Comparative evaluation of QSAR-based thiazole compounds as SDHI inhibitors

Sunita Gupta and Ashok Nagore
Department of Chemistry, PMCOE Govt. Science College, Rewa (M.P.)-486003

Abstract

Using QSAR techniques, a database containing compounds based on thiazole derivatives with values of succinate dehydrogenase inhibitors against *S. sclerotiorum* (SDHI) pEC_{50} was utilized to establish a structure-activity linkeighty percent of the data set was utilized as a learning set and the remaining twenty percent was used as a test set to evaluate the models' external predictions. Several techniques, including internal validation, external validation, and the randomized Y test, were used to assess the predictive capacity and dependability of the established models. The Williams plot was utilized in the applicability domain to identify outliers.According to the 2D QSAR the best model was created utilizing four descriptors: J, Log P, NRB and MD. The multiple linear regression approach (MLR) produced $R^2 = 0.80$ and $Q^2 = 0.63$, while the partial least squares regression method (PLS) produced $R^2 = 0.78$ and $Q^2 = 0.64$. All internal and external validation criteria have been satisfactorily met by these models. According to the 3D-QSAR data, the optimal model was chosen using the molecular field analysis approach (CoMSIA), yielding $R^2 = 0.957$, $Q^2 = 0.614$ and R^2-test = 0.80.The nature and location of some structural indicators crucial for enhancing the investigated biological activity, such as steric, electrostatic, and hydrophobic substituents as well as hydrogen bond donor substitutes, are revealed by the analysis of the CoMSIA contour maps. The creation of novel thiazole compounds with extremely high pEC_{50} values will also benefit from these findings.

Keywords: Succinate dehydrogenase inhibitors, thiazolederivates 2D-QSAR, 3D-QSAR.

1. Introduction

A combination of respiratory proteins is called succinate dehydrogenase (SDH). Numerous studies demonstrate the toxicity of its SDHI inhibitors on specific organisms, including fish, frogs, bees and even people, where stopping mitochondrial respiration causes the pathogen to die [1-2].Since 1966, a number of chemical families, including oxathilinecarboxamides (carboxin), benzamides (fluopyram), paired carboxamides, carboxamides furan (fenfuram), thiophenecarboxamide (isofetamide) and pyrazolecarboxamides (fluxapyroxad, pydiflumetofen), have been marketed as succinate dehydrogenase (SDHI) inhibitors to protect plants.

SDHI has sparked interest in the development of a new fungicide due to its broad range of activity against several fungi [2]. However, many fungi have developed resistance and the capacity to stop their activity as a result of the extensive and regular use of SDHI fungicide [4-5]. Finding novel inhibitors that react and close these gaps has thus become essential. Recent developments in computational chemistry, specifically the QSAR approach, have given fresh and qualitative value to drug discovery and biological activity assessment prior to laboratory testing [4].

We have used a variety of thiazole derivatives in our work because of their diverse biological activity [5]. For instance, ethaboxam can cure a number of illnesses caused by oomycetes [6], and thiacloprid can disorient pollinators and insects by attacking their nerve systems [7].The goal of creating QSAR models is to enable the creation of novel drugs with high succinate dehydrogenase inhibitor activity against *S. sclerotiorum* (ssSDH) utilizing two methods: 3D and 2D QSAR.

2. Methodology

2.1 Database

Twenty thiazole derivative compounds with a range of biological activity reported by succinate dehydrogenase antifungal inhibitors against S were included in the database we used for our QSAR investigation. *S. sclerotiorum* were extracted from the literature [8]. Effective concentration (EC), which represents biological activity, was transformed into pEC_{50} using the formula $pEC_{50} = -\log (EC_{50})$ (Table 1). 80% of the learning set and 20% of the test set were randomly selected from this database.

