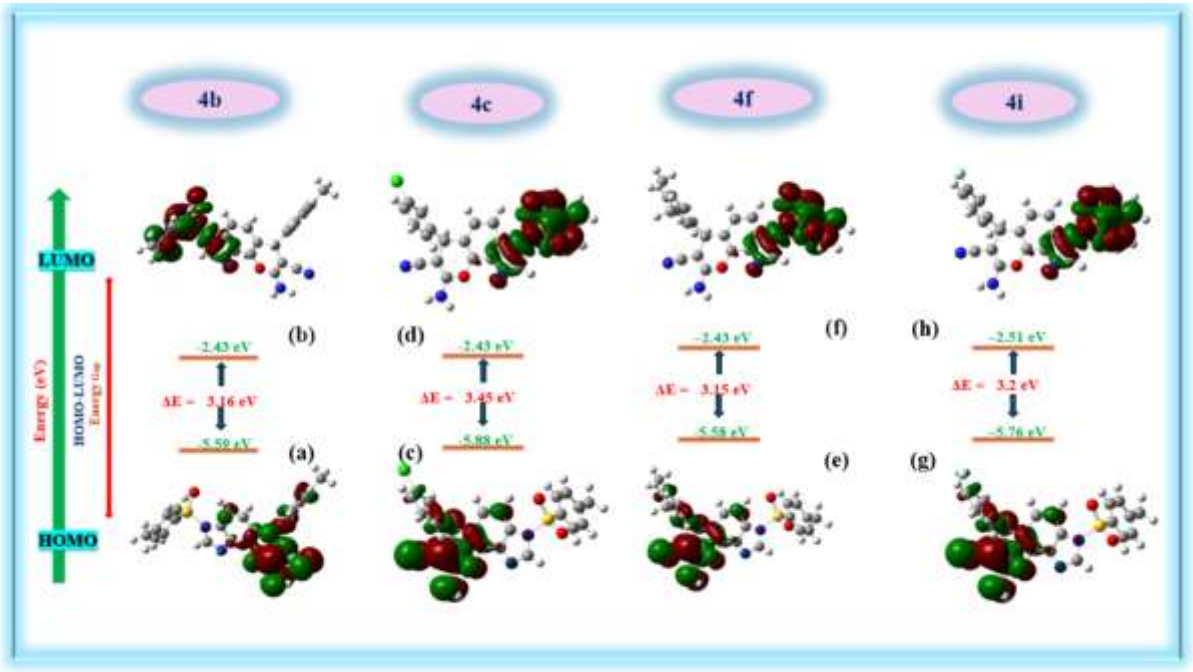


Figure 7: Illustration of energy gap of top compounds **(4b, 4c, 4f,** and **4i)** after frontier orbital study

3.6 Machine Learning Based QSAR (ML-QSAR)

To identify the most promising AChE inhibitors among the designed benzimidazole-dihydropyran derivatives, multiple machine-learning classification models were developed and evaluated **(Table 7)**. Among these, the **Random Forest model demonstrated the highest predictive power**, achieving an accuracy of **84.91%**, a Kappa statistic of **0.721**, and the lowest RMSE (**0.328**) and MAE (**0.245**) values, indicating its superior reliability in distinguishing active from inactive compounds. This performance surpassed that of the K-Star (81.36% accuracy) and IBK (78.42% accuracy) models, while significantly outperforming simpler classifiers such as Naïve Bayes and Logistic Regression. The lower RAE and RRSE values associated with Random Forest further emphasize its robustness and reduced prediction error relative to other models. Based on these performance metrics, the Random Forest classifier was selected as the optimized predictive model and subsequently used to evaluate the activity of the designed molecular library. The model consistently identified **compounds 4b, 4c, 4f, and 4i** as active AChE inhibitors, supporting the results obtained from docking, FMO analysis, and ADMET profiling. These findings highlight the value of integrating machine-learning predictions into the early stages of Alzheimer's drug-discovery workflows to strengthen decision-making and prioritize the most promising therapeutic candidates.

Table 7: Performance comparison of various machine-learning models used for predicting AChE inhibitory activity

Model	Accuracy (%)	Kappa Statistic	MAE	RMSE	RAE (%)	RRSE (%)
Random Forest	**84.91**	**0.721**	**0.245**	**0.328**	**55.42**	**69.10**
K-Star	81.36	0.669	0.163	0.362	38.25	76.83
IBK (k-NN)	78.42	0.618	0.174	0.389	40.72	82.48
Decision Tree (J48)	73.95	0.548	0.217	0.419	50.18	89.94
Multilayer Perceptron	68.73	0.463	0.276	0.402	64.81	86.33
Naïve Bayes	60.54	0.351	0.318	0.447	73.95	95.81

Logistic Regression	58.11	0.298	0.354	0.431	81.24	92.94
SMO (SVM)	58.48	0.304	0.339	0.439	77.93	94.46

3.7 Dataset Characterization and Physicochemical Property Analysis

The curated AChE dataset retrieved from BindingDB comprised several thousand bioactive small molecules spanning a wide range of chemical scaffolds and structural complexity. Following curation, the dataset exhibited a continuous pIC_{50} distribution ranging from approximately -3 to +8, with the majority of compounds clustered in the mid-potency range (pIC_{50} 2-6), consistent with a heterogeneous bioactivity landscape typical of a multi-scaffold kinase/enzyme inhibitor dataset. Eleven physicochemical descriptors were computed to characterize the structural space, and their pairwise relationships with experimental bioactivity were examined through scatter plots **(Figure 8)** and a correlation heatmap **(Figure 9).**

Activity vs. HBArevealed a broad distribution of hydrogen bond acceptor counts (0-22) with peak activities observed in the 4-8 HBA range, suggesting that moderate polarity is associated with enhanced AChE inhibitory potency. A similar trend was observed for HBD, where compounds with 0-5 donors dominated the high-activity cohort. The relationship between biological activity and nHET demonstrated that compounds with 2-6 heteroatoms tend to exhibit higher pIC_{50} values, consistent with the necessity for polar interactions within the AChE gorge active site. MW and nAtom relationships indicated an optimal mass range of 300-600 Da for peak inhibitory activity, with heavier compounds (MW > 700 Da) showing diminishing potency-a pattern congruent with Lipinski's constraints on oral bioavailability. logPand SlogP analysis demonstrated that intermediate lipophilicity (logP 1-4) correlates with superior activity, reflecting the amphipathic nature of the AChE binding gorge. nRot, TPSA, nRing, and QED collectively reinforced the notion that conformationally restricted, moderately polar, ring-rich compounds constitute the optimal pharmacophoric profile for AChE inhibition.

