

# Novel Thiosemicarbazone Quantum Dots in the Treatment of Alzheimer's Disease Combining In Silico Models Using Fingerprints and Physicochemical Descriptors

Nguyen Minh Quang, Hoa Tran Thai,\* Hoa Le Thi, Nguyen Duc Cuong, Nguyen Quoc Hien, DongQuy Hoang, Vu Thi Bao Ngoc, Vo Ky Minh, and Pham Van Tat\*



Cite This: *ACS Omega* 2023, 8, 11076–11099



Read Online

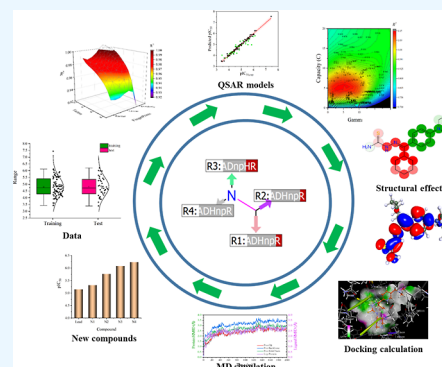
ACCESS |

Metrics & More

Article Recommendations

Supporting Information

**ABSTRACT:** Searching for thiosemicarbazone derivatives with the potential to inhibit acetylcholinesterase for the treatment of Alzheimer's disease (AD) is an important current goal. The QSAR<sub>KPLS</sub>, QSAR<sub>ANN</sub>, and QSAR<sub>SVR</sub> models were constructed using binary fingerprints and physicochemical (PC) descriptors of 129 thiosemicarbazone compounds screened from a database of 3791 derivatives. The  $R^2$  and  $Q^2$  values for the QSAR<sub>KPLS</sub>, QSAR<sub>ANN</sub>, and QSAR<sub>SVR</sub> models are greater than 0.925 and 0.713 using dendritic fingerprint (DF) and PC descriptors, respectively. The in vitro pIC<sub>50</sub> activities of four new design-oriented compounds N1, N2, N3, and N4, from the QSAR<sub>KPLS</sub> model using DFs, are consistent with the experimental results and those from the QSAR<sub>ANN</sub> and QSAR<sub>SVR</sub> models. The designed compounds N1, N2, N3, and N4 do not violate Lipinski-5 and Veber rules using the ADME and BoiLED-Egg methods. The binding energy, kcal mol<sup>-1</sup>, of the novel compounds to the 1ACJ-PDB protein receptor of the AChE enzyme was also obtained by molecular docking and dynamics simulations consistent with those predicted from the QSAR<sub>ANN</sub> and QSAR<sub>SVR</sub> models. New compounds N1, N2, N3, and N4 were synthesized, and the experimental in vitro pIC<sub>50</sub> activity was determined in agreement with those obtained from in silico models. The newly synthesized thiosemicarbazones N1, N2, N3, and N4 can inhibit 1ACJ-PDB, which is predicted to be able to cross the barrier. The DFT B3LYP/def-SV(P)-ECP quantization calculation method was used to calculate  $E_{\text{HOMO}}$  and  $E_{\text{LUMO}}$  to account for the activities of compounds N1, N2, N3, and N4. The quantum calculation results explained are consistent with those obtained in in silico models. The successful results here may contribute to the search for new drugs for the treatment of AD.



## 1. INTRODUCTION

In recent years, there has been much interest in exploring and developing new active drugs, particularly for generating prospective small-molecular-weight pharmaceuticals interacting with various biological targets.<sup>1</sup> This can also be a streamlined drug design approach. Researching and finding potential new drugs will take years and be quite expensive.<sup>2</sup> Furthermore, successful results are achievable with only a tiny fraction of novel compounds getting regulatory approval.<sup>3</sup> Finally, the medicine will be manufactured and commercialized. This procedure frequently includes numerous preclinical and clinical experiments. Clinical studies seek to evaluate whether a novel treatment is safer and more effective than an existing one.<sup>4,5</sup> The primary tasks in the preclinical testing phase are connected to discovering possible novel drugs. Potentially active compounds can eventually transform into a model compound that can be improved later.<sup>6,7</sup>

One of the key screening strategies resulted in the performance of new compound searches. A set of test chemicals may be screened against biochemical or cellular targets.<sup>1,7</sup> The candidate compounds identified during the screening process are deemed the most appropriate. Following

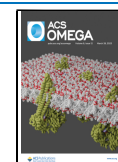
the screening, both PC and pharmacological parameters were examined concurrently.<sup>8</sup> They were then assessed for their potential to synthesize a lead compound.<sup>9</sup> Because of the various effector effects, the screening data acquired may have limitations in general. Positive results can frequently pave the way for further essential tests. Negative screening results, on the other hand, may become less important during testing. We can employ computational methods as an essential option to overcome the limits of the screening process.<sup>10,11</sup>

Furthermore, the computational method enables both time- and cost-effective solutions. Virtual screening procedures can be a viable option. Virtual screening uses a vast data library of compounds created and screened using an in silico model to provide a selection of potentially valuable drugs. This is

**Received:** December 13, 2022

**Accepted:** March 7, 2023

**Published:** March 17, 2023



regarded as an effective tool in modern drug discovery.<sup>12</sup> Only when X-ray crystallography data are available is the docking method preferred. In the absence of X-ray protein crystal structures, ligand-based approaches such as pharmacological group mapping, shape-based screening, and fingerprint similarity may be the only viable and appropriate option. If there are many test compounds, the ligand-based technique is favored.<sup>13</sup>

In practice, the virtual screening process depends on the data set. The fingerprinting approach will become an extremely effective tool for detecting new drugs. Furthermore, this approach efficiently retrieves chemicals that are bound to protein targets. It can provide a solid foundation for fragmentation investigations, such as associative atomic diagrams.<sup>14</sup> There can be tens of thousands of different atom arrangements. Using variable combinations cleared the path for producing realistic directions in searches using fingerprinting techniques. The data set size and the intricacy of the correlation between variables were reduced. This inquiry could be the best instrument for forecasting a molecule's property–expression interactions. Beyond its active properties, this exhibits chemical diversity and other features.<sup>15</sup>

Alzheimer's disease (AD) affects around 44 million individuals globally. There are currently no cures for AD and just symptomatic treatments. Given the scarcity of effective treatments for AD due to its multifactorial pathophysiology, developing new multitargeted medicines is desirable. AD is one of the most serious neurological disorders affecting the elderly.<sup>22</sup> Donepezil, an acetylcholinesterase inhibitor, was used to create a new class of thiosemicarbazone derivatives.<sup>23</sup> These thiosemicarbazones are intended to target five important AD markers: low acetylcholinesterase levels, autophagy dysfunction, metal homeostasis disruptions, protein aggregation, and oxidative stress. PBPT (pyridoxal 4-N-(1-benzylpiperidin-4-yl) thiosemicarbazone) emerged as the leading compound among these thiosemicarbazones.<sup>23,24</sup> This drug demonstrated the most promising multifunctional activity, with very low antiproliferative activity, considerable chelation efficiency, oxidative stress inhibition, and acetylcholinesterase inhibitory activity. These varied qualities show PBPT's potential as a multifunctional drug for treating AD. The biological activity of the thiosemicarbazone derivatives was investigated using acetylcholinesterase (AChE).<sup>25</sup> Acetylcholinesterase is engaged in the termination of impulse transmission by rapidly hydrolyzing the neurotransmitter acetylcholine in several cholinergic pathways in the central and peripheral nervous systems. Enzyme inactivation caused by different inhibitors results in acetylcholine buildup, overstimulation of nicotinic and muscarinic receptors, and neurotransmission disturbance. As a result, as related medicines and toxins, acetylcholinesterase inhibitors interact with the enzyme as their principal target.<sup>26</sup>

This introduction discusses the toxicity and pharmacology of reversible and irreversible acetylcholinesterase-inactivating drugs. In the case of commonly used reversible inhibitors in treating neurodegenerative illnesses, special attention should be paid to the currently approved drugs (donepezil, rivastigmine, and galantamine) in the pharmacopeia. Pesticides that cause AD and include toxic carbamates are utilized.<sup>26,27</sup> Following that, the irreversible mechanism of acetylcholinesterase inhibition generated by organophosphorus chemicals and their specific and nonspecific harmful effects, as well as irreversible inhibitors of pharmaceutical actions, were

described.<sup>25</sup> In addition, the pharmacological therapy of organophosphate poisoning is discussed, emphasizing oxime reactants with decreased enzymatic activity employed as causative medicines following poisoning.<sup>28</sup>

Furthermore, organophosphorus and carbamate pesticides can be detoxified in mammals by enzymatic hydrolysis before reaching their targets in the nervous system. Carboxylesterases are the most efficient carbamate degraders, whereas phosphotriesterases are the most effective organophosphate detoxifiers.<sup>29</sup> Cholinesterase inhibitors (also known as acetylcholinesterase inhibitors) are a type of medication that prevents acetylcholine from being broken down usually. Acetylcholine is the body principal neurotransmitter, acting in the peripheral and central nervous systems. Acetylcholine, for example, is released by motor neurons to stimulate muscles; acetylcholinesterase is also involved in arousal, attention, learning, memory, and motivation. Cholinesterase inhibitors inhibit the cholinesterase enzyme, which is responsible for the breakdown of acetylcholine. The concentration of acetylcholine in the synaptic cleft rises as a result (the space between two nerve endings).<sup>30</sup> Cholinesterase inhibitors are mostly used to treat dementia in AD patients. Acetylcholine levels in the brain are lower in those with AD.<sup>27</sup> Cholinesterase inhibitors have only had a minor impact on dementia symptoms such as cognition.<sup>31</sup> The protein acetylcholinesterase is found in the organism homo sapiens. Acetylcholinesterase stops signal transduction at the synapse by rapidly hydrolyzing the released acetylcholine into the synaptic cleft.<sup>27</sup> Cholinesterase inhibitors can produce vasodilation, pupil constriction, increased sweat, saliva, tear secretion, decreased heart rate, and mucus secretion in the respiratory system.<sup>30,31</sup> Biological or chemical properties can be precisely predicted using either a continuum or a categorization.<sup>32,33</sup>

Nowadays, with the strong development of computer techniques, machine learning techniques have become important tools applied in different scientific fields. With the development of supervised machine learning methods such as support vector regression, partial least-squares (PLS) regression, and artificial neural networks, these methods have now been significantly improved (SVR). PC descriptors of the structures and molecular fingerprints are used in these techniques.<sup>34,35,38</sup> The potential for solving issues has been dramatically increased as a result. Using the methods of artificial neural network QSAR<sub>ANN</sub>, support vector regression QSAR<sub>SVR</sub>, and kernel-based PLS regression QSAR<sub>KPLS</sub>, models were created from the relation between molecular descriptors and activity.<sup>17,18,38</sup> A trustworthy predictive QSAR model has been created that can be utilized to direct the creation of new chemicals and routine synthesis or acquisition of additional compounds. Applying a specific QSAR model to a set of possible compounds can help identify which ones have the necessary qualities.<sup>18,38</sup> This could help determine which aspect of chemical structure contributes favorably and adversely to the expected behavior. Instead of merely listing and testing a considerably more significant number of compounds, this method enables the logical design of the chemical structure by altering the structure by changing the next R functional group.<sup>17,38</sup> To achieve this goal, a variety of QSAR models can be applied. Recently, supervised machine learning algorithms have advanced significantly. Important elements of this approach include recursive partitioning and the QSAR method using 2D descriptors or chemical fingerprints. QSAR can also be applied to chemical modeling.

A highly efficient QSAR<sub>KPLS</sub> method based on unrestricted chemical fingerprinting in combination with kernel-based (KPLS) is also used.<sup>17,18,38</sup>

This study reports the design and synthesis of the acetylcholinesterase inhibitory potential quantum dot thiosemicarbazone derivatives based on in silico models, including QSAR<sub>KPLS</sub>, QSAR<sub>ANN</sub>, and QSAR<sub>SVR</sub> models, and docking and molecular dynamics (MD) simulations. Binary fingerprints and molecular physicochemical (PC) descriptors were used to construct the QSAR models, prediction compound activities, and decipher compound activities. Binary fingerprint distinctiveness is reduced to a single chemical component. The molecular structure is readily solved by the QSAR<sub>KPLS</sub> model using the atomic contributions of each R substituent. This explains the structural characteristics that are advantageous and disadvantageous while developing novel substances. A sufficient amount of data is provided by the neural network model QSAR<sub>ANN</sub> to predict the molecular activity effectively. With a molecular topology that is inspired, the QSAR<sub>ANN</sub> model is applied. The thiosemicarbazone class, which inhibits acetylcholinesterase for pharmacological objectives, was studied using the QSAR<sub>ANN</sub> model. In this paper, we also report employing docking calculation approaches, QSAR<sub>ANN</sub> and QSAR<sub>SVR</sub> models, and models to simultaneously predict in vitro activity and binding energy (BE).

## 2. DATA SETS AND METHODS

**2.1. Biodataset Selection.** The data were obtained from PubChem and ChEMBL databases with 3791 thiosemicarbazone compounds.<sup>39</sup> This study performed data screening based on the Agglomerative Hierarchical Clustering (AHC) statistical technique.<sup>40</sup> Agglomerative Hierarchical Clustering was used to make up homogeneous groups of objects (classes) based on their description by a set of variables or from a matrix describing the similarity or dissimilarity between the objects. After the categorical screening, 129 compounds, including inhibitory activity on acetylcholinesterase, were selected, as given in Table S1.<sup>39</sup> The 1ACJ-PDB protein is acetylcholinesterase selected from the PDB database of *Tetronarce californica*.<sup>41</sup> The binding sites of acetylcholinesterase for the quaternary ligands were investigated by X-ray crystallography and photosynthetic labeling. The crystal structure of the complex with the ligand was determined.<sup>41</sup>

The data set was randomly divided into a training set and a test set with a proportionality of 70:30, respectively. The IC<sub>50</sub> (nM) concentration activity values were converted to pIC<sub>50</sub> (−log IC<sub>50</sub>) values.<sup>39</sup> The descriptive statistics for the training and test data set are shown in Table 1. Before building the QSAR model, the distribution of the activity data sets was

checked,<sup>44</sup> as shown in Figure 1. Compounds were selected to be suitable for the QSAR study. The distribution of activities of

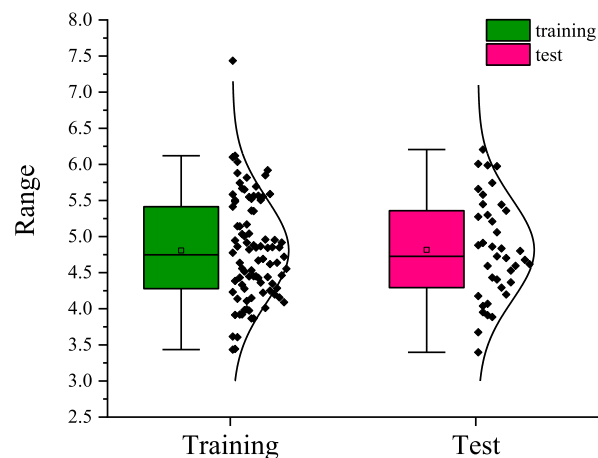


Figure 1. Distribution of training and test data.

the compounds met the requirement of the standard distribution. The pIC<sub>50</sub> activities of two paired data sets had high similarity in the range of 3.435–7.435 for the training set and 3.397–6.205 for the test set. All structures are optimized down to minimal energy using the UFF force field. This force field is capable of describing electrostatic interactions well.<sup>45</sup>

**2.2. Selection of Molecular Descriptors.** The molecular descriptors need to be computed to build the QSAR models and are mainly combined with binary fingerprints and PC descriptors.<sup>17,18</sup> The binary fingerprints can be obtained from molecular structures. There are different groups of fingerprints. The selection of each type of fingerprint depends on the molecular structure. Fingerprint types are often selected based on the error of the model. There are three commonly used fingerprint types: linear fingerprints (LFs), radial fingerprints (RFs), and dendritic fingerprints (DFs) that consist of linear and branched fragments.<sup>19</sup>

Besides, there are also PC descriptors such as molecular weight (MW), octanol–water partition (A log *P*), hydrogen bond acceptors (HBAs), hydrogen bond donors (HBDs), molar refraction (MR), polar surface area (PSA), total polar surface area (TPSA), molecular polarizability (Polar), number of rotatable bonds (nRB), electron topological states (Estate), chiral center count (CCC), heavy atom count (HAC), and ring count (RC).<sup>17,36</sup>

**2.3. QSAR<sub>KPLS</sub> Model.** The QSAR<sub>KPLS</sub> model is built on the same essential principle as when building a QSAR model from a training set of compounds. A group of independent *x* variables are initially assembled from a combination of fingerprints that “occurred” in at least one compound in the training set. Therefore, each variable *x* is a descriptor with a binary value (0/1) representing a particular chemical fragment found at least once. A given variable *x* will be discarded if its mean is 7 or more standard deviations (SDs) from the mean of all *x* variables in the group.<sup>46,47</sup> This process succeeds in removing fingerprints that are close to zero or almost always very close to 1. Additional *x* variables are removed by removing one variable from each pair with a more significant correlation than 0.95.<sup>46,47</sup> The remaining *x* variables are then automatically scaled (zero means and 1 SD), with all scaling factors saved to apply to any further test set data.<sup>17,47</sup> For a pair

Table 1. Descriptive Statistics for the Training and Test Sets for Acetylcholinesterase (AChE)

statistical values	training set	test set
mean	4.807	4.815
standard error	0.076	0.112
SD	0.720	0.702
sample variance	0.519	0.493
range	4.000	2.808
minimum	3.435	3.397
maximum	7.435	6.205
observations	90	39

of auto-scaled variables, represented by the column vectors  $x_i$  and  $x_j$ , the kernel matrix  $K$  is defined as

$$K_{ij} = \exp(-d_{ij}^2/\sigma_j^2) \text{ with } d = \|x_i - x_j\| \quad (1)$$

Here, the  $\sigma$  parameter is nonlinear and can be adjusted to optimize the accuracy of the missed predictions. The kernel matrix of the training and test sets is centralized, and the QSAR<sub>KPLS</sub> model using the five latent factors is optimal. This results in significantly fluctuating  $\sigma$  values upon random seed selection. However, the predictions of the QSAR<sub>KPLS</sub> model are insensitive to  $\sigma$  values of 20 or greater.<sup>17</sup> Therefore,  $\sigma$  is kept constant at 20 for all calculations.

