

Predicting And Modeling Of Anti-Aggregation Activities For Some Piperazinyl-Glutamate-Pyridine/Primidin Derivatives P2Y₁₂ Antagonists Through Multidimensional QSAR And Molecular Docking

Hiba Alarandas¹, M.H. Fatemi^{2*}, Zahra Pahlavan Yali³

^{1,2,3}Chemometrics Laboratory, Department of Chemistry, University of Mazandaran, Babolsar, IRAN.
EMAIL: hibaalarandas666@gmail.com¹, mhfatemi@umz.ac.ir², zpahlevanyali@yahoo.co³

Abstract:

In this work hybrid molecular docking quantitative structure activity relationship (QSAR) methodology is used to modeling and predict the inhibitory activities of some piperazinyl-glutamate-pyridine/primidin derivatives toward P2Y₁₂ protein. Data set consist of inhibitory activities (as IC₅₀ in μM^{-1}) of 52 piperazinyl-glutamate-pyridine/primidin derivatives, which can be used in treatment of thrombocytopenia. After docking of these derivative's to P2Y₁₂ protein, the most stable structure of ligands is chosen and frozen, to calculate molecular descriptors. In the next step prescreening of descriptors are done and stepwise feature selection methods was used to select the most relevant descriptors. Then the selected descriptors are used to developing multiple linear regression (MLR) and support vector machine (SVM) models. The statistical parameters of these model are; the outperformed SVM $r=0.84$, 0.87 ; RMSE= 0.42 , 0.82 for training and test sets, respectively, compared to $r=0.72$, 0.82 , RMSE= 0.72 , 0.77 for MLR). Comparison between these values of and other statistics reveals the superiority of SVM over MLR models. In the next step virtual screening based on the lead derivatives is outperformed to identify new efficient candidate based on ADME properties and docking studies.

Keywords: QSAR, Pyridine/Pyrimidine, P2Y₁₂ antagonist, GPCRs, Support vector regression, Molecular docking.

1. INTRODUCTION

A blood clot is a specific volume of blood that changes from a liquid state to a semi-solid or gel-like state. The main components of a blood clot are; dense platelets, red blood cells and a woven network of fibrin protein. Blood clot formation is a natural physiological response to wound healing, preventing excessive bleeding. Blood clots are harmless when stationary, but they can become dangerous when moving [1, 2]. When a clot breaks free, it may travel through veins to the heart and lungs. They can be trapped in the cardiopulmonary circulation which disturbs blood flow and creating acute medical emergencies when they obstruct healthy vasculature [3].

The P2Y₁₂ protein plays a crucial role in platelet aggregation, making it an ideal target for antithrombotic drugs [4, 5]. This surface platelet protein serves as a key regulator of coagulation. Within the central nervous system (CNS), P2Y₁₂ is predominantly found in microglia, where it mediates essential neuroprotective functions and physiological monitoring [6]. As a G protein-coupled receptor (GPCR) specific to platelets, the P2Y₁₂ receptor represents an attractive therapeutic target for the selective modulation of adenosine diphosphate (ADP)-induced platelet activation. Irreversible P2Y₁₂ antagonists, such as clopidogrel, prasugrel, ticagrelor and cangrelor, inhibit platelet ADP activation via the P2Y₁₂ receptor. This class of oral antiplatelet medication treats peripheral artery disease and acute coronary syndrome through selective, irreversible inhibition of the P2Y₁₂ receptor [7]. Following activation, clopidogrel binds irreversibly to platelets, resulting in a slow onset and offset of pharmacodynamics. This property reduces acute effectiveness and complicates management during bleeding events, trauma, or emergency surgeries. Direct-acting or P2Y₁₂ inhibitors may overcome these limitations while demonstrating improved efficacy [8].

P2Y₁₂ receptor inhibitors constitute a critical class of antithrombotic drugs that prevent platelet activation through receptor blockade. Among these, piperazinyl-glutamate-pyridine/pyrimidine derivatives show particular promise due to enhanced pharmacological properties stemming from three key components: The piperazinyl group improves solubility and bioavailability of drug consolidates [9]. The glutamate moiety facilitates hydrogen bonding and ionic interactions with key residues of targeted proteins (e.g., His187, Lys179). The pyridine/pyrimidine ring mediates π - π stacking with proteins and improves the hydrophobic interactions [10].

Many researches employs integrated computational approaches that combine molecular docking, molecular dynamics (MD) simulations, and quantitative structure-activity relationship (QSAR) modeling to elucidate antagonist binding modes to design new and efficient drugs [11]. QSAR methodology offers an efficient drug development approach that reduces time and costs, aiming to establish mathematical relationships between molecular structural features and biological activity. Numerous reports document successful QSAR modeling of drug candidate activity, for example drug –protein complexes [12]. A strong correlation between 3D contour maps and molecular docking results has identified several critical features of the binding mechanism. Molecular docking is a computational technique used to predict the preferred structure and orientation of one molecule, typically a small ligand or drug, when it binds to a target protein or receptor. This process helps in understanding the interactions between molecules for example drug-protein complexes and is widely used in drug discovery and design [13].

In this work some QSAR models are developed by using molecular docking-derived molecular descriptors to predict inhibitory activities of piperazinyl-glutamate-pyridine/pyrimidine derivatives as P2Y₁₂ inhibitors. The developed models are used to design of new efficient drug candidates [14].

2. MATERIAL AND METHODS

2.1. Data set

To date, some piperazinyl-glutamate-pyridine/pyrimidine derivatives have been collected as potent, orally bioavailable P2Y₁₂ antagonists for the inhibition of platelet aggregation, the present study utilizes a dataset consisting of 52 piperazinyl-glutamate-pyridine/pyrimidine derivatives, with their effective concentrations as P2Y₁₂ antagonists reported by Xu W. in 2006. The chemical structures of the dataset are shown in Table 1 and their half-maximal inhibitory concentrations (IC₅₀) are indicated, with values ranging from 0.1 to 8.1 μ M for compounds 1 and 42, respectively [15]. The IC₅₀ values, or (half-maximal inhibitory concentrations), reflect the potency of a substance to inhibiting a specific biological or biochemical function. In this study, the biological process being inhibited is antithrombotic activity (specifically platelet aggregation). The focus is on the inhibition of the P2Y₁₂ receptor's biological function [16].

To divide the dataset into training and test sets, all compounds were sorted according to their IC₅₀ values. The test set was then selected from this list based on a desired distance from each other. Using this procedure, 42 molecules were considered as the training set for model development, and 10 compounds were selected as the test set to evaluate the predictability of the developed models (see Table 1). In the next step, the chemical structures of all molecules were drawn using the HyperChem package (version 7.5) and optimized employing molecular mechanics and semi-empirical (AM1) methods. The optimized structures were then converted from *.hin format to *.pdb using the Open Babel program and further transformed to *.pdbqt format using the PYRX package (<https://pyrx.sourceforge.io/>; <https://sourceforge.net/projects/pyrx/>) for use as inputs in molecular docking studies [17].

2.2. Molecular docking

Molecular docking is a computational technique used to predict the preferred structure and orientation of one molecule, typically a small ligand or drug, when it binds to a target protein or receptor. This process helps to understanding the interactions between molecules and is widely used in drug discovery and design [18]. In this work the molecular docking computations were performed using Auto Dock 4.2 by using flexible ligand - rigid protein docking strategy [19]. Several X-ray crystal structures illustrating the binding of P2Y₁₂ with inhibitors are available in the protein data bank (PDB). For molecular docking, the X-ray crystal structure of PDK1 (PDB ID: 4NTJ) was retrieved from the Uniporter protein database (<http://www.uniprot.org>) and utilized. The Cartesian coordinates for the molecular docking boxes were set at 30 Å for each dimension (X, Y, and Z). Additionally, the coordinates for the box center were specified as -4.5, 60, and 109 for X, Y, and Z, respectively. At the end of calculation 10 different modes were generated, which their optimal conformation was selected for further analysis [20].

2.3. Descriptors calculation

In cheminformatics, a molecular descriptor is a quantitative measure derived from a systematic and standardized process that encodes various molecular structural information into a symbolic representation of a molecule. These descriptors are instrumental in conveying a wide array of molecular properties and are essential for developing robust QSAR models [21]. In hybrid molecular docking QSAR, the molecular descriptors are calculated from the optimal conformational structures of interested compounds of interest after docking them to the target protein (see Figure. 1).

In this work, the frozen structures derived from docking studies were then subjected to descriptor calculation [22]. To achieve this, the optimized docking structures of piperazinyl-glutamate-pyridine/pyrimidine derivatives were converted from *.pdb to *.hin format. The Dragon 7.0 software (<https://vcclab.org/lab/edragon/>) was employed in conjunction with the ChemDes program (<http://www.scbdd.com/chemdes/>) and PaDEL to calculate a pool of descriptors based on the optimal three-dimensional structures of organic compounds obtained from docking studies. A comprehensive set of descriptors was generated, encompassing 1,440 descriptors by Dragon, 910 descriptors by ChemDes, and 510 descriptors from the PaDEL descriptor tool. These descriptors span a broad spectrum of molecular characteristics, providing an extensive dataset for subsequent modelling efforts [23].