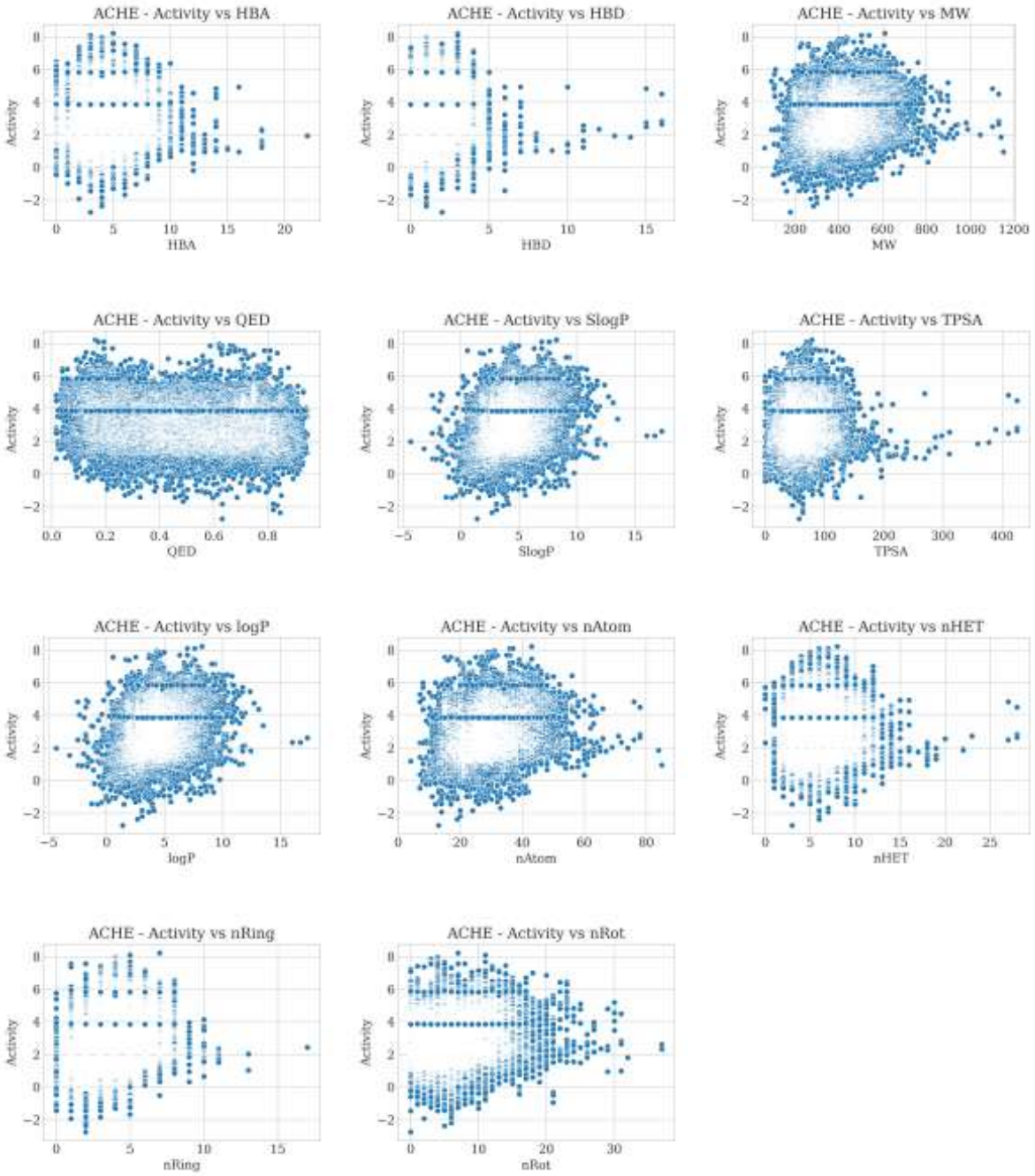


Figure 8. Scatter plot illustrating the relationship between experimental pIC_{50} activity and molecular descriptors

3.8 Molecular Descriptor Intercorrelation Analysis

Pairwise Pearson correlation coefficients among the eleven computed descriptors were computed and visualized as a color-coded heatmap (**Figure 9**). The analysis revealed several noteworthy collinearities: MW and nAtom exhibited near-perfect positive correlation (r =

0.98), confirming their redundancy as independent features. HBA and nHET were strongly correlated (r = 0.82), reflecting the chemical reality that heteroatoms predominantly serve as hydrogen bond acceptors. Similarly, TPSA and HBA showed strong covariance (r = 0.81), consistent with TPSA being a composite function of polar surface contributions from nitrogen and oxygen atoms. logP and SlogP were perfectly collinear (r = 1.00) since both are calculated via the Wildman-Crippen method in RDKit.

QED demonstrated consistent negative correlations with MW (r = -0.75), nAtom (r = -0.75), nRot (r = -0.70), logP (r = -0.64), and TPSA (r = -0.44), indicating that as molecular complexity increases, overall drug-likeness as assessed by QED tends to decrease. These intercorrelation patterns informed the interpretation of ML feature importance analyses and highlight the importance of employing fingerprint-based (rather than descriptor-based) featurization for the primary QSAR models, as raw binary fingerprints are substantially less collinear.