**2.4. QSAR<sub>ANN</sub> Model.** The overall model of the built and tested neural network can simultaneously predict the pIC<sub>50</sub> activity and BE for the receptor active site of the AChE protein using a combination of selected fingerprints and PC descriptors.<sup>18,37</sup> The neural network model has a standard structure: an input layer, a hidden layer, and an output layer. The neural network models are trained using the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm.<sup>48</sup>

We choose the multilayer neural network model. The building process of the QSAR<sub>ANN</sub> model is done by the automated searching technique SANN of the neural network architecture.<sup>18</sup> The neurons on the hidden layer are automatically searched for between 8 and 25 neurons. The data set in Table S1 is divided into 93 (70%) compounds in the training set, 18 (15%) compounds in the validation set, and 18 (15%) compounds in the testing set. The automatically selected training algorithm is related to the learning rate, momentum, and target epochs. The specified learning rate is used to adjust the neural network weights. A gradient descent algorithm is also used. It is a first-order optimization algorithm that tries to move step by step to successively lower points in the search space to determine the minimum position. The training algorithm BFGS<sup>48,49</sup> is a powerful quadratic training algorithm with fast convergence. The epochs 10000 are used during the neural network training to adjust the weights to reduce the error.

The model QSAR<sub>ANN</sub> I( $m$ )-HL( $n$ )-O( $k$ ) includes the input layer I( $m$ ) with  $m$  input neurons as dendritic binary fingerprints, the PC descriptors, and a mixture of DF and PC descriptors; the hidden layer HL( $n$ ) with  $n$  hidden neurons; and the output layer O( $k$ ) with  $k$  output neurons that consists of the pIC<sub>50</sub> biological activity and BE, kcal mol<sup>-1</sup>, of the ligands to the receptor active site.

**2.5. Building the QSAR<sub>SVR</sub> Model.** The BE and pIC<sub>50</sub> activity data sets are nonlinearly distributed data. The QSAR<sub>ANN</sub> model simultaneously predicts the pIC<sub>50</sub> values and BEs for the compounds in the test set and the newly designed compounds. This study also builds a QSAR model based on a support vector regression technique. This is based on the concept of a decision plane defining decision boundaries. Machine learning techniques also include the support vector regression method. This method uses the separation plane between a data set with different distributed pIC<sub>50</sub> and BE values. This method is widely used for nonlinear data sets in different areas. Compounds can be initially mapped, i.e., rearranged, using a set of mathematical functions known as the kernel. The process of rearranging compounds is called mapping. Instead of making a complicated curve, optimal linearity can use the linear separability of compounds with different pIC<sub>50</sub> and BE values to separate them.

The functional dependence of the variables pIC<sub>50</sub> and BE on a set of molecular descriptors in the support vector regression model (QSAR<sub>SVR</sub>) needs to be estimated.<sup>50,51</sup> Similar to other regression methods, the relationship between the molecular descriptors and the pIC<sub>50</sub> activities and BE values are given by a deterministic function  $f(x)$  plus the addition of some additive noise

$$y = f(x) + \text{noise} \quad (2)$$

A process can consist of sequential optimization of an error function. This process includes classification and sequential optimization of the error function. A function  $f(x)$  needs to be defined to allow accurate prediction of new cases. This can be achieved by training the QSAR<sub>SVR</sub> model on a training set.<sup>50,51</sup> Depending on the definition of the error function  $f(x)$ , the function type can be identified. The first error function used is known as the epsilon-SVM regression

$$0.5w^T w + C \sum_{i=1}^N \xi_i + C \sum_{i=1}^N \xi_i^* \quad (3)$$

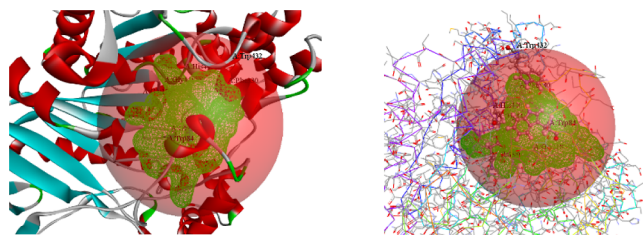
Kernel radial basis function (RBF) can be chosen for constructing the Support Vector Machines Regression model.<sup>50,51</sup> Here,  $C$  is the capacity constant and should be chosen carefully to avoid overfitting,  $w$  is the vector of coefficients,  $\xi_i$  represents the parameters for handling non-separable data (inputs), and the  $i$  index labels the  $N$  training cases. It should be noted that the larger the  $C$ , the more the error is penalized.<sup>18,51</sup>

$$K(X_i, X_j) = \{\exp(-\gamma |X_i - X_j|^2)\} \quad (4)$$

where  $K(X_i, X_j) = \phi X_i \cdot \phi X_j$ ; i.e., the kernel function represents a dot product of input data points mapped into the higher dimensional feature space by transformation  $\phi$ . The kernel  $\phi$  is used to transform data from the independent input to the feature space. Value gamma  $\gamma$  is an adjustable parameter of certain kernel functions.

**2.6. Molecular Docking Calculations.** **2.6.1. Protein Preparation.** To perform docking of protein receptor interactors, we considered the component-target pathway and protein interaction network and analyzed their importance in the KEGG pathway. The 1ACJ-PDB protein construct was selected as a molecular binding target from the RCSB PDB database (<https://www.pdb.org/>).<sup>41</sup> The ligands and solvents were removed from the protein structures—these structures were then stored in PDB format after hydrogenation and optimization. The optimal energy of the protein after optimization is  $-6151.25$  kcal mol<sup>-1</sup>. The active ingredients are screened and compared with the chemical components. The processed target protein and effective compound structures are added into AutoDock for molecular binding.<sup>52,53</sup> After binding tests, the most suitable structure is selected based on the BE value. The 3D structure of the 1ACJ-PDB protein with the active target site is prepared for binding, as shown in Figure 2.

The receptor is first mapped to a grid to identify a binding site. The receptor occupies grid points within a given distance of the marked receptor atoms and, therefore, is not considered the site for ligand atoms. The common practice for determining a binding site is based on a known ligand position already in an active site in coordinate  $x = 3.214$ ,  $y = 65.854$ , and  $z = 65.167$ , as illustrated in Figure 2. This coordinate site is used for binding 129 compounds and novel compounds to



**Figure 2.** The active target site of the AChE receptor includes amino acids TRP84, GLY118, GLU199, PHE330, TRP432, HIS440, and GLY441 from receptor cavities.

determine the significant interaction that inhibits the 1ACJ-PDB protein. Receptor–ligand interactions allow the investigation of the weak binding and nonbinding interactions between small ligands and protein receptors. Attaching ligands to proteins is a complex and challenging task. Binding is usually approximate in that the structure of the ligand and acceptor does not change during binding. The most common approximation for binding is to hold the protein as a rigid structure and attach a series of ligands to the active site. This study implements docking calculation in this way.

**2.6.2. Ligand Preparation.** A database of 129 compounds in Table S1 and newly designed compounds was created in research. The optimal energy of these compounds is minimized using the force field (UFF) with a total number of steps of 2000.<sup>45</sup> The optimization process is stopped to obtain stable optimal energy if the energy difference is less than 0.05. The structures of the ligand molecules were reconstructed using ChemDraw Ultra and saved in SDF file format. The energy-minimized compounds were then transferred to virtual screening by docking using the Autodock technique to simulate the ligand docking into the active site of the 1ACJ-PDB protein.<sup>41</sup>

**2.7. MD Simulations.** Simulations of MD often include many phases, each of which is carried out at a different pressure or temperature. After performing the docking calculation, we can carry out the MD simulation to determine the ability of ligand binding to the active site of the 1ACJ-PDB protein of AChE. The protein–ligand complex was placed in an aqueous solvent box containing 13,603 molecules of H<sub>2</sub>O along with NaCl at a concentration of 0.15 M. MD simulation<sup>42,43</sup> was performed under the NPT condition with  $N = 49,250$  atoms, and the system is heated from 0 to  $T = 300$  K, pressure  $P = 1.0325$  bar. The 1ACJ-PDB protein consists of 350 residuals and 4218 heavy atoms. The equilibration and production simulation times are given a run duration of 200 ns.

**2.8. Synthesis Path of Potential Inhibitors.** The newly designed thiosemicarbazones with the potential to inhibit the 1ACJ-PDB protein of AChE in the treatment of AD have been predicted by QSAR models and demonstrated by docking and MD simulations. The QSAR<sub>KPLS</sub> and PharmaRQSAR models were combined to assess the effects of substituents R<sub>1</sub>, R<sub>2</sub>, R<sub>3</sub>, and R<sub>4</sub> in each compound with an eye toward designing new compounds. The synthesis ways of newly designed thiosemicarbazones can be started from the primary compounds. We used a mixture of mainly thiosemicarbazone derivatives and a few suitable compounds such as piperidine-4-amine, di-(pyridine-2-yl)methanone, *N*-ethylcyclohexanamine, and isothiocyanic acid in EtOH, NaOH, or HCl.<sup>65</sup> The mixtures were implemented in the range of 4 h. After finishing the reaction, the solvents were evaporated under vacuum, and the solids formed were crystallized from a suitable DMF solvent<sup>65,66</sup> to form the solid thiosemicarbazones, respectively.

The spectral data of <sup>1</sup>H NMR and <sup>13</sup>C NMR in DMSO-*d*<sub>6</sub>, <sup>15</sup>N NMR in chloroform-*d* solvent, and chemical shifts expressed in single ppm taste and the results of the elemental analysis demonstrated the chemical structures of new thiosemicarbazones. The mass spectrometry (MS) data of new thiosemicarbazones were also used for identification.<sup>10</sup> The results of the spectral and elemental analysis of the compounds were consistent with the designed structures. In addition, we also determined the melting points on the thermoelectric device, which are not calibrated. The analyzed results were performed at different analysis centers. We suggest four newly designed thiosemicarbazone compounds, N1, N2, N3, and N4, which can be synthesized from the following reaction pathways.

The following reaction synthesizes the newly designed compound N1 (Scheme 1).

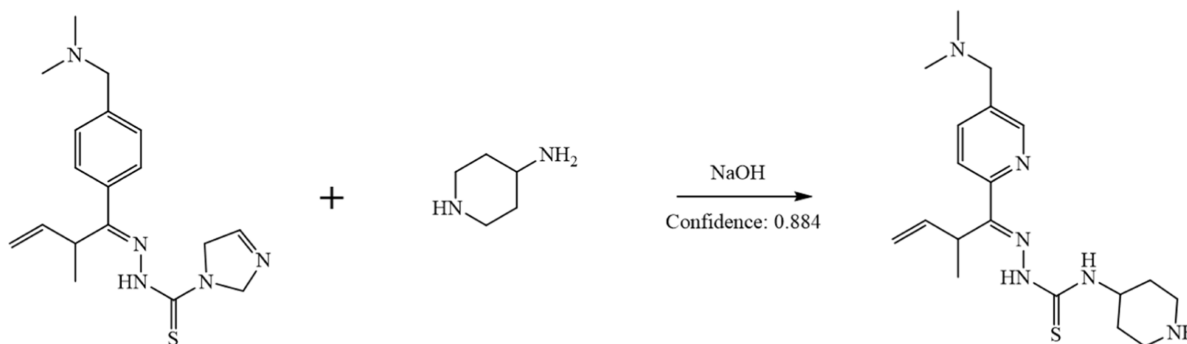
The following reaction synthesizes the newly designed compound N2 (Scheme 2).

The following reaction synthesizes the newly designed compound N3 (Scheme 3).

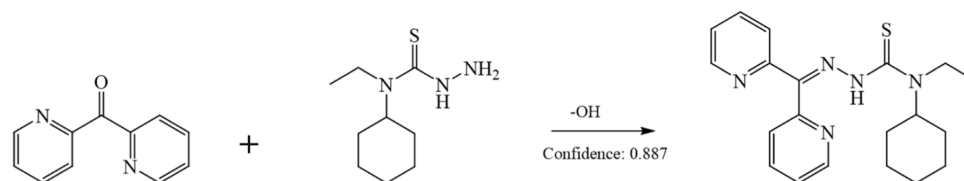
The following reaction synthesizes the newly designed compound N4 (Scheme 4).

**2.9. Quantum Calculations.** Quantum chemical calculations can be used to find out if the newly made molecules can interact with amino acids at the active receptor sites in proteins. It may be interesting to consider the mechanism of potential compounds that inhibit the 1ACJ-PDB protein acetylcholinesterase (AChE) against AD. Theoretical level DFT calculations are carried out on a Turbomole 2016 using the B3LYP method and the basis set def-SVP.<sup>67,68</sup> For optimizing the geometry of new hypothetical potential

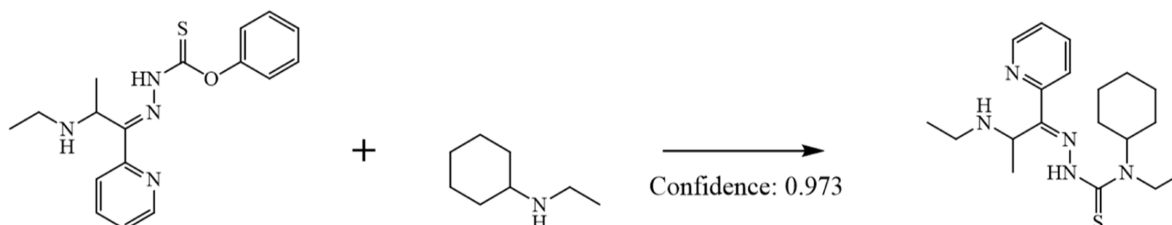
**Scheme 1.** Synthesis Diagram of the New Compound (N1)



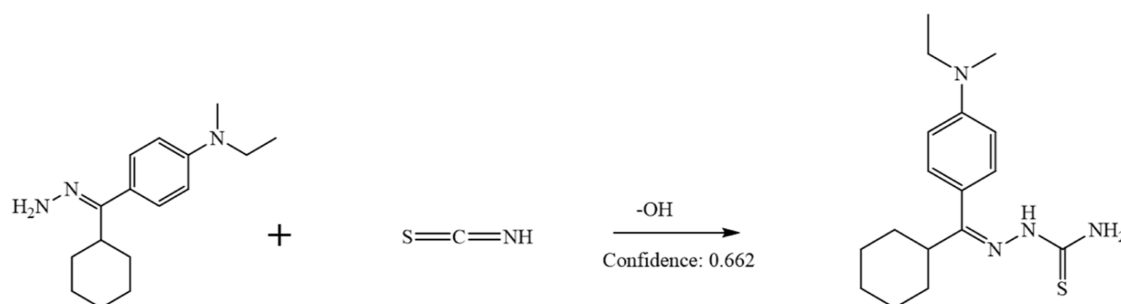
## Scheme 2. Synthesis Diagram of the New Compound (N2)



## Scheme 3. Synthesis Diagram of the New Compound (N3)



## Scheme 4. Synthesis Diagram of the New Compound (N4)



compounds, there are no symmetry constraints. Additionally, the relative effective core potential (ECP) is also included in the calculation. To assess the strength of chemical bonds in compounds, the ideal molecular energies at the B3LYP/def-SVP level<sup>67,68</sup> are employed. To examine how their intermolecular interactions could affect their capacity to generate donor–acceptor bonds, natural binding orbitals (NBOs) were conducted. Based on the energy highest occupied molecular orbital (HOMO), or  $E_{\text{HOMO}}$ , which describes a molecule's propensity to give up electrons, while the lowest unoccupied molecular orbital (LUMO), or  $E_{\text{LUMO}}$ , describes a molecule's capacity to take electrons, the electron density distribution is established.<sup>69</sup> The energy gap  $\Delta E = E_{\text{LUMO}} - E_{\text{HOMO}}$  reflects the proclivity of the organic compounds to effectively bond the receptor interaction surface.  $I = -E_{\text{HOMO}}$  and  $A = -E_{\text{LUMO}}$  are the HOMO and LUMO energies, respectively, that are used to compute the ionization potential ( $I$ ) and electron affinity ( $A$ ) of the possible inhibitors.

### 3. RESULTS AND DISCUSSION

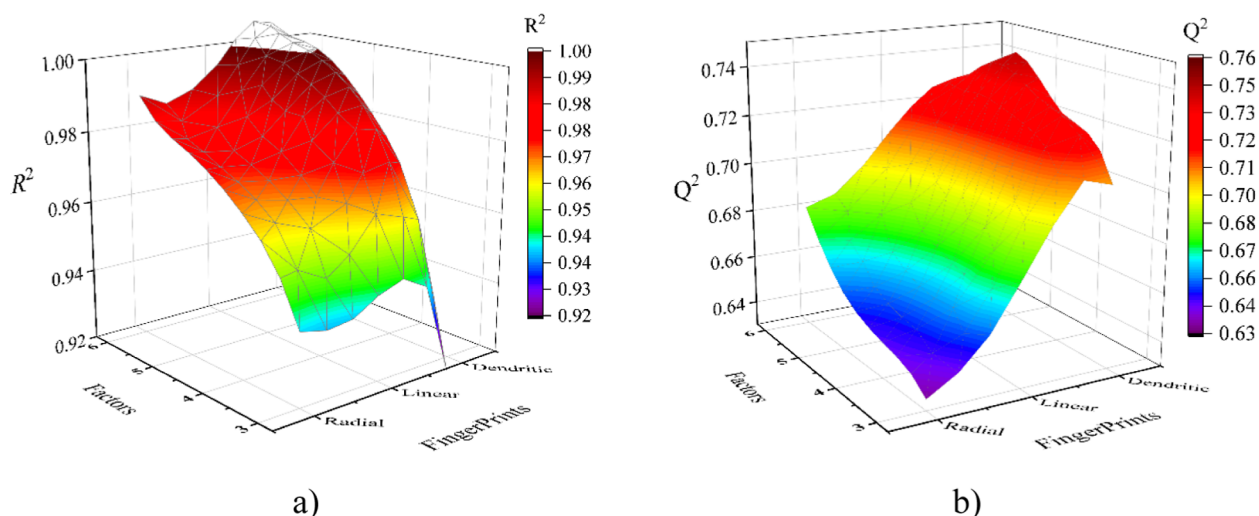
**3.1. QSAR<sub>KPLS</sub> Model.** Binary fingerprints are used to build QSAR<sub>KPLS</sub> models because it has many key advantages. The QSAR<sub>KPLS</sub> model is more favorable than other methods. The three types of fingerprints used are linear, radial, and dendritic binary. They are also used for virtual screening. In addition, latent factors in the range of 3–6 are also used to build the QSAR<sub>KPLS</sub> models. These latent factor numbers can be changed automatically to prevent the SD of the QSAR<sub>KPLS</sub> model from falling below 0.4. We can choose a nonlinear kernel level of 0.05. The results of building the QSAR<sub>KPLS</sub>

model based on the exploration and selection of the number of latent factors and fingerprint types are shown in Table 2.