2.4. Descriptors selection

Prior to variable selection, calculated descriptors were meticulously pre-processed to ensure quality and reliability. This step involved the elimination of constant variables, near-constant variables, and descriptors with zero values across all samples. Furthermore, to prevent multicollinearity issues, descriptors with Pearson correlation coefficients exceeding 0.90 were also removed. The remaining 504 descriptors were used in the variable selection step [24].

The descriptors selection step is very important step in QSAR model development with the purpose of introducing a concise, interpretable, and accurate prediction model [25-27]. The dominant paradigm of thought estimation of relationships between variables of molecular structures $X(x_1, x_2, \dots, x_n)$ and desired biological activity (Y) and shaped a set of patterns for a distinct data-matrix set [28-29]. Stepwise, forward and backward multiple linear regression (MLR) analysis can be applied for selection of relevant variables $X(x_1, x_2, \dots, x_m)$ [30]. The analysis of variance (ANOVA) using F-test and Pearson correlation for checking the statistically significant correlation between matrixes of selected molecular descriptors and the value of IC_{50} is confirmed the validity of the model input predictors. It should be mentioned that Figure 2, also, the “optimal number of descriptors” was evaluated by the correlation coefficient (r) and standard error (SE) relative to descriptor in a breakpoint procedure.

2.5. Models development

In order to developing QSAR models selected molecular descriptors and IC_{50} are considered as independent and dependent variables, respectively. Equations for predicting IC_{50} values were derived from the training data using MLR and subsequently validated using the test data. Multiple linear regression (MLR) analysis applied for the 3D-QSAR paradigm of thought predict the linear relationships between relevant selected descriptors of molecular structures $X(x_1, x_2, \dots, x_m)$ and desired biological activity (Y). Also, support vector regression (SVR) algorithm employed to nonlinearly map the data into a feature (selected descriptor) space. The performance of regression processes presented and compared with analysis of correlation and error parametric statistics to show the excellent of modeling process. All statistical inferences are then made based on interpretable developed MLR and SVR models [31-32]; parametric statistics is a set of statistical methods that use a parametric model for a probabilistic phenomenon. The correlation coefficient (r) was subjected to statistical analysis to evaluate the performance of the goodness-of-fit for the model. Additionally, the standard error (SE) and mean-square error (MSE) were computed for each model [33]. These statically parameters are calculated from the following equations

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

$$SE = \sqrt{\frac{1}{DF} \sum_{i=1}^n ((y_i - \hat{y}_i)^2)} \quad (\text{eq. 2})$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n ((y_i - \hat{y}_i)^2)} \quad (\text{eq. 3})$$

here y_i is the actual value, $\hat{y}_i$ is the predicted value, and $\bar{y}$ is the average value of IC_{50} . The recommended values of these statistics to ensure the reliability of predictions from the QSAR model are; a correlation coefficient ($R \geq 0.8$), along with a coefficient of determination ($R^2 \geq 0.6$) for in vivo data.

Support vector machine can used to investigate the relation between dependent variable (end point) with independent variable (molecular descriptor) in QSAR modeling. The performance of SVM in some case is better than MLR due to choice of kernel function and their nonlinear capabilities [34-35]. The term “kernel” is referred to a set of mathematical functions used in support vector machines that allow interaction with the data. The kernel function generally transforms the training data set in such a way that a nonlinear decision surface becomes a linear equation in a higher-dimensional space. In simpler

terms, this function returns the result of the inner product between two points in a standard feature space [36].

3. RESULTS AND DISCUSSION

3.1 Docking results

To investigate that how piperazinyl-glutamate-pyridine / primidine derivatives bind to the P2Y₁₂ protein's active site docking studies are performed (Figure. 3). Obtained results indicate that hydrogen bond interactions, both as acceptors and donors, emerged as the most lucrative ligand-protein interactions [37]. Following the docking analysis, compounds 12, 22, and 41 were identified as the top-performing compounds with binding constant values of -7.7, -7.4, and -7.4 μM^{-1} , respectively that indicate, their potential as active compounds. Interestingly, the representative compound 41 exhibits the highest potency as anticancer agent against platelet Lys179 cells with the lowest IC₅₀ values, that indicating its potential to use for treatment of chemo resistant platelet. This compound can acts as a hydrogen bond acceptor from Lys179 and forms two interactions with His187and Lys179 as a hydrogen bond donor. Conversely, compound 9, was categorized as inactive compound; based on both docking results and the QSAR models results.

3.2. QSAR models

The dataset, consisting of 52 compounds, was divided into a training set of 42 compounds (80%) for model construction and a test set of 10 compounds (20%) for model validation, utilizing the Kennard-Stone algorithm. Initially, stepwise multiple linear regression (SW-MLR) was employed to develop the quantitative structure-activity relationships model using a distinct set of descriptors. The resulting MLR QSAR model, which exhibits predictive capability for the training set, is as follows:

$$\text{IC}_{50} = 321.09 - 82.11 * \text{bcutm8} + 135.76 * \text{bcutv13} - 1.06 * \text{PEOEVS A1} - 47.06 * \text{bcutp4} + 0.08 * \text{EstateVSA8} - 76.71 * \text{ATSe7} + 5.73 * \text{Smin6} + 33.11 * \text{Smax35} - 3.49 * \text{PEOEVS A9} \quad (\text{eq. 4})$$

$$n_{\text{train}} = 42; r_{\text{train}} = 0.84; \text{RMSE}_{\text{train}} = 0.42; n_{\text{test}} = 10; r_{\text{test}} = 0.87; \text{RMSE}_{\text{test}} = 0.82$$

here, n_{train} is the number of compounds in the training set, r_{train} is the correlation coefficient for the training set, $\text{RMSE}_{\text{train}}$ is the root mean square error for the training set, n_{test} is the number of compounds in the test set, r_{test} is the correlation coefficient for the test set, $\text{RMSE}_{\text{test}}$ is the root mean square error for the test set. The details of statistical parameters of equation (4) are indicated in Table 2 In order to improve QSAR model, support vector regression method is used to correlate the variation of independent variables (selected molecular descriptors) to their respected dependent variable IC₅₀. The details of statistical parameters of equation (4) are indicated in Table 3.

The SVM model was implemented using STATISTICA software (Version 14.5.0.12). The values of SVM parameters (including type of kernel function γ , ϵ , and c) are optimized by continuous changing of them and monitoring the error of model for training and test sets to minimize SVM error. The optimized SVMs parameters are, kernel functions RBF, $C=100$, $\gamma=0.700$, and $\epsilon=0.900$. Then the optimized SVM is used to estimate the values of IC₅₀ for training and test sets. The experimental and SVM predicted IC₅₀ values and their corresponding residuals are shown in Table 3. The outperformed SVM ($r=0.84$, 0.87 ; $\text{RMSE}=0.42$, 0.82 for training and test sets, respectively, compared to $r=0.72$, 0.82 , $\text{RMSE}=0.72$, 0.77 for MLR). Important statistical parameters for both SVM and MLR are shown in Table 4, which can be used to compare the performance of these models. Compression between these parameters and those indicated in Table 4, reveals the superiority of SVM over MLR model. The SVM calculated values of IC₅₀ for both training and test sets are plotted against their experimental values in Figure. 4(a), which reveals good correlation between them. Also, their residuals are platted in Figure. 4(b), random distribution of residuals around the zero line indicate that there is no any biases in developed SVM model.

3.3. Interpretation of descriptors

In this study, stepwise multiple linear regression (SW-MLR) method identified ninth descriptors (Table 5) crucial for predicting IC₅₀ which offering insights into potential drug discovery for platelet aggregation. The first descriptor, is bcutm8 which is the 8th highest eigenvalue of a mass-weighted Burden matrix 1. This chemo-informatics descriptors can used to capture molecular structural features of molecules, which has the positive effects on P2Y₁₂ inhibition actives of studied chemicals [37].

The second descriptor, is bcutv13 (mean Burden descriptors based on atomic volumes), which is the 13th highest eigenvalue of a modified Burden matrix weighted by atomic van der Waals volumes [38]. The next descriptors is PEOEVSA1 (mean MOE-type descriptors using Estate indices and surface area

contributions) which is a valuable descriptor in computational chemistry and combines the benefits of E-State indices with surface area considerations, providing a comprehensive view of molecular properties [39]. The fourth descriptor is bcutp4 (Highest eigenvalue.4 of Burden matrix/weighted by atomic polarizabilities) optimization [40]. The next descriptor is EstateVSA8 (mean MOE-type descriptors using Estate indices and surface area contributions) which is a molecular descriptor used in computational chemistry and cheminformatics to characterize molecular properties based on van der Waals surface area (VdWSA) contributions from different atom types or pharmacophore features [41]. It is part of the VolSurf and Estate (Electro topological State) descriptor families, combining surface area calculations with atom-type classifications. EstateVSA8 can be used to filter molecules based on desirable surface properties.