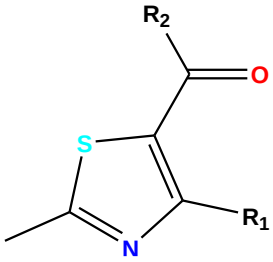
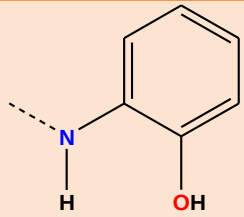
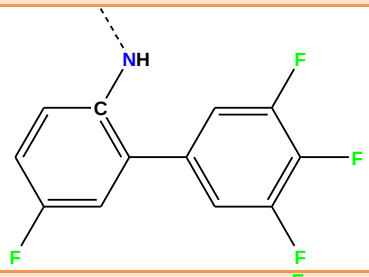
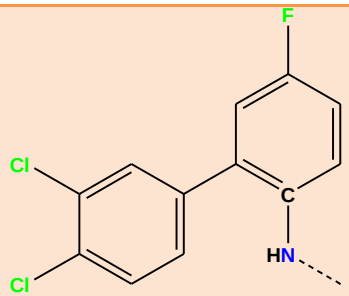
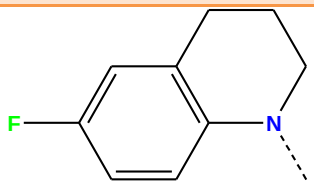
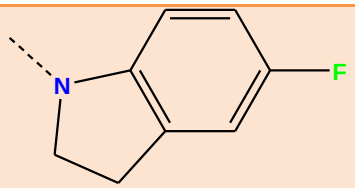
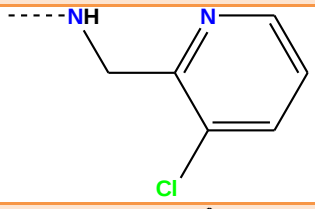
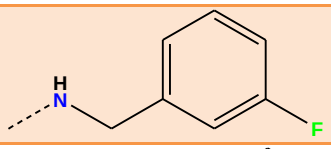
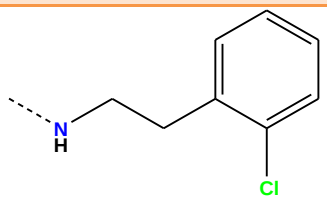
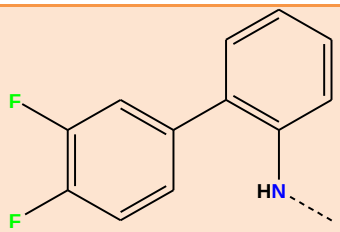
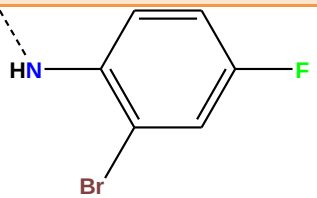
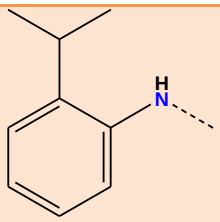
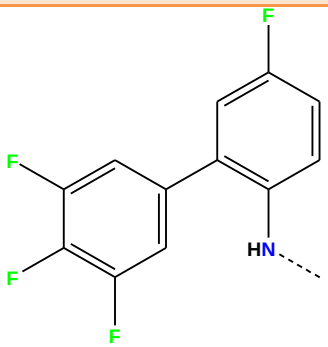
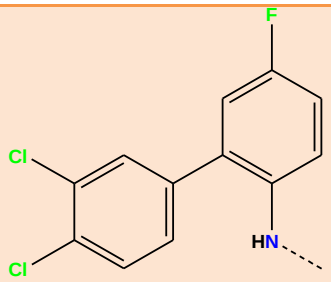
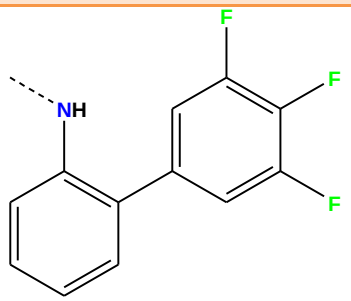
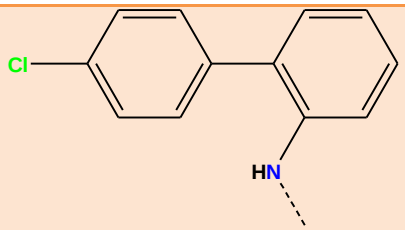
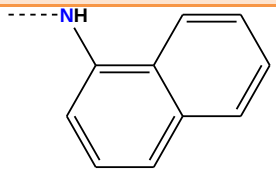
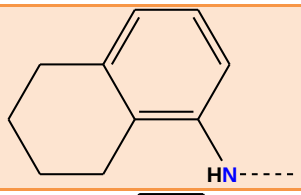
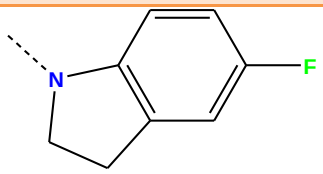
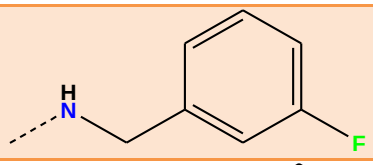
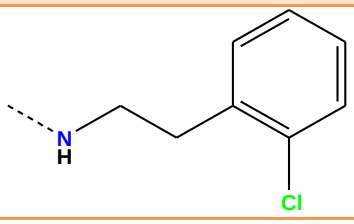