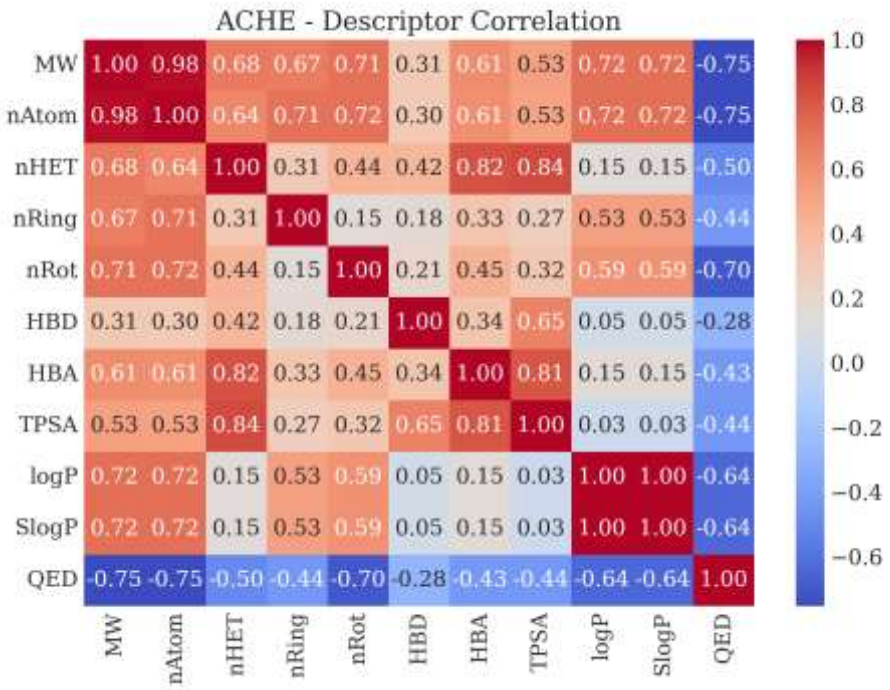


Figure 9. Pearson correlation heatmap of eleven two-dimensional physicochemical descriptors computed for the AChE dataset.

3.9 Chemical Space Mapping

To probe the structural diversity and activity landscape of the AChE compound library, dimensionality reduction was applied to the standardized ECFP4 fingerprint matrix. t-SNE embedding **(Figure 10)** revealed a broadly dispersed but generally interconnected chemical

space, consistent with a dataset assembled from diverse chemical scaffolds across multiple literature sources. Compounds color-coded by pIC_{50} demonstrated a heterogeneous distribution of activity across the chemical space, with high-activity compounds ($pIC_{50} > 6$, yellow) dispersed throughout multiple clusters rather than confined to a single structural region. This observation suggests that AChE inhibitory activity arises from diverse pharmacophoric solutions, underscoring the importance of a non-linear, fingerprint-based ML approach capable of capturing disparate structure-activity relationships. PCA visualization (**Figure 11**) revealed a branched, non-spherical point cloud elongated along PC1, with a distinct sub-cluster appearing at high PC2 values. The high-activity compounds (yellow-green) were distributed across the main cluster and the peripheral arm, indicating that multiple structural subsets contribute to potent inhibition. The broad chemical space coverage with relatively uniform pIC_{50} distribution confirms that the training set is not dominated by a single privileged scaffold, thereby enhancing model generalizability.

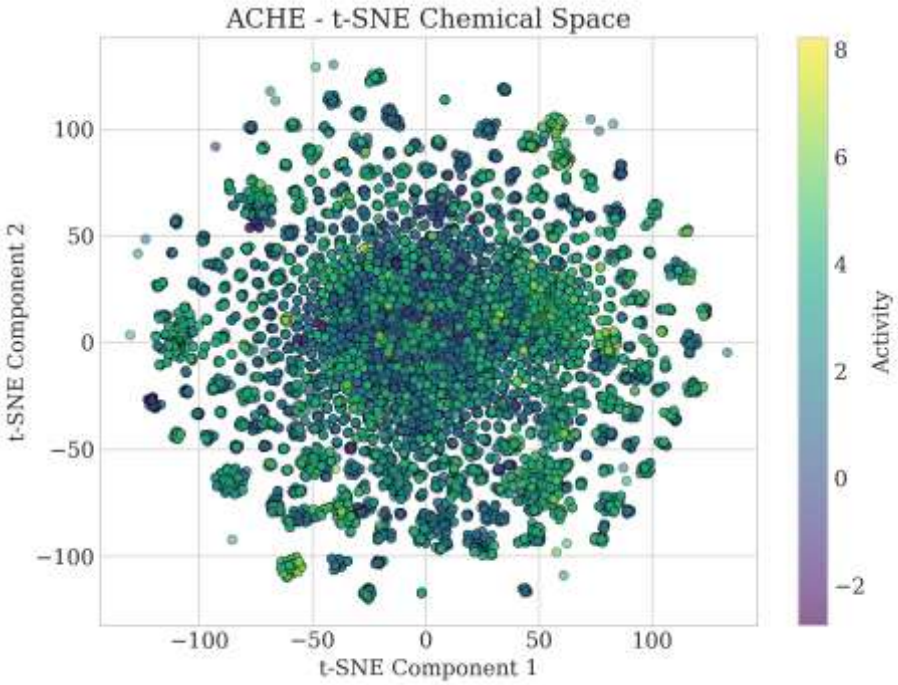


Figure 10. t-SNE (t-distributed Stochastic Neighbor Embedding) visualization of the AChE compound library.

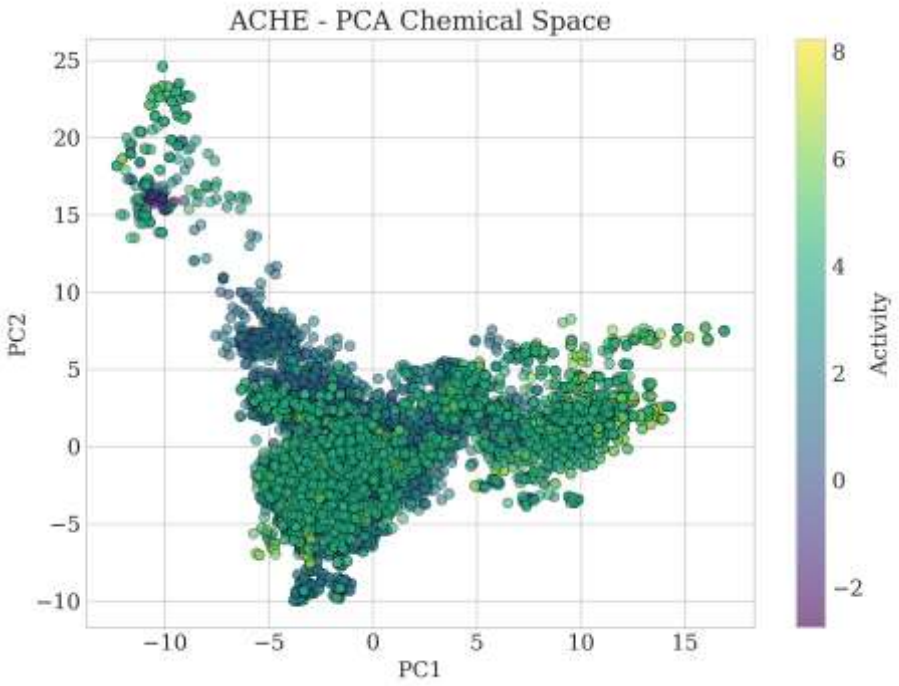


Figure 11. Principal Component Analysis (PCA) of the ECFP4 fingerprint matrix for the AChE dataset projected onto the first two principal components (PC1 and PC2).