The sixth descriptor is ATSe7 (Broto-Moreau type descriptor). Which is autocorrelation of a topological structure-lag7/weighted by atomic Sanderson electro negativity [42]. The subsequent descriptors, Smin6 and Smax35, represent the minimum and maximum E-state values for specific atom types, accounting for 3D structural influences on molecular binding interactions [43]. The final descriptors are PEOEVSA9, which can represent the polar surface area of molecules. These descriptors essential for understanding how the spatial arrangement of atoms influences the molecular properties and behaviors and is a powerful tool for quantifying the effect of different steric and electronic interaction between drugs and proteins [44].

3.4. Applicability domain analysis

The applicability domain (AD) of a QSAR model is critical for validating the model's predictions and ensuring their reliability. To define the AD, a Williams plot is employed, which displays standardized residuals versus leverage values (h). This visualization aids in identifying outliers and influential compounds and can providing insights into the robustness of the model. The leverage equation is calculated by: $h_i = x_i (X^T X)^{-1} x_i^T$, where x_i represents the descriptor vector for interested compound and X is the descriptor matrix derived from the training set. The warning leverage value (h^*) calculated is as follow:

$$h^* = 3(d + 1)/n \quad (\text{eq. 5})$$

In this equation, d is the number of predictor variables, and n is the number of compounds in the training set. According to the above explanation Williams plots for both the MLR and SVM models were generated using a warning leverage value of $h^* = 0.45$ that are shown in Figure 5. As can be seen in these figures, the majority of compounds fall within the applicability domain.

3.5. ADMET analysis

A virtual screening procedure was applied to a large commercial chemical database, resulting in 17 hits. These hits were further screened using the QSAR model for P2Y12 inhibitory activity prediction, resulting in hits that were subsequently evaluated for their Absorption, Distribution, Metabolism, and Excretion (ADME) properties. The pharmacokinetic parameters for the five identified hits were determined to fall within the acceptable range intended for human use which are shown in Table 6 and Figure 6. Highlighted by bold chemicals in are hits new candidate which have potential to consider as new drugs according to their pharmacokinetic and ADME results.

4. CONCLUSION

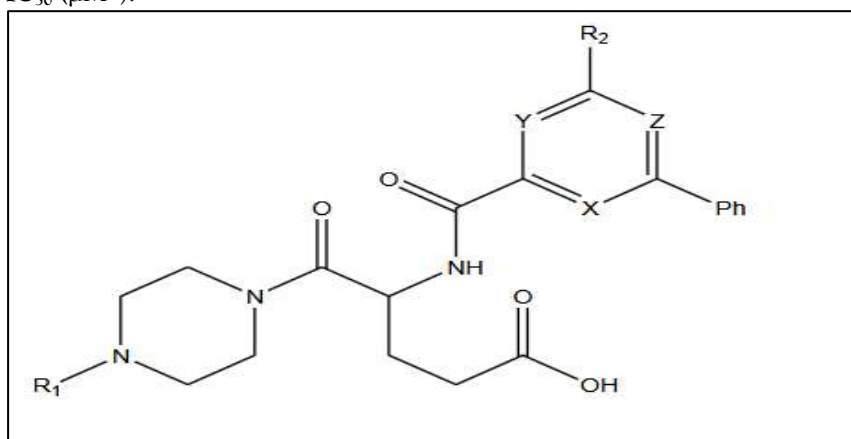
In this study some SVM and MLR models are developed based on molecular descriptors that are calculated from docking derived structures of interested piperazinyl-glutamate-pyridine/primidin derivative's as P2Y12 protein inhibitors. Analyzing of docking data and selected molecular descriptors indicate that steric and electronic interaction together with H-bond donor/acceptor ability of drugs candidate play important role on inhibitory activities (as IC50) of studied piperazinyl-glutamate-pyridine/primidin derivatives. The models' predictive ability and robustness were assessed using various statistical parameters, such as RMSE and r. for training the results indicate that the SVM method, when coupled with appropriate descriptors, can effectively predict the activity of new derivatives in the treatment of platelets. The visualization of the QSAR model and the docking mode into the target protein provided insights into the structure-activity relationship, offering explicit indications for designing improved piperazinyl-glutamate-pyridine/primidin derivatives. Additionally, the outcomes of this study provide valuable insights into the development of novel and potent P2Y12 inhibitors, holding promise for the creation of new drugs for type 2 platelets.