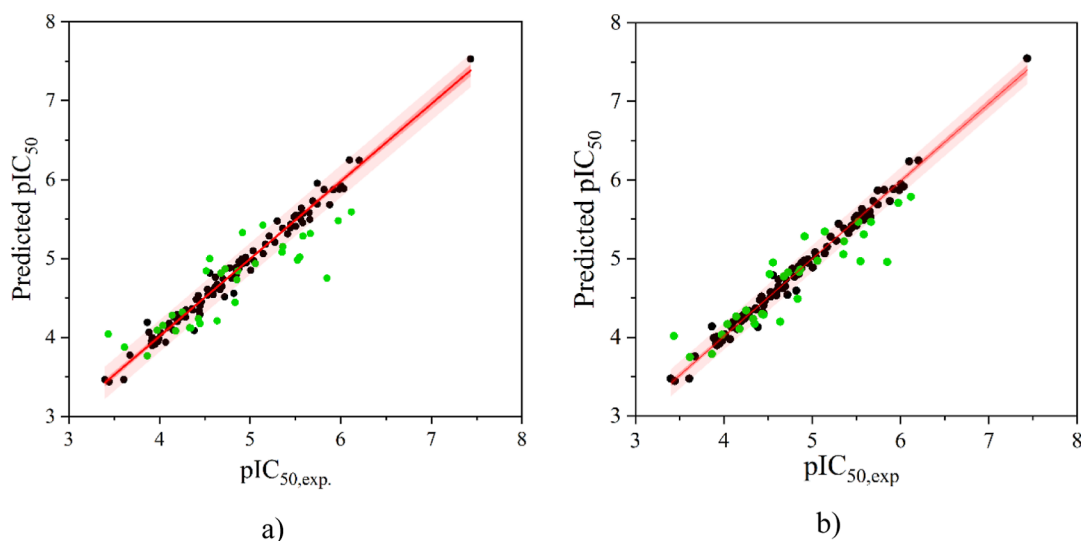
**Table 2. Comparison of Training and Prediction Ability of QSAR<sub>KPLS</sub> Models Based on Groups of Fingerprints and Latent Factors of Change**

no.	latent factor	fingerprints	$R^2_{\text{training}}$	$Q^2_{\text{test}}$
1	4	dendritic	0.961	0.720
2	3	dendritic	0.925	0.713
3	6	dendritic	0.996	0.722
4	5	dendritic	0.991	0.750
5	5	linear	0.986	0.688
6	6	linear	0.994	0.710
7	4	linear	0.971	0.675
8	3	linear	0.945	0.674
9	5	radial	0.978	0.653
10	4	radial	0.965	0.644
11	3	radial	0.942	0.636
12	6	radial	0.987	0.680

The QSAR<sub>KPLS</sub> model is built based on the kernel-based PLS regression to select attributes on the structural data set and apply this model to other data sets. The QSAR<sub>KPLS</sub> model is available with the fingerprint as the independent variable  $x$ . The binary fingerprints in the QSAR<sub>KPLS</sub> model are built as a linear combination of the input  $x$  variables by correlating them with the  $y$  variable. This method is helpful for large numbers of variables  $x$ , with both under- and overdetermined systems. Table 2 shows the statistical values  $R^2$  and  $Q^2$  of the training and test set for the QSAR<sub>KPLS</sub> models built using the three



**Figure 3.** Optimal surfaces  $R^2$  and  $Q^2$  depend on latent factors and binary fingerprint groups: (a) training set and (b) test sets.



**Figure 4.** Correlation between experimental and predicted  $pIC_{50}$  activity from  $QSAR_{KPLS}$  using a latent factor of 5 and independent variables: (a) DF descriptor and (b) mixed DF and PC descriptors; symbols: (●) training set; (green ●) test set; (orange ■) 95% training set; and (peach ■) 95% prediction.

best-performing fingerprint types and latent factors, respectively, as shown in Figure 3.

We can see that the dendritic binary fingerprint and latent factor value of 5 gave optimal  $R^2$  and  $Q^2$  values calculated by eqs 5 and 6. In addition, Table 2 and Figure 3 show that the DF was better than the others, on average, for training and test sets. Thus, we can choose the optimal latent factor in the predictive results of the  $QSAR_{KPLS}$  model.

Values  $R^2$  and  $Q^2$  are used to evaluate the QSAR models for the fitting training set and the test set. The values  $R^2$  and  $Q^2$ <sup>44,48,59</sup> are calculated by the following formula.

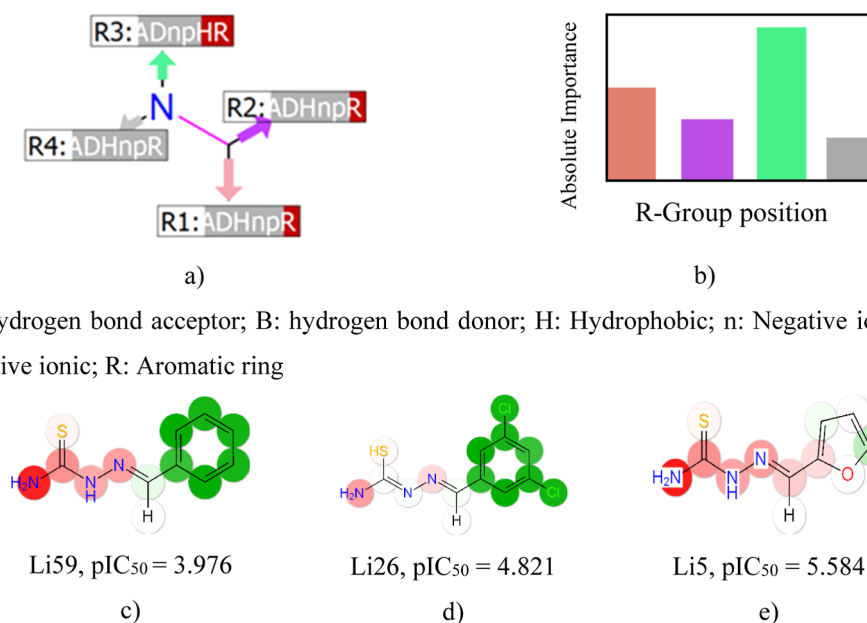
$$R^2 = 1 - \frac{\sum_i^{\text{training}} (y_i - \hat{y}_i)^2}{\sum_i^{\text{training}} (y_i - \bar{y}_{\text{training}})^2} \quad (5)$$

$$Q^2 = 1 - \frac{\sum_i^{\text{test}} (y_i - \hat{y}_i)^2}{\sum_i^{\text{test}} (y_i - \bar{y}_{\text{test}})^2} \quad (6)$$

Here,  $\bar{y}_{\text{training}}$  and  $\bar{y}_{\text{test}}$  are the average observed activities of the training and test sets, respectively. Moreover,  $y_i$  and  $\hat{y}_i$  are the experimental and calculated activities from the QSAR models.

Interestingly, the advantage of DFs in diverse applications has been demonstrated in many recent studies. Many scientists have supported this by using fingerprints to build QSAR models that have been published in various journals.<sup>16–21</sup> We can also see that this fingerprint is very flexible and powerful for structural characterization. For this reason, DFs can be used in all subsequent work here, as exhibited in Figure 4.

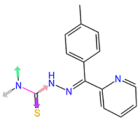
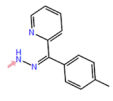



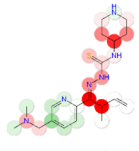
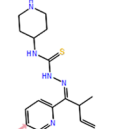



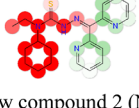
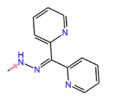


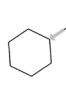
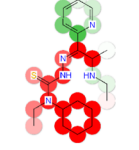
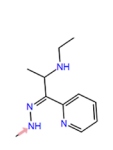


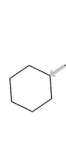
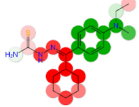
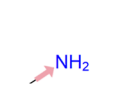

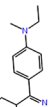
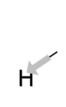
The QSAR model can easily infer structure–activity relationships (SARs) by carefully analyzing well-designed compounds. A certain complexity and subtlety of this relationship can be challenging to unravel without a reliable model. Furthermore, the QSAR model can generally help visualize the advantages and disadvantages of a molecular structure. This could help bridge the gap between simply figuring out what properties are essential for bioactivity. From this, it is possible to identify and design new compounds that



A: hydrogen bond acceptor; B: hydrogen bond donor; H: Hydrophobic; n: Negative ionic; p: Positive ionic; R: Aromatic ring

**Figure 5.** The significant effects of R substituents for inhibition activity 1ACJ-PDB of AChE are visualized using the PharmaQSAR and QSAR<sub>KPLS</sub> models with the molecular descriptors DF: (a) effects of R substituents R<sub>1</sub>, R<sub>2</sub>, R<sub>3</sub>, and R<sub>4</sub>; (b) absolute importance R substituents; (c) weak activity; (d) moderate activity; and (e) vigorous activity.

**Table 3. Predicted Results of pIC<sub>50</sub> activity of Newly Designed Compounds N1, N2, N3, and N4 Resulting from the QSAR<sub>KPLS</sub> Model Using the DF and PC Descriptors and a Mixture of DF and PC**

Compound	R substituent groups				predicted values pIC <sub>50</sub>		
	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	R <sub>4</sub>	DF	PC	DF and PC
 Lead compound Li39					5.166	5.178	5.153
 New compound 1 (N1)					5.254	4.680	5.266
 New compound 2 (N2)					5.691	5.623	5.723
 New compound 3 (N3)					6.067	5.615	6.045
 New compound 4 (N4)					6.091	5.271	6.121

can be synthesized or obtained as part of a potential optimal compound.

The QSAR<sub>KPLS</sub> model can help us visualize the significant effects of structural factors on bioactivity. The QSAR<sub>KPLS</sub>



**Table 4. Validation Results of the QSAR<sub>ANN</sub> Models I(m)-HL(n)-O(k) Using the DF and PC Descriptors and a Mixture of These Descriptors**

QSAR <sub>ANN</sub> model I(m)-HL(n)-O(k)	R <sup>2</sup> <sub>training</sub>	R <sup>2</sup> <sub>test</sub>	Q <sup>2</sup> <sub>valid</sub>	training error	test error	validation error	training algorithm
			DFs				
I(200)-HL(15)-O(2)	0.961	0.832	0.809	0.069	0.306	0.507	BFGS
			PC Descriptors				
I(13)-HL(6)-O(2)	0.956	0.937	0.929	0.095	0.185	0.260	BFGS
			DF and PC Descriptors				
I(213)-HL(12)-O(2)	0.988	0.907	0.919	0.024	0.262	0.259	BFGS

model was built from DFs and has shown that the red position indicates vigorous activity, and the green position describes low activity when changing the R substituents, as illustrated in Figure 5a. We assessed the effect of a substituent using the PharmaRQSAR technique,<sup>20,58</sup> as depicted in Figure 5. The R group analysis defines the substituents attached to a core and allows the properties of the R substituent to be displayed. Cores can be specified using the smarts pattern or defined by the maximum common substructure obtained from a PharmaRQSAR run.

Based on the results of exploring the influence of R substituents using the PharmaRQSAR models, it is easy to orient the design of new substances, which can envision the design of new compounds with more vigorous activity. We chose the ligand Li39 in Table S1 as the lead compound to experimentally design four new compounds, N1, N2, N3, and N4, and were able to synthesize them, as shown in Table 3. The significant influence of the R substituents at the four respective positions is depicted in Figure 5b.

Furthermore, Figure 5c,d,e shows the QSAR<sub>KPLS</sub> model using the mixture of DF and PC descriptors that allow complete visualization of the atomic effect. The QSAR<sub>KPLS</sub> model built from DFs quickly detects the changes in compounds with weak, medium, and vigorous activities.

The atomic sites that increase the predicted activity are colored red, and the atoms that decrease the activity are shown in green. The color intensity already reflects the magnitude of the effect. Of particular note is the magnitude representing the activity signals of the atomic effect at a particular location in the chemical structure skeleton. The fingerprint results indicate that the molecular fragments consist of atoms from a fixed structural skeleton and atoms from the change region. Changing a substituent can add and remove a part from fixed or change regions. It can also change from weak to moderate and vigorous activity.

The structural effect is achieved in a good direction, resulting in an atom of the molecular structural skeleton bound to such sites that can turn from green to red, as shown in Figure 4 and Table 3. We can see the molecular skeleton when there is an impact that induces a conjugation, and such molecular modeling suggests that parts of the molecular skeleton may be unrelated or unfavorable except when the correct substituent is attached.

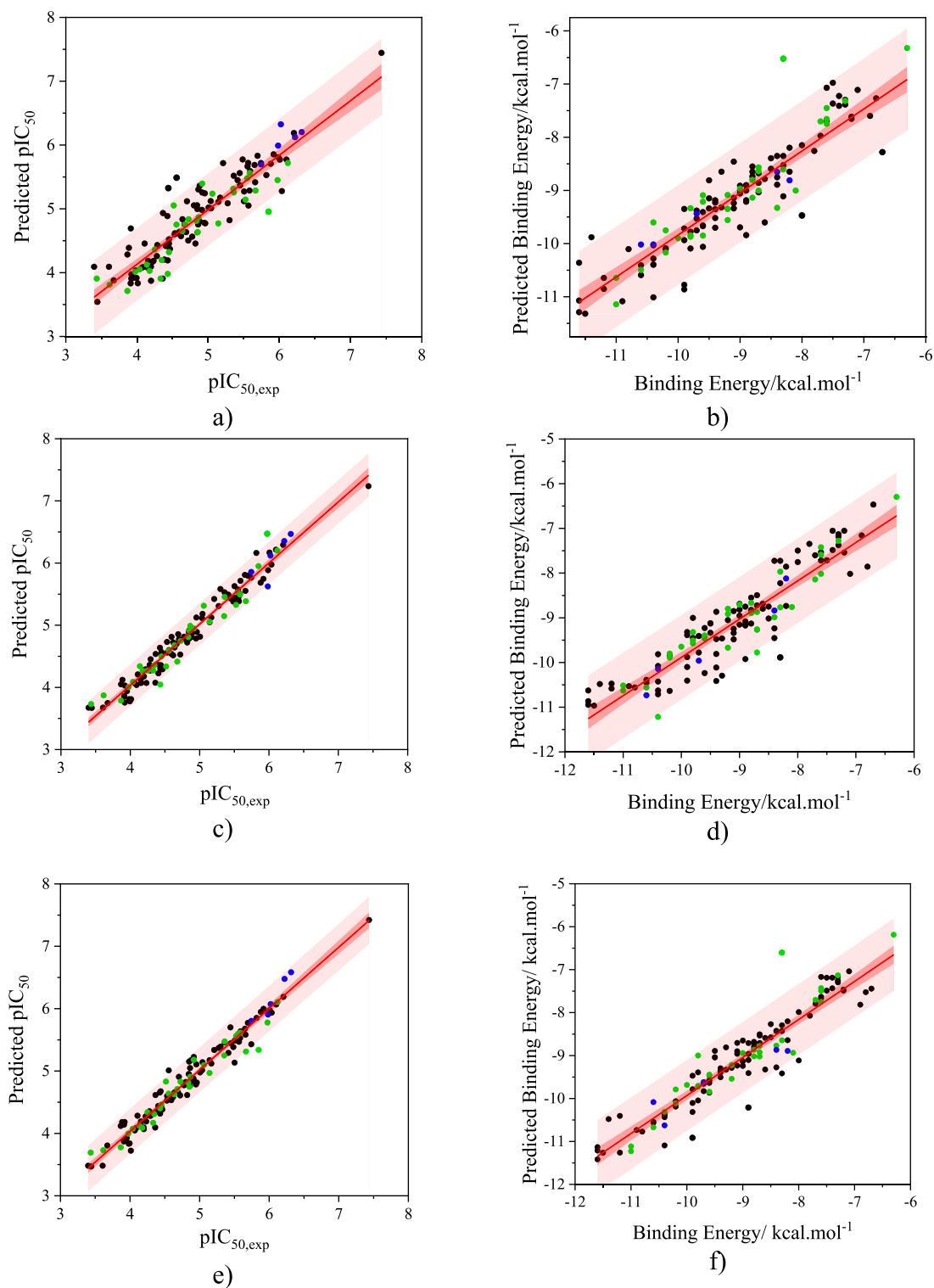
We can exploit these transformations to synthesize or acquire new compounds that change from one sequence to the next. However, the general procedure is to identify favorable features of moderate and vigorous bioactivity of bound R substituents and transformations. This can infer and lead to favorable combinations of properties. Of course, there is nothing new in such an approach, but it is made much easier by QSAR<sub>KPLS</sub> models. Thus, we can get the influence of the R substituent on the part of the structural skeleton. We see that

four major substituents can be attached to the molecular skeleton with the effect depicted by colored arrows such as green, gray, purple, and red, as shown in Figure 4 and Table 3. The R substituent bonding to each site is depicted by color corresponding to the biological effect upon attaching to the molecular skeleton.

The molecular structure of the lead compound Li39 is a relatively active substance, containing R<sub>1</sub>, R<sub>2</sub>, R<sub>3</sub>, and R<sub>4</sub> substituents bonded to thiosemicarbazone skeleton sites with different color effects, as described in Table 3. Aromatic ring substituents are important groups R<sub>1</sub> and R<sub>3</sub>, which significantly increase the activity of pIC<sub>50</sub>. New compounds can be designed toward this change of R<sub>1</sub> and R<sub>3</sub> substituents. The pIC<sub>50</sub> activities of the four newly designed compounds N1, N2, N3, and N4 were predicted by the QSAR<sub>KPLS</sub> model using DF, PC, and mixed DF and PC descriptors. The QSAR<sub>KPLS</sub> model successfully demonstrated that the activity of the novel substances changed appropriately according to the rules shown in Table 3.

**3.2. QSAR<sub>ANN</sub> Model.** The results of building the neural network architecture I(m)-HL(n)-O(k) are obtained by using different input neurons. 200 dendritic binary fingerprint descriptors are used as neurons on the input layer I(m); and biologically active pIC<sub>50</sub> and BE, kcal·mol<sup>-1</sup> are used as neurons on the output layer with k = 2; the optimal number of neurons on the hidden layer with n = 15 was determined by the genetic algorithm. In this case, the neural network architecture I(200)-HL(15)-O(2) is defined. Similarly, 13 PC descriptors are used as neurons on the input layer I(m), and the number of hidden neurons n = 6 is optimized, so the neural network architecture is I(13)-HL(6)-O(2). Using mixed descriptors, including DF and PC descriptors as input neurons on the input layer I(m), the optimal number of hidden neurons, in this case, is n = 12, so the neural network architecture is I(213)-HL(12)-O(2).

The training and test sets given in Table S1 were used to build and validate the QSAR<sub>ANN</sub> models. The applicability of these neural network architectures is cross-validated by training according to the learning parameters, and the training algorithm is fixed, as given in Section 2.4. In the training process of these neural networks, we found that higher learning rates can converge faster but exhibit greater instability. A learning rate of 0.7 or less is reasonable. Higher learning rates can cause weight differences. The learning rate was selected using a gradient descent algorithm. Momentum 0.7 is also an important parameter used to compensate for slow convergence, and if the weighted adjustment is consistent in one direction, "speed up". Momentum usually dramatically increases the convergence rate, and higher rates can reduce the learning rate to increase the stability without much loss in the convergence rate. The validation results of these network



**Figure 6.** Correlation between experimental and predicted  $pIC_{50}$  and BE values from the  $QSAR_{ANN}$  models for the training and test set using input neurons: (a,b) DF descriptor; (c,d) PC descriptor; and (e,f) mixed descriptors of the DF and PC; symbols: (●) training set; (green ●) test set; (blue ●) new compound; (orange ■) 95% training set; and (peach ■) 95% prediction.