Fig. 1 Structure of studied thiazole derivatives

Table 1 Chemical structures and inhibitory activities of the compounds studied

Compound	R_1	R_2	pEC_{50}
1*	OCH_3		1.456
2*	OCH_3		1.770
3	OCH_3		1.833
4*	OCH_3		1.538

5	OCH_3		1.910
6	OCH_3		1.472
7	OCH_3		1.631
8	OCH_3		1.742
9	Propargyl		1.398
10	OH		3.000
11*	OH		2.886
12	OH		2.010

13	OH		1.975
14	OH		2.638
15	OH		1.854
16	OH		1.886
17	OH		2.097
18	OH		2.721
19	OH		2.319
20	OH		1.959

*-Test compounds

2.2 Molecular descriptors

Two programs have optimized the set of molecules: the Gaussian 09 program using the DFT method and the three-way hybrid function parameters (B3LYP)/6-31G (d) [12–13] and the ChemBioOffice program using the molecular mechanics (MM) method using the force field MM2 [9]. Additionally, the ACP/ChemSketch software was used to derive additional descriptors [11]. Table 2 compiles the results of all the descriptor calculations:

Table 2 Descriptors values used

pEC_{50}	MR	MV	MW	J	LogP	VDW	NRB	WI	Pc	αe	E_{LUMO}	E_{HUMO}	Gap	MD
1.456*	70.68	189.7	264.057	1.78	2.11	221.9	4	620	533.5	28.02	0.621	3.502	2.881	7.163
1.770*	93.37	273.4	396.056	1.61	4.63	303.41	5	1831	720.3	37.01	1.092	3.628	2.536	6.745
1.833	103.18	284.6	410.006	1.64	5.41	316.83	5	1670	770.7	40.9	1.266	3.695	2.429	6.756
1.538*	79.61	231.5	306.084	1.47	2.91	258.84	3	883	622.5	31.56	1.327	5.585	4.258	5.492
1.910	75	213.7	292.068	1.51	2.46	241.12	3	784	582.5	29.73	1.347	5.545	4.198	2.537
1.472	72.33	217.6	297.034	1.6	1.59	239.94	5	759	571.3	28.67	1.214	6.034	4.821	4.888
1.631	81.87	241.9	280.068	1.58	2.27	235.15	5	771	639.8	32.45	1.281	6.127	4.846	4.249
1.742	75.88	216.4	310.054	1.7	3.02	261.19	6	917	591.3	30.08	1.232	6.063	4.831	2.815
1.398	71.14	206.1	308.043	1.76	2.93	244.45	6	995	557.1	28.2	1.476	5.931	4.455	6.004
3.000	71.69	186.1	329.947	1.75	3.18	218.84	3	628	534.4	28.42	1.571	5.880	4.309	4.651
2.886*	78.19	215.7	276.093	1.72	3.52	246.8	4	713	591.9	30.99	1.364	4.167	2.803	3.783
2.010	88.57	247.8	382.040	1.66	4.49	285.78	4	1649	678.7	35.11	1.553	6.159	4.605	5.874
1.975	98.38	259.1	395.990	1.45	5.27	299.04	4	1499	729	39	1.723	5.902	4.179	4.296
2.638	88.58	243.6	364.049	1.45	4.35	280.92	4	1501	671.5	35.11	1.602	5.914	4.312	2.175
1.854	93.49	242.9	344.039	1.51	4.52	280.24	4	1219	686	37.06	14.241	5.958	-8.284	4.719
1.886	81.85	199.8	284.062	1.44	3.26	238.88	3	812	580.6	32.44	1.618	5.466	3.848	5.480
2.097	80.79	212.5	288.093	1.44	3.72	252.32	3	812	603.1	32.03	1.216	6.049	4.834	3.767
2.721	70.2	188.2	278.053	1.58	2.32	223.46	2	680	539.1	27.83	1.383	5.578	4.194	2.268
2.319	67.53	192	266.053	1.63	2.77	217.57	4	668	529.6	26.77	1.327	6.284	4.957	3.860
1.959	77.07	216.3	296.039	1.5	3.52	243.6	5	801	598.1	30.55	1.280	6.217	4.936	2.487