REFERENCES

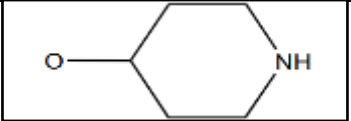
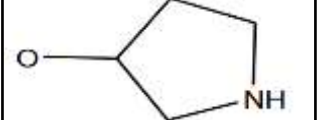
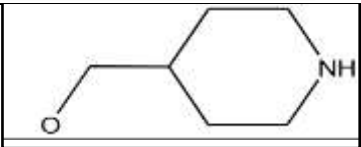
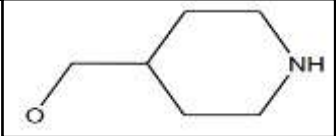
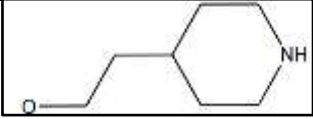
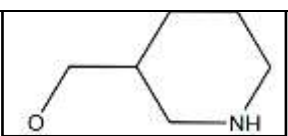
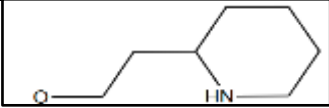
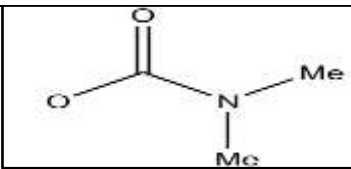
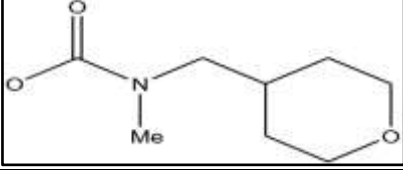
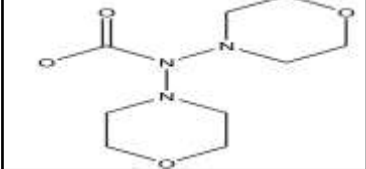
- Kennedy, C. (2021). The P2Y/P2X divide: How it began. *Biochemical Pharmacology*, 187, 114408. <https://doi.org/10.1016/j.bcp.2021.114408>
- Müller, C. E., & Jacobson, K. A. (2011). Recent developments in adenosine receptor ligands and their potential as novel drugs. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1808(5), 1290–1308. <https://doi.org/10.1016/j.bbamem.2010.12.017>
- Müller, C. E., Baqi, Y., & Namasivayam, V. (2020). Agonists and antagonists for purinergic receptors. *Methods in Molecular Biology*, 2041, 45–64. https://doi.org/10.1007/978-1-4939-9717-6_4
- Thimm, D., Schiedel, A. C., Peti-Peterdi, J., Kishore, B. K., & Müller, C. E. (2015). The nucleobase adenine as a signalling molecule in the kidney. *Acta Physiologica*, 213(4), 808–818. <https://doi.org/10.1111/apha.12451>
- Zimmermann, H. (2021). Ectonucleoside triphosphate diphosphate hydrolases and ecto-5'-nucleotidase in purinergic signaling: How the field developed and where we are now. *Purinergic Signalling*, 17(1), 117–125. <https://doi.org/10.1007/s11302-020-09745-8>
- Illes, P., Müller, C. E., Jacobson, K. A., Grutter, T., Nicke, A., Fountain, S. J., Kennedy, C., Schmalzing, G., Jarvis, M. F., Stojilkovic, S. S., King, B. F., & Di Virgilio, F. (2021). Update of P2X receptor properties and their pharmacology: IUPHAR Review 30. *British Journal of Pharmacology*, 178(3), 489–514. <https://doi.org/10.1111/bph.15299>
- Jacobson, K. A., & Müller, C. E. (2016). Medicinal chemistry of adenosine, P2Y and P2X receptors. *Neuropharmacology*, 104, 31–49. <https://doi.org/10.1016/j.neuropharm.2015.12.001>
- Lambertucci, C., Dal Ben, D., Buccioni, M., Marucci, G., Thomas, A., & Volpini, R. (2015). Medicinal chemistry of P2X receptors: Agonists and orthosteric antagonists. *Current Medicinal Chemistry*, 22(8), 915–928. <https://doi.org/10.2174/0929867322666150114163438>
- Müller, C. E. (2015). Medicinal chemistry of P2X receptors: Allosteric modulators. *Current Medicinal Chemistry*, 22(8), 929–941. <https://doi.org/10.2174/0929867322666150114163447>
- Jacobson, K. A., Iljerman, A. P., & Müller, C. E. (2021). Medicinal chemistry of P2 and adenosine receptors: Common scaffolds adapted for multiple targets. *Biochemical Pharmacology*, 187, 114311. <https://doi.org/10.1016/j.bcp.2021.114311>
- Soto, F., Lambrecht, G., Nickel, P., Stühmer, W., & Busch, A. E. (1999). Antagonistic properties of the suramin analogue NF023 at heterologously expressed P2X receptors. *Neuropharmacology*, 38(1), 141–149. [https://doi.org/10.1016/S0028-3908\(98\)00159-5](https://doi.org/10.1016/S0028-3908(98)00159-5)
- Tian, M., Abdelrahman, A., Baqi, Y., Fuentes, E., Azazna, D., Spanier, C., Densborn, S., Hinz, S., Schmid, R., & Müller, C. E. (2020). Discovery and structure relationships of salicylanilide derivatives as potent, non-acidic P2X1 receptor antagonists. *Journal of Medicinal Chemistry*, 63(11), 6164–6178. <https://doi.org/10.1021/acs.jmedchem.0c00396>
- Wolf, C., Rosefort, C., Fallah, G., Kassack, M. U., Hamacher, A., Bodnar, M., Wang, H., Illes, P., Kless, A., Bahrenberg, G., Schmalzing, G., & Hausmann, R. (2011). Molecular determinants of potent P2X2 antagonism identified by functional analysis, mutagenesis, and homology docking. *Molecular Pharmacology*, 79(4), 649–661. <https://doi.org/10.1124/mol.110.069153>
- Baqi, Y., Hausmann, R., Rosefort, C., Rettinger, J., Schmalzing, G., & Müller, C. E. (2011). Discovery of potent competitive antagonists and positive modulators of the P2X2 receptor. *Journal of Medicinal Chemistry*, 54(3), 817–830. <https://doi.org/10.1021/jm1012193>
- Schneider, R., Leven, P., Glowka, T., Kuzmanov, I., Lysson, M., Schneiker, B., Miesen, A., Baqi, Y., Spanier, C., Grants, I., Mazzotta, E., Villalobos-Hernandez, E., Kalff, J. C., Müller, C. E., Christofi, F. L., & Wehner, S. (2021). A novel P2X2-dependent purinergic mechanism of enteric gliosis in intestinal inflammation. *EMBO Molecular Medicine*, 13(3), e12724. <https://doi.org/10.15252/emmm.202012724>
- Lee, H. J., Joung, S. K., Kim, Y. G., Yoo, J. Y., & Han, S. B. (2004). Bioequivalence assessment of ambroxol tablet after a single oral dose administration to healthy male volunteers. *Pharmacological Research*, 49(1), 93–98. <https://doi.org/10.1016/j.phrs.2003.07.009>
- Bouscary, A., Quessada, C., René, F., Spedding, M., Henriques, A., Ngo, S., & Loeffler, J. P. (2020). Drug repositioning in neurodegeneration: An overview of the use of ambroxol in neurodegenerative diseases. *European Journal of Pharmacology*, 884, 173446. <https://doi.org/10.1016/j.ejphar.2020.173446>
- Malerba, M., & Ragnoli, B. (2008). Ambroxol in the 21st century: Pharmacological and clinical update. *Expert Opinion on Drug Metabolism & Toxicology*, 4(8), 1119–1129. <https://doi.org/10.1517/17425255.4.8.1119>
- Krajewski, J. L. (2020). P2X3-containing receptors as targets for the treatment of chronic pain. *Neurotherapeutics*, 17(3), 826–838. <https://doi.org/10.1007/s13311-020-00894-7>
- Ford, A. P. (2012). In pursuit of P2X3 antagonists: Novel therapeutics for chronic pain and afferent sensitization. *Purinergic Signalling*, 8(1), 3–26. <https://doi.org/10.1007/s11302-011-9271-6>
- Marucci, G., Dal Ben, D., Buccioni, M., Marti Navia, A., Spinaci, A., Volpini, R., & Lambertucci, C. (2019). Update on novel purinergic P2X3 and P2X2/3 receptor antagonists and their potential therapeutic applications. *Expert Opinion on Therapeutic Patents*, 29(12), 943–963. <https://doi.org/10.1080/13543776.2019.1693542>
- Müller, C. E. (2010). Emerging structures and ligands for P2X3 and P2X4 receptors—Towards novel treatments of neuropathic pain. *Purinergic Signalling*, 6(2), 145–148. <https://doi.org/10.1007/s11302-010-9181-z>
- Morice, A. H., Birring, S. S., Smith, J. A., McGarvey, L. P., Schelfhout, J., Martin Nguyen, A., Xu, Z. J., Wu, W. C., Muccino, D. R., & Sher, M. R. (2021). Characterization of patients with refractory or unexplained chronic cough participating in a phase 2 clinical trial of the P2X3-receptor antagonist gefapixant. *Lung*, 199(2), 121–129. <https://doi.org/10.1007/s00408-021-00429-7>
- Dicpinigaitis, P. V. (2021). Coming soon: The first-ever drug(s) for refractory chronic cough. *Lung*, 199(2), 83–84. <https://doi.org/10.1007/s00408-021-00430-0>
- Bahia, M. S., et al. (2023). A comparison between 2D and 3D descriptors in QSAR modeling based on bioactive conformations. *Molecular Informatics*, 42(4), 2200186. <https://doi.org/10.1002/minf.202200186>
- Grisoni, F., Consonni, V., & Todeschini, R. (2018). Impact of molecular descriptors on computational models. In *Computational Chemogenomics* (pp. 171–209). Springer.

