

2D and 3D-QSAR Modeling of Hydrazone Derivatives as Antiplatelet Agents

Azad Soltani^a, Farzad Kobarfard^a, Zahra Hajimahdi^{*a}, Salimeh Amidi^{*a,b}

^a Department of Medicinal Chemistry, School of Pharmacy, Shahid Beheshti University of Medical Sciences, Tehran, Iran.

^b Pharmaceutical Sciences Research Center, Shahid Beheshti University of Medical Sciences, Tehran, Iran.

Received: February 14, 2026 **Last Revision:** June 06, 2026 **Accepted:** July 19, 2026 **Available online:** October 11, 2026.

Abstract

Cardiovascular diseases (CVDs) remain the leading cause of mortality worldwide, with platelet aggregation playing a key role in thrombus formation. Hydrazone derivatives have emerged as promising antiplatelet agents, yet a systematic understanding of their structure–activity relationships remains limited. In this study, we developed predictive 2D and 3D QSAR models for a series of hydrazone derivatives as inhibitors of arachidonic acid-induced platelet aggregation.

In 2D-QSAR analysis, the stepwise multiple linear regression (SW-MLR) and genetic algorithm-based MLR (GA-MLR) models exhibited satisfactory predictive accuracy for both the training and test sets. Notably, the GA-MLR model showed better external predictivity, with an R^2_{test} value of 0.81, compared to the SW-MLR model.

Based on the 3D-QSAR analysis, the CoMFA model showed better performance in predicting pIC_{50} values ($R^2_{\text{train}} = 0.920$, $Q^2_{\text{LOO}} = 0.531$, and $R^2_{\text{test}} = 0.677$) than the CoMSIA model ($R^2_{\text{train}} = 0.890$, $Q^2_{\text{LOO}} = 0.502$, and $R^2_{\text{test}} = 0.600$). Furthermore, 3D-QSAR studies using CoMFA and CoMSIA further revealed that steric, electrostatic, hydrophobic, and hydrogen-bond donor/acceptor fields influence antiplatelet activity. Overall, the 2D and 3D QSAR models provide a reliable predictive framework for the antiplatelet potency of hydrazone derivatives and can guide the design of more potent and selective inhibitors, supporting future cardiovascular drug discovery.

Keywords: Hydrazone derivatives; Platelet aggregation; QSAR; CoMFA; CoMSIA; Cardiovascular disease.

1. Introduction

Cardiovascular diseases (CVDs) such as coronary artery disease, myocardial infarction, stroke, and peripheral vascular diseases remain the leading cause of mortality worldwide and constitute a major source of long-term disability. The pathogenesis of cardiovascular diseases is due to dysregulation of different pathways, such as inflammatory and thrombotic pathways [1-3].

Platelet aggregation and thrombus formation play a central role in the pathogenesis of CVDs. Following the platelet aggregation induced by agonists such as adenosine diphosphate (ADP) or Collagen and thrombus formation, acute vascular occlusion and ischemic events may ensue. In addition, platelets contribute to vascular inflammation and atherosclerotic progression through the release of bioactive mediators and cellular

* Corresponding Authors:

Zahra Hajimahdi, Department of Medicinal Chemistry, School of Pharmacy, Shahid Beheshti University of Medical Sciences, Tehran, Iran. E-mail: z.hajimahdi@sbmu.ac.ir.

Salimeh Amidi, Department of Medicinal Chemistry, School of Pharmacy, Shahid Beheshti University of Medical Sciences, Tehran, Iran. E-mail: s.amidi@sbmu.ac.ir.

Cite this article as: Soltani A., Kobarfard F., Hajimahdi Z., Amidi S. 2D and 3D-QSAR Modeling of Hydrazone Derivatives as Antiplatelet Agents. Iran. J. Pharm. Sci., 2026, 22 (1): 362-375.

DOI: <https://doi.org/10.22037/ijps.v22i1.51576>

interactions [1, 2]. The arachidonic acid (AA) cascade is one of the biochemical mechanisms of platelet activation. Upon stimulation, phospholipase A₂ liberates arachidonic acid from cell membrane phospholipids, which is subsequently converted by cyclooxygenase-1 (COX-1) into thromboxane A₂ (TXA₂), a potent promoter of platelet aggregation and vasoconstriction. Dysregulation of this pathway plays a decisive role in thrombosis and cardiovascular complications [3-5]. Various therapeutic approaches have been developed to inhibit platelet activation and aggregation [3]. Despite significant progress in prevention and treatment, the development of improved and innovative therapeutic strategies remains an ongoing necessity.

Among these therapeutic efforts, hydrazone derivatives have emerged as a promising class of antiplatelet agents. This scaffold has also been reported to exhibit additional biological activities, including anti-inflammatory, antioxidant, and enzyme inhibitory effects. Several experimental studies have demonstrated that substituted hydrazones effectively inhibit platelet aggregation induced by arachidonic acid, collagen, and ADP. Indole-based arylhydrazones have shown significant inhibition of arachidonic acid-induced platelet aggregation [6]. N-acylhydrazone derivatives and coumarin-hydrazone hybrids have also demonstrated strong *in vitro* antiplatelet activity [7, 8]. These findings emphasize the relevance of hydrazone scaffolds in modulating platelet function and provide a basis for further optimization.

In recent years, computational drug design, particularly quantitative structure-activity relationship (QSAR) modeling, has provided a powerful tool for identifying and optimizing lead compounds in antiplatelet research with reduced development costs [8]. QSAR modeling provides powerful tools for correlating molecular descriptors with biological activity and pharmacokinetic behavior. This approach has been successfully applied to different targets such as COX-1 inhibitors, thromboxane synthase blockers, and P2Y₁₂ receptor antagonists [9-12]. The convergence of hydrazone-based medicinal chemistry and QSAR offers a rational strategy for discovering next-generation platelet inhibitors. By identifying key structural features

of activity, QSAR-guided design enables systematic optimization of hydrazone derivatives before biological evaluation.

In this study, we aim to develop and validate 2D/3D QSAR models for hydrazone derivatives as inhibitors of AA-induced platelet aggregation, to elucidate the structural features of antiplatelet activity, and provide a predictive framework for designing more potent and selective antithrombotic agents.

2. Materials and Methods

2.1. Dataset

A dataset of 71 hydrazone derivatives as inhibitors of arachidonic acid (AA)-induced platelet aggregation with their corresponding activity data was compiled from reported and previously unpublished studies by Kobarfard *et al.* [6, 13-16]. The IC₅₀ values were converted to -log IC₅₀ and expressed as pIC₅₀ (Table 1). The dataset was divided into training and test sets using the Kennard-Stone algorithm implemented in the Dataset Division 1.2 software [17].