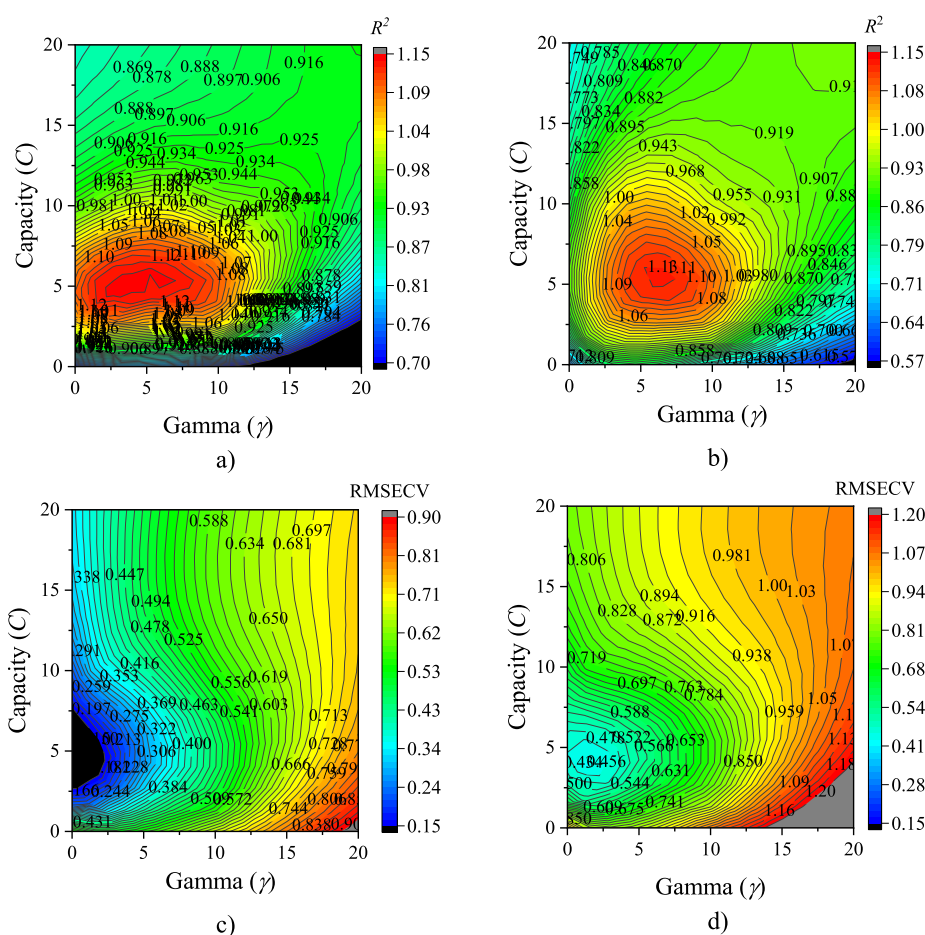
architectures with different input and hidden neurons are given in Table 4.

The  $QSAR_{ANN}$   $I(m)$ - $HL(n)$ - $O(k)$  models were selected based on the standard least-squares technique to confirm the responsiveness of the data in the case of random effects. Therefore, the test results of available observations for each case are efficient and valid. The predicted quality of the

$QSAR_{ANN}$  models can be compared with those from the experimental data in Table S1 and the  $QSAR_{KPLS}$  models using the DF, PC, and mixed descriptors of the DF and PC, respectively, as illustrated in Table S2. Scatter correlation plots for the  $QSAR_{ANN}$  models correspond to the molecular descriptors, as shown in Figure 6.

**Table 5. Optimal Parameters Gamma ( $\gamma$ ) and Capacity ( $C$ ) of the QSAR<sub>SVR</sub> Models Were Selected for Predicting the pIC<sub>50</sub> Values and BE, kcal mol<sup>-1</sup>, by Using the Grid Surface Technique**

		pIC <sub>50</sub> , mM					BE, kcal mol <sup>-1</sup>						
R <sup>2</sup>	C	0.01	0.1	1	10	100	R <sup>2</sup>	C	0.01	0.1	1	10	100
gamma		1	2	3	4	5	gamma		1	2	3	4	5
0.01	1	0.641	0.735	0.893	<b>0.945</b>	0.960	0.01	1	0.567	0.611	0.676	0.773	0.827
0.1	2	0.611	0.869	0.945	0.965	0.964	0.1	2	0.684	0.726	0.809	0.869	0.897
1	3	0.632	0.825	0.964	0.966	0.966	1	3	0.649	0.790	<b>0.908</b>	0.938	0.953
10	4	0.580	0.766	0.976	0.975	0.975	10	4	0.622	0.741	0.964	0.965	0.965
100	5	0.696	0.788	0.979	0.978	0.978	100	5	0.730	0.792	0.970	0.970	0.970
RMSECV	C	0.01	0.1	1	10	100	RMSECV	C	0.01	0.1	1	10	100
gamma		1	2	3	4	5	gamma		1	2	3	4	5
0.01	1	0.727	0.590	0.296	<b>0.204</b>	0.186	0.01	1	1.168	1.008	0.726	0.623	0.610
0.1	2	0.637	0.345	0.231	0.221	0.203	0.1	2	1.033	0.690	0.615	0.618	0.742
1	3	0.615	0.400	0.323	0.329	0.327	1	3	1.002	0.701	<b>0.606</b>	0.740	0.807
10	4	0.733	0.624	0.546	0.554	0.554	10	4	1.168	1.015	0.846	0.873	0.849
100	5	0.747	0.724	0.687	0.708	0.683	100	5	1.182	1.158	1.052	1.088	1.048

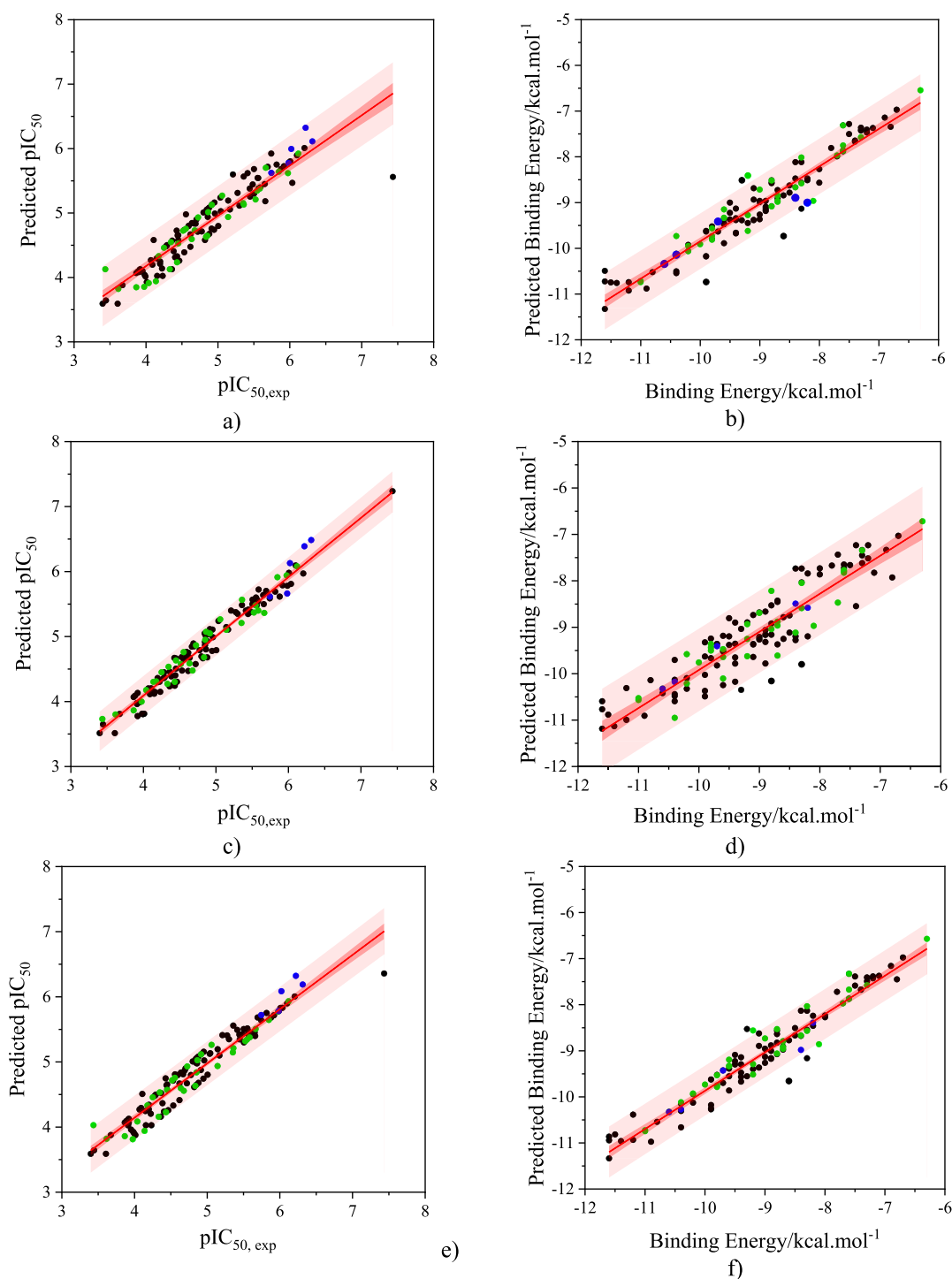


**Figure 7.** Contour plots for searching optimal parameters gamma,  $\gamma$ , and capacity,  $C$ ; the optimal region of the  $R^2$  values: (a) pIC<sub>50</sub> activity and (b) BE; the optimal region of the RMSECV values: (c) pIC<sub>50</sub> activity and (d) BE.

The QSAR<sub>ANN</sub> models seem to give better predicted results than the QSAR<sub>KPLS</sub> models. Significantly, using the DFs combined with the PC descriptors yield more accurate predicted results, as shown in Tables 4 and S2. Furthermore, the  $R^2_{\text{training}}$  values range from 0.956 to 0.988; the  $R^2_{\text{test}}$  values range from 0.832 to 0.937; the  $Q^2_{\text{valid}}$  values range from 0.809 to 0.929. All these statistical values demonstrate that the QSAR<sub>ANN</sub> models have accurate predictive quality. The predicted results from the QSAR<sub>KPLS</sub> and QSAR<sub>ANN</sub> models

built with the mixed descriptors of DF and PC descriptors have proven to be better. Most importantly, the  $R^2_{\text{test}}$  and  $Q^2_{\text{valid}}$  values exceed 0.8.

Thus, we can see that the QSAR<sub>ANN</sub> models using a combination of DF and PC descriptors predict the average accuracy between the two approaches. The prediction results of these QSAR<sub>ANN</sub> models are much closer to the experimental value than those using only DFs. In this case, the RMSE values obtained for pIC<sub>50</sub> and BE are more strongly observed for the



**Figure 8.** Correlation between experimental and predicted  $pIC_{50}$  values from the  $QSAR_{SVR}$  models. (a,b) DF descriptor; (c,d) PC descriptor; and (e,f) mixed descriptors of the DF and PC; symbols: (●) training set; (green ●) test set; (blue ●) new compound; (orange ■) 95% training set; and (peach ■) 95% prediction.

compounds in the test set and the newly designed substances, as described in Tables 4 and S2. However, the results can be obtained using different  $QSAR_{ANN}$  models. The  $QSAR_{ANN}$  model that combines the DF and PC molecular descriptors performs better than the fingerprint-only model. Indeed, the predicted results are surprising for the test and newly designed compounds. The combined models yield a significantly higher  $Q^2$  than the other two models.

**3.3.  $QSAR_{SVR}$  Model.** The  $QSAR_{SVR}$  model is built based on the optimal choice of kernel function types, the kernel parameter capacity ( $C$ ), and gamma ( $\gamma$ ). The multiplication functions include linear function, polynomial function, radial function RBF, and sigmoid function. By far, the kernel function-type RBF is the most common choice used in support vector machine regression. This is mainly based on their finite and localized responses over the entire range of the real  $x$ -axis. The RMSEC (mean squared error from calibration)

Table 6. Results of the Drug-Likeness Assessment of Newly Designed Compounds

ligands	Lipinski rule				Veber rule		synthetic accessibility		drug-likeness
	log <i>P</i>	HBD	HBA	MW	nRB	TPSA	score (easy to difficult)		
	≤5	≤5	≤10	≤500	≤10	≤140	1 ≤ score ≤10		
Li39	2.811	1	4	298.13	5	72.08	2.90		yes
N1	1.962	3	6	374.23	9	96.14	4.06		yes
N2	3.586	1	5	367.18	7	84.44	3.39		yes
N3	4.289	2	5	361.23	9	84.11	4.23		yes
N4	4.098	2	4	318.19	6	85.74	3.01		yes

and RMSECV (from cross-validation) values were used to confirm the predictive quality of the QSAR<sub>SVR</sub> model relative to the reference. This process is performed so that the optimal parameter capacity (*C*) and gamma ( $\gamma$ ) can be found, as given in Table 5 and Figure 7. The difference between calibration and cross-validation is expressed as equal to the RMSEC and RMSECV values. For QSAR<sub>SVR</sub> models in general, RMSECV should preferably be close to RMSEC, which indicates that the QSAR<sub>SVR</sub> model is not overfitted. This is calculated by

$$\text{RMSECV} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (7)$$

Here,  $y$  and  $\hat{y}_i$  are the activity values calculated from the QSAR models.

We can utilize the optimal parameter of gamma ( $\gamma$ ) in the range 0.01–0.10 for pIC<sub>50</sub> activity and gamma of 0.005–0.010 for BE, kcal mol<sup>-1</sup>, using input variables DF, PC, and combined descriptors of DFs and PC. The correlation coefficients  $R^2$  selected must be the largest, and RMSECV values were the smallest, respectively, as shown in Table 5. The QSAR<sub>SVR</sub> model was constructed using the DFs, the number of support vectors is 78 for pIC<sub>50</sub> bioactivity, and the number of support vectors is 90 for BE. The correlation coefficients  $R^2_{\text{training}}$  are 0.940 and 0.958 for pIC<sub>50</sub> activity and BE, respectively. Using the PC descriptors for constructing the QSAR<sub>SVR</sub> model, the number of support vectors for pIC<sub>50</sub> bioactivity and BE correspond to 28 and 82, respectively. The correlation coefficient  $R^2_{\text{training}}$  was equal to 0.972 and 0.953 for pIC<sub>50</sub> bioactivity and BE, respectively. We also used the combined descriptors of DFs and PC for constructing the QSAR<sub>SVR</sub> model. Here, the number of support vectors for pIC<sub>50</sub> activity and BE correspond to 70 and 85, respectively. The correlation coefficients  $R^2_{\text{training}}$  were 0.940 and 0.937 for pIC<sub>50</sub> and BE, respectively. All results and predictive errors of pIC<sub>50</sub> and BE, kcal mol<sup>-1</sup>, for ligands in the test set in Table S1 are determined using the QSAR<sub>SVR</sub> models, as given in Tables S2 and S3. We have successfully built the QSAR<sub>KPLS</sub>, QSAR<sub>ANN</sub>, and QSAR<sub>SVR</sub> models. All QSAR models have demonstrated significant advantages in supporting the design of new compounds and predicting the pIC<sub>50</sub> bioactivity and BE. The QSAR models did well on training and test sets resulting from the screening process of the ChEMBL Database, as illustrated in Table S1. Thus, these constructed QSAR models also perform possibly well for diverse and complex data sets. Indeed, a diverse range of thiosemicarbazones might be well suited for these QSAR models.

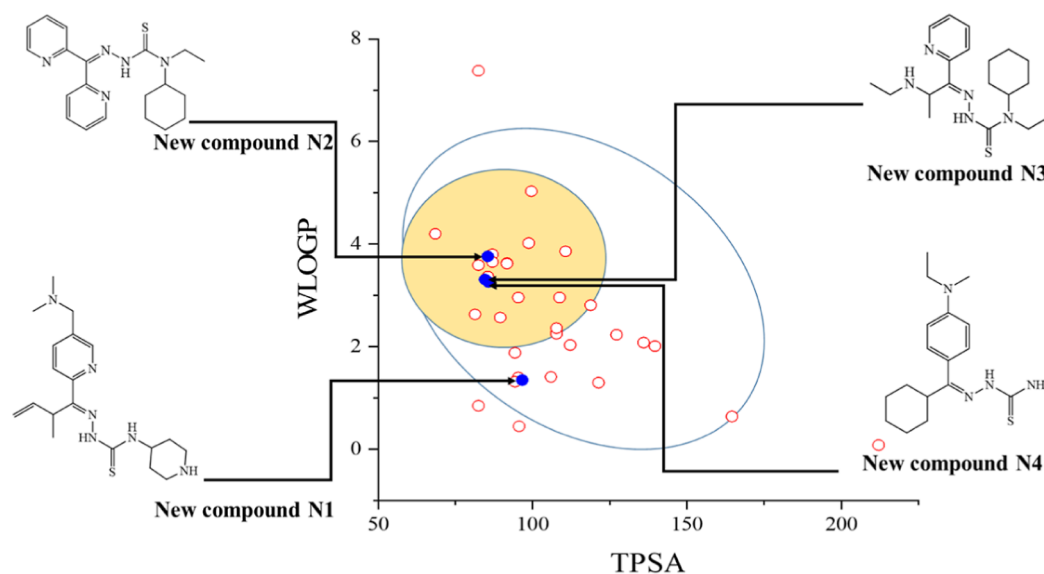
This study performed rigorous testing of hypotheses regarding the QSAR model and thiosemicarbazone molecular structures. The QSAR models built here reconciled the overall difference in chemical structure and the ability of different bonding types of molecular structures. Therefore, QSAR

models can have any implications that are still unclear and can also be elucidated. The QSAR models are well predicted across the relevant test sets; those QSAR models can likely reliably tolerate a wide variety of compounds. This gives the critical value that the statistics show. The fingerprint-based QSAR<sub>KPLS</sub> model can effectively support molecular design on diverse data sets. This can be accomplished by combining training sets and building a ligand–receptor binding model that includes all active targets of the 1ACJ-PDB protein AChE.

The results of building QSAR models to predict pIC<sub>50</sub> biological activity and BE have shown the quality of QSAR models, as shown in Figure 8. The error of the predictions for the test sets was received to be positive, as shown in Tables S2 and S3. To test the inhibitory activity of the newly designed thiosemicarbazones, we performed further docking calculation stages to validate and predict the binding ability of the newly designed compounds on the active target site of the 1ACJ-PDB receptor. Here, the machine learning models QSAR<sub>ANN</sub> and QSAR<sub>SVR</sub> using combined descriptors of DF and PC with a latent factor of 5 predict pIC<sub>50</sub> bioactivity and BE to the target site on 1ACJ-PDB. Before performing the docking calculation for the newly designed compounds in Table 3, we tested the drug-likeness properties of four newly designed compounds, N1, N2, N3, and N4, by a virtual screening using the Lipinski-5, Ghose, Veber, Egan, and Muegge rules.