Where: MR- Molar refreactivity Å^3/mol); WI- Wiener Index; MV- Molecular volume cm^3); PC- Parachor (cm^3); Mw- Molecular weight; αe- Polarizability (cm^3); J- Balaban Index; E_{LUMO}- Lowest Unoccupied Molecular Orbital Energy (eV); Log P- Octanol/Water partition coefficient; E_{HOMO}- Highest Occupied Molecular Orbital Energy (eV); VDW- VANDER WALES Volume (cm^3); Gap- Different between HOMO and LUMO (eV); NRB- Number of Rotatable Bonds; DM- Dipole moment (debye)

2.3 2D-QSAR studies

The literature offers a variety of statistical mathematical techniques for conducting a QSAR analysis. The two most popular approaches for creating predictive QSAR models in the field of biological activity modeling will be sufficient for this investigation [12]: XLSTAT 2009 (www.xlstat.com) was used for the statistical analyses of multiple linear regression (MLR) and partial least squares method (PLS).

2.3.1 Multiple linear regression analysis (MLR)

The most popular method for determining a linear equation of a variable Y (pEC_{50}) as a function of the explanatory variables X (descriptors) is multiple regression analysis (MLR). The following is the expression for the multiple regression equation [13]:

$$Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \cdots\cdots\cdots + \alpha_p X_p$$

Where, Y: Output (pEC_{50}); α: Regression coefficient and X: Input variables (descriptors)

2.3.2 Partial least squares (PLS) analysis

Even when the inputs are numerous and highly collinear, the PLS regression approach is a well-liked regression methodology that may be utilized to evaluate the linear relationship between the dependent variable (Y) and several independent variables (X) [14].

2.4 3D-QSAR studies

For the 3D-QSAR study, the same dataset of 20 chemicals from the 2D-SQAR investigations was utilized once more. The biological activity of the chosen series was predicted using the CoMSIA approach based on the steric, electrostatic, hydrophobic, hydrogen bond donor, and hydrogen bond acceptor fields. In order to create a 3D-QSAR model, a linear relationship between the CoMSIA descriptors and the activity under study was established using the PLS approach [15].

2.4.1 Minimization and alignment procedure

The SYBYL-X2.0 program [16] was used to construct and minimize the structures of the studied compounds by using the conjugate gradient under the standard Tripos Field force field [17] with GasteigerHückel atomic partial charges [18] by the Powell method with a convergence criterion of 0.01 Kcal/mol and the maximum number of iterations is 10000. Molecular alignment is an important step in 3D-QSAR, linked to the orientation of molecules in 3D space, by the alignment technique in SYBYL-X2.0 all the compounds were aligned by considering compounds 10 as a model.