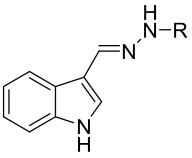
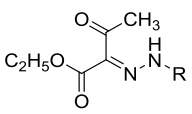
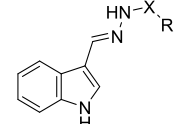
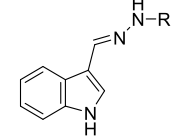
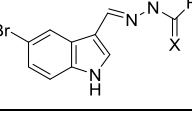
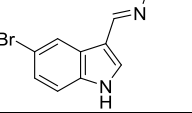
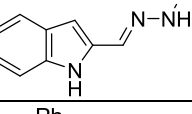
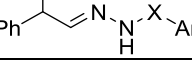
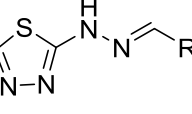
2.2. Descriptors Calculation

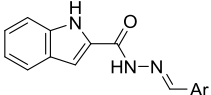
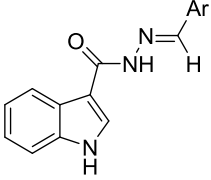
The structure of all compounds was sketched by ChemDraw 18.2. The AM1 semi-empirical force field was used to minimize the energy of compounds in Hyperchem 8.0 software. A total of 1,444 molecular descriptors (1D, 2D, and physicochemical) were calculated for optimized compounds using PaDEL-Descriptor 2.20 software [18].

2.3. Data Pretreatment

Objective feature selection was subsequently performed by the v-WSP algorithm implemented in v-WSP 1.2 software to reduce descriptor redundancy. Constant descriptors, descriptors with more than 90% identical values across compounds, and those exhibiting high intercorrelation were identified and removed from the dataset. Correlations among descriptors and between descriptors and biological activity were evaluated, retaining only those descriptors with the highest correlation to pIC₅₀ for model development [19, 20].

Table 1. Molecular structures of hydrazone derivatives and their experimental and predicted antiplatelet activity (pIC₅₀) by 2D and 3D QSAR models.

General Structure	No	X	R	Exp pIC ₅₀ ^a	Predicted pIC ₅₀ by			
					SW-MLR	GA-MLR	CoMFA	CoMSIA
	1	-	phenyl	5.00	4.97	4.76	5.09	4.92
	2*	-	4-fluorophenyl	4.54	5.11	4.92	4.91	5.01
	3	-	2-chlorophenyl	4.84	4.55	4.70	4.78	4.80
	4	-	3-chlorophenyl	4.55	4.66	4.64	4.53	4.66
	5	-	4-chlorophenyl	4.65	4.42	4.55	4.76	4.97
	6	-	4-bromophenyl	4.67	4.68	4.64	4.92	4.89
	7*	-	2-methylphenyl	4.25	4.55	4.70	4.72	4.67
	8*	-	4-methylphenyl	4.24	4.42	4.55	4.76	4.97
	9	-	4-methoxyphenyl	5.15	4.75	4.76	4.98	4.91
	10	-	phenyl	3.38	3.45	3.49	3.43	3.34
	11	-	4-methylphenyl	3.34	3.27	3.50	3.24	3.36
	12*	-	2,6-dimethylphenyl	3.44	2.99	3.56	3.63	3.39
	13	-	2-methylphenyl	3.52	3.27	3.58	3.58	3.40
	14*	-	2,3-dimethylphenyl	3.46	2.84	3.61	3.86	3.65
	15*	-	4-hydroxyphenyl	3.93	3.45	3.95	3.33	3.33
	16	-	4-methoxyphenyl	3.57	3.45	3.46	3.48	3.55
	17	C=O	-NH ₂	3.71	3.73	3.61	3.88	3.73
	18	C=S	NH ₂	3.85	3.94	4.17	3.86	4.00
	19	C=O	-NHNH ₂	3.72	3.77	3.63	3.85	3.79
	20*	C=O	4-pyridyl	3.66	4.19	3.99	4.14	4.75
	21	-	phenyl	5.00	5.17	4.91	5.09	4.89
	22	-	thiadiazolyl	4.28	4.38	4.18	4.24	4.21
	23	O	NH ₂	3.80	3.85	3.56	3.61	3.79
	24	S	NH ₂	3.76	3.73	4.08	3.62	3.87
	25*	-	thiadiazolyl	4.02	4.15	3.24	4.16	3.97
	26	-	phenyl	5.52	5.29	5.60	5.55	5.63
	27	C=O	4-pyridyl	4.36	4.61	3.47	4.28	4.21
	28	-	phenyl	3.74	3.78	4.16	3.60	3.90
	29	-	Phenyl	4.07	4.29	4.43	4.58	4.76
	30	-	2-Fluorophenyl	4.74	5.17	4.70	4.85	4.68
	31	-	3-Fluorophenyl	4.67	4.67	4.76	4.70	4.73
	32	-	4-Fluorophenyl	4.10	-	4.35	4.69	4.75
	33	-	2-Chlorophenyl	5.24	4.88	4.82	4.79	4.64
	34	-	3-Chlorophenyl	5.14	5.04	4.78	4.96	4.69
	35	-	4-Chlorophenyl	4.20	4.69	4.59	4.66	4.72
	36	-	2-Bromophenyl	4.81	5.13	4.86	4.62	4.61
	37*	-	3-Bromophenyl	5.11	5.64	4.93	5.08	4.66
	38	-	4-Bromophenyl	5.08	5.02	5.23	5.17	4.69
	39	-	2-nitrophenyl	4.74	4.67	4.76	4.77	4.76
	40	-	3-nitrophenyl	5.59	5.35	5.57	5.50	5.62
	41	-	4-nitrophenyl	5.11	5.02	4.87	5.13	5.07

General Structure	No	X	R	Exp pIC ₅₀ ^a	Predicted pIC ₅₀ by			
					SW-MLR	GA- MLR	CoMFA	CoMSIA
	42	-	3-pyridyl	4.68	4.55	4.67	4.77	4.98
	43*	-	2-furyl	5.66	4.92	5.56	5.15	5.26
	44*	-	2-thienyl	5.42	4.50	5.27	5.43	5.05
	45	-	3-indolyl	4.28	4.21	4.37	4.41	4.47
	46	-	5-bromo-3-indolyl	4.02	3.76	4.10	3.88	4.06
	47*	-	phenyl	4.03	3.94	3.73	3.73	3.44
	48	-	2,4-dichlorophenyl	3.06	2.99	3.10	3.47	3.39
	49	-	2,6-dichlorophenyl	3.08	3.39	3.33	3.19	3.21
	50	-	4-methoxyphenyl	3.85	3.70	3.59	3.63	3.53
	51	-	3-nitro-2-thienyl	3.14	3.35	3.49	3.15	3.21
	52	-	2,4-dihydroxyphenyl	3.38	3.77	3.42	3.77	3.82
	53	-	2,5-dihydroxyphenyl	4.29	4.09	4.01	3.94	3.91
	54	-	2,3-dihydroxyphenyl	4.35	4.22	4.18	4.06	4.08
	55	-	3-nitro-2-furyl	3.21	3.18	3.51	3.23	3.12
	56*	-	3-methoxyphenyl	3.86	3.54	4.10	3.97	3.62
	57	-	phenyl	3.74	3.66	3.69	3.85	3.52
	58	-	2,6-dichlorophenyl	3.60	3.67	3.51	3.79	3.61
	59*	-	4-methoxyphenyl	3.30	3.60	3.54	3.80	3.78
	60	-	2-hydroxy-5-methoxyphenyl	3.66	3.69	3.62	3.83	3.91
	61	-	4-nitrophenyl	3.53	3.61	3.69	3.45	3.48
	62	-	3-indolyl	3.49	3.41	3.31	3.48	3.59
	63*	-	2-thienyl	3.72	4.02	4.12	4.29	3.70
	64	-	5-chloro-2-nitrophenyl	3.53	3.25	3.41	3.35	3.71
	65	-	2,3-dihydroxyphenyl	3.77	4.05	4.05	3.56	3.76
	66	-	2,5-dihydroxyphenyl	3.99	3.99	3.94	3.97	3.93
	67	-	2,4-dihydroxyphenyl	3.64	3.64	3.59	3.56	3.60
	68	-	2,3-dimethoxyphenyl	3.63	3.78	4.12	3.76	3.61
	69*	-	3,4,5-trimethoxyphenyl	3.28	3.68	3.45	3.98	3.34
	70	-	3-nitrophenyl	3.51	3.60	3.21	3.55	3.55
	71	-	3-methoxyphenyl	3.46	3.45	3.93	3.48	3.27

*Test set of 2D QSAR model

^a Experimental pIC₅₀

2.4. Variable Selection and Building a 2D QSAR Model

Stepwise regression (SW) using SPSS 26 software and Genetic Algorithm (GA) implemented in Genetic Algorithm 4.0 software were applied for subjective feature selection. Subsequently, multiple Linear Regression (MLR) implemented in MLRPlus Validation 1.3 software was used to model a relationship between the selected molecular descriptors by SW or GA and pIC₅₀.