In addition, we can also see in Figures 4, 6, and 8 that some anomalies appear. They are outside the prediction and training regions at the 95% confidence level. Some of these scores represent too small a percentage of the total sample in the training and control groups. This also does not significantly affect the predictive power of QSAR models related to new compounds, as shown in Figure 12. The prediction results for new compounds are all within the prediction domain at the 95% confidence level (dark pink region). Moreover, if the training process is prolonged, it can also get results for most compounds in the 95% confidence region. The obtained QSAR<sub>ANN</sub> and QSAR<sub>SVR</sub> models produce more distributed prediction results and, in particular, predict new compounds, which can cause inaccuracy if the training process is prolonged. We may find that our prediction results are thus optimal. Thus, these QSAR models are not overfit. The obtained QSAR models are optimally capable of reliably predicting new compounds.

These rules aid in assessing whether new substances have the most drug-like activity. Compounds that respond to drug-like activity should not violate the Lipinski-5 rule for molecular mass, HBD, HBA, and log *P* coefficient. Furthermore, satisfying the Ghose, Veber, Egan, and Muegge rules can also support a more thorough and comprehensive evaluation of drug-like features based on additional properties such as MR, surface area (SA), PSA, or TPSA.<sup>61,62</sup> Compounds that satisfy the above rules may also exhibit pharmacokinetic properties.



**Figure 9.** Validation of pharmacokinetic properties using the model BOILED-Egg; symbols: (orange ○) 36 compounds in the test set and (blue ●) newly potential compounds N1, N2, N3, and N4.

This pharmacokinetic property may support the assessment of the ability of new substances to be absorbed from the gastrointestinal tract and penetrate the blood–brain barrier, as given in Table 6.

The biological activity of a drug substance is related to its pharmacokinetics and bioavailability properties. Drugs requiring gastrointestinal absorption and crossing the blood–brain barrier are two important pharmacokinetic criteria for drug development. Therefore, using the Brain Or Intestinal EstimateD (BOILED-Egg) permeation model is also proposed as an accurate prediction method based on the polarity of the molecule.<sup>63</sup>

The FEED-Egg model can filter and validate potential novel compounds N1, N2, N3, and N4, as depicted in Figure 9. Compounds in the BOILED-Egg yellow region are predicted to permeate the passive blood–brain barrier. Compounds in the BOILED-Egg white region are predicted to be passively absorbed from the gastrointestinal tract.

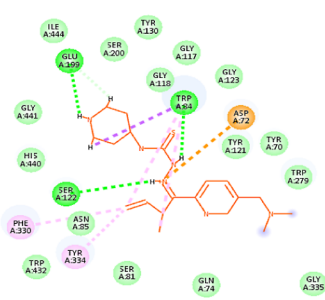
Furthermore, the CYP450, CYP1A2, CYP3A4, CYP2C9, CYP2C19, and CYP2D6 enzyme systems belong to the group of monooxygenases found in most living organisms. This enzyme system plays an important role in human physiology. Newly designed compounds have appropriate inhibition of the CYP1A2, CYP2C19, CYP2C9, and CYP3A4 enzyme systems, and the activities of these enzymes vary widely among individuals; although studies have shown a high heritability, the underlying genetic factors are still unknown. We found novel transpositional associations between regulatory genes and the CYP1A2, CYP2C19, and CYP2C9 genes and CYP3A4 expression and function. The drug-like feature is also presented in pharmacokinetics to evaluate the individual ADME behaviors of molecules.<sup>62</sup>

The purpose of this synthetic accessibility prediction is to assist pharmaceutical chemists in their efforts to discover drug synthesis. It is possible to chemically describe the fragments to be considered in a given molecule. This may assist in developing a necessary warning that the structure is chemically reactive, metabolically unstable, or has poor pharmacokinetic properties. We can apply these and other PC filters to design compounds that meet pharmacochemical synthesis criteria.

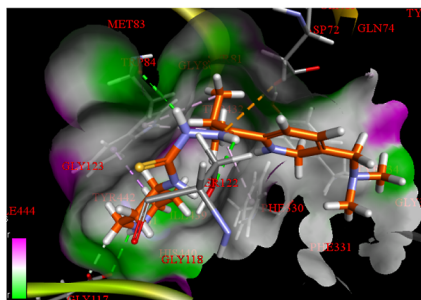
Table 6 shows that the synthesis accessibility of the new compounds is less than 5. This suggests that these newly designed compounds can also be chemically synthesized easily.

**3.4. Molecular Docking Calculation.** Simulation execution processes are carried out in primary stages such as protein preparation, finding active sites on proteins by genetic algorithm, and preparing optimal structures of ligands, details described in Section 2.6. The ligands are converted from mol format to PDBQT and perform the receptor hydrogenation and charge calculation. Finally, Autodock performs molecular binding activities on the 1ACJ-PDB AChE receptor.<sup>53,54</sup> Among them, we can visualize the results of molecular connectivity, as exhibited in Figure 10. The protein receptor active site can be represented by a sphere as shown in Figure 2, consisting of residues SER122, TRP84, GLU199, GLY118, ASP72, PHE330, TRP334, TRP432, HIS440, GLY80, and GLY441, as shown in Figure 10. The active site prediction was significantly based on a genetic algorithm with amino acids TRP84, GLY118, GLU199, PHE330, TRP432, HIS440, and GLY441.

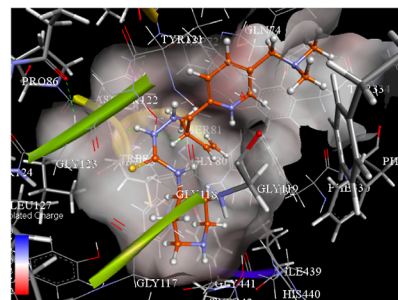
The BE is calculated from a set of points located in a cavity. This definition allows specifying the size and shape of the active target location. The newly designed compounds N1, N2, N3, and N4 were screened based on their ability to bind to the active target site at coordinates  $x = 3214$ ,  $y = 65,854$ , and  $z = 65,167$ . Among these, the amino acids of the 1ACJ-PDB receptor were identified using the Autodock virtual screening technique.<sup>55</sup> The ligand–protein interactions are depicted in 2D and 3D, and as shown, the estimated binding energies range from  $-9.90$  to  $-7.50$  kcal mol<sup>-1</sup>, as shown in Table 7. The ligands with the lowest binding energies include compounds N2 and N4. This indicates that these molecules have a high binding affinity for the target protein. Appropriate ligand-binding algorithms can be used here to screen for incompatible ligands and rapidly generate conjugates based on the conformation of the ligands. The ligand interacts with the indole of TRP84, and its *m*-hydroxyl exhibits electronic hydrogen bonding with SER200 and HIS440. The unique structure of this complex lies in the orientation of the phenyl ring of PHE330. The residues TRP84, GLY118, GLU199,

New compound-1 (N1), Binding Energy:  $-7.50 \text{ kcal.mol}^{-1}$ 

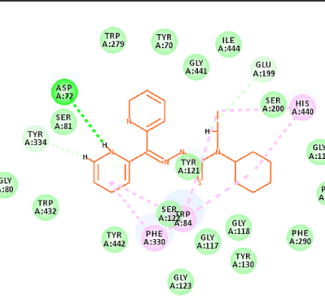
Interaction 2D



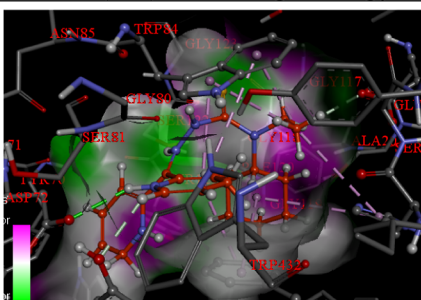
H-bond interaction



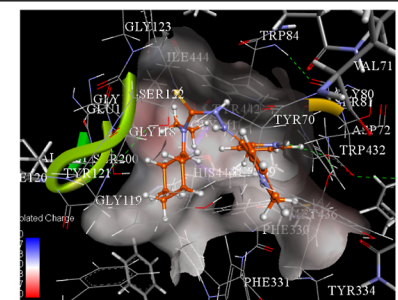
Charge interaction

New compound-2 (N2), Binding Energy:  $-9.40 \text{ kcal.mol}^{-1}$ 

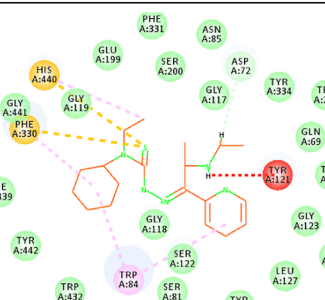
Interaction 2D



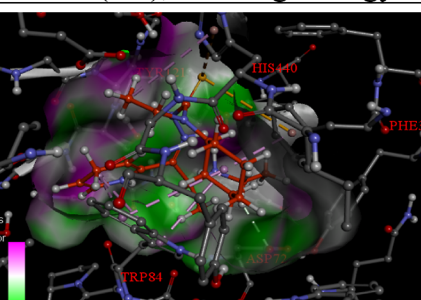
H-bond interaction



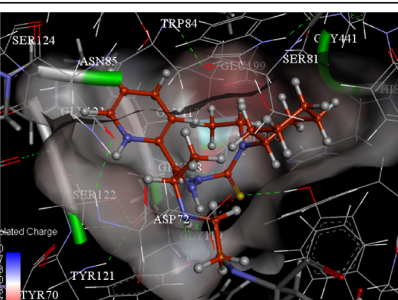
Charge interaction

New compound-3 (N3), Binding Energy:  $-8.30 \text{ kcal.mol}^{-1}$ 

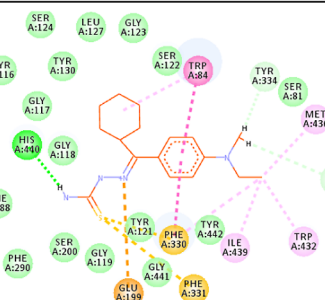
Interaction 2D



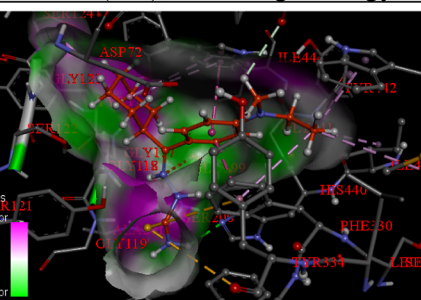
H-bond interaction



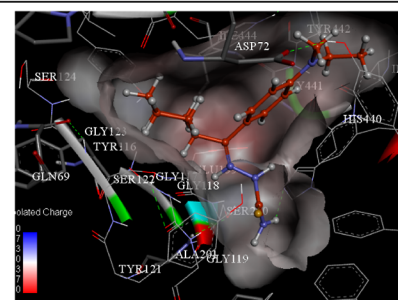
Charge interaction

New compound-4 (N4), Binding Energy:  $-9.70 \text{ kcal.mol}^{-1}$ 

Interaction 2D



H-bond interaction



Charge interaction

(a)

(b)

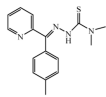
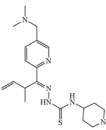
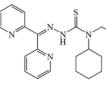
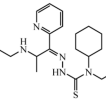
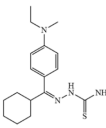
(c)

**Figure 10.** (a) 2D interactions and (b,c) 3D interactions between newly designed compounds N1, N2, N3, and N4 and the receptor of 1ACJ-PDB AChE. The green and violet regions are for hydrogen acceptor and donor; the red and gray regions are charge regions.

PHE330, TRP432, HIS440, and GLY441 are located at the active target site, and the shape of this site determines the hydrogen bonding and charge. In addition, we can see that the

substituents  $R_1$ ,  $R_2$ ,  $R_3$ , and  $R_4$  attached to the molecular skeleton have been analyzed for their important influence on biological activity.

Table 7. Ligands in the Pose from the Docking Calculation Are Analyzed for Amino Acid–Ligand Interactions

Ligands	Binding Energy (kcal.mol <sup>-1</sup> )	Amino acid–Ligand interactions	RMSD-L (Å)	Bond Length (Å)
 Lead structure Li39	-9.90	H-bond interactions receptor-ligand Li39	1.11	
		GLY117:O-H36 ligand Li39		2.99
		TYR70:O-H39 ligand Li39		3.07
		ASP72:OD2-H41 ligand Li39		2.76
 New compound N1	-7.50	H-bond interactions receptor-ligand N1	1.09	
		SER122:OG-HN ligand N1		2.71
		TRP84:O-H ligand N1		2.24
		GLU199:OE2-H ligand N1		2.63
		GLU199:OE1-H28 ligand N1		2.09
 New compound N2	-9.40	H-bond interactions receptor-ligand N2	1.18	
		ASP72:OD2-HN ligand N2		2.49
		GLU199:OE1-H5 ligand N2		3.01
		TYR334:OH-H21 ligand N2		2.59
 New compound N3	-8.30	H-bond interactions receptor-ligand N2	1.26	
		ASP72:OD2-H31 ligand N3		2.55
		Hydrophobic interactions Pi-Orbitals		
		TRP84 - N ligand N3		4.08
		PHE330 - N ligand N3		3.65
 New compound N4	-9.70	H-bond interactions receptor-ligand N4	0.68	
		HIS440:NE2-H26 ligand N4		2.31
		TYR334:OH-H32 ligand N4		2.84
		GLY80:O- H33 ligand N4		3.10

From the obtained results, the compounds exhibit a significant fit to the protein-binding pose through a non-covalent interaction network such as hydrogen bonding and unsaturated orbital bonding. Compound **N2** displays a bond energy of  $-9.4$  kcal mol<sup>-1</sup> and has hydrogen bonds between the amino acid and the ligand ASP72:OD2-HN, GLU199:OE1-H5, and TYR334:OH-H21. The distance of these bonds is equal to 2.49, 3.01, and 2.59 Å. Compound **N4** has a BE of  $-9.70$  kcal mol<sup>-1</sup> and has similar interactions to hydrogen bonding between amino acids and ligands HIS440:NE2-H26, TYR334:OH-H32, and GLY80:O-H33. The lengths of these bonds are equal to 2.31, 2.84, and 3.10 Å, respectively, as shown in Figure 10.

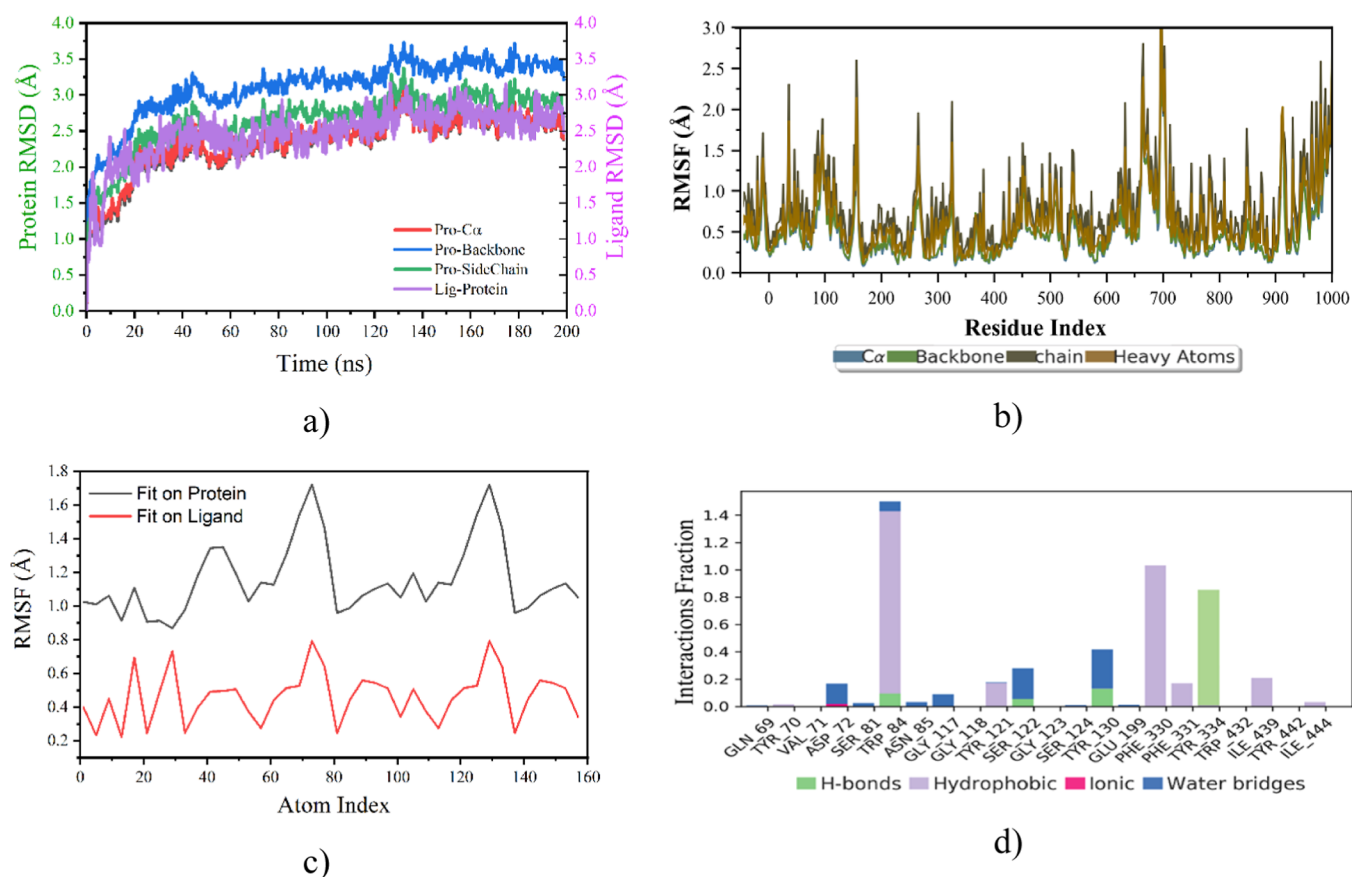
The rmsd (Å) values of the ligand and pose can be calculated relative to the pose using eq 8.<sup>55</sup> This is represented by a set of dendrograms created to display a dendrogram of interactions. The results of the rmsd (Å) estimation of the constructs against the specified reference structure are shown in Table 7. The rmsd values of the 1ACJ-PDB–ligand complexes were calculated. This rmsd value may be referred to as rmsd-L.

Furthermore, we were also able to find that compounds **N1** and **N3** had interactions with the target site through residues ASP72, SER122, TRP84, and GLU199 at distances of 2.55, 2.71, 2.24, and 2.63 Å, respectively. Thus, compounds **N1**, **N2**, and **N4** exhibit hydrogen bonding interactions with amino acids. This demonstrates that essential amino acids at the active target site also play a crucial role in ligand binding to the target site, as described in Table 7.