2.4.2 Calculation of CoMSIA descriptors

The CoMSIA method [19] is an extension of the CoMFA method [20] and consists of calculating descriptors of steric, electrostatic, hydrophobic, H-bond donor and H-bond acceptor fields on a 3D grid with the default grid spacing of 2, 0 Å along with the three Cartesian directions. The attenuation factor has been set to 0.3 [21].

2.5 Validation of the QSAR models

A number of criteria, including the coefficient of determination (R^2), the mean square error (MSE), the t-test, and the Fisher test, were used to assess the quality of the chosen models.

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y}_i)^2}$$

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2$$

Internal validation looked at the established QSAR model, whereas external validation evaluated the prediction capacity.

Internal verification- To verify the quality of the established models, the cross-validation approach was used to evaluate the internal validation of the models created using the Leave One-Out (LOOCV) (Q^2 LOO). Q^2 was computed using the following formula:

$$Q^2 = 1 - \frac{\sum(y_i - y'_i)^2}{\sum(y_i - y_{mean})^2}$$

Randomization test- Next, we conducted a randomization test to assess the resulting model. In this test, the values of pEC_{50} are randomly permutated multiple times without altering the matrix of the values of descriptors X. Following each iteration, the model is rebuilt.

External validation- The molecules in the test set, identified by R^2test, were used to assess the prediction potential of the chosen model.

The metricsΔrm^2,$\hat{r}^2m$ which have been computed using the following formulas [22], provide an extra, stricter external validation criterion, according to numerous investigations.

$$r_m^2 = r^2(1 + \sqrt{r^2 - r_0^2})$$

$$r'^2_m = r^2(1 + \sqrt{r^2 - r'^2_0})$$

Where,

- r^2 is the square correlation coefficient of the predicted and observed activity values of the test phase.

- r'^2_0 and r'^2_0 are respectively the coefficients for determining predicted values and those observed and the observed values and those predicted for the test set with zero interception.

Tropsha and his associates' criterion [23] state that a QSAR model has an appropriate predictive quality if the test set satisfies the following requirements.

- Q^2>0.5;
- R^2>0.6;
- (R^2- R_0^2) /R^2<0.1 and 0.85 < k < 1.15;
- (R^2- R'^2_0) /R^2<0.1 and 0.85 < k’ < 1.15.

Where,

k and k′ are respectively the slope of the predicted values to those observed and the observed values to those predicted for the test set with zero intercepts.

2.5 Applicability domain (AD)

The QSAR models were constructed on a set number of compounds, therefore do not encompass all the chemical space, for this applicability domain (DA) is an important tool allowing to test the predicted reliability of such a model on novel compounds. The application domain in our work was determined by calculating each compound's leverage (hi) as a function of standardized residues in the form of a William trace [24].

$$h_i = x_i^T (X^T X)^{-1} x_i$$

Where X expresses the descriptor matrix of the training set and xi is the descriptor vector of compound i, the critical value h* is represented as follows:

$$h^* = \frac{3(p+1)}{n}$$

With n is the number of compounds in the training set, p is the number of descriptors selected in the developed model.

3. Results and discussion

3.1 Principal component analysis (PCA)

Using the evaluation of the descriptor-descriptor and descriptor-activity correlation pEC50, we have reduced and summarized the number of descriptors in our database by analyzing its primary components. The nature of the relationship between the fourteen descriptors (variables) determined using coefficients of Pearson correlation (n). Table 3 shows that all of the descriptors aside from the Balaban Index J, Partition Coefficient Log P, Lowest Unoccupied Molecular Orbital Energy E_{LUMO} and Gap have a negative connection with pEC_{50}. The correlation coefficient often does not exceed 0.5%.