2.5. 2D QSAR Model Validation and Evaluation

The developed 2D QSAR models were validated by various statistical tests including the correlation coefficient (R²), Fisher's F- test (F-value), leave-one-out (LOO), cross-validation (Q²_{cv}), Y randomization (cR²_p), and Tropsha and Golbraikh's criteria by MLRPlus Validation 1.3 software [21].

The LOO cross-validation regression coefficient (Q²_{cv}) is calculated as follows:

$$Q^2_{cv} = 1 - \frac{\sum_{i=1}^{train} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{train} (y_i - \bar{y})^2}$$

Where y_i is the experimental pIC₅₀ value and $\hat{y}_i$ is the corresponding predictive pIC₅₀ value. The n train value is the total number of compounds in the dataset and $\bar{y}$ is the mean pIC₅₀ value of all compounds.

For the Y-randomization test, the coefficient of determination, cR²_p, was calculated according to the following formula:

$$cR^2_p = R \times \sqrt{R^2 - \text{average}R^2_r}$$

Where R² and average R²_r are the squared correlation coefficients of the non-randomized and average squared

correlation coefficients of the randomized models respectively.

The Tropsha and Golbraikh's parameters were also calculated for assessing the QSAR model.

$$(r^2 - r_0^2)/r^2 < 0.1 \text{ or } (r^2 - r_0'^2)/r^2 < 0.1$$
$$0.85 < k < 1.15 \text{ or } 0.85 < k' < 1.15$$

Where r^2 represents the squared correlation coefficient between the predicted and observed pIC_{50} values. r_0^2 is the squared correlation coefficient and k is the slope obtained from a linear least-squares regression of predicted versus observed pIC_{50} values performed without an intercept whereas $r_0'^2$ is the squared correlation coefficient and k' is the slope of observed versus predicted pIC_{50} values, also performed without an intercept. The absolute difference between r_0^2 and $r_0'^2$ was also calculated [21].

2.6. 3D QSAR Model Development

3D QSAR studies were conducted using Sybyl-X 1.2 software. The 3D structure of 71 compounds was converted to the MOL2 format by OpenBabel software and subsequently imported into Sybyl-X as a dataset. Compound alignment was performed using the most active compound as the template. Comparative Molecular Field Analysis (CoMFA) and Comparative Molecular Similarity Indices Analysis (CoMSIA) were applied. CoMFA evaluated steric (S) and electrostatic (E) fields surrounding the molecules, whereas CoMSIA additionally examined hydrophobic (H), hydrogen-bond donor (D), and hydrogen-bond acceptor (A) fields. To develop a reliable 3D QSAR model, partial least squares (PLS) analysis was used to correlate CoMFA and CoMSIA descriptors with pIC_{50} . Grid and Probe Parameters were adjusted according to previous studies [22].

2.7. 3D QSAR Model Validation and Evaluation

The developed 3D QSAR models were validated internally by the LOO and leave-many-out (LGO) cross-validated correlation coefficients (Q^2_{LOO} and Q^2_{LGO}), the correlation coefficient (R^2), the standard error of estimation (SEE) and the F-value. External validation was performed by the determination coefficient (R^2_{test}) [10, 22].

3. Results

3.1. 2D-QSAR Model Building and Validation

Among 1444 calculated descriptors, 922 descriptors were detected as constant and collinear descriptors and removed, leaving 522 descriptors that were used for subjective feature selection.

Figure 1 demonstrates the relationship between number of descriptors, the *R-squared*, and the standard error of the estimate (SEE) obtained during stepwise regression feature selection. As the number of descriptors increased, *R-squared* increased, and SEE decreased. The number of descriptors yielding the highest *R-squared* and the lowest SEE was considered optimal. Considering that the number of selected descriptors should not be more than 20% of the number of compounds in the training set, 9 descriptors were finally selected for model development. Similarly, during GA based feature selection, 9 descriptors were also selected.

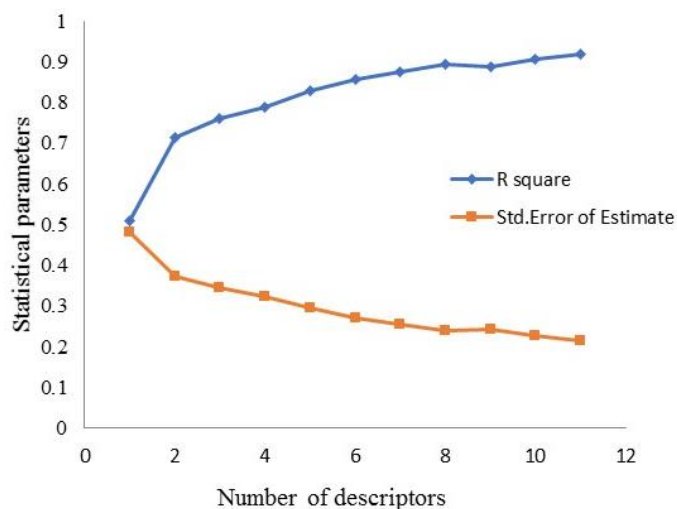


Figure 1. Influences of the number of descriptors on the statistical parameters (*R-squared* and standard error of the estimate) in the stepwise regression feature selection.

The selected molecular descriptors by SW and GA, which were used as independent variables in the model development, are summarized in Table 2 [23]. The models were then built using an MLR approach. The best linear equations obtained from the SW-MLR and GA-MLR models were presented in Table 3.

Table 2. Molecular descriptors used in the 2D QSAR models.