The **N2**– and **N4**–1ACJ-PDB complexes have the lowest BE and have three H-bonding interactions. These complexes successfully bind to the active site of the 1ACJ-PDB AChE protein, including van der Waals, electrostatic,  $\pi$ –sulfur, and  $\pi$ – $\pi$  or  $\pi$ –alkyl interactions. These interactions provide additional information about interactions that may support the binding of molecules at the active site more firmly and powerfully to the receptor active site of the 1ACJ-PDB AChE protein.

Combining the docking calculation results and the QSAR models, we found that the R<sub>1</sub> and R<sub>3</sub> substituents are –H, or phenyl rings, which can essentially increase the activity of the compound. The R<sub>2</sub> and R<sub>4</sub> substituents show little effect. Thus, we have explained why the focus is mainly on changing the R<sub>1</sub>





**Figure 11.** Results of MD simulation for the new compound N4: (a) rmsd-L evolution of 1ACJ–ligands N4; (b) peaks indicate areas of 1ACJ protein that fluctuate the most during the simulation; (c) ligand RMSF shows the ligand N4 fluctuations broken down by atom; and (d) 1ACJ–ligand simulation can monitor N4 interactions.

and R<sub>3</sub> substituents, as shown in Table 3. We obtained positive results for the pIC<sub>50</sub> activity of the novel designed compounds.

**3.5. MD Simulation.** The MD simulation process is performed under the conditions given in Section 2.7. The rmsd-L and RMSF values<sup>64</sup> calculated by eqs 8 and 9 were used to characterize the simulated 1ACJ–ligand N4 complex. The rmsd-L values of the 1ACJ–ligand N4 complexes were used to consider all 1ACJ profiles, and then the rmsd-L values were determined using atomic selection, as depicted in Figure 11a. The oscillations of this complex at the end of the simulation are thermal-average structural oscillations. The rmsd-L changing values can suggest that the simulation has reached equilibrium. A simulation time of 200 ns was allotted for the simulation, which was divided into two stages for system equilibrium and production. The system conducts the production simulation for the remainder of the simulation time after equilibrium has been reached at 20 ns, as shown in Figure 11a. Variations in rmsd-L values between 1 and 3 for 1ACJ–PDB protein are acceptable. The MD simulation must also be able to converge because the rmsd-L data shows a continuous change around a fixed value.

The measured values (rmsd-L) are calculated during simulation. It is calculated for all frames in the simulation trajectory.<sup>56,57,59</sup> The rmsd-L for frame  $x$  is

$$\text{RMSD}_x = \sqrt{\frac{1}{N} \sum_{i=1}^N (r_i^x(t_x) - r_i(t_{\text{ref}}))^2} \quad (8)$$

where  $N$  is the number of atoms,  $t_{\text{ref}}$  is the reference time, and reference  $r_x$  is the position of the selected atoms in a frame  $x$ , where frame  $x$  is recorded at time  $t_x$ . The discrepancy ( $r_i^x(t_x) - r_i(t_{\text{ref}})$ ) is the distance between  $r^x$  at time  $t_x$  and  $r_i$  at reference time  $t_{\text{ref}}$ .

Furthermore, the rmsd-L value of the 1ACJ–PDB–ligand complex also showed no significant increase or decrease at the end of the simulation. Thus, the system equilibrium has been reached, and the MD simulation time can last long enough to have the results of a study. The fluctuating RMSF in Figure 11b represent the most fluctuating protein regions during the simulation. We found that the N- and C-termini of the protein are more variable than any other region. Because they are stiffer than the unstructured protein, other secondary structural components, such as alpha helices and beta fibers, have fewer fluctuations.

In Figure 11c, it can be seen that the RMSF values of the 1ACJ–PDB–ligand N4 complex show atomically subdivided ligand changes. The RMSF value for the ligand N4 sheds light on the entropy contribution to the binding event and their interaction with the 1ACJ protein. The protein-related ligand fluctuations are shown on the ligand-fitted complex on the protein line. The 1ACJ–PDB–ligand complex was first seen on the protein backbone, and the RMSF value for the ligand was then calculated using the ligand-heavy atoms. The line of ligand oscillations is determined for ligand-heavy atoms, describing the fluctuations in which the N4 ligand lies in each frame of reference. These RMSF values accurately describe the intrinsic atomic vibrations of N4 ligand. The

fit–protein and fit–ligand curves in Figure 11c vary similarly and are steady when the oscillation process of the ligand to fit into the active receptor site gradually changes. The RMSF characterizes local changes in residues along the protein chain<sup>60</sup>

$$\text{RMSF}_i = \sqrt{\frac{1}{T} \sum_{t=1}^T (r_i^x(t_x) - r_i(t_{\text{ref}}))^2} \quad (9)$$

where  $T$  is the trajectory time over which the RMSF is calculated,  $t_{\text{ref}}$  is the reference time,  $r_i$  is the position of residue  $i$ ,  $r^x$  is the position of atoms in residue  $i$  after superposition on the reference, and the angle brackets  $i$  indicate that the average of the square distance is taken over the selection of atoms in the residue. The distance between position  $r_i$  at the reference time  $t_{\text{ref}}$  and  $r^x$  at time  $t_x$  is the discrepancy ( $r_i^x(t_x) - r_i(t_{\text{ref}})$ ).

During the simulation, protein interactions with ligands can be observed. These interactions can be categorized and summarized according to their nature, as can be seen in Figure 11d. The four protein–ligand interactions are hydrogen bonding, aqueous alkaline, ionic, and water bridged. There are more detailed subcategories inside each interaction type. Because some protein residues can cause several interactions of the same subtype to the ligand, values above 1.0 are feasible. The unique interaction is maintained for 70% of the simulation duration. This can be seen by the stacked plots normalized throughout the trajectory at 0.7. In the bonding of ligands, hydrogen bonds, or H-bonds, are crucial.

Hydrogen bonding qualities must be considered while developing drugs since they significantly impact drug selectivity, metabolism, and adsorption. Four subtypes of hydrogen bonds can also be distinguished between proteins and ligands: backbone acceptor, backbone donor, side-chain acceptor, and side-chain sponsors. The amino acids TYR334, TYR130, SER124, TYR121, SER122, SER81, GLU199, and SER84 are hydrogen bonds, hydrophobic bonds, and hydrogen bridge bonds. These essential bonds provide a strong binding between the ligand and the receptor of the IACJ-PDB protein. The dynamic simulation results here for the N4 ligand are relatively consistent with the docking calculation results. In addition to hydrogen bonding between amino acids and ligands, other bonding interactions, such as hydrophobic bonds, also play an essential role. It induces sustained binding between the ligand N4 and the active target site of the IACJ-PDB AChE protein. The hydrophobic bonds consist of  $\pi$ -cation,  $\pi$ - $\pi$ , and other nonspecific interaction types. These interactions involve hydrophobic amino acid fractions and aromatic or aliphatic groups on the ligand. We also found that this type of binding also includes  $\pi$ -cation interactions. These hydrophobic interactions are in the range of 4.5 Å. Nonspecific bonds are in the range of 3.6 Å.

### 3.6. Potential Inhibitors and In Vitro Bioactivity.

There was no additional purification performed on any of the chemical reagents or solvents used in the experiment; they were all bought from commercial vendors. The reaction mixture was cooled before being filtered, and the filtrate was evaporated under reduced pressure. We synthesized the designed compounds capable of AD treatment based on the reaction schemes in Section 2.8 and suitable reaction conditions as designed. To obtain the molecular structures predicted to inhibit AChE, compounds with structures corresponding to the potentiometric orientations in Table 3 were synthesized. Based on various spectral analysis

techniques, information on the molecular structures of compounds N1, N2, N3, and N4 is summarized below.

**3.6.1. New Compound (N1).** The compound (*E*)-*N'*-(1-(5-((dimethylamino)methyl)pyridin-2-yl)-2-methylbut-3-en-1-ylidene)-1*H*-imidazole-1-carbothiohydrazide combined with the compound piperidin-4-amine produces a new compound (N1) (*E*)-2-(1-(5-((dimethylamino)methyl)pyridin-2-yl)-2-methylbut-3-en-1-ylidene)-*N*-(piperidin-4-yl)hydrazine-1-carbothioamide. This compound is a dark solid, yield = 75%, mp 167.67 °C, (dioxane); chemical formula C<sub>19</sub>H<sub>30</sub>N<sub>6</sub>S; molecular weight: 374.225; <sup>1</sup>H NMR (500 MHz, DMSO-*d*<sub>6</sub>):  $\delta$  7.73 (d,  $J$  = 6.0 Hz, 1H), 5.94–5.82 (m, 1H), 5.29 (dt,  $J$  = 16.3, 2.3 Hz, 1H), 5.21 (dt,  $J$  = 11.0, 2.3 Hz, 1H), 4.47 (p,  $J$  = 4.1 Hz, 1H), 3.90 (dp,  $J$  = 6.1, 5.2 Hz, 1H), 3.19–3.09 (m, 1H), 2.99 (dddd,  $J$  = 14.3, 5.1, 4.1, 2.5 Hz, 2H), 2.86–2.77 (m, 2H), 2.48 (s, 2H), 2.43 (s, 5H), 1.89 (dtd,  $J$  = 13.9, 5.1, 2.4 Hz, 2H), 1.83 (d,  $J$  = 5.5 Hz, 1H), 1.66–1.57 (m, 3H), 1.13 (dd,  $J$  = 6.6, 0.9 Hz, 3H); <sup>13</sup>C NMR (125 MHz, DMSO-*d*<sub>6</sub>):  $\delta$  144.13, 50.49, 49.19, 46.79, 45.17, 28.12, 26.74, 17.86; <sup>15</sup>N NMR (51 MHz, chloroform-*d*):  $\delta$  331.05, 201.05, 141.20, 36.10, 34.45; MS  $m/z$  (%): 374.2253 (100.0%), 376.235 (23.92%), 377.232 (7.20%), 378.233 (1.246%), 379.234 (0.147%); Elemental Analysis: C, 60.93%; H, 8.07%; N, 22.44%; S, 8.56%.

**3.6.2. New Compound (N2).** The compound di(pyridine-2-yl)methanone combines with the compound *N*-cyclohexyl-*N*-ethylhydrazinecarbothioamide to form a new compound (N2) (*N*-cyclohexyl-2-(di(pyridine-2-yl)methylene)-*N*-ethyl hydrazine-1-carbothioamide). This compound is a dark solid, yield = 85%, mp 218.02 °C, (dioxane); chemical formula: C<sub>20</sub>H<sub>25</sub>N<sub>5</sub>S; molecular weight: 368.190; <sup>1</sup>H NMR (500 MHz, DMSO-*d*<sub>6</sub>):  $\delta$  7.07 (ddd,  $J$  = 4.7, 2.9, 1.5 Hz, 2H), 3.80 (p,  $J$  = 5.9 Hz, 1H), 3.76 (d,  $J$  = 8.0 Hz, 1H), 3.73 (d,  $J$  = 7.9 Hz, 1H), 2.73 (d,  $J$  = 1.5 Hz, 1H), 2.02–1.92 (m, 2H), 1.74 (ddd,  $J$  = 8.2, 4.1, 2.0 Hz, 1H), 1.74–1.66 (m, 3H), 1.66–1.60 (m, 1H), 1.63–1.56 (m, 3H), 1.46 (dq,  $J$  = 8.2, 3.3 Hz, 3H), 1.46–1.38 (m, 1H), 1.26 (t,  $J$  = 7.9 Hz, 3H); <sup>13</sup>C NMR (125 MHz, DMSO-*d*<sub>6</sub>):  $\delta$  58.19, 45.84, 29.51, 25.73, 23.06, 12.40; <sup>15</sup>N NMR (51 MHz, chloroform-*d*):  $\delta$  331.06, 201.05, 48.01; MS  $m/z$  (%): 368.190 (100.0%), 369.193 (24.527%), 370.190 (7.490%), 371.209 (1.244%), 372.202 (0.048%); Elemental Analysis: C, 65.36%; H, 6.86%; N, 19.06%; S, 8.72%.

**3.6.3. New Compound (N3).** Compound *O*-phenyl (*E*)-2-(2-(ethylamino)-1-(pyridin-2-yl)propylidene)hydrazine-1-carbothioate combines with compound *N*-ethylcyclohexanamine to form a new compound (N3) ((*E*)-*N*-cyclohexyl-*N*-ethyl-2-(2-(ethylamino)-1-(pyridine-2-yl)propylidene)hydrazine-1-carbothioamide). This compound is a brown solid, yield = 80%, mp 193.45 °C, (dioxane); chemical formula: C<sub>19</sub>H<sub>31</sub>N<sub>5</sub>S; molecular weight: 361.55; <sup>1</sup>H NMR (500 MHz, DMSO-*d*<sub>6</sub>):  $\delta$  4.09 (dq,  $J$  = 6.8, 4.9 Hz, 1H), 3.84–3.71 (m, 2H), 2.78–2.67 (m, 1H), 2.02–1.92 (m, 1H), 1.77–1.56 (m, 4H), 1.51–1.38 (m, 2H), 1.32–1.23 (m, 3H), 1.18 (t,  $J$  = 5.9 Hz, 2H); <sup>13</sup>C NMR (125 MHz, DMSO-*d*<sub>6</sub>):  $\delta$  58.19, 56.17, 45.84, 29.56, 25.73, 23.06, 18.70, 14.96, 12.40; <sup>15</sup>N NMR (51 MHz, chloroform-*d*):  $\delta$  331.06, 201.05, 48.01, 35.94; MS  $m/z$  (%): 362.237 (100.0%), 363.240 (23.565%), 364.237 (7.178%), 365.237 (1.220%), 366.239 (0.143%), 367.240 (0.013%); Elemental Analysis: C, 63.12%; H, 8.64%; N, 19.37%; S, 8.87%.

**3.6.4. New Compound (N4).** (*E*)-4-(Cyclohexyl(hydrazineylidene)methyl)-*N*-ethyl-*N*-methyl aniline compound combines with compound isothiocyanic acid to form a newly designed compound (N4) ((*Z*)-2-(cyclohexyl(4-

(ethyl(methyl)amino)phenyl)methylene)hydrazine-1-carbothioamide). This compound is a brown solid, yield = 85%, mp 198.20 °C (dioxane); chemical formula:  $C_{17}H_{26}N_4S$ ; molecular weight: 318.48;  $^1H$  NMR (500 MHz, DMSO- $d_6$ ):  $\delta$  8.98 (s, 1H), 2.93–2.82 (m, 2H), 1.96 (ddt,  $J = 12.2, 7.9, 5.7$  Hz, 1H), 1.84–1.76 (m, 2H), 1.74–1.65 (m, 1H), 1.68–1.57 (m, 1H), 1.55–1.46 (m, 1H), 1.50–1.38 (m, 1H), 1.13 (t,  $J = 7.9$  Hz, 2H);  $^{13}C$  NMR (125 MHz, DMSO- $d_6$ ):  $\delta$  41.87, 41.66, 28.86, 28.58, 12.58;  $^{15}N$  NMR (51 MHz, chloroform- $d$ ):  $\delta$  331.06, 290.40, 210.86, 201.05;  $m/z$ : 319.195 (100.0%), 320.198 (20.975%), 321.194 (6.614%), 322.195 (1.044%), 323.196 (0.114%), 324.197 (0.0097%); Elemental Analysis: C, 64.11%; H, 8.23%; N, 17.59%; S, 10.07%.

Following the synthesis of novel thiosemicarbazones, these newly synthesized compounds were tested for in vitro biological activity against acetylcholinesterase. They were dissolved in DMSO at room temperature, and the enzyme solution was recorded and compared with the appropriate standard inhibitor, NDGA.<sup>70</sup> The classical Ellman method can be used to determine if these newly synthesized compounds can prevent CHE from working in vitro. The inhibitory activity of a new compound has been assigned an  $IC_{50}$  value, which is the concentration of the compound that inhibits 50% of enzyme activity, as shown in Table 8. The thiol ester

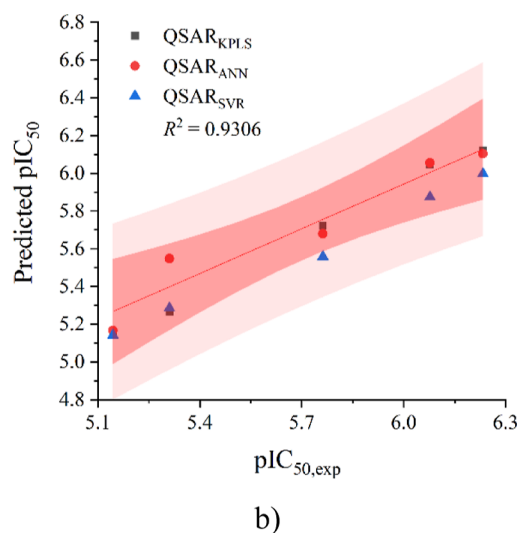
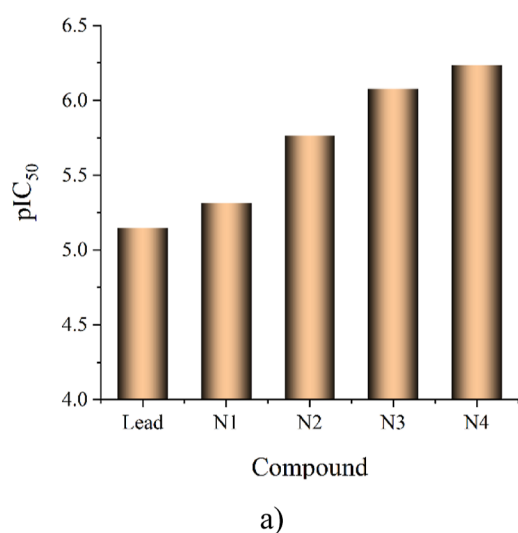
acetylthiocholine, used as the substrate AChE, is hydrolyzed to afford thiocholine and acetate. After adding the stock solution, it was appropriately diluted to the desired final concentration (100 M) in a pH 8 phosphate buffer (0.15 M), and the absorbance was measured at 405 nm. Tacrine was also used as the reference compound for this AChE enzyme that has been implicated in the treatment of AD.<sup>71</sup> The experimental results show that four new thiosemicarbazone derivatives have inhibitory effects on the 1ACJ-PDB protein target site of the AChE enzyme, consistent with the drug-likeness properties that meet the Lipinski-5 and Veber rules. Thiosemicarbazone compounds N1, N2, N3, and N4 have better bioactivity than the lead compound Li39. The experimental bioactivities  $pIC_{50}$  of new compounds were consistent with the activity predicted by the QSAR<sub>KPLS</sub>, QSAR<sub>ANN</sub>, and QSAR<sub>SVR</sub> models ( $pIC_{50} > 5.0$ ), higher than that of the lead compound Li39. New compounds N2, N3, and N4 show higher experimental activity than compounds N1 and the lead Li39. Of these, compound N4 seems to have the best bioactivity.