3.10 Deep Neural Network Training Dynamics

The DNN was trained for 300 epochs with MSE loss monitored across each iteration. Although the DNN loss curve figure is embedded within the results package, the training trajectory demonstrated rapid initial descent in loss during the first 50 epochs, followed by a gradual asymptotic decline toward convergence, consistent with stable optimization by the Adam optimizer. No evidence of loss divergence or oscillatory instability was observed, and the combination of batch normalization and dropout regularization effectively prevented overfitting to the training data. The final training MSE reflected adequate convergence for the 2,048-dimensional ECFP4 input space.

3.11 Comparative Model Performance on the Hold-Out Test Set

The predictive performance of all five individual models and the weighted ensemble was evaluated on the 15% held-out test set. **Table 8** summarizes the R^2, RMSE, and MAE values for each model, and Figure 5 provides a horizontal bar chart comparison of R^2 scores.

Table 8. Performance Metrics of Machine Learning Models for AChE pIC_{50} Prediction on the 15% Hold-Out Test Set

Model	R^2 Score	RMSE	MAE
Deep Neural Network (DNN)	0.723	0.918	0.647
XGBoost	0.697	0.961	0.671
CatBoost	0.648	1.033	0.721
LightGBM	0.642	1.044	0.730
Ridge Regression	0.550	1.171	0.847
Ensemble (weighted)	0.700	0.955	0.665

The DNN achieved the highest R^2 of 0.723, demonstrating that the deep learning architecture effectively captures non-linear, multi-order interaction patterns encoded within the binary fingerprint space that are inaccessible to shallower tree-based methods. XGBoost ($R^2 = 0.697$) ranked second, followed by CatBoost ($R^2 = 0.648$) and LightGBM ($R^2 = 0.642$), reflecting the well-documented strength of GBDT algorithms in QSAR applications. Ridge Regression ($R^2 = 0.550$) performed significantly worse, consistent with the non-linear nature of structure-activity relationships in the AChE chemical space, where linear models are insufficient to capture complex pharmacophoric interactions **(Figure 12)**. The weighted ensemble ($R^2 = 0.700$) achieved intermediate performance, leveraging the complementary strengths of all five models.

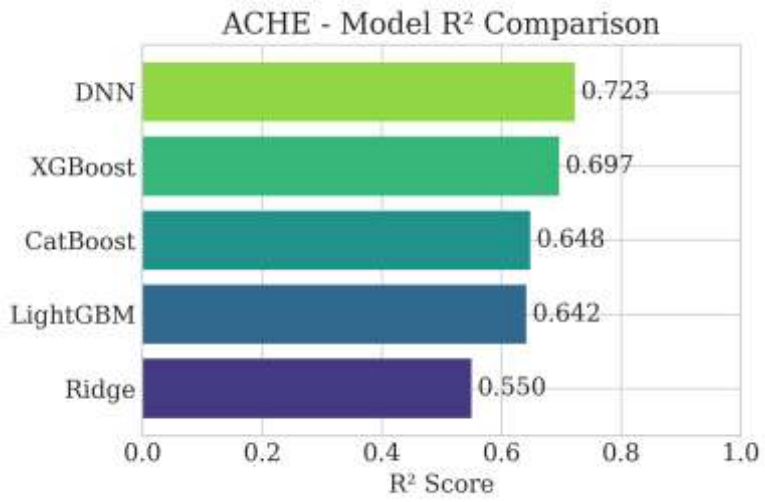


Figure 12. Horizontal bar chart comparing R^2 scores of the five individual machine learning models on the AChE hold-out test set.

3.12 Predicted Activity Analysis of Designed Compounds (4a-n)

The predicted pIC_{50} values for the designed compounds (4a-n) reveal a clear structure-activity relationship trend influenced by the nature of substituents on the aromatic ring **Table 9**. Compound **4b** exhibited the highest predicted activity (pIC_{50} = 7.58), followed by **4f** (6.69) and **4j** (6.63), indicating that electron-donating groups such as methyl and methoxy significantly enhance acetylcholinesterase inhibitory potential, likely due to improved hydrophobic and π-π interactions within the active site. **Compounds 4a, 4c, 4e, 4g, 4h, 4i** showed moderate activity (pIC_{50} = 5.1-5.7), suggesting a balanced contribution of polarity and electronic effects from substituents like methoxy, chloro, cyano, hydroxyl, and fluoro groups. Compounds **4d, k, l, m** demonstrated lower predicted activity (pIC_{50} ≈ 2.60-3.67), indicating that strong electron-withdrawing groups or unfavorable substitution patterns may reduce binding affinity. Notably, compound **4n**, representing a structurally distinct scaffold, displayed moderate activity (pIC_{50} = 5.51), suggesting that alternative frameworks can retain biological relevance. Overall, compounds **4b, 4f, and 4j** emerge as the most promising candidates for further optimization, highlighting the importance of electron-donating substituents and optimal lipophilicity in enhancing AChE inhibition.