QSAR Models	Descriptors	Type	Description	Class
SW-MLR	GATS4e	2D	Geary autocorrelation of lag 4 weighted by Sanderson electronegativity	Auto correlation descriptor
	GATS7s	2D	Geary autocorrelation of lag 7 weighted by I-state	
	ATSC8p	2D	Centered Broto-Moreau autocorrelation of lag 8 weighted by polarizability	
	ATSC5c	2D	Broto-Moreau autocorrelation of lag 8 weighted by charges	Burden Modified Eigen values descriptor
	SpMax5_Bhm	2D	Smallest absolute eigenvalue of Burden modified matrix-n 5 / weighted by relative mass	
	minsBr	2D	Minimum atom-type E-State: -Br	
	SsssCH	2D	Sum of sssCH E-states	
	CrippenLogP	2D	Crippen's LogP	
GA-MLR	BCUTp-1l	2D	nhigh lowest polarizability weighted BCUTS	BCUT descriptor
	SpMax5_Bhm	2D	Smallest absolute eigenvalue of Burden modified matrix-n 5 / weighted by relative mass	Burden Modified Eigen values descriptor
	GATS7m	2D	Geary autocorrelation of lag 7 weighted by mass	Auto correlation descriptor
	ATSC3v	2D	Centred Broto-Moreau autocorrelation of lag 3 weighted by van der Waals volume	
	GATS6e	2D	Geary autocorrelation of lag 6 weighted by Sanderson electronegativity	
	SIC4	2D	Structural Information Content index (neighborhood symmetry of 4-order)	Information Content descriptor
	BCUTc-1h	2D	nlow highest partial charge weighted BCUTS	BCUT descriptor
	nHBint7	2D	Count of E-State descriptors of strength for potential Hydrogen Bonds of path length 7	Electrotopological State Atom Type descriptor
	nHBint9	2D	Count of E-State descriptors of strength for potential Hydrogen Bonds of path length 9	
	VR1_Dzp	2D	Randic-like eigenvector-based index from Barysz matrix weighted by van der Waals polarizability	Barysz Matrix descriptor

Table 3. The developed stepwise multiple regression (SW-MLR) and Genetic Algorithm Multiple Linear Regression (GA-MLR) models.

Model	Equation
SW-MLR	$pIC_{50} = 14.27662(\pm 0.9348) + 1.05085(\pm 0.26006) \text{ GATS4e} - 0.53366(\pm 0.09934) \text{ GATS7s} - 3.72474(\pm 0.29686) \text{ SpMax5_Bhm} + 0.1803(\pm 0.04188) \text{ ATSC8p} + 1.10531(\pm 0.31971) \text{ minsBr} - 9.55178(\pm 1.70069) \text{ SsssCH} + 0.34474(\pm 0.06431) \text{ CrippenLogP} - 0.14368(\pm 0.04262) \text{ BCUTp-1l} + 0.73162(\pm 0.27368) \text{ ATSC5c}$
GA-MLR	$pIC_{50} = 12.85335(\pm 1.78214) - 2.91697(\pm 0.33601) \text{ SpMax5_Bhm} - 0.73202(\pm 0.20037) \text{ GATS7m} + 1.15984(\pm 1.35359) \text{ SIC4} - 2.50658(\pm 0.84) \text{ BCUTc-1h} - 0.30919(\pm 0.05284) \text{ nHBint7} - 0.00059(\pm 0.00032) \text{ ATSC3v} + 0.12683(\pm 0.0588) \text{ nHBint9} - 0.52007(\pm 0.16937) \text{ GATS6e} + 0.00232(\pm 0.00048) \text{ VR1_Dzp}$

The equations of these 2D-QSAR models were separately used for calculating the pIC_{50} (Table 1). The validation parameters of these models were also represented in Table 4. The plot of predicted pIC_{50} against experimental pIC_{50} by SW-MLR and GA-MLR is shown in Figure 2.

The SW-MLR and GA-MLR models exhibited

satisfactory predictive accuracy for both the training and test sets, with $R^2 > 0.8$ for the training set and $R^2 > 0.6$ for the test set. Notably, the GA-MLR model showed better external predictive ability with a higher R^2_{test} value compared to the SW-MLR model (Figure 2). LOO cross-validation for both 2D QSAR models shows acceptable predictive accuracy ($Q^2_{\text{cv}} > 0.5$).

Table 4. Statistical parameters in two developed QSAR models.

Validation parameters	2D-QSAR models		Remark	Threshold value
	SW-MLR	GA-MLR		
Correlation coefficient (R^2_{train})	0.92	0.86	Passed	$R^2_{\text{train}} > 0.8$
Root mean square error (RMSE)	0.21	0.28	-	Low
F-value	54.79	30.37	-	High
Cross-validation coefficient (Q^2_{cv})	0.66	0.78	Passed	$Q^2_{\text{cv}} > 0.5$
Correlation coefficient (R^2_{test})	0.61	0.81	Passed	$R^2_{\text{test}} > 0.6$
Y-randomization (cR^2_{p})	0.83	0.77	Passed	$\text{cR}^2_{\text{p}} > 0.5$
$ r_0^2 - r'^2_0 $	0.03	0.02	Passed	$ r_0^2 - r'^2_0 < 0.3$
k	1.00	0.98	Passed	$0.85 < k < 1.15$
$[(r^2 - r_0^2)/r^2]$	0.11	0.00	Passed	$(r^2 - r_0^2)/r^2 < 0.1$
k'	0.99	1.02	Passed	$0.85 < k' < 1.15$
$[(r^2 - r'^2_0)/r^2]$	0.11	0.03	Passed	$(r^2 - r'^2_0)/r^2 < 0.1$

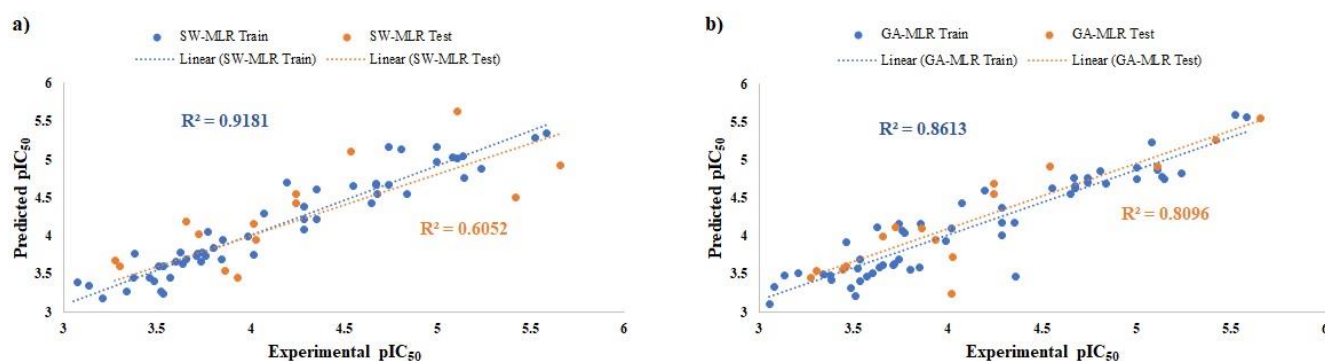
r^2 : Squared correlation coefficients between the predicted and observed values of the activity of the test group (R^2_{test}).

r_0^2 : Squared correlation coefficients from regression through the origin (Predicted vs Observed).

r'^2_0 : Squared correlation coefficients from regression through the origin (Observed vs Predicted).

k: Slopes of regression lines through the origin (Predicted vs Observed).

k': Slopes of regression lines through the origin (Observed vs Predicted).

**Figure 2.** The plot of experimental pIC_{50} against predicted pIC_{50} based on a) SW-MLR and b) GA-MLR in train and test sets.

The Y-randomization test is used to confirm that the developed 2D-QSAR models are not built by chance and the correlation between descriptors and the corresponding pIC_{50} is not obtained randomly. For this test, pIC_{50} values were randomly distributed fifty times without changing the descriptors, and the models were rebuilt. The results of the Y-randomization test for the first fifteen iterations performed on the SW-MLR and GA-MLR models are presented in Table 5. As shown in Tables 4 and 5, the randomized 2D QSAR models exhibited lower R^2 and Q^2_{LOO} values compared with the original models. In addition, a cR^2_{p} coefficient > 0.5 demonstrated that the original SW-MLR and GA-MLR models (Table 3) are

acceptable and were not obtained by chance.