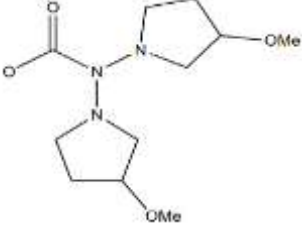
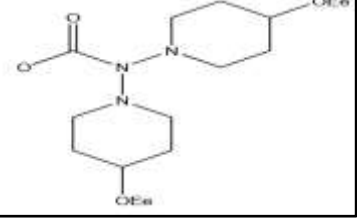
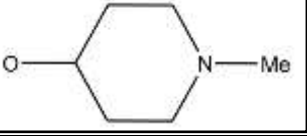
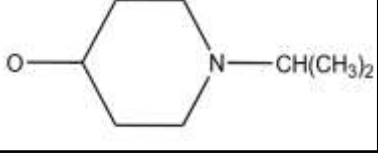
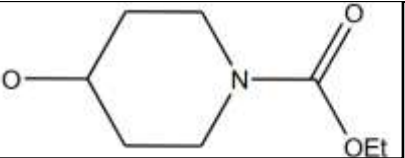
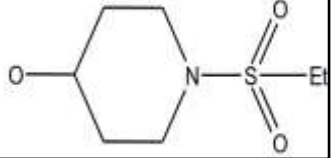
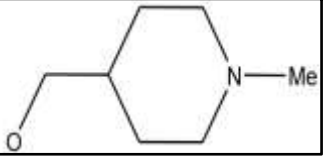
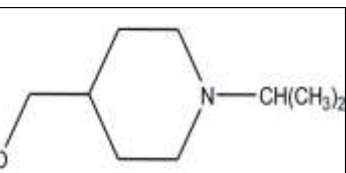
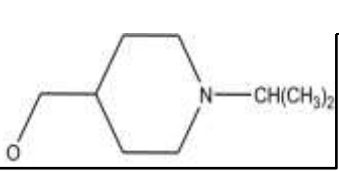
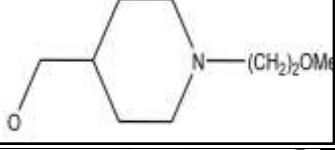
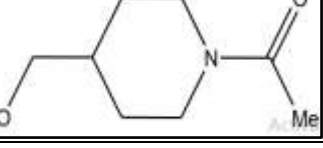
31. Raghunathan, S., & Priyakumar, U. D. (2022). Molecular representations for machine learning applications in chemistry. *International Journal of Quantum Chemistry*, 122(7), e26870. <https://doi.org/10.1002/qua.26870>
32. Kleandrova, V. V., & Speck-Planche, A. (2020). The QSAR paradigm in fragment-based drug discovery: From the virtual generation of target inhibitors to multi-scale modeling. *Mini-Reviews in Medicinal Chemistry*, 20*(14), 1357–1374. <https://doi.org/10.2174/1389557520666200429102336>
33. Wang, T., et al. (2017). The advancement of multidimensional QSAR for novel drug discovery—Where are we headed? *Expert Opinion on Drug Discovery*, 12(8), 769–784. <https://doi.org/10.1080/17460441.2017.1336156>
34. Shahlaei, M. (2013). Descriptor selection methods in quantitative structure–activity relationship studies: A review study. *Chemical Reviews*, 113(10), 8093–8103. <https://doi.org/10.1021/cr3004339>
35. Bolboaca, S. D., & Jäntschi, L. (2013). Quantitative structure-activity relationships: Linear regression modelling and validation strategies by example. *Biomath*, 2(1), 1309089. <https://doi.org/10.11145/j.biomath.2013.09.089>
36. Brereton, R. G., & Lloyd, G. R. (2010). Support vector machines for classification and regression. *Analyst*, 135(2), 230–267. <https://doi.org/10.1039/B918972F>
37. Sebaugh, J. (2011). Guidelines for accurate EC50/IC50 estimation. *Pharmaceutical Statistics*, 10(2), 128–134. <https://doi.org/10.1002/pst.426>
38. Mei, H., et al. (2005). Support vector machine applied in QSAR modelling. *Chinese Science Bulletin*, 50(20), 2291–2296. <https://doi.org/10.1007/BF03183741>
39. Koo, J., et al. (2018). Quantitative structure-activity relationship (QSAR) study on P2Y12 receptor antagonists. *European Journal of Medicinal Chemistry*, 143, 1190–1200. <https://doi.org/10.1016/j.ejmech.2017.11.036>
40. Burden, F. R., & Winkler, D. A. (1996). The Burden matrix: A new approach to the analysis of molecular structure. *Journal of Chemical Information and Computer Sciences*, 36(6), 1255–1260. <https://doi.org/10.1021/ci960080g>
41. Todeschini, R., & Consonni, V. (2000). *Molecular descriptors for chemoinformatics*. Wiley-VCH.
42. Roy, K., & Kar, S. (2013). A comprehensive review on molecular descriptors. *Molecular Informatics*, 32(1), 20. <https://doi.org/10.1002/minf.201200053>
43. Consonni, V., & Todeschini, R. (2008). The VolSurf approach to the characterization of molecular properties. *Molecular Informatics*, 27(8), 1095–1112. <https://doi.org/10.1002/minf.200800079>
44. Hall, L. H., & Kier, L. B. (1995). Electrotopological state indices for atom types: A novel approach to the characterization of molecular structure. *Journal of Chemical Information and Computer Sciences*, 35(4), 682–693. <https://doi.org/10.1021/ci00001a012>
46. Gasteiger, J., & Marsili, M. (1980). A new model for calculating atomic charges. *Tetrahedron Letters*, 21(16), 1525–1528. [https://doi.org/10.1016/S0040-4039\(00\)82290-7](https://doi.org/10.1016/S0040-4039(00)82290-7)
47. Wang, Y., et al. (2015). Understanding the role of polar surface area in drug discovery. *European Journal of Medicinal Chemistry*, 90, 371–381. <https://doi.org/10.1016/j.ejmech.2014.12.018>
48. Bolboaca, S. D., & Jäntschi, L. (2013). Quantitative structure-activity relationships: Linear regression modelling and validation strategies by example. *Biomath*, 2(1), Article 1309089. <https://doi.org/10.11145/j.biomath.2013.09.089>
49. Bolboaca, S. D., & Jäntschi, L. (2013). Quantitative structure-activity relationships: Linear regression modelling and validation strategies by example. *Biomath*, 2(1), Article 1309089. <https://doi.org/10.11145/j.biomath.2013.09.089>

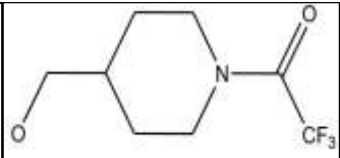
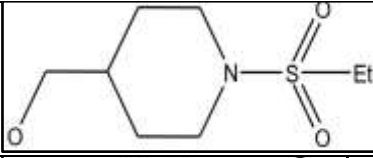
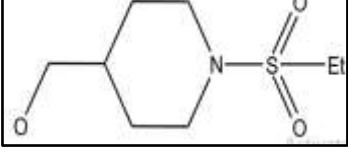
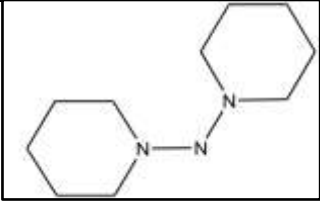
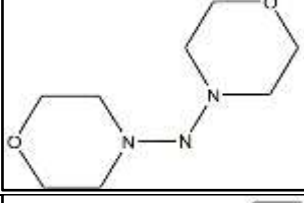
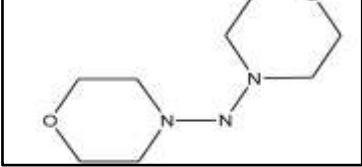
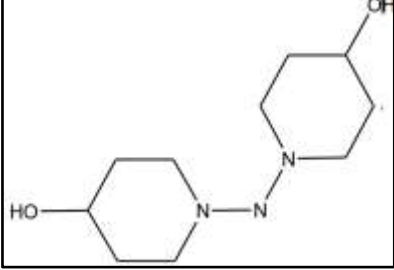
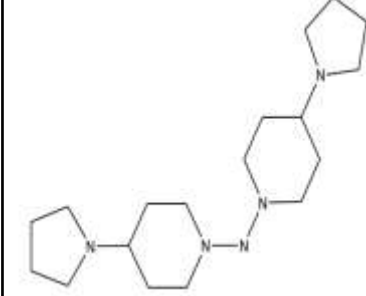
Table1. Structural of studied piperazinyl-glutamate-pyridine / pyrimidine derivatives and their experimental IC₅₀ (μM⁻¹).



No.	R ₁	X	Y	Z	R ₂	IC ₅₀ (μM ⁻¹)
1	EtOCO	N	N	CH	H	3.90
2	HeptOCO	N	CH	CH	H	0.78
3	CH ₂ ° Pent	N	CH	CH	H	0.91
4	(CH ₂) ₂ CH(CH ₃) ₂	N	CH	CH	H	0.71
5	EtOCO	N	CH	CH	OH	1.30
6	BuOCO	N	CH	CH	OH	0.95

7	PentOCO	N	CH	CH	OH	0.85
8	EtOCO	N	CH	CH	OMe	1.00
9	BuOCO	N	CH	CH	OBu	2.40
10	EtOCO	N	CH	CH	O(CH ₂) ₂ OH	2.20
11	EtOCO	N	CH	CH	O(CH ₂) ₃ OH	0.80
12	EtOCO	N	CH	CH	O(CH ₂) ₂ OMe	0.74
13	BuOCO	N	CH	CH	O(CH ₂) ₂ OMe	2.70
14	EtOCO	N	CH	CH	O(CH ₂) ₃ OMe	0.86
15	PentOCO	N	CH	CH		0.39
16	BuOCO	N	CH	CH		0.75
17	BuOCO	N	CH	CH		6.80
18	BuOCO	N	CH	CH		1.10
19	PentOCO	N	CH	CH		2.40
20	PentOCO	N	CH	CH		1.60
21	BuOCO	N	CH	CH		1.20
22	BuOCO	N	CH	CH		0.65
23	PentOCO	N	CH	CH		0.62
24	PentOCO	N	CH	CH		0.75
25	PentOCO	N	CH	CH		0.63

26	PentOCO	N	CH	CH		0.60
27	PentOCO	N	CH	CH		2.00
28	PentOCO	N	CH	CH		0.43
29	PentOCO	N	CH	CH		0.52
30	PentOCO	N	CH	CH		0.82
31	PentOCO	N	CH	CH		0.41
32	PentOCO	N	CH	CH		0.36
33	BuOCO	N	CH	CH		0.40
34	PentOCO	N	CH	CH		2.00
35	PentOCO	N	CH	CH		2.40
36	BuOCO	N	CH	CH		4.40

37	PentOCO	N	CH	CH		5.20
38	BuOCO	N	CH	CH		11.00
39	PentOCO	N	CH	CH		1.90
40	EtOCO	N	CH	CH	NHMe	2.30
41	EtOCO	N	CH	CH	NHPr	5.00
42	PentOCO	N	CH	CH	NH(CH ₂) ₂ OH	1.80
43	EtOCO	N	CH	CH		1.10
44	EtOCO	N	CH	CH		7.10
45	PentOCO	N	CH	CH		0.66
46	PentOCO	N	CH	CH		0.33
47	EtOCO	N	CH	CH		29.00

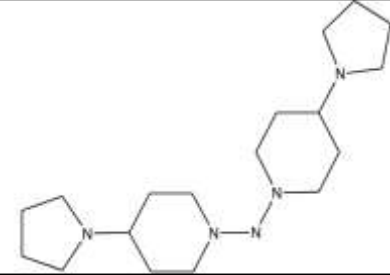
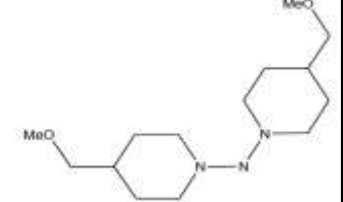
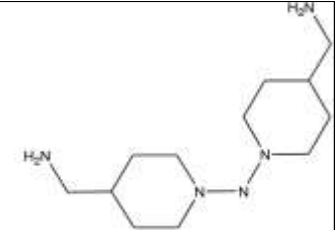
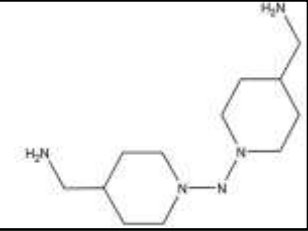
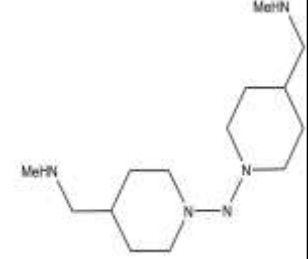
48	BuOCO	N	CH	CH		8.60
49	EtOCO	N	CH	CH		9.90
50	EtOCO	N	CH	CH		2.70
51	EtOCO	N	CH	CH		1.40
52	BuOCO	N	CH	CH		1.30

Table 2. Statistical parameters of developed MLR model.