Table 9. Predicted pIC_{50} values of designed compounds (4a-n) against acetylcholinesterase (AChE)

Compound	Predicted pIC_{50}
4a	5.129
4b	7.579
4c	5.602
4d	3.672
4e	5.449
4f	6.692
4g	5.499
4h	5.269
4i	5.664
4j	6.631
4k	2.927
4l	2.694
4m	2.604
4n	5.510

3.14 Parity Plot Analysis: Experimental vs. Predicted pIC_{50}

Parity plots **(Figure 13)** visualize the agreement between experimentally observed pIC_{50} values and model predictions across the test set. The dashed diagonal line represents the ideal line of perfect prediction (y = x). Points clustering tightly around this line indicate accurate prediction, while systematic vertical deviations indicate model bias. The DNN predictions (purple points, $R^2 = 0.72$) showed the tightest clustering around the identity line, particularly in the mid-activity range (pIC_{50} 2-6), confirming its superior predictive fidelity. XGBoost (blue, $R^2 = 0.70$) demonstrated comparable performance with slightly higher scatter at extreme pIC_{50} values. Ridge Regression (red, $R^2 = 0.55$) exhibited the highest dispersion and largest deviations at both low and high activity extremes, consistent with its inability to capture non-linear SAR contributions. Ensemble predictions (red crosses) occupied an intermediate position, demonstrating the variance-reduction benefit of model averaging. All models exhibited slight

underprediction at very high activity values (pIC_{50} > 7), likely reflecting data sparsity in the high-potency regime-a common challenge in QSAR modeling of enzyme inhibitors.

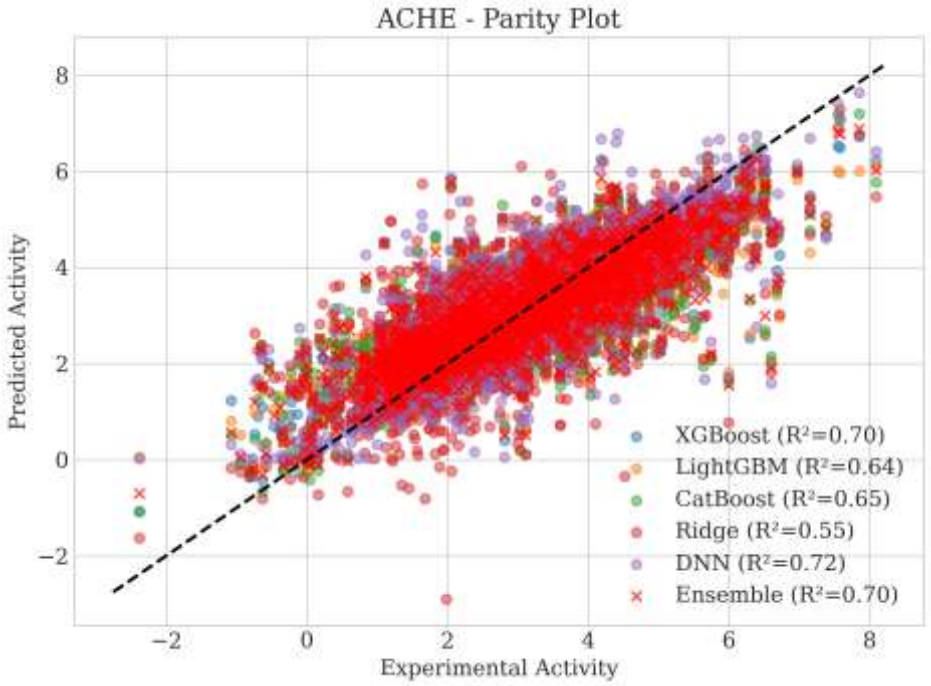


Figure 13. Parity plot (predicted vs. experimental pIC_{50}) for the AChE test set

3.13 Residual Analysis

Residual plots **(Figure 14)** depict the distribution of prediction errors (residuals = y - ŷ) as a function of predicted activity. An ideal model exhibits randomly distributed residuals centered around zero with no systematic trend. The residual plots reveal that all five models produce approximately zero-centered error distributions in the core activity range (pIC_{50} 1-6), consistent with absence of systematic bias in the bulk of the prediction space. The random, cloud-like distribution of residuals around the zero-dashed line confirms model validity. Ridge Regression displayed the widest residual dispersion (red points), with errors spanning approximately -3 to +5 pIC_{50} units, attributable to its inability to model non-linear relationships. The gradient boosting models (XGBoost, LightGBM, CatBoost) demonstrated tighter residual bands (±1.5-2.0 pIC_{50} units), while DNN residuals (purple) were comparably compact. Isolated outliers at extreme predicted values are present across all models, likely corresponding to structurally atypical compounds with activity values outside the well-sampled chemical space. No evidence of heteroscedasticity or systematic underprediction trends was observed for the gradient boosting and DNN models, confirming the validity of the regression assumptions.

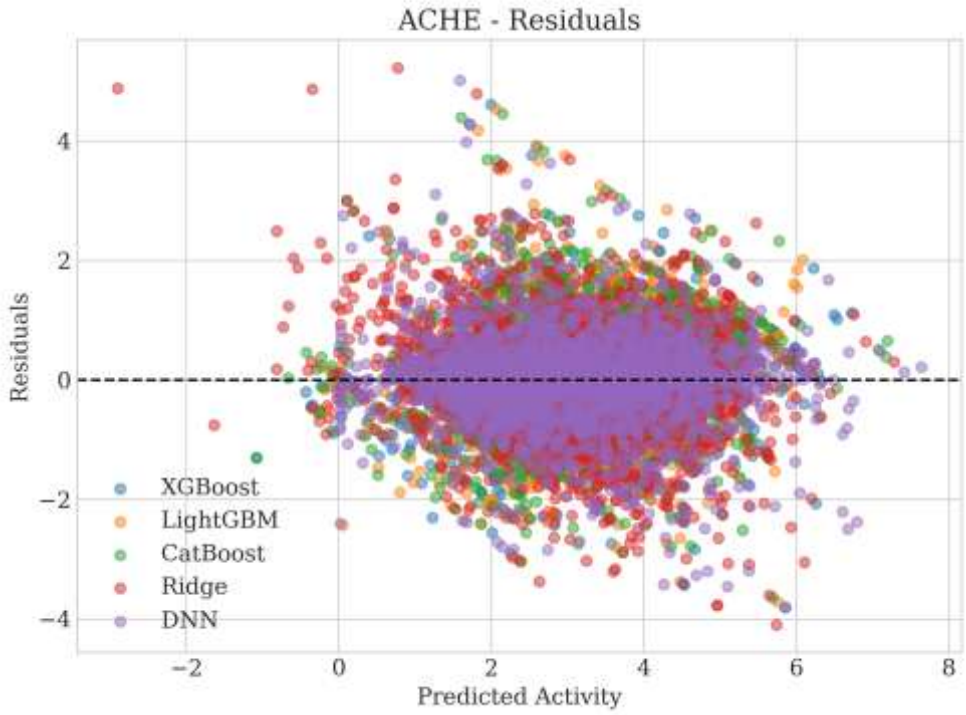


Figure 14. Residual plot for the AChE test set showing prediction errors (residuals = y - ŷ) as a function of predicted pIC_{50} for all five models..