Furthermore, the bioactivities of the thiosemicarbazone derivatives N2, N3, and N4 were more significant than those of the thiosemicarbazone compounds Li39 and N1 due to the strong influence of the R1 and R3 substituents with aromatic rings, mainly the influence of the group with the electron-withdrawing aromatic ring of the phenyl ring. The correlation comparisons between the experimental  $pIC_{50}$  values of the newly synthesized substances N1, N2, N3, and N4 with Li39 and those from QSAR models are depicted in Figure 12a,b.

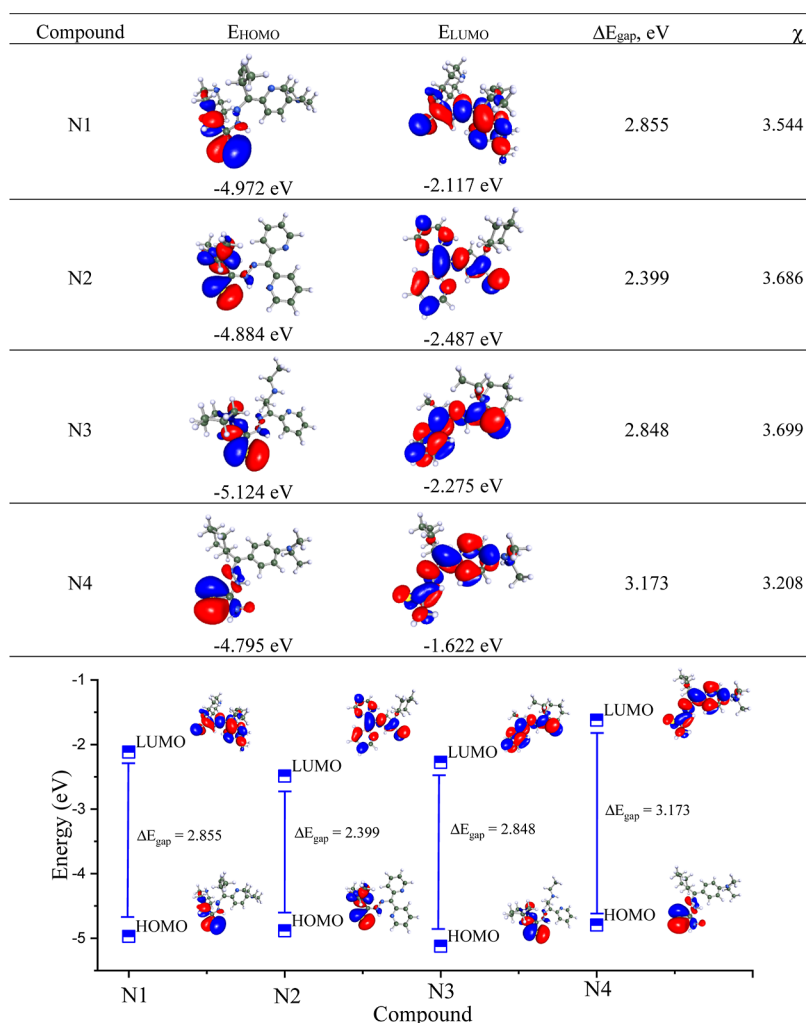
Figure 12 shows that the in vitro activity of the newly engineered compounds is consistent with the results predicted from QSAR, docking, and MD simulations. The predicted results from the QSAR model correlate well with the experimentally determined results. In general, the prediction domains of QSAR models are in the 95% confidence area. However, the QSAR<sub>ANN</sub> model seems to give the best prediction results.

**Table 8. Comparison of Experimental In Vitro  $pIC_{50}$  Activity and Those from QSAR Models Using DF and PC Descriptors for the Newly Synthesized Compounds N1, N2, N3, and N4**

tested compounds	$pIC_{50,exp}$	predicted $pIC_{50}$ using mixed descriptors DF and PC		
		QSAR <sub>KPLS</sub>	QSAR <sub>ANN</sub>	QSAR <sub>SVR</sub>
Li39	5.144 (ref 40)	5.153	5.166	5.142
N1	5.310 (this work)	5.266	5.548	5.286
N2	5.762 (this work)	5.723	5.680	5.558
N3	6.077 (this work)	6.045	6.055	5.874
N4	6.233 (this work)	6.121	6.104	5.997



**Figure 12.** (a) Comparison of in vitro  $pIC_{50}$  bioactivity of newly synthesized compounds N1, N2, N3, and N4 with the lead compound Li39 and (b) correlation between experimental in vitro  $pIC_{50}$  bioactivity and those from QSAR models.



**Figure 13.** Molecular orbitals and orbital energy levels of p-type for new compounds N1, N2, N3, and N4 at the B3LYP/def-SVP level.  $\Delta E_{\text{gap}} = E_{\text{HOMO}} - E_{\text{LUMO}}$  and  $\chi = |E_{\text{HOMO}} + E_{\text{LUMO}}|/2$ .

#### 4. DISCUSSION

In this work, the in silico models demonstrated the nature of the interactions between the new ligands and the active target site of 1ACJ-PDB AChE. The predicted results from the theory models agree with those from the in vitro experimental measurements.

Besides constructing the QSAR<sub>KPLS</sub>, QSAR<sub>ANN</sub>, and QSAR<sub>SVR</sub> models, we utilized the PharmaQSAR model to assess the significant susceptibility of the R<sub>1</sub>, R<sub>2</sub>, R<sub>3</sub>, and R<sub>4</sub> groups. The considered results proved that the substituents at position R<sub>1</sub> (*E*)-[(2-pyridyl)(*p*-tolyl)methylidene]hydrazine, [bis(2-pyridyl)methylidene]hydrazine, 2-(ethylamino)-1-hydrazono-1-(2-pyridyl)propane, and (*E*)-1-[2-methyl-1-(2-pyridyl)propylidene]-4-(4-piperidyl)thiosemicarbazide and the substituent at position R<sub>3</sub> 4-[cyclohexyl(imino)methyl]-*N*-ethyl-*N*-methylcyclohexan-1-amine were more important than the other positions because these R<sub>1</sub> and R<sub>3</sub> groups carry aromatic rings and are easy to form conjugates throughout the molecular system. This resulted in a more significant difference than for pIC<sub>50</sub> activity. The significance of substituents R<sub>1</sub> and R<sub>3</sub> reflects the influence on the electrophilic or nucleophilic groups at that position. The significance of groups R<sub>1</sub>, R<sub>2</sub>, R<sub>3</sub>, and R<sub>4</sub> is shown in the histogram color of Table 3. Many different R substituents can

be used to change what a site-specific R substituent is expected to do to a molecule. We expect sites with R groups exhibiting more significant chemical variation to exhibit a larger range of property values upon transformation, thus of greater importance. We may have reason to believe that placing the groups for the hydrogen bonding in a certain position is important. These possibilities should be kept in mind when interpreting the measure of importance.

Through these research results, we found that the PC properties of the target protein are determined by the important amino acid residues TRP, GLY, GLU, PHE, TYP, HIS, SER, and GLY. In addition, proteins have properties such as acidity and isoelectric point. Other essential protein properties, such as aliphatic index and mean hydrophobicity, also contribute to the inhibitory potential of 1ACJ-PDB. This proves to be necessary to combine new drug discovery research that must combine theoretical and experimental pathways to realize the mechanism of action of the 1ACJ-PDB AChE protein inhibitor compound. Although there are more than 50 good protein structures of human AChE in the UNIPROTDB database, we chose the 1ACJ protein for this study based on structural studies and exploration of its receptor active site. Aromatic residues are present at the active receptor site of acetylcholinesterase. It is also part of the amino acid structures that can interact with thiosemicarbazone ligands carrying

nitrogen-containing aromatic heterocycles. These rings will be able to form donor and acceptor bonds when the ligand enters the active site of the indole-carrying receptors Trp-84 and *m*-hydroxyl. It exhibits branched hydrogen bonding with two members of the catalytic triad, Ser-200 and His-440. On this basis, we also wanted to investigate and evaluate the effect of the thiosemicarbazone ligand in the case of this protein, 1ACJ. We also found that compounds with similar structures interact well with similar structures. Furthermore, the binding sites of Torpedo acetylcholinesterase for quaternary ligands have been investigated by X-ray crystallography. In our newly designed ligand structures, the C=S and N–N bonds are in reverse order. Our novel ligands can interact with the indoles of Trp-84 and Trp-279, respectively. This close interaction has also been confirmed. As a result of the docking study, we found that the structure and chemistry together indicate an important role for aromatic groups at the binding sites of ligands carrying aromatic rings with nitrogen atoms and evidence additional information on the roles of Trp-84 and Phe-330.

The results obtained for the newly designed thiosemicarbazone ligands agree with the results of investigations by X-ray crystallography.<sup>41</sup> The average of the thiosemicarbazone ligand with the amino acids of the AChE receptor is in the range of 3.005 Å. This result is consistent with the interaction between the crystal structure and the receptor at a resolution of 2.8 Å. In the protein–ligand complex, the nitrogen atom interacts with the indole TRP84 and exhibits binding to TYR33, SER122, and HIS440. A thiosemicarbazone ligand–indole TRP84 complex is directed to the active site; another group is also attached to the structural part of TRP84. The structural difference is the protein binding and the determining ligand, which is the orientation of the phenyl ring of PHE330. These interactions are consistent with the active site 3H photo-sensitive probe results. The chemical structures of the novel thiosemicarbazones suggest an essential role for aromatic groups, such as 2D binding site interactions shown in Figures 5a and 10a. This provides solid additional evidence for TRP84 and PHE330 interaction with the active site.

Furthermore, to obtain a more accurate assessment of the mechanism and influence of molecular structure on the inhibitory activity of 1ACJ-PDB, we perform DFT quantum mechanical calculations at the theoretical level of B3LYP/def-SVP.<sup>67,68</sup> We used quantum mechanical calculations to discuss the potential activity of novel substances based on the molecular orbital structures of the compounds N1, N2, N3, and N4. These new theoretically optimized compounds all converge well.

The structure of the compounds will be less affected by various hydrogen bonding, van der Waals, and electrostatic interactions. As a result, we can confirm that the structures N1, N2, N3, and N4 can exist in a stable state at all times. It can be confirmed that these molecular structures can also be stable when moving through different body environments and interact well with the 1ACJ-PDB protein receptor. The energies of the HOMO and LUMO molecular orbitals of the compounds N1, N2, N3, and N4 were calculated, as shown in Figure 13. These quantum calculation results can explain the interactions of the compounds with the 1ACJ-PDB receptor based on the electron exchange. The electron density on the molecular orbitals HOMO and LUMO is different, so the electron exchange is also different.

This may result in compounds N1, N2, N3, and N4 having different interactions upon access to the active and inhibitory

target sites for 1ACJ-PDB because the electrons in each molecule appear to have a different localization. Therefore, the molecular orbitals HOMO and LUMO are also localized in different  $\pi$  electron regions of the molecular structure, i.e., these orbitals are concentrated mainly in C=S bonds and heterocyclic groups. Compounds N1, N2, N3, and N4 were all subjected to NBO analysis, as shown in Figure 13. The  $E_{\text{HOMO}}$  values of the molecules are all below  $-4.5$  eV. This is the level required for the stability of a compound. In general, the structures of compounds N1, N2, N3, and N4 are relatively stable in terms of electron structure; the  $E_{\text{HOMO}}$  energy difference between compounds is not much. Compound N3 has a value  $E_{\text{HOMO}} = -5.1239$  eV, indicating that the electronic structure of this compound is the most stable.

Furthermore, the band gap energies  $\Delta E = E_{\text{HOMO}} - E_{\text{LUMO}}$  of compounds N1, N2, N3, and N4 are 2.399–3.173 eV. This energy range readily exchanges electrons to form bonds. This means that they are all quantum favorable and can form bonds between N1, N2, N3, and N4 molecules with the active target site of the 1ACJ-PDB protein structure.

In addition, electrostatic interactions between electro-negative ( $\chi$ ) ligands can also be a reliable indicator of compound inhibition. Ligands tend to form bonds with the residues. Greater electronegativity means stronger electron acceptability. The most electronegative compounds, N2 and N3, can be electron acceptors.

In general, the results of quantum mechanical calculations also show that the inhibitory activities of substances N1, N2, N3, and N4 are consistent with the results predicted from the prediction of the QSAR models, docking, MD simulation, and experimental in vitro  $\text{pIC}_{50}$  activities. The discrepancy between the inhibibility of the new compounds for 1ACJ-PDB AChE was not significant. Through quantum calculations, bonds are formed by changing electrons in the HOMO and LUMO orbitals. Thus, the mechanism of action of potential AChE inhibitors on the 1ACJ-PDB receptor may include inactivation of the enzyme, leading to the accumulation of ACh at the junction nerve. Thus, the mode of action of potential inhibitors of AChE could become a key agent in the pharmacotherapy of the symptoms of AD. Therefore, thiosemicarbazone compounds may potentiate cholinergic neurotransmission in brain regions. The use of these potential thiosemicarbazone class inhibitors relies on interaction with AChE as the primary target.

## 5. CONCLUSIONS

From the results of this study, we can draw the following conclusions:

We have presented the research process from the virtual screening of the database to the synthesis of new compounds efficiently using the in silico model for studying the thiosemicarbazone series. Four new compounds designed and synthesized from the orientation of the  $\text{QSAR}_{\text{KPLS}}$ ,  $\text{QSAR}_{\text{ANN}}$ , and  $\text{QSAR}_{\text{SVR}}$  models meet the Lipinski-5 and Veber virtual screening rules for drug-likeness properties and docking calculation results. The experimental in vitro activity results against AChE also agreed with those predicted from in silico models. The results obtained for the four novel compounds inhibiting AChE showed that the N2, N3, and N4 derivatives had  $\text{pIC}_{50}$  activities of 5.762, 6.077, and 6.233 (promising  $\text{IC}_{50}$  1.730, 0.838, and 0.585 nM), respectively, compared with the  $\text{pIC}_{50} = 5.144$  activity of lead compound Li39 (promising  $\text{IC}_{50} = 7.1779$  nM). Molecular connectivity

analysis revealed that the engineered and synthesized compounds were predicted to be consistent with the positions of the R<sub>1</sub> and R<sub>3</sub> substituents and their ability to bind to the AChE target receptor. In conclusion, the newly synthesized thiosemicarbazone compounds N2, N3, and N4 could be potential inhibitors of AChE in AD treatment.

The integration of analyses of the QSAR<sub>KPLS</sub>, QSAR<sub>ANN</sub>, and QSAR<sub>SVR</sub> models allows for prediction and design direction to create an important reference data set in new compound development. The information on the substituents generating new important structures has been outlined as valuable information for orienting the design of new derivatives with improved activity and properties.

## ■ ASSOCIATED CONTENT

### SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acsomega.2c07934>.

Structural data of the group of compounds used in this work; QSAR<sub>ANN</sub> and QSAR<sub>SVR</sub> models predicting pIC<sub>50</sub> and BE in kcal mol<sup>-1</sup> for the experimental group and the newly designed compound group; error data evaluating the predictability for pIC<sub>50</sub> and BE for the compounds; and important calculation results (PDF)

## ■ AUTHOR INFORMATION

### Corresponding Authors

**Hoa Tran Thai** – Faculty of Chemistry, Hue University of Sciences, Hue University, Hue City 530000, Viet Nam; [orcid.org/0000-0003-4814-4364](https://orcid.org/0000-0003-4814-4364); Email: [tthaihoa@hueuni.edu.vn](mailto:tthaihoa@hueuni.edu.vn)

**Pham Van Tat** – Department of Sciences and Journal Management, Hoa Sen University, Ho Chi Minh 700000, Viet Nam; [orcid.org/0000-0002-7497-7741](https://orcid.org/0000-0002-7497-7741); Email: [phvantat@gmail.com](mailto:phvantat@gmail.com)

### Authors

**Nguyen Minh Quang** – Faculty of Chemical Engineering, Industrial University of Ho Chi Minh City, Ho Chi Minh 700000, Viet Nam; [orcid.org/0000-0002-2418-9253](https://orcid.org/0000-0002-2418-9253)

**Hoa Le Thi** – Faculty of Chemistry, Hue University of Sciences, Hue University, Hue City 530000, Viet Nam; [orcid.org/0000-0002-1359-187X](https://orcid.org/0000-0002-1359-187X)

**Nguyen Duc Cuong** – Faculty of Chemistry, Hue University of Sciences, Hue University, Hue City 530000, Viet Nam; School of Hospitality and Tourism, Hue University, Hue City 530000, Viet Nam; [orcid.org/0000-0002-7341-3661](https://orcid.org/0000-0002-7341-3661)

**Nguyen Quoc Hien** – Vietnam Atomic Energy Institute, Hanoi City 100000, Viet Nam

**DongQuy Hoang** – Faculty of Materials Science and Technology, University of Science, Vietnam National University, Ho Chi Minh 700000, Viet Nam; Vietnam National University, Ho Chi Minh City 700000, Viet Nam; [orcid.org/0000-0002-5494-6796](https://orcid.org/0000-0002-5494-6796)

**Vu Thi Bao Ngoc** – Faculty of Chemistry and Environment, University of Dalat, Dalat City 660000, Viet Nam; [orcid.org/0000-0002-8485-7874](https://orcid.org/0000-0002-8485-7874)

**Vo Ky Minh** – Franklin High School, Elk Grove, California 95757, United States

Complete contact information is available at: <https://pubs.acs.org/doi/10.1021/acsomega.2c07934>

## Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

Hue University supported this work partially under the Core Research Program, grant no. NCM.DHH.2022.04.