Variable	Unstandardized Coefficients		t	Sig.
	B	Std. Error		
(Constant)	321.09	117.84	2.72	0.01
bcutm8	82.11	28.18	2.91	0.00
bcutv13	-135.76	60.37	-2.24	0.03
PEOEVSA1	1.06	0.44	2.36	0.02
bcutp4	-47.06	44.50	-1.05	0.00
EstateVSA8	0.08	0.02	2.92	0.00
ATSe7	-76.71	40.64	-1.88	0.07
Smin6	5.73	4.24	1.35	0.01
Smax35	33.11	11.76	2.81	0.01
PEOEVSA9	-3.49	0.73	-4.75	0.00

Table 3. The experimental and SVM predicted IC₅₀ values and their residuals.

No.	IC ₅₀ exp	IC ₅₀ pre	residual
1	3.90	5.80	-1.90
2	0.78	0.60	0.18
3	0.91	0.63	0.28
4	0.71	1.11	-0.40
5	1.30	1.56	-0.26
6	0.95	1.56	-0.61
7	0.85	0.76	0.09
8	1.00	1.63	-0.63
9	2.40	2.22	0.18
10	2.20	1.44	0.76
11	0.80	3.53	-2.73
12	0.74	-0.28	1.02
13	2.70	2.23	0.47
14	0.86	1.98	-1.12
15	0.39	1.35	-0.96
16	0.75	-0.17	0.92
17	6.80	2.96	3.84
18	1.10	4.84	-3.74
19	2.40	2.33	0.07
20	1.60	2.23	-0.63
21	1.20	1.78	-0.58
22	0.65	2.07	-1.42
23	0.62	1.19	-0.57
24	0.75	1.24	-0.49
25	0.63	1.69	-1.06
26	0.60	2.28	-1.68
27	2.00	2.37	-0.37
28	0.43	0.70	-0.27
29	0.52	1.20	-0.68
30	0.82	4.28	-3.46
31	0.41	7.54	-7.13
32	0.36	3.50	-3.14
33	0.40	0.52	-0.12
34	2.00	1.49	0.51
35	2.40	-0.61	3.01
36	4.40	-0.01	4.41
37	5.20	1.31	3.89
38	11.00	1.01	9.99
39	1.90	0.92	0.98
40	2.30	1.34	0.96
41	5.00	13.00	-8.00
42	1.80	2.28	-0.48
43	1.10	2.22	-1.12

44	7.10	1.42	5.68
45	0.66	0.94	-0.28
46	0.33	0.93	-0.60
47	29.00	5.84	23.16
48	8.60	0.66	7.94
49	9.90	0.63	9.27
50	2.70	1.11	1.59
51	1.40	1.56	-0.16
52	1.30	1.56	-0.26

Table 4. The Statistical parameters of SVM and MLR models.

Parameter	Set	MLR	SVM
r	Training	0.72	0.84
r	Test	0.82	0.87
RMES	Training	0.42	0.82
RMES	Test	0.72	0.77
SE	Training	1.44	0.44
SE	Test	1.33	0.79

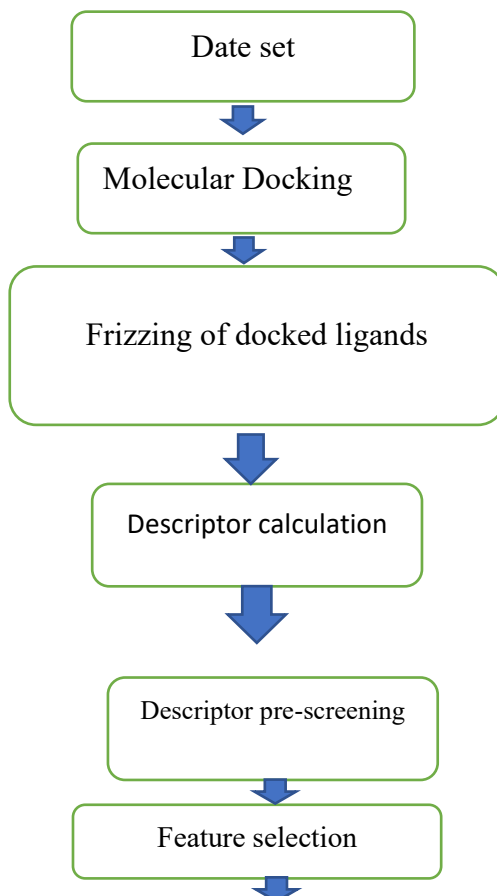
Table 5: A summary of the molecular descriptors utilized in model construction.

No	Symbol	Class	Meaning	Role in Model
1	bcutm8	Burden descriptors	Lowest eigenvaluen.8 of Burden matrix/weighted by atomic masses	Negative coefficient → Higher values reduce IC50 (improve potency).
2	bcutv13	Burden descriptors	Lowest eigenvaluen.13 of Burden matrix/weighted by atomic vander Waals volumes	Positive coefficient → May relate to bulky groups favoring activity.
3	PEOEVSA1	Partial charge descriptor	MOE-type descriptors using partial charges and surface area contributions	Negative impact → Polar interactions may hinder binding.
4	bcutp4	Burden descriptors	Highest eigenvaluen.4 of Burden matrix/weighted by atomic polarizabilities	Negative coefficient → Suggests specific steric/electronic features boost potency.
5	EstateVSA8	Electro topological	. MOE-type descriptors using Estate indices and surface area contributions	Minimal positive effect.
6	ATSe7	Atom-type E-state	Broto-Moreau autocorrelation of a topological structure-lag7/weighted by atomic Sanderson electro negativities	Strong negative coefficient → Critical for activity (e.g., H-bond acceptors).
7	Smin6	Spatial minima/maxima	Minimum of E-State value of specified atom type	Adjusts 3D shape effects on binding.
8	Smax35	Spatial minima/maxima	Maximum of E-State value of specified atom type	Adjusts 3D shape effects on binding.

9	PEOEVS9	Partial charge descriptor	MOE-type descriptors using partial charges and surface area contributions	Polar surface area may reduce potency.
---	---------	---------------------------	---	--

pharmacokinetic parameters of identified hits (Heavy atoms, Aromatic heavy atoms, Fraction Csp³, Rotatable bonds, H-bond acceptors, H-bond donors, MR, TPSA and XLOGP³).

Code	MW	#Heavy atoms	#Aromatic heavy atoms	Fraction Csp ³	#Rotatable bonds	#H-bond acceptors	#H-bond donors	MR	TPSA	XLOGP ³
a	636.78	46	12	0.56	17	8	2	187.08	135.62	1.44
b	636.78	46	12	0.56	17	8	2	187.08	135.62	1.44
c	637.77	46	12	0.56	18	9	2	185.15	141.61	1.48
d	569.65	41	12	0.48	18	8	4	160.4	161.4	2.31
e	667.75	48	12	0.53	19	10	2	186.83	167.91	3.44



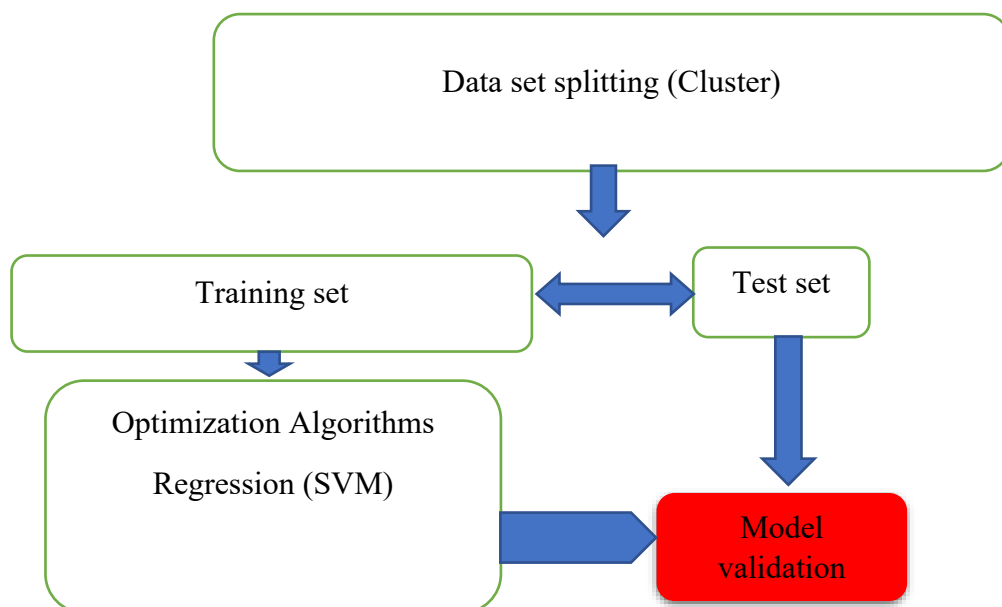


Figure.1. QSAR workflow for modeling P2Y₁₂ inhibitors.