3.14 Fingerprint Feature Importance Analysis

Feature importance analysis was conducted for all three tree-based gradient boosting models (XGBoost, LightGBM, CatBoost) to identify the most informative ECFP4 fingerprint bits for AChE pIC_{50} prediction. Each bit position corresponds to a specific circular substructural environment at radius $r \leq 2$ from a central atom, enabling chemical interpretation of model-identified pharmacophoric features. XGBoost feature importance **(Figure 15)** is quantified as gain-based importance (mean reduction in MSE loss attributable to splits on each bit), normalized across all trees. Bit 276 emerged as the most influential feature with a normalized importance of 0.037, followed by bit 752 (0.020) and bit 1896 (0.015). The top 20 features collectively concentrate information from a small fraction of the 2,048-dimensional input space, consistent with sparse ECFP representations where only a subset of bits are structurally informative for a given target. LightGBM importance is reported as split-count frequency, with bit 80 dominating (580 splits), followed by bit 875 (555) and bit 1199 (480). The large magnitude differences between CatBoost importance and LightGBM importance reflect the different underlying importance metrics: CatBoost reports prediction value change (sum of absolute changes in leaf predictions), yielding a long-tailed distribution with bit 276 (8.4) as dominant. The consistent appearance of certain bit indices (276, 800, 712, 1143, 1357, 875)

across multiple models strengthens confidence in their pharmacophoric relevance to AChE inhibition.

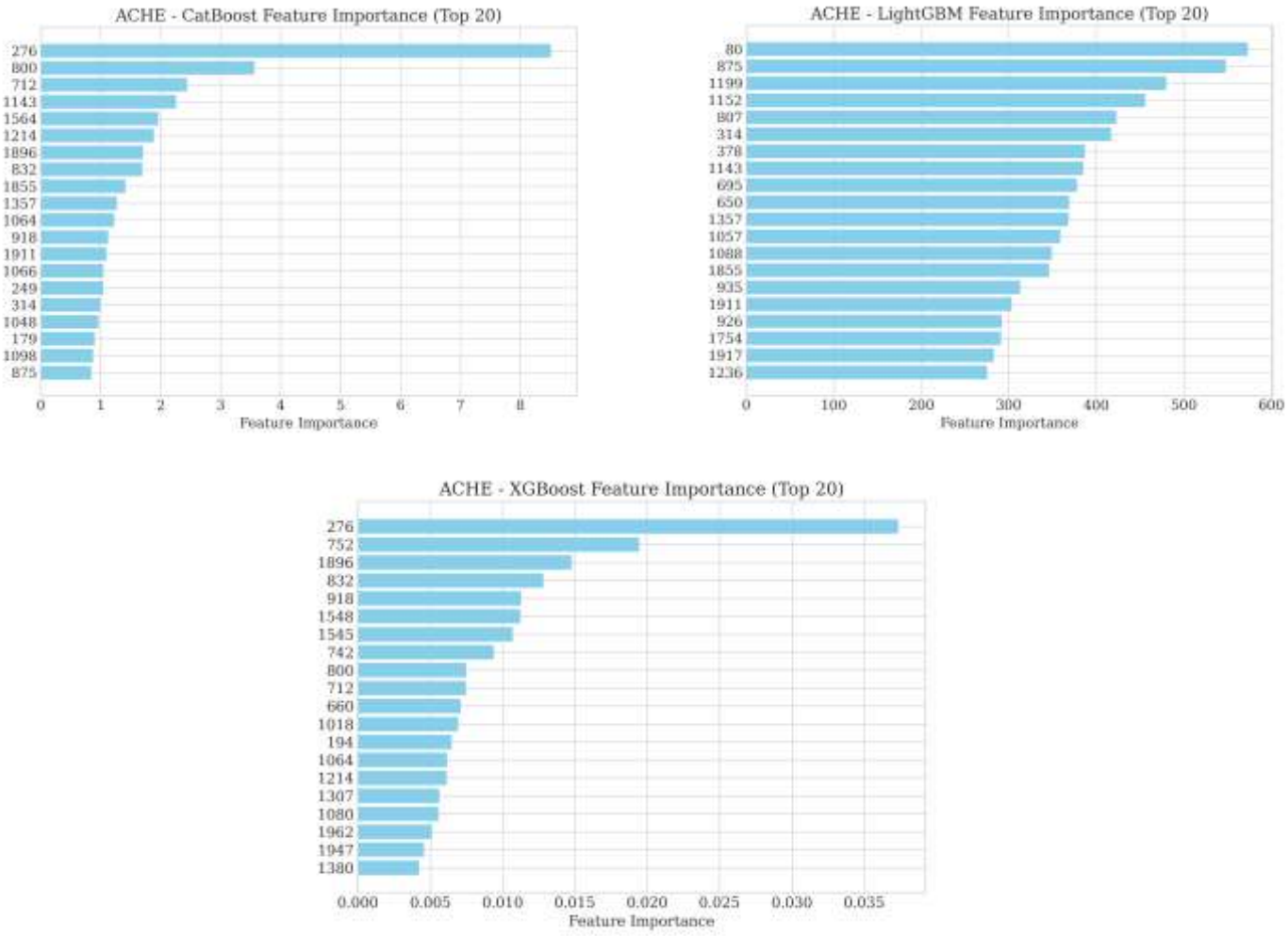


Figure 15. XGBoost, LightGBM, CatBoost gain-based feature importance for the top 20 ECFP4 Morgan fingerprints