The absolute difference between r_0^2 and r'^2_0 was less than 0.3, indicating acceptable symmetry between predicted and observed activities. According to the calculated Tropsha and Golbraikh's parameters, the external predictive reliability of these models was confirmed.

The intercorrelation among the selected descriptors was evaluated by Pearson's correlation analysis. As shown in Table 6, the correlation coefficient values for all descriptor pairs were below 0.57, indicating that the descriptors selected by the stepwise regression and genetic algorithm methods were completely independent.

Table 5. Y- randomization parameters for SW-MLR and GA-MLR models.

Model Type	SW-MLR		GA-MLR	
	R ²	Q ² _{Loo}	R ²	Q ² _{Loo}
Original	0.918076	0.655163	0.861339	0.781019
Random 1	0.226244	-0.47705	0.159086	-0.56408
Random 2	0.148596	-0.47227	0.177314	-0.24999
Random 3	0.143841	-0.61756	0.189946	-0.4796
Random 4	0.183869	-1.12941	0.151265	-0.24611
Random 5	0.125036	-0.31246	0.122901	-0.43855
Random 6	0.084233	-0.50086	0.210727	-0.11717
Random 7	0.123477	-0.2919	0.195431	-0.19296
Random 8	0.301001	-0.00097	0.167945	-0.38489
Random 9	0.333539	-0.89609	0.145309	-0.21242
Random 10	0.29498	-1.31699	0.115918	-0.30966
Random 11	0.23649	-0.08174	0.24017	-0.1766
Random 12	0.223276	-0.17356	0.159188	-0.22905
Random 13	0.152143	-0.38747	0.110369	-0.4281
Random 14	0.147617	-0.97525	0.207733	-0.16051
Random 15	0.051058	-0.85439	0.1117	-0.31267
Random model parameters for fifty iterations				
Average R²	0.193065		0.187523	
Average Q²_{Loo}	-0.98732		-0.29682	

Table 6. Correlation coefficient matrix of the selected descriptors in the developed SW-MLR and GA-MLR models.

SW-MLR	GATS4e	GATS7s	SpMax5_Bhm	ATSC8p	minsBr	SsssCH	CrippenLogP	BCUTp-1l	ATSC5c
GATS4e	1.000								
GATS7s	-0.213	1.000							
SpMax5_Bhm	0.090	0.436	1.000						
ATSC8p	-0.118	0.428	0.663	1.000					
minsBr	-0.069	-0.177	-0.002	-0.123	1.000				
SsssCH	0.148	-0.069	0.000	-0.030	-0.040	1.000			
CrippenLogP	0.550	-0.133	0.405	0.203	0.199	0.310	1.000		
BCUTp-1l	0.239	-0.348	-0.230	-0.220	0.413	0.156	0.481	1.000	
ATSC5c	0.453	-0.055	0.116	0.004	0.142	0.065	0.321	0.207	1.000

GA-MLR	SpMax5_Bhm	GATS7m	SIC4	BCUTc-1h	nHBint7	ATSC3v	nHBint9	GATS6e	VR1_Dzp
SpMax5_Bhm	1.000								
GATS7m	-0.519	1.000							
SIC4	-0.208	0.012	1.000						
BCUTc-1h	0.254	0.104	-0.120	1.000					
nHBint7	0.099	-0.240	0.320	0.210	1.000				
ATSC3v	-0.172	0.240	-0.412	0.507	-0.155	1.000			
nHBint9	0.285	-0.070	0.148	0.320	0.232	-0.042	1.000		
GATS6e	0.318	-0.334	0.033	-0.064	0.030	-0.376	0.154	1.000	
VR1_Dzp	0.191	0.024	0.071	-0.035	0.148	-0.168	0.046	0.168	1.000

The standardized regression coefficients of the selected descriptors in SW-MLR and GA-MLR models, along with their magnitude, were demonstrated graphically in [Figure 3](#) and [Figure 4](#). In these figures,

descriptors with positive standardized coefficients contribute positively to antiplatelet activity, whereas descriptors with negative coefficients exert a negative influence on activity.

The descriptors of the developed SW-MLR equation are governed by the autocorrelation descriptors (GATS4e, GATS7s, ATSC8p, ATSC5c), electrotopological state atom type descriptors (minsBr, SsssCH), eigenvalue-based parameters (SpMax5_Bhm), along with the Crippen descriptor and BCUT descriptor. In this model, the Crippen descriptor (CrippenlogP) shows the strong positive contribution, while the Burden modified eigenvalue descriptor (SpMax5_Bhm) exhibits the strong negative contribution toward antiplatelet activity. The SpMax5_Bhm (Smallest absolute eigenvalue of Burden modified matrix-n 5/weighted by relative mass), suggesting that increased molecular branching and connectivity patterns introduced by Burden modified eigenvalues are unfavorable for antiplatelet activity. The CrippenLogP descriptor emphasizes the role of molecular lipophilicity in improving biological response, likely by facilitating membrane permeability or hydrophobic interactions with the biological target [9].

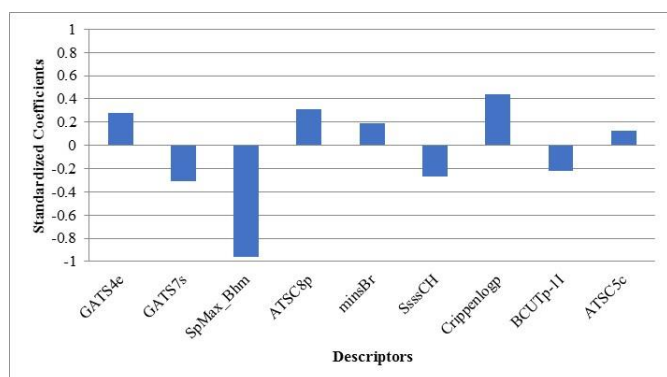


Figure 3. Standardized coefficients versus descriptor values in SW-MLR.

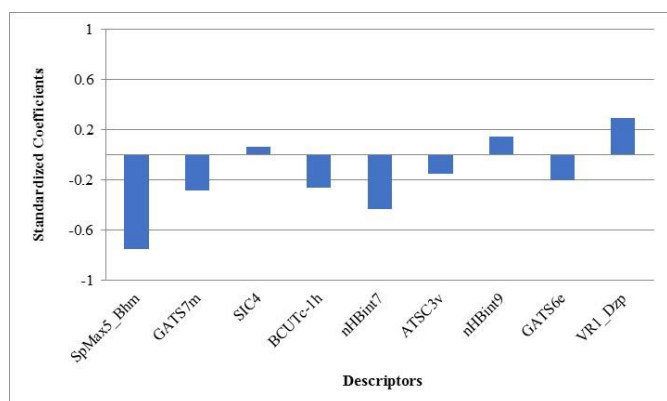


Figure 4. Standardized coefficients versus descriptor values in GA-MLR.

Overall, this SW-MLR equation reveals that antiplatelet activity is governed by a balance of electronic, topological, lipophilic, and structural features, with autocorrelation descriptors, electrotopological state atom type descriptors, eigenvalue-based parameters, and lipophilicity playing key roles in determining the biological response.