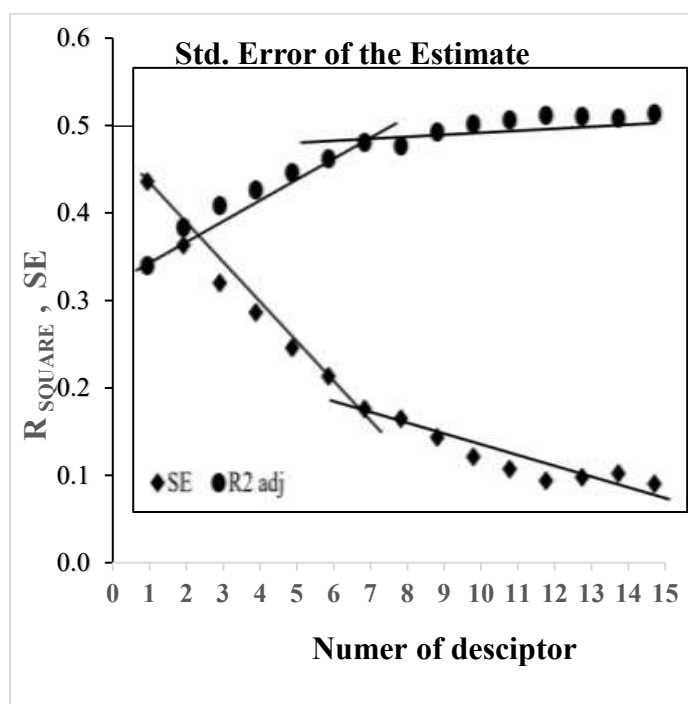
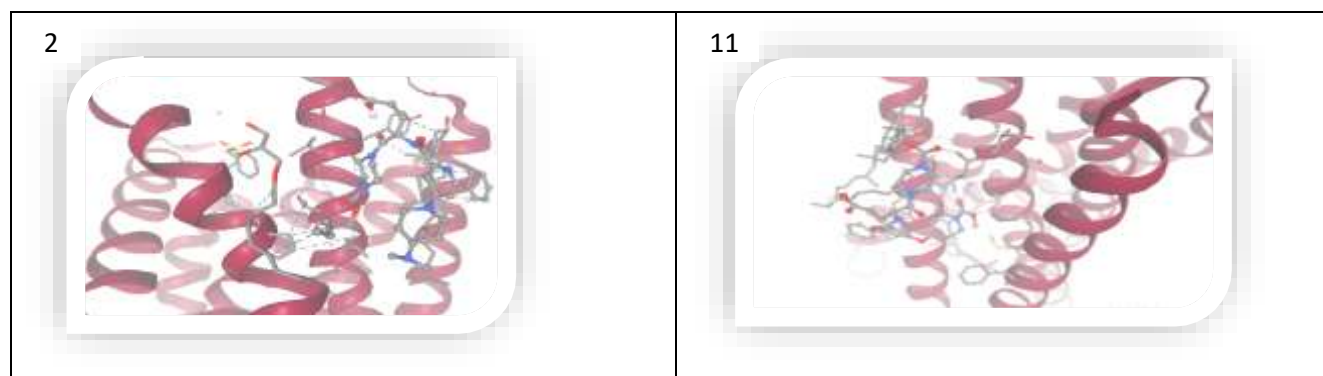


Figure.2. Variations in the correlation coefficient (R) and standard error (SE) relative to descriptor count.



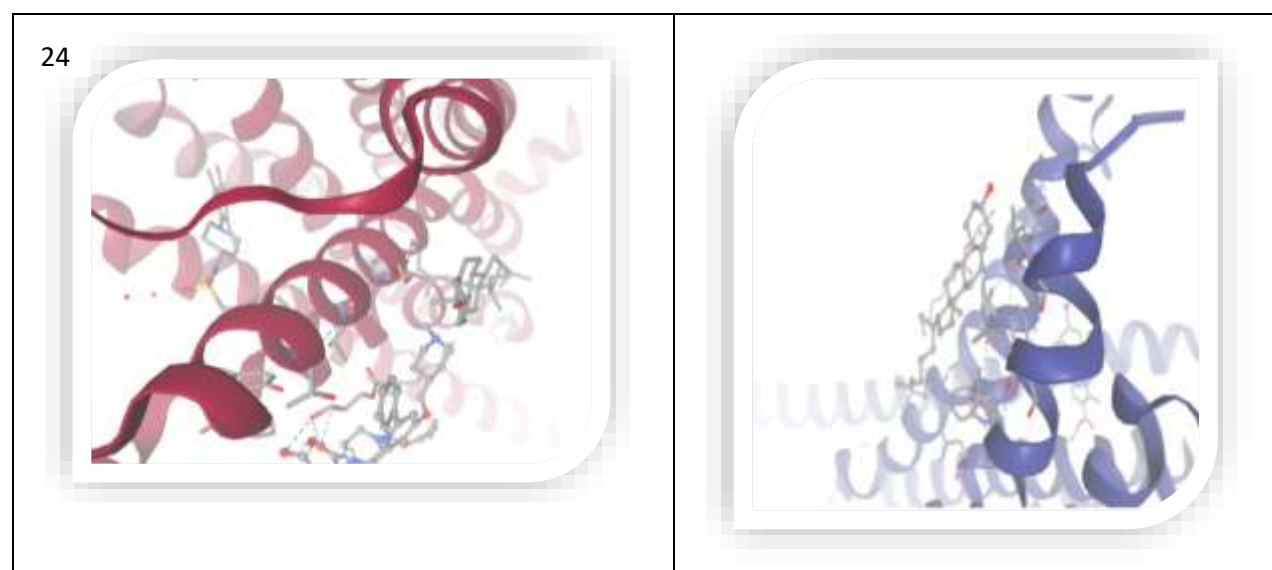


Figure. 3. Docking result between P2Y₁₂ and some piperazinyl-glutamate-pyridine / pyrimidine derivatives. (No. of chemicals are identical with Table 1)