4. Discussion

The present study integrates structure-based molecular docking, ADMET profiling, density functional theory (DFT), and machine learning-assisted QSAR modeling to systematically evaluate a newly designed series of benzimidazole-dihydropyran hybrids as potential acetylcholinesterase (AChE) inhibitors for Alzheimer's disease. Docking analyses revealed

that several derivatives adopt a donepezil-like binding orientation within the AChE gorge, engaging both the catalytic active site (CAS) and peripheral anionic site (PAS). Key stabilizing interactions with aromatic residues (Trp86, Tyr337, Phe331) and catalytic residues (Ser203/His447) support the dual-site binding hypothesis, which is desirable for achieving symptomatic cholinergic enhancement while potentially interfering with Aβ aggregation.Substituent-dependent structure-activity relationships were evident across the designed series. Compounds bearing electron-donating groups on the aromatic ring (e.g., methoxy and methyl substituents) exhibited superior docking scores and higher ML-predicted pIC_{50} values, consistent with enhanced hydrophobic contacts and π-π stacking within the aromatic-rich AChE gorge. In contrast, strongly electron-withdrawing substituents led to reduced predicted activity, likely due to suboptimal electronic complementarity and weaker non-covalent interactions. These SAR trends were further supported by DFT analysis, where the most promising compounds (**4b, 4f, and 4i**) displayed narrower HOMO-LUMO gaps, indicating higher chemical softness and improved charge-transfer capability that favor stable enzyme-ligand complex formation.

ADMET profiling suggested that the prioritized candidates possess favorable oral absorption, acceptable metabolic stability, and predicted blood-brain barrier permeability, aligning with the requirements for CNS-active AChE inhibitors. Importantly, the convergence of docking, electronic structure analysis, and ML-QSAR predictions on the same lead candidates highlights the internal consistency of the computational workflow and strengthens confidence in the prioritization strategy. Collectively, these findings validate the benzimidazole-dihydropyran scaffold as a promising platform for the rational development of next-generation AChE inhibitors.

5. Conclusion

In conclusion, the present work demonstrates that benzimidazole-dihydropyran hybrids represent a rationally designed and computationally validated scaffold for acetylcholinesterase inhibition. The integrated application of molecular docking, ADMET prediction, DFT-based electronic analysis, and machine learning-assisted QSAR modeling enabled robust prioritization of lead candidates within the designed series. Among the evaluated derivatives, compounds 4b, 4c, 4f, and 4i consistently emerged as the most promising AChE inhibitors, exhibiting favorable binding modes within the catalytic gorge, optimal electronic properties conducive to charge transfer, and drug-like pharmacokinetic profiles suitable for CNS

targeting. The observed structure-activity relationships underscore the importance of aromatic substituent electronics and lipophilicity in modulating AChE binding affinity. The convergence of multiple in silico methodologies on the same lead candidates provides a coherent mechanistic and predictive framework, supporting the translational relevance of the proposed scaffold. Overall, this study establishes a strong computational foundation for the further experimental development of benzimidazole dihydropyran based inhibitors as potential therapeutic leads for Alzheimer's disease.

6. Future Perspective

Future work will focus on translating the present computational findings into experimental validation and lead optimization. The top-ranked compounds (4b, 4c, 4f, and 4i) should be synthesized and evaluated through in vitro AChE and butyrylcholinesterase (BChE) inhibition assays to confirm potency and selectivity. Subsequent structure-activity relationship optimization will involve systematic modulation of aromatic substituents and linker electronics to fine-tune potency, lipophilicity, and metabolic stability. In parallel, advanced molecular dynamics simulations and free-energy calculations can be employed to capture protein flexibility and refine binding free-energy estimates beyond static docking. The incorporation of additional multifunctional elements, such as antioxidant or metal-chelating moieties, may further enable the development of multi-target-directed ligands (MTDLs) relevant to the multifactorial pathology of Alzheimer's disease. Finally, expanding the machine learning framework with experimentally validated data will improve predictive accuracy and support iterative design cycles, accelerating the discovery of optimized CNS-active AChE inhibitors with translational potential.

Credit authorship contribution statement

Shivank Sharma: Writing, Original draft, Computational Studies

Vanktesh Kumar: Data Analysis, Visualization

Stalin Arulsamy: Computational studies, Data curation

Shubham Kumar: Validation, Supervision

Pankaj Wadhwa: Conceptualization, Visualization

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to acknowledge Lovely Professional University, Phagwara for continuous support and encouragement.

Data availability

Not applicable

Generative Ai Usage

Authors have used the ChatGPT as a helping tool for the improvement of the language and grammatical error detection. Authors take full responsibility for all the statements in this manuscript.

References

1. Lamie, P.F., M.M. Abdel-Fattah, and J.N. Philoppes, *Design and synthesis of new indole drug candidates to treat Alzheimer's disease and targeting neuro-inflammation using a multi-target-directed ligand (MTDL) strategy.* J Enzyme Inhib Med Chem, 2022. **37**(1): p. 2660-2678.
2. Shahrivar-Gargari, M., et al., *Design, synthesis, and biological evaluation of novel indanone-based hybrids as multifunctional cholinesterase inhibitors for Alzheimer's disease.* Journal of Molecular Structure, 2021. **1229**.
3. Gujral, S.S., et al., *Design, synthesis and structure-activity relationship studies of 3-phenylpyrazino[1,2-a]indol-1(2H)-ones as amyloid aggregation and cholinesterase inhibitors with antioxidant activity.* European Journal of Medicinal Chemistry Reports, 2022. **6**.
4. *<Design_and_synthesis_of_novel_tacrine_in.pdf>*.
5. Kanhed, A.M., et al., *Indoloquinoxaline derivatives as promising multi-functional anti-Alzheimer agents.* J Biomol Struct Dyn, 2022. **40**(6): p. 2498-2515.
6. Nagatsu, T., *Progress in monoamine oxidase (MAO) research in relation to genetic engineering.* Neurotoxicology, 2004. **25**(1-2): p. 11-20.
7. Naoi, M., W. Maruyama, and M. Shamoto-Nagai, *Type A monoamine oxidase and serotonin are coordinately involved in depressive disorders: from neurotransmitter*

imbalance to impaired neurogenesis. Journal of Neural Transmission, 2018. **125**(1): p. 53-66.