According to data presented in [Figure 4](#), the descriptors in the best GA-MLR equation are mainly governed by the autocorrelation descriptors (GATS7m, GATS6e, ATSC3v), electrotopological state atom type descriptors (nHBint7, nHBint9), eigenvalue-based parameters (SpMax5_Bhm), information content descriptor, BCUT descriptors, and Barysz matrix descriptors (VR1_Dzp). The VR1_Dzp (Randic-like eigenvector-based index from Barysz matrix weighted by van der Waals polarizability) and SpMax5_Bhm contribute the strongest positive and negative effects on antiplatelet activity, respectively. Similar to the SW-MLR equation, the SpMax5_Bhm descriptor is unfavorable for antiplatelet activity. The VR1_Dzp descriptor emphasizes the beneficial role of molecular polarizability and electronic distribution in enhancing antiplatelet activity.

In both SW-MLR and GA-MLR models, autocorrelation descriptors, electrotopological state atom type, eigenvalue-based parameters, and BCUT descriptor contribute. Both models consistently highlight a strong negative contribution to the importance of SpMax5_Bhm. Overall, despite methodological differences in descriptor selection, both models demonstrated that antiplatelet activity arises from a balance among molecular topology, electronic structure, and steric effects.

3.2. 3D-QSAR Model Building and Validation

A series of 71 hydrazone derivatives was used for 3D-QSAR model development using CoMFA and CoMSIA methods. The linear relationship between the CoMFA and CoMSIA descriptors and the pIC_{50} values was evaluated by the PLS method. The calculated pIC_{50} values of the compounds by 3D-QSAR models were presented in Table 1. Statistical parameters and validation results of CoMFA and CoMSIA models were presented in [Table 7](#).

Table 7. Statistical parameters and validation results of CoMFA and CoMSIA models using PLS analysis.

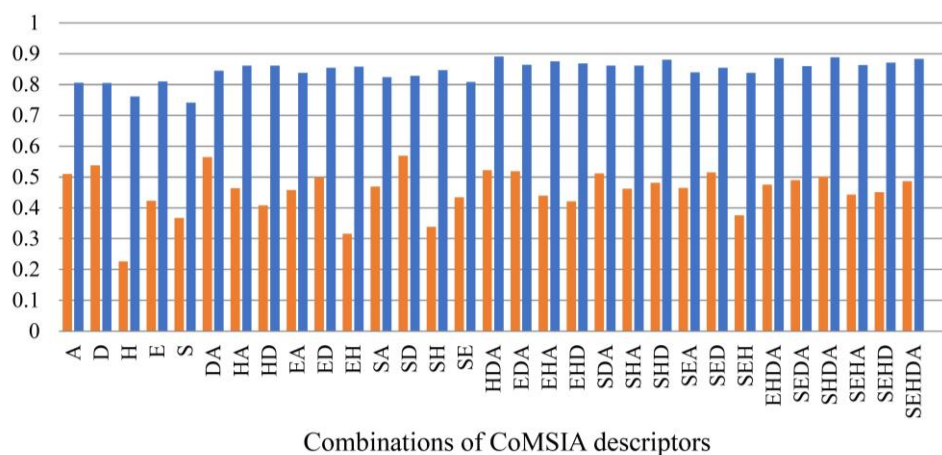
PLS statistics	CoMFA	CoMSIA
Standard error of estimate (SEE)	0.208	0.247
Correlation coefficient of train set (R^2_{train})	0.921	0.890
leave-one-out (LOO) correlation coefficient (Q^2_{LOO})	0.531	0.502
leave-group-out correlation coefficient (Q^2_{LGO})	0.572	0.518
F value	114.563	78.305
correlation coefficient of test set (R^2_{test})	0.677	0.600
Number of PLS components	4	3
Steric (S)	32.4	11.4
Electrostatic (E)	67.6	-
Hydrophobic (H)	-	25.2
Donor (D)	-	32.5
Acceptor (A)	-	31.0

The R^2 of the developed CoMFA model with four component analysis was 0.921 with a high F value of 114.56 and a low SEE of 0.208. Furthermore, this model demonstrated the acceptable Q^2 values of LOO and LGO greater than 0.5 (0.531 and 0.572). Analysis of field contributions revealed that the steric (S) and electrostatic (E) fields contributed 32.4% and 67.6%, respectively, to the CoMFA model.

In the constructed CoMSIA model by three principal components, the obtained R^2 was 0.890, accompanied by an F value of 78.305 and a low SEE of 0.247. The Q^2_{LOO} and Q^2_{LGO} showed acceptable values (0.502 and 0.518). The developed CoMSIA model, similar to the CoMFA

model, demonstrated acceptable internal predictive ability.

Different combinations of the five CoMSIA fields, including hydrogen bond acceptor (A), hydrogen bond donor (D), hydrophobic (H), electrostatic (E), and steric (S), were constructed and evaluated to obtain the best CoMSIA model. Based on the Q^2_{LOO} (0.502) and R^2_{train} (0.890) values, the SHDA combination was selected as the best model. Field contribution analysis indicated the contributions of 11.4% steric, 25.2% hydrophobic, 32.5% hydrogen-bond donor, and 31.0% hydrogen-bond acceptor (Figure 5).

**Figure 5.** Comparison of different combinations of CoMSIA descriptors. S, E, H, D, and A represent the steric, electrostatic, hydrophobic, H-bond donor, and acceptor property fields, respectively. Blue color = R^2_{train} and orange color = Q^2_{LOO} .

The validation results of these two models were similar and acceptable. The correlation between CoMFA and CoMSIA models' prediction and experimental pIC_{50} for training and test sets was demonstrated in Figure 6. According to the results, the CoMFA model has better ability for pIC_{50} prediction ($R^2_{\text{train}} = 0.92$, and $R^2_{\text{test}} = 0.677$).

3.2.1. Contour Map Analysis

The CoMFA contour maps analysis for compound 43 was demonstrated in Figure 7. In the steric field contour map, the green contours indicate that the presence of a bulky group in these areas enhances the biological activity, whereas the yellow contours lead to a reduction in activity. For instance, bearing a bulky 4-bromophenyl

substituent attached to the imine moiety showed a higher pIC_{50} value than compound 32 containing a small 4-fluorophenyl group (pIC_{50} of 5.08 vs 4.10). Moreover, in compounds 10 - 16, where linear substituents are attached to the imine group, the pIC_{50} values decreased within the range of 3.34 to 3.93. In the electrostatic field CoMFA contour map, the presence of electron-donating groups in the blue contour regions enhances the antiplatelet activity. For example, compound 26 with an electron-rich indole ring attached to the $\text{HC}=\text{N}$ group exhibited a high pIC_{50} of 5.52. Similarly, compound 44 containing an electron-rich thienyl ring at the same $\text{HC}=\text{N}$ position showed a pIC_{50} of 5.42. In contrast, compound 42, which has an electron-deficient 3-pyridyl ring at this position, demonstrated reduced antiplatelet activity ($\text{pIC}_{50} = 4.68$).