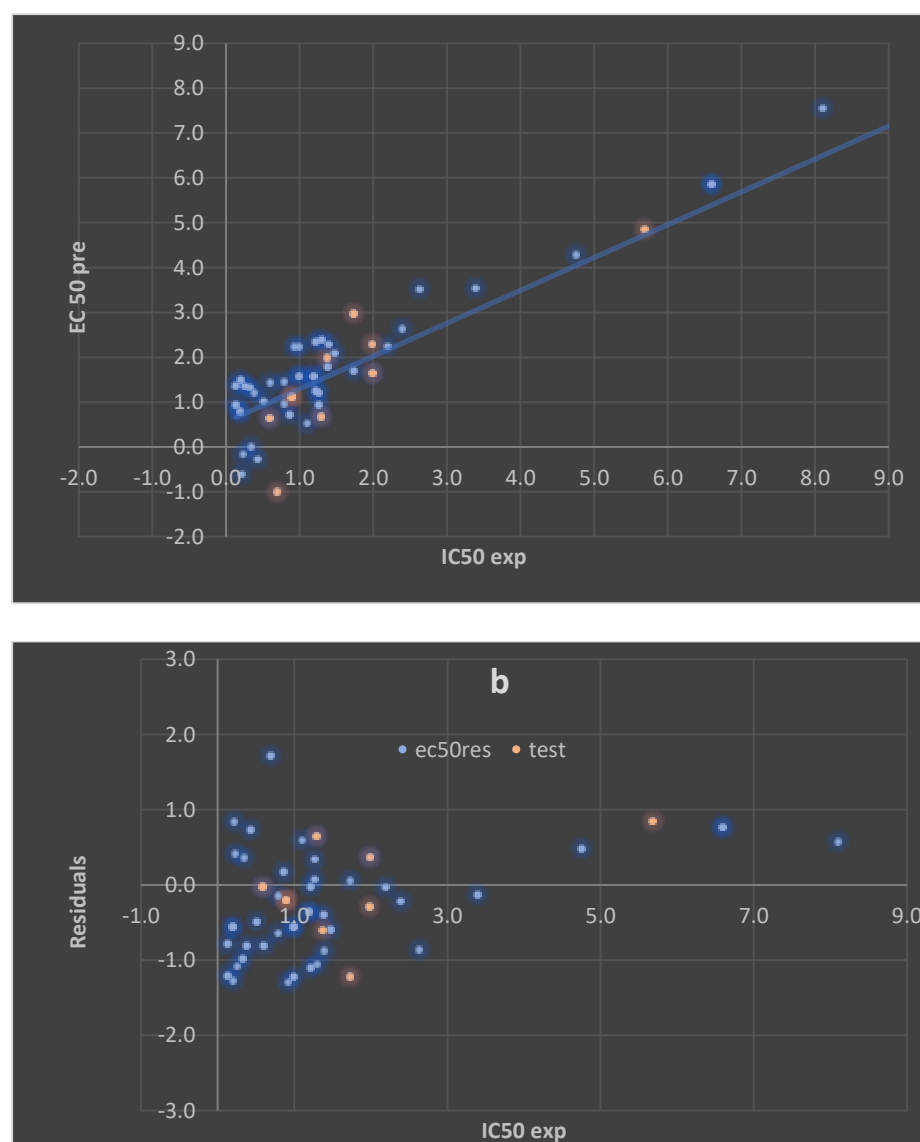


Figure. 4. The plot of SVM predicted (a) and residuals (b) against the experimental values of IC₅₀.

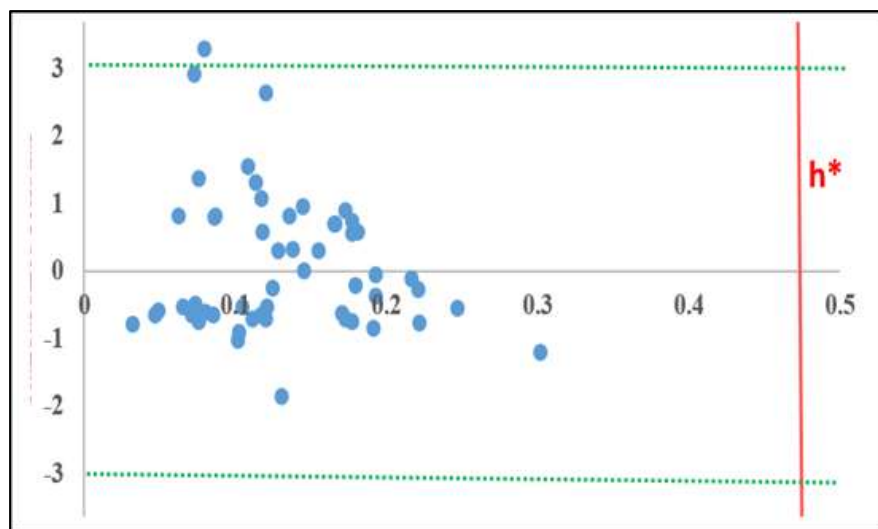
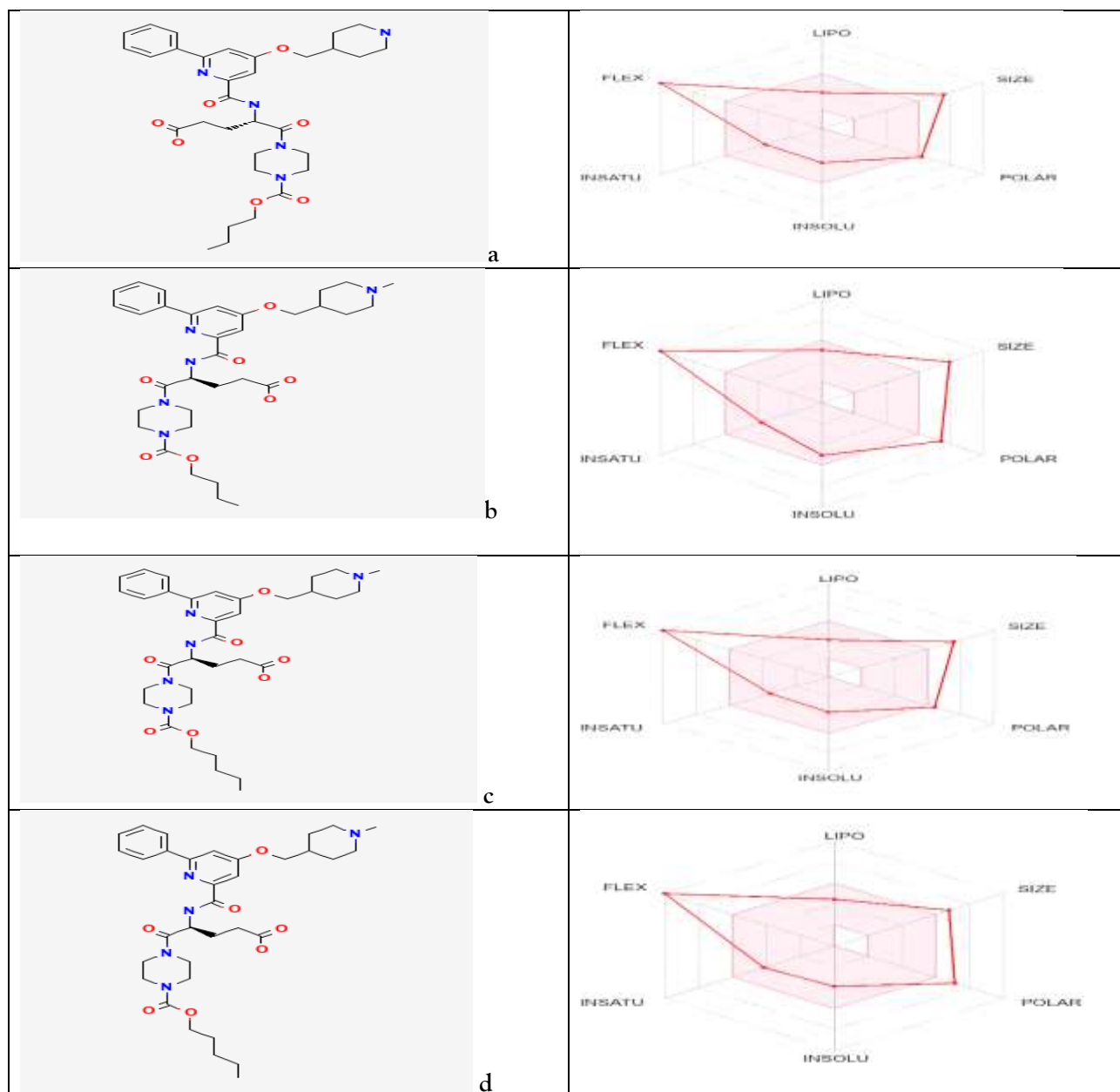


Figure. 5. The results of applicability domain analysis (Williams plot).



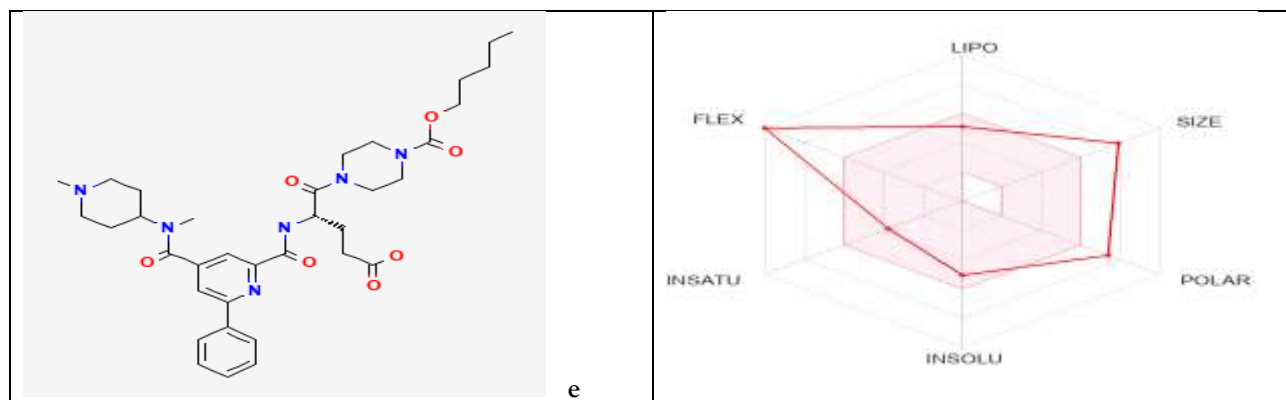


Figure. 6. Results of ADME analysis for hit drug candidates.