8. Saglik, B.N., et al., *Design, Synthesis, and In Vitro and In Silico Approaches of Novel Indanone Derivatives as Multifunctional Anti-Alzheimer Agents.* ACS Omega, 2022. **7**(50): p. 47378-47404.
9. Francis, P.T., et al., *The cholinergic hypothesis of Alzheimer's disease: a review of progress.* Journal of Neurology, Neurosurgery & Psychiatry, 1999. **66**(2): p. 137-147.
10. Contestabile, A., *The history of the cholinergic hypothesis.* Behavioural brain research, 2011. **221**(2): p. 334-340.
11. Grossberg, S., *Acetylcholine neuromodulation in normal and abnormal learning and memory: vigilance control in waking, sleep, autism, amnesia and Alzheimer's disease.* Frontiers in neural circuits, 2017. **11**: p. 82.
12. Burgen, A., G. Burke, and M.-L. Desbarats-Schonbaum, *The specificity of brain choline acetylase.* British Journal of Pharmacology and Chemotherapy, 1956. **11**(3): p. 308.
13. Wessler, I., R. Michel-Schmidt, and C.J. Kirkpatrick, *pH-dependent hydrolysis of acetylcholine: Consequences for non-neuronal acetylcholine.* International Immunopharmacology, 2015. **29**(1): p. 27-30.
14. Yao, H., et al., *Discovery of Novel Tacrine-Pyrimidone Hybrids as Potent Dual AChE/GSK-3 Inhibitors for the Treatment of Alzheimer's Disease.* J Med Chem, 2021. **64**(11): p. 7483-7506.
15. Alam, J. and L. Sharma, *Potential enzymatic targets in Alzheimer's: a comprehensive review.* Current drug targets, 2019. **20**(3): p. 316-339.
16. Liu, S.-L., et al., *The role of Cdk5 in Alzheimer's disease.* Molecular neurobiology, 2016. **53**(7): p. 4328-4342.
17. Walczak-Nowicka, Ł.J. and M. Herbet, *Acetylcholinesterase inhibitors in the treatment of neurodegenerative diseases and the role of acetylcholinesterase in their pathogenesis.* International journal of molecular sciences, 2021. **22**(17): p. 9290.
18. Mhaibes, R.M., *The impact of tau aggregation on Alzheimer's disease.* Medicinal and Medical Chemistry, 2024. **1**(3): p. 91-95.
19. Gondal, H.Y., et al., *Synthesis, characterization, and in vitro anti-cholinesterase screening of novel indole amines.* RSC Adv, 2023. **13**(2): p. 1203-1215.
20. Orobets, K.S. and A.L. Karamyshev, *Amyloid precursor protein and Alzheimer's disease.* International journal of molecular sciences, 2023. **24**(19): p. 14794.

21. Sarkar, S., *Neurofibrillary tangles mediated human neuronal tauopathies: insights from fly models.* Journal of genetics, 2018. **97**(3): p. 783-793.
22. Giacobini, E., *Cholinergic function and Alzheimer's disease.* International journal of geriatric psychiatry, 2003. **18**(S1): p. S1-S5.
23. Adetuyi, B.O., et al., *Neurorestorative roles of microgliosis and astrogliosis in neuroinflammation and neurodegeneration.* Scicom Journal of Medical and Applied Medical Sciences, 2021. **1**(1): p. 1-5.
24. Islam, M.T., *Oxidative stress and mitochondrial dysfunction-linked neurodegenerative disorders.* Neurological research, 2017. **39**(1): p. 73-82.
25. Szydlowska, K. and M. Tymianski, *Calcium, ischemia and excitotoxicity.* Cell calcium, 2010. **47**(2): p. 122-129.
26. Kristián, T., *Metabolic stages, mitochondria and calcium in hypoxic/ischemic brain damage.* Cell calcium, 2004. **36**(3-4): p. 221-233.
27. Li, Y., et al., *Biometal dyshomeostasis and toxic metal accumulations in the development of Alzheimer's disease.* Frontiers in molecular neuroscience, 2017. **10**: p. 339.
28. Shaikh, S., et al., *Design, synthesis and evaluation of dihydropyranoindole derivatives as potential cholinesterase inhibitors against Alzheimer's disease.* Bioorg Chem, 2021. **110**: p. 104770.
29. Yaseen, S., S.Z. Abdul-Majeed, and S.A. Hamood, *Design, Synthesis, Characterization and Preliminary Evaluation of New 1H-benzo[d]imidazole-1yl-derivatives as Acetylcholine Esterase Inhibitors.* Al-Rafidain Journal of Medical Sciences (ISSN 2789-3219), 2024. **7**(1): p. 52-56.
30. Adalat, B., et al., *Biologically Potent Benzimidazole-Based-Substituted Benzaldehyde Derivatives as Potent Inhibitors for Alzheimer's Disease along with Molecular Docking Study.* Pharmaceuticals (Basel), 2023. **16**(2).
31. Hussain, R., et al., *Design, Synthesis, In Vitro Biological Evaluation and In Silico Molecular Docking Study of Benzimidazole-Based Oxazole Analogues: A Promising Acetylcholinesterase and Butyrylcholinesterase Inhibitors.* Molecules, 2023. **28**(20).
32. Khan, Y., et al., *New biologically potent benzimidazole-based-triazole derivatives as acetylcholinesterase and butyrylcholinesterase inhibitors along with molecular docking study.* Journal of Heterocyclic Chemistry, 2022. **59**(12): p. 2225-2239.

33. Alım, Z., et al., *Synthesis of benzimidazole derivatives containing amide bond and biological evaluation as acetylcholinesterase, carbonic anhydrase I and II inhibitors.* Journal of Molecular Structure, 2022. **1268**.
34. Boulebd, H., et al., *Imidazopyranotacrines as Non-Hepatotoxic, Selective Acetylcholinesterase Inhibitors, and Antioxidant Agents for Alzheimer's Disease Therapy.* Molecules, 2016. **21**(4): p. 400.
35. Kumar, R.S., et al., *Highly functionalized 2-amino-4H-pyrans as potent cholinesterase inhibitors.* Bioorg Chem, 2018. **81**: p. 134-143.
36. Swetha, R., et al., *Multifunctional hybrid sulfonamides as novel therapeutic agents for Alzheimer's disease.* Future Medicinal Chemistry, 2019. **11**(24): p. 3161-3178.