Table 3 Correlation matrix (Pearson (n)) between different obtained descriptors

	pEC_{50}	MR	MV	MW	J	LogP	VDW	NRB	WI	Pc	αe	E_{LUMO}	E_{HUMO}	Gap	MD
pEC_{50}	1														
MR	-0.083	1													
MV	-0.231	**0.925**	1												
MW	-0.011	0.847	0.834	1											
J	0.017	-0.426	-0.321	-0.153	1										
LogP	0.159	0.891	0.765	0.859	-0.301	1									
VDW	-0.160	**0.937**	**0.939**	**0.905**	-0.311	0.878	1								
NRB	-0.494	0.201	0.400	0.259	0.324	0.148	0.340	1							
WI	-0.118	0.859	0.880	**0.940**	-0.209	0.844	**0.939**	0.319	1						
Pc	-0.146	**0.980**	**0.979**	0.866	-0.389	0.855	**0.960**	0.293	0.893	1					
αe	-0.083	**1.000**	**0.925**	0.847	-0.426	0.891	**0.937**	0.201	0.859	**0.980**	1				
E_{LUMO}	-0.041	0.274	0.144	0.145	-0.176	0.253	0.190	-0.045	0.132	0.218	0.274	1			
E_{HUMO}	0.072	-0.225	-0.230	-0.178	-0.360	-0.174	-0.239	-0.038	-0.225	-0.211	-0.225	0.162	1		
Gap	0.065	-0.346	-0.217	-0.202	0.061	-0.309	-0.266	0.033	-0.204	-0.286	-0.346	**-0.949**	0.159	1	
MD	-0.499	0.179	0.199	0.259	0.411	0.122	0.195	0.180	0.272	0.171	0.179	0.014	-0.562	-0.194	1

However, if we look at the relationship between the descriptors, we can see that some of them have a strong association, such as MR, MW, MV, VDW, IW, αe and Pc. As a result, we can keep one descriptor and remove the rest. Gap and ELUMO have a strong inverse relationship (r = -0.949). These descriptors are shown as a circle of correlations in Fig. 2.

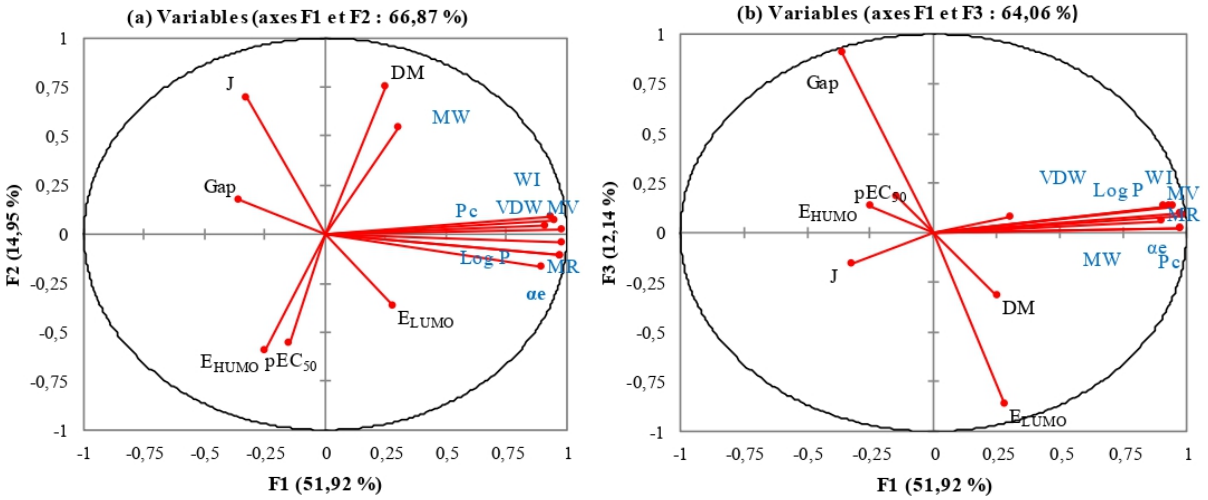


Fig. 2 Correlation circle for axes F1, F2 (a) and F1, F3 (b)