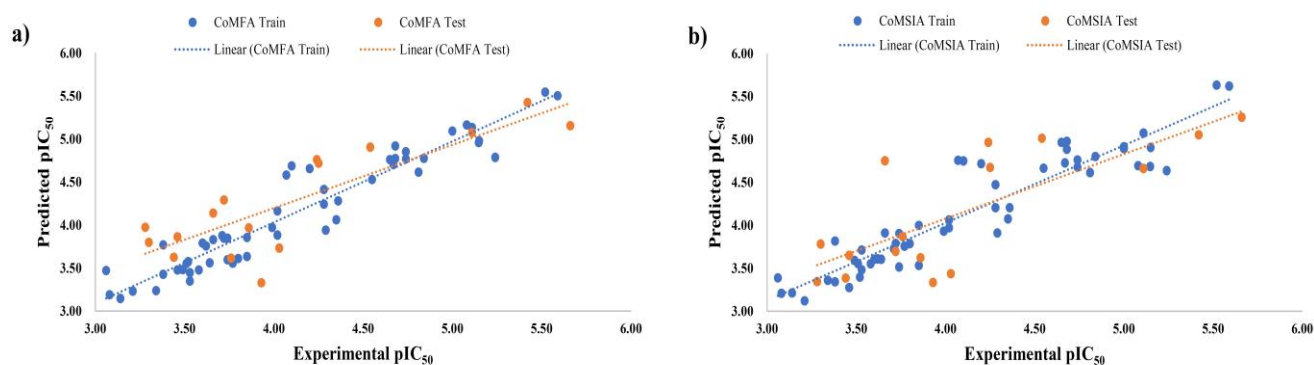


Figure 6. The plot of experimental pIC_{50} against predicted pIC_{50} based on a) CoMFA and b) CoMSIA in train and test sets.

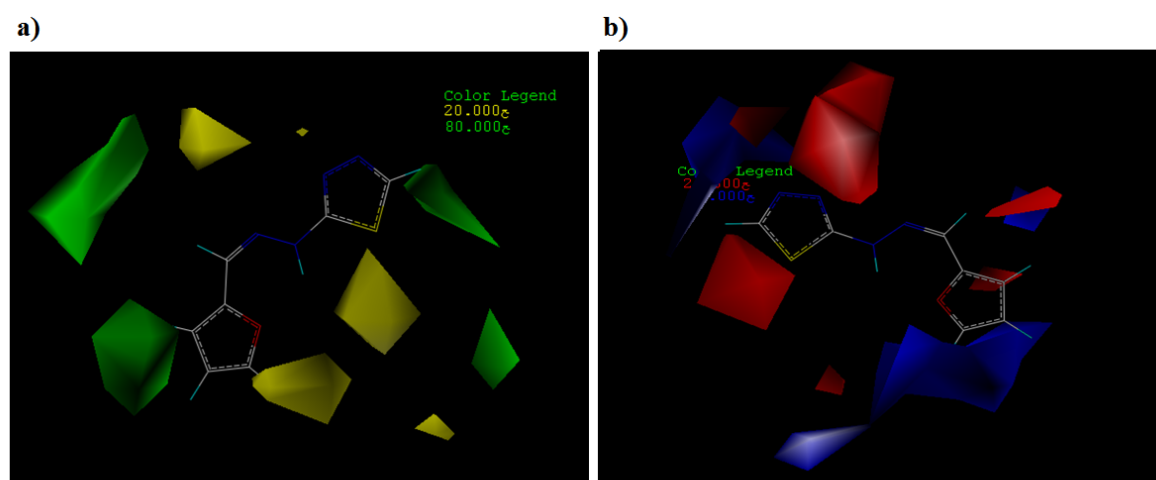


Figure 7. Contour maps of CoMFA: Steric field (a) and Electrostatic field (b).

The CoMSIA contour maps of compound 43 are shown in **Figure 8**. In the steric field contour map, the green contours indicate regions where the presence of bulky substituents enhances biological activity, whereas the yellow contours represent areas where bulky groups lead to a reduction in activity. In the hydrophobic contour map, the yellow contours indicate that introducing hydrophobic groups in these areas, such as the thiadiazolyl group, may lead to enhancing the biological activity. In the hydrogen bond donor contour map, the cyan contours around the NH of the hydrazone group indicate favorable regions, whereas the purple contours represent unfavorable regions. In the hydrogen-bond acceptor contour map, the light purple contours around the nitrogen atom of the hydrazone group indicate favorable regions, while the red contours represent unfavorable hydrogen-bond acceptor regions.

4. Conclusion

The convergence of hydrazone-based compounds and QSAR modeling offers a particularly attractive strategy for the discovery of next-generation antiplatelet agents. In this study, we developed robust 2D and 3D QSAR

models to elucidate the structural determinants of hydrazone derivatives as inhibitors of arachidonic acid-induced platelet aggregation.

The 2D-QSAR models constructed using SW-MLR and GA-MLR demonstrated that molecular descriptors related to autocorrelation, electrotopological states, eigenvalue-based parameters, and lipophilicity significantly influence antiplatelet activity. The models were validated using leave-one-out cross-validation, Y-randomization, and test sets, confirming their predictive reliability.

3D-QSAR analysis using CoMFA and CoMSIA further revealed that steric, electrostatic, hydrophobic, and hydrogen-bond donor/acceptor fields around the hydrazone scaffold play critical roles in modulating activity. Contour map interpretations highlighted key regions where specific substituents can enhance or diminish activity.

Overall, the developed 2D and 3D QSAR models offer a powerful strategy for predicting the antiplatelet potential of hydrazone derivatives and provide a predictive framework for future hydrazone-based antithrombotic drug discovery and optimization.

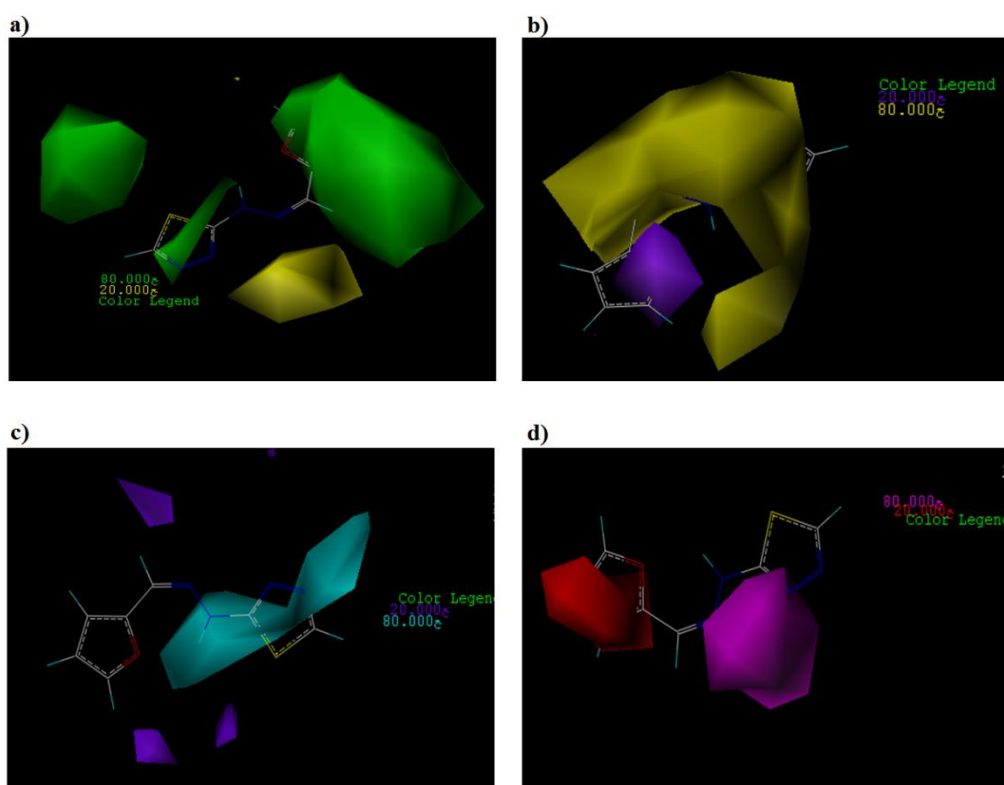


Figure 8. Contour maps of CoMSIA: Steric field (a), Hydrophobic field (b), Hydrogen bond donor (c), and Hydrogen bond acceptor (d).