## ■ REFERENCES

- (1) Hughes, J. P.; Rees, S.; Kalindjian, S. B.; Philpott, K. L. Principles of early drug discovery. *Br. J. Pharmacol.* **2011**, *162*, 1239–1249.
- (2) Reddy, A. S.; Zhang, S. Polypharmacology: drug discovery for the future. *Expert Rev. Clin. Pharmacol.* **2013**, *6*, 41–47.
- (3) Fogel, D. B. Factors associated with clinical trials that fail and opportunities for improving the likelihood of success: A review. *Contemp. Clin. Trials Commun.* **2018**, *11*, 156–164.
- (4) Wang, Q.; Dwivedi, Y. Advances in novel molecular targets for antidepressants. *Prog. Neuro-Psychopharmacol. Biol. Psychiatry* **2021**, *104*, 110041.
- (5) Oxford, A. E.; Stewart, E. S.; Rohn, T. T. Clinical Trials in Alzheimer Disease: A Hurdle in the Path of Remedy. *Int. J. Alzheimer's Dis.* **2020**, *2020*, 5380346.
- (6) Schneider, L. S.; Sano, M. Current Alzheimer disease clinical trials: Methods and placebo outcomes. *Alzheimer's Dementia* **2009**, *5*, 388–397.
- (7) Huang, L.-K.; Chao, S.-P.; Hu, C.-J. Clinical trials of new drugs for Alzheimer disease. *J. Biomed. Sci.* **2020**, *27*, 18.
- (8) Lecoutey, C.; Legay, R.; Davis, A.; Sopková-de Oliveira Santos, J.; Dallemagne, P.; Rochais, C. Development of Novel Potential Pleiotropic Compounds of Interest in Alzheimer's Disease Treatment through Rigidity Strategy. *Molecules* **2021**, *26*, 2536.
- (9) Azam, F.; Amer, A. M.; Abulifa, A. R.; Elzwawi, M. M. Ginger components as new leads for the design and development of novel multi-targeted anti Alzheimer's drugs: a computational investigation. *Drug Des., Dev. Ther.* **2014**, *8*, 2045–2059.
- (10) Guo, J.; Xu, A.; Cheng, M.; Wan, Y.; Wang, R.; Fang, Y.; Jin, Y.; Xie, S.-S.; Liu, J. Design, Synthesis and Biological Evaluation of New 3,4-Dihydro-2(1H)-Quinolinone-Dithiocarbamate Derivatives as Multifunctional Agents for the Treatment of Alzheimer's Disease. *Drug Des., Dev. Ther.* **2022**, *16*, 1495–1514.
- (11) Arrué, L.; Cigna-Méndez, A.; Barbosa, T. ; Borrego-Muñoz, P.; Struve-Villalobos, S.; Oviedo, V.; Martínez-García, C.; Sepúlveda-Lara, A.; Millán, N.; Márquez Montesinos, J. C. E.; et al. New Drug Design Avenues Targeting Alzheimer's Disease by Pharmacoinformatics-Aided Tools. *Pharmaceutics* **2022**, *14*, 1914.
- (12) Bojić, T.; Sencanski, M.; Perovic, V.; Milicevic, J.; Glisic, S. In Silico Screening of Natural Compounds for Candidates 5HT<sub>6</sub> Receptor Antagonists against Alzheimer's Disease. *Molecules* **2022**, *27*, 2626.
- (13) Maveyraud, L.; Mourey, L. Protein X-ray Crystallography and Drug Discovery. *Molecules* **2020**, *25*, 1030.
- (14) Gimeno, A.; Ojeda-Montes, M. J.; Tomás-Hernández, S.; Cereto-Massagué, A.; Beltrán-Debón, R.; Mulero, M.; Pujadas, G.; Garcia-Vallvé, S. The Light and Dark Sides of Virtual Screening: What Is There to Know? *Int. J. Mol. Sci.* **2019**, *20*, 1375.
- (15) Gentile, F.; Yaacoub, J. C.; Gleave, J.; Fernandez, M.; Ton, A.-T.; Ban, F.; Stern, A.; Cherkasov, A. Artificial intelligence-enabled virtual screening of ultra-large chemical libraries with deep docking. *Nat. Protoc.* **2022**, *17*, 672–697.
- (16) Li, Z.; Zhang, X.; Liao, J.; Fan, X.; Cheng, Y. An ultra-robust fingerprinting method for quality assessment of traditional Chinese medicine using multiple reaction monitoring mass spectrometry. *J. Pharm. Anal.* **2021**, *11*, 88–95.
- (17) An, Y.; Sherman, W.; Dixon, S. L. Kernel-Based Partial Least Squares: Application to Fingerprint-Based QSAR with Model Visualization. *J. Chem. Inf. Model.* **2013**, *53*, 2312–2321.
- (18) Myint, K.-Z.; Wang, L.; Tong, Q.; Xie, X.-Q. Molecular Fingerprint-Based Artificial Neural Networks QSAR for Ligand

Biological Activity Predictions. *Mol. Pharmaceutics* **2012**, *9*, 2912–2923.

(19) Fernández-de Gortari, E.; García-Jacas, R.; Martínez-Mayorga, K.; Medina-Franco, J. L. Database fingerprint (DFP): an approach to represent molecular databases. *J. Cheminf.* **2017**, *9*, 9.

(20) Duan, J.; Dixon, S. L.; Lowrie, J. F.; Sherman, W. Analysis and comparison of 2D fingerprints: Insights into database screening performance using eight fingerprint methods. *J. Mol. Graphics Modell.* **2010**, *29*, 157.

(21) Sadik, K.; Byadi, S.; Hachim, M. E.; El Hamdani, N.; Podlipnik, C.; Aboulmouhajir, A. Multi-QSAR approaches for investigating the relationship between chemical structure descriptors of Thiadiazole derivatives and their corrosion inhibition performance. *J. Mol. Struct.* **2021**, *1240*, 130571.

(22) Breijyeh, Z.; Karaman, R. Comprehensive Review on Alzheimer's Disease: Causes and Treatment. *Molecules* **2020**, *25*, 5789.

(23) Akhoun, B. A.; Choudhary, S.; Tiwari, H.; Kumar, A.; Barik, M. R.; Rathor, L.; Pandey, R.; Nargotra, A. Discovery of a new donepezil-like acetylcholinesterase inhibitor for targeting Alzheimer's disease: computational studies with biological validation. *J. Chem. Inf. Model.* **2020**, *60*, 4717–4729.

(24) Zaib, S.; Munir, R.; Younas, M. T.; Kausar, N.; Ibrar, A.; Aqsa, S.; Shahid, N.; Asif, T. T.; Alsaab, H. O.; Khan, I. Hybrid Quinoline-Thiosemicarbazone Therapeutics as a New Treatment Opportunity for Alzheimer's Disease-Synthesis, In Vitro Cholinesterase Inhibitory Potential and Computational Modeling Analysis. *Molecules* **2021**, *26*, 6573.

(25) Mavroidi, B.; Kaminari, A.; Matiadis, D.; Hadjipavlou-Litina, D.; Pelecanou, M.; Tzinia, A.; Sagnou, M. The Prophylactic and Multimodal Activity of Two Isatin Thiosemicarbazones against Alzheimer's Disease In Vitro. *Brain Sci.* **2022**, *12*, 806.

(26) Colovic, M. B.; Krstic, D. Z.; Lazarevic-Pasti, T. D.; Bondzic, A. M.; Vasic, V. M. Acetylcholinesterase Inhibitors: Pharmacology and Toxicology. *Curr. Neuropharmacol.* **2013**, *11*, 315–335.

(27) Marucci, G.; Buccioni, M.; Ben, D. D.; Lambertucci, C.; Volpini, R.; Amenta, F. Efficacy of acetylcholinesterase inhibitors in Alzheimer's disease. *Neuropharmacology* **2021**, *190*, 108352.

(28) Eddleston, M.; Chowdhury, F. R. Pharmacological treatment of organophosphorus insecticide poisoning: the old and the (possible) new. *Br. J. Clin. Pharmacol.* **2015**, *81*, 462–470.

(29) Wang, D.; Zou, L.; Jin, Q.; Hou, J.; Ge, G.; Yang, L. Human carboxylesterases: a comprehensive review. *Acta Pharm. Sin. B* **2018**, *8*, 699–712.

(30) Masondo, N. A.; Stafford, G. I.; Aremu, A. O.; Makunga, N. P. Acetylcholinesterase inhibitors from southern African plants: An overview of ethnobotanical, pharmacological potential and phytochemical research including and beyond Alzheimer's disease treatment. *S. Afr. J. Bot.* **2019**, *120*, 39–64.

(31) Zhang, H.; Wang, Y.; Wang, Y.; Li, X.; Wang, S.; Wang, Z. Recent advance on carbamate-based cholinesterase inhibitors as potential multifunctional agents against Alzheimer's disease. *Eur. J. Med. Chem.* **2022**, *240*, 114606.

(32) Schneider, M.; Pons, J.-L.; Labesse, G.; Bourguet, W. In Silico Predictions of Endocrine Disruptors Properties. *Endocrinology* **2019**, *160*, 2709–2716.

(33) Liu, H.; Wang, L.; Lv, M.; Pei, R.; Li, P.; Pei, Z.; Wang, Y.; Su, W.; Xie, X.-Q. AlzPlatform: An Alzheimer's Disease Domain-Specific Chemogenomics Knowledgebase for Polypharmacology and Target Identification Research. *J. Chem. Inf. Model.* **2014**, *54*, 1050–1060.

(34) Neves, B. J.; Braga, R. C.; Melo-Filho, C. C.; Moreira-Filho, J. T.; Muratov, E. N.; Andrade, C. H. QSAR-Based Virtual Screening: Advances and Applications in Drug Discovery. *Front. Pharmacol.* **2018**, *9*, 1275.

(35) Dhamodharan, G.; Mohan, C. G. Machine learning models for predicting the activity of AChE and BACE1 dual inhibitors for the treatment of Alzheimer's disease. *Mol. Diversity* **2022**, *26*, 1501–1517.

(36) Viswanadhan, V. N.; Ghose, A. K.; Revankar, G. R.; Robins, R. K. Atomic Physicochemical Parameters for Three Dimensional

Structure Directed Quantitative Structure-Activity Relationships. 4. Additional Parameters for Hydrophobic and Dispersive Interactions and Their Application for an Automated Superposition of Certain Naturally Occurring Nucleoside Antibiotics. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 163–172.

(37) Sadeghi, F.; Afkhami, A.; Madrakian, T.; Ghavami, R. QSAR analysis on a large and diverse set of potent phosphoinositide 3-kinase gamma (PI3K $\gamma$ ) inhibitors using MLR and ANN methods. *Sci. Rep.* **2022**, *12*, 6090.

(38) de Moura, E. P.; Fernandes, N. D.; Monteiro, A. F. M.; de Medeiros, H. I. R.; Scotti, M. T.; Scotti, L. Machine Learning, Molecular Modeling, and QSAR Studies on Natural Products Against Alzheimer's Disease. *Curr. Med. Chem.* **2021**, *28*, 7808–7829.

(39) An open data resource of binding, functional and ADMET bioactivity data. ChEMBL Dataset of thiosemicarbazone compounds for inhibiting Acetylcholinesterase. [https://www.ebi.ac.uk/chembl/g/#search\\_results/compounds/query=thiosemicarbazone](https://www.ebi.ac.uk/chembl/g/#search_results/compounds/query=thiosemicarbazone); EMBL-EBI: Hinxton, Cambridgeshire, CB10 1SD, UK.

(40) Liu, N.; Xu, Z.; Zeng, X.-J.; Ren, P. An agglomerative hierarchical clustering algorithm for linear ordinal rankings. *Inf. Sci.* **2012**, *557*, 170–193.

(41) Harel, M.; Schalk, I.; Ehret-Sabatier, L.; Bouet, F.; Goeldner, M.; Hirth, C.; Axelsen, P. H.; Silman, I.; Sussman, J. L. Quaternary ligand binding to aromatic residues in the active-site gorge of acetylcholinesterase. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 9031–9035.

(42) D. E. Shaw Research. Desmond Molecular Dynamics System. *Maestro-Desmond Interoperability Tools*; Schrödinger: New York, NY, 2021.

(43) Bowers, K. J.; Chow, D. E.; Xu, H.; Dror, R. O.; Eastwood, M. P.; Gregersen, B. A.; Klepeis, J. L.; Kolossvary, I.; Moraes, M. A.; Sacerdoti, F. D.; et al. Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters. *Proceedings of the ACM/IEEE Conference on Supercomputing (SC06), Tampa, Florida, 2006, November 11-17, 2006*. DOI: 10.1109/SC.2006.54.

(44) Holmes, D. T. Chapter 2—Statistical methods in laboratory medicine, *Contemporary Practice in Clinical Chemistry Contemporary Practice in Clinical Chemistry*, 4th ed., 2020; pp 15–35. DOI: 10.1016/B978-0-12-815499-1.00002-8.

(45) Jaillet, L.; Artemova, S.; Redon, S. IM-UFF: extending the Universal Force Field for interactive molecular modeling. *J. Mol. Graphics Modell.* **2017**, *77*, 350–362.

(46) Rosipal, R.; Trejo, L. J. Kernel Partial Least Squares Regression in Reproducing Kernel Hilbert Space. *J. Mach. Learn. Res.* **2001**, *2*, 97–123.

(47) Zhu, Q.; Zhao, Z.; Liu, F. Developing new products with kernel partial least squares model inversion. *Comput. Chem. Eng.* **2021**, *155*, 107537.

(48) Xia, J.-H.; Rusli; Kumta, A. S. Feedforward Neural Network Trained by BFGS Algorithm for Modeling Plasma Etching of Silicon Carbide. *IEEE Trans. Plasma Sci.* **2010**, *38*, 142–148.

(49) Bishop, C. M. *Neural Networks for Pattern Recognition*; Oxford University Press, Inc: 198 Madison Ave. New York, NY United States, 1995. DOI: 10.5555/525960.

(50) Prakash, O.; Khan, F. Development of Method for Three-Point Data Estimation and SVR-QSAR Model to Screen Anti Cancer Leads. *Comb. Chem. High Throughput Screening* **2013**, *16*, 425–434.

(51) Quang, N. M.; Mau, T. X.; Ai Nhung, N. T.; Minh An, T. N.; Van Tat, P. Novel QSPR modeling of stability constants of metal-thiosemicarbazone complexes by hybrid multivariate technique: GA-MLR, GA-SVR and GA-ANN. *J. Mol. Struct.* **2019**, *1195*, 95–109.

(52) Morris, G. M.; Goodsell, D. S.; Pique, M. E.; et al. User Guide, AutoDock, Automated Docking of Flexible Ligands to Flexible Receptors, AutoDock, AutoGrid, AutoDockTools, Copyright 1991–2009, Modification date: July 28, 2014.

(53) Venkatchalam, C. M.; Jiang, X.; Oldfield, T.; Waldman, M. LigandFit: a novel method for the shape-directed rapid docking of ligands to protein active sites. *J. Mol. Graphics Modell.* **2003**, *21*, 289–307.

- (54) Krammer, A.; Kirchhoff, P. D.; Jiang, X.; Venkatachalam, C. M.; Waldman, M. LigScore: a novel scoring function for predicting binding affinities. *J. Mol. Graphics Modell.* **2005**, *23*, 395–407.
- (55) Tirado-Rives, J.; Jorgensen, W. L. Contribution of conformer focusing to the uncertainty in predicting free energies for protein-ligand binding. *J. Med. Chem.* **2006**, *49*, 5880–5884.
- (56) Knapp, B.; Frantal, S.; Cibena, M.; Schreiner, W.; Bauer, P. Is an Intuitive Convergence Definition of Molecular Dynamics Simulations Solely Based on the Root Mean Square Deviation Possible? *J. Comput. Biol.* **2011**, *18*, 997–1005.
- (57) Maiorov, V. N.; Crippen, G. M. Significance of Root-Mean-Square Deviation in Comparing Three-dimensional Structures of Globular Proteins. *J. Mol. Biol.* **1994**, *235*, 625–634.
- (58) Coumar, M. S. *Molecular Docking for Computer-Aided Drug Design Fundamentals, Techniques, Resources and Applications*, 2021; Elsevier Inc. Academic Press
- (59) Safarizadeh, H.; Garkani-Nejad, Z. Molecular docking, molecular dynamics simulations and QSAR studies on some of 2-arylethenylquinoline derivatives for inhibition of Alzheimer's amyloid-beta aggregation: Insight into mechanism of interactions and parameters for design of new inhibitors. *J. Mol. Graphics Modell.* **2019**, *87*, 129–143.
- (60) Farmer, J.; Kanwal, F.; Nikulsin, N.; Tsilimigras, C.; Jacobs, B.; Jacobs, D. J. Statistical Measures to Quantify Similarity between Molecular Dynamics Simulation Trajectories. *Entropy* **2017**, *19*, 646.
- (61) Ertl, P.; Rohde, B.; Selzer, P. Fast Calculation of Molecular Polar Surface Area as a Sum of Fragment-Based Contributions and Its Application to the Prediction of Drug Transport Properties. *J. Med. Chem.* **2000**, *43*, 3714.
- (62) Klon, A. E.; Lowrie, J. F.; Diller, D. J. Improved Naive Bayesian Modeling of Numerical Data for Absorption, Distribution, Metabolism and Excretion (ADME) Property Prediction. *J. Chem. Inf. Model.* **2006**, *46*, 1945–1956.
- (63) Daina, A.; Zoete, V. A BOILED-Egg To Predict Gastrointestinal Absorption and Brain Penetration of Small Molecules. *ChemMedChem* **2016**, *11*, 1117–1121.
- (64) Bowers, K. J.; Chow, D. E.; Xu, H.; Dror, R. O.; Eastwood, M. P.; et al. Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters. *Proceedings of the ACM/IEEE Conference on Supercomputing (SC06), Tampa, Florida, 2006, Nov. 11-17, 2006*. DOI: 10.1145/1188455.1188544
- (65) Hussein, M. A.; Iqbal, M. A.; Umar, M. I.; Haque, R. A.; Guan, T. S. Synthesis, structural elucidation and cytotoxicity of new thiosemicarbazone derivatives. *Arabian J. Chem.* **2019**, *12*, 3183.
- (66) Gomha, S. M.; Abdelhady, H. A.; Hassain, D. Z. H.; Abdelmonsef, A. H.; El-Naggar, M.; Elaasser, M. M.; Mahmoud, H. K. Thiazole-Based Thiosemicarbazones: Synthesis, Cytotoxicity Evaluation and Molecular Docking Study. *Drug Des., Dev. Ther.* **2021**, *15*, 659–677. <https://www.dovepress.com/on 01-Jun-2022>
- (67) Schuchardt, K. L.; Didier, B. T.; Elsethagen, T.; Sun, L.; Gurumoorthi, V.; Chase, J.; Li, J.; Windus, T. L. Basis Set Exchange: A Community Database for Computational Sciences. *J. Chem. Inf. Model.* **2007**, *47*, 1045–1052.
- (68) Ahlrichs, R.; Armbruster, M. K.; Bachorz, R. A.; Bär, M.; Baron, H.-P.; Bauernschmitt, R.; Bischoff, F. A.; Böcker, S.; Burow, A. M.; Crawford, N.; et al.; *User's Manual for Turbomole. Program Package for Ab initio Electronic Structure Calculations*, Turbomole Version 7.1, May 17, 2016. <http://www.turbomole.com>.
- (69) Balasubramani, G. P.; Chen, S.; Coriani, M.; Diedenhofen, M. S.; Frank, Y. J.; Franzke, F.; Furche, R.; Grotjahn, M. E.; Harding, C.; Hättig, A.; et al. Turbomole: Modular program suite for ab initio quantum-chemical and condensed-matter simulations. *J. Chem. Phys.* **2020**, *152*, 184107.
- (70) Hadjipavlou-Litina, D.; Garnelis, T.; Athanassopoulos, C. M.; Papaioannou, D. Kukoamine A analogs with lipoxygenase inhibitory activity. *Enzyme Inhib. Med. Chem.* **2009**, *24*, 1188–1193.
- (71) Liargkova, T.; Eleftheriadis, N.; Dekker, F.; Voulgari, E.; Avgoustakis, C.; Sagnou, M.; Mavroidi, B.; Pelecanou, M.; Hadjipavlou-Litina, D. Small Multitarget Molecules Incorporating the Enone Moiety. *Molecules* **2019**, *24*, 199.