Ethical Approval

This study was conducted in accordance with the ethical principles with approval code of IR.SBMU.RETECH.REC.1402.814.

Acknowledgment

This work was supported by the Pharmaceutical Sciences Research Center of Shahid Beheshti University of Medical Sciences.

Conflict of interest

All authors declare there is no conflict of interest.

Data availability

The data are available upon reasonable request from the corresponding authors.

Authors Contributions

A.S. contributed to Software Analysis, Data Analysis, Validation, and Visualization. F.K. contributed to Conceptualization, Supervision, and Revising the Manuscript. S.A. contributed to Conceptualization, Supervision, Writing the original draft, and revising the manuscript. Z.HM. contributed to Conceptualization, Supervision, Methodology, Data Interpretation, and Revising the Manuscript.

Authors Orcid numbers:

Farzad Kobarfard: [0000-0001-6679-3275](https://orcid.org/0000-0001-6679-3275)

Zahra Hajimahdi: [0000-0001-5866-4895](https://orcid.org/0000-0001-5866-4895)

Salimeh Amidi: [0000-0002-6032-3237](https://orcid.org/0000-0002-6032-3237)

Funding

This work was supported by a grant from the Pharmaceutical Sciences Research Center of Shahid Beheshti University of Medical Sciences (Grant Number 43006575).

Using artificial intelligence chatbots

ChatGPT (GPT-5) was used only for language editing.

References

1. Papapanagiotou, A., et al., The Role of Platelets in Cardiovascular Disease: Molecular Mechanisms. *Curr Pharm Des*, 2016. 22(29): p. 4493-4505.
2. Renga, B. and F. Scavizzi, Platelets and cardiovascular risk. *Acta Cardiol*, 2017. 72(1): p. 2-8.
3. Stanger, L., A. Yamaguchi, and M. Holinstat, Antiplatelet strategies: past, present, and future. *J Thromb Haemost*, 2023. 21(12): p. 3317-3328.
4. Badimon, L., et al., The key contribution of platelet and vascular arachidonic acid metabolism to the pathophysiology of atherothrombosis. *Cardiovasc Res*, 2021. 117(9): p. 2001-2015.
5. van der Meijden, P.E.J. and J.W.M. Heemskerk, Platelet biology and functions: new concepts and clinical perspectives. *Nat Rev Cardiol*, 2019. 16(3): p. 166-179.
6. Haj Mohammad Ebrahim Tehrani, K., et al., Synthesis, Antiplatelet Activity and Cytotoxicity Assessment of Indole-Based Hydrazone Derivatives. *Iran J Pharm Res*, 2015. 14(4): p. 1077-86.
7. Haffouz, A., et al., Synthesis and structural characterization of coumarin-hydrazone derivatives as antiplatelet agents: in vitro and in silico evaluation of activity and toxicity. *3 Biotech*, 2025. 15(9): p. 322.
8. Katritzky, A.R., et al., QSAR study of antiplatelet agents. *Bioorg Med Chem*, 2006. 14(22): p. 7490-500.
9. Ramzan, A., et al., Antiplatelet activity, molecular docking and QSAR study of novel N'-arylmethylidene-3-methyl-1-phenyl-6-p-chlorophenyl-1H-pyrazolo[3,4-b]pyridine-4-carbohydrazides. *Med Chem Res*, 2017. 27(2): p. 388-405.
10. Hao, M., et al., Combined 3D-QSAR, molecular docking, and molecular dynamics study on piperazinyl-glutamate-pyridines/pyrimidines as potent P2Y12 antagonists for inhibition of platelet aggregation. *J Chem Inf Model*, 2011. 51(10): p. 2560-72.
11. Shunlai, L., et al., Essential Structural Profile of Novel Adenosine Derivatives as Antiplatelet Aggregation Inhibitors Based on 3D-QSAR Analysis Using CoMFA, CoMSIA, and SOMFA. *Russ J Bioorg Chem*, 2020. 46(3): p. 448-457.
12. Yan, Y., et al., Studies of tricyclic piperazine/piperidine furnished molecules as novel integrin $\alpha\beta_3/\alpha\text{IIb}\beta_3$ dual antagonists using 3D-QSAR and molecular docking. *J Mol Graph Model*, 2011. 29(5): p. 747-62.
13. Farhady, S., et al., Synthesis and Antiplatelet Activity Evaluation of a Group of Novel Ethyl Acetoacetate Phenylhydrazone Derivatives. *Iran J Pharm Res*, 2021. 20(2): p. 307-315.
14. Haj Mohammad Ebrahim Tehrani, K., et al., One pot synthesis and biological activity evaluation of novel Schiff bases derived from 2-hydrazinyl-1,3,4-thiadiazole. *Chem Pharm Bull (Tokyo)*, 2013. 61(2): p. 160-6.
15. Mashayekhi, V., et al., Synthesis of novel indole hydrazone derivatives and evaluation of their antiplatelet aggregation activity. *Chem Pharm Bull (Tokyo)*, 2013. 61(2): p. 144-50.
16. Mirfazli, S.S., et al., Design, synthesis and QSAR study of arylidene indoles as anti-platelet aggregation inhibitors. *Med Chem Res*, 2016. 25(1): p. 1-18.
17. Martin, T.M., et al., Does Rational Selection of Training and Test Sets Improve the Outcome of QSAR Modeling? *J Chem Inf Model*, 2012. 52(10): p. 2570-2578.

18. Yap, C.W., PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints. *J Comput Chem*, 2011. **32**(7): p. 1466-74.
19. Ballabio, D., et al., A novel variable reduction method adapted from space-filling designs. *Chemometr Intell Lab Syst*, 2014. **136**: p. 147-154.
20. Ambure, P., et al., "NanoBRIDGES" software: Open access tools to perform QSAR and nano-QSAR modeling. *Chemometr Intell Lab Syst*, 2015. **147**: p. 1-13.
21. Shayanfar, A. and S. Shayanfar, Is regression through origin useful in external validation of QSAR models? *Eur J Pharm Sci*, 2014. **59**: p. 31-35.
22. Dowlati Beirami, A., Z. Hajimahdi, and A. Zarghi, Docking-based 3D-QSAR (CoMFA, CoMFA-RG, CoMSIA) study on hydroquinoline and thiazinan-4-one derivatives as selective COX-2 inhibitors. *J Biomol Struct Dyn*, 2019. **37**(11): p. 2999-3006.
23. Todeschini, R. and V. Consonni, *Handbook of Molecular Descriptors. Methods and Principles in Medicinal Chemistry*, ed. R. Mannhold, H. Kubinyi, and H. Timmerman. Vol. 11. 2000, Weinheim: Wiley.