3.2 2D-QSAR models by regression analysis

The regression techniques developed were used to perform the 2D-QSAR modeling; the chosen models are shown by the following equations:

- Multiple linear regression analysis (MLR):

$$pEC_{50} = -0.526 + 2{,}502 * J + 0.178 * \log P - 0.311 * NRB - 0.172 * DM$$

- Partial least square method (PLS):

$$pEC_{50} = -0.113 + 2.281 * J + 0.134 * \log P - 0.339 * NRB - 0.123 * DM$$

Table 4 lists the values of the MLR and PLS models' statistically significant parameters:

Table 4 Statistical results of the 2D-QSAR equation generated by MLR, PCR and PLS methods

Statistical indicators	Results	
	MLR	PLS
n	16	16
R	0.89	0.88
R^2	0.80	0.78
DDL	11	13
F	13.344	12.23
P-value	0.0001	0.0001
MSE	0.054	0.041

The variation of the biological activity under study is explained by a coefficient of determination R^2 in the QSAR equations above. With improved coefficients of determination (R^2) that account for 83% and 82% of the activity variation, respectively, and mean square errors of roughly 0.054 and 0.039, the linear regression models MLR and PLS are statistically significant.

Table 6 Multi-colinearity test

Variable	VIP	
	MLR	PLS
J	1.252	0.098
Log P	1.602	0.247
NRB	1.454	1.702
DM	1.338	1.017

As indicated in Table 6, we employed the VIF variance inflation factors to determine whether or not there is collinearity between the chosen descriptors. The models created by MLR and PLS exhibit good stability, and the VIF values of four descriptors are less than 4.00, indicating that there is no collinearity between the chosen descriptors.

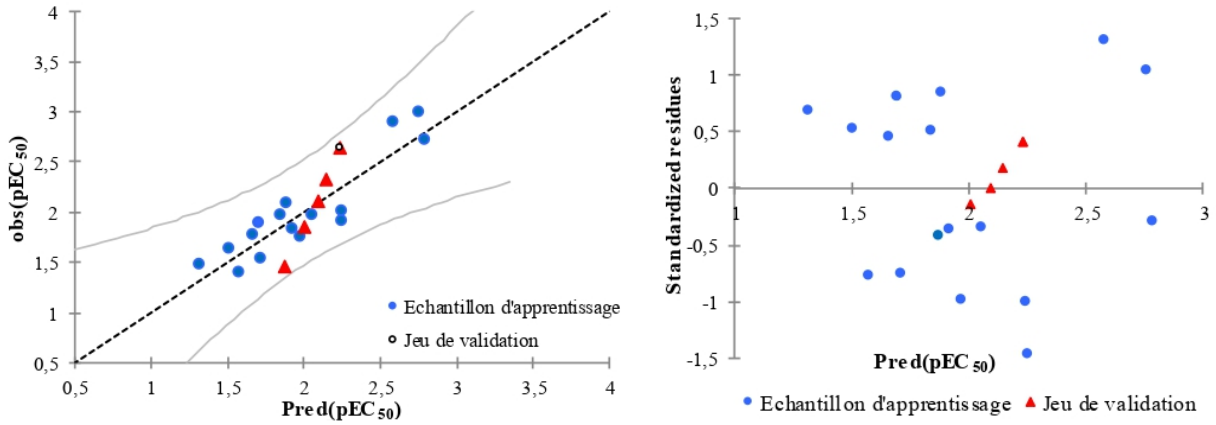


Fig. 3 Relationship between the observed pEC_{50} values and those predicted by the MLR model

The aforementioned findings indicate that four topological descriptors J, Log P and NRB as well as an electronic descriptor, DM, accounted for the biological activity of thiazole derivatives under study. Positive values of the regression coefficients of the models created by MLR and PLS indicate that the variables J and Log P contribute positively to the value of pEC50, whereas negative values indicate that the variables NRB and DM contribute negatively to the value of pEC_{50}. Figs. 3 and 4, respectively, demonstrate the correlation between the observed pEC50 values and those predicted by the MLR and PLS algorithms during the training and test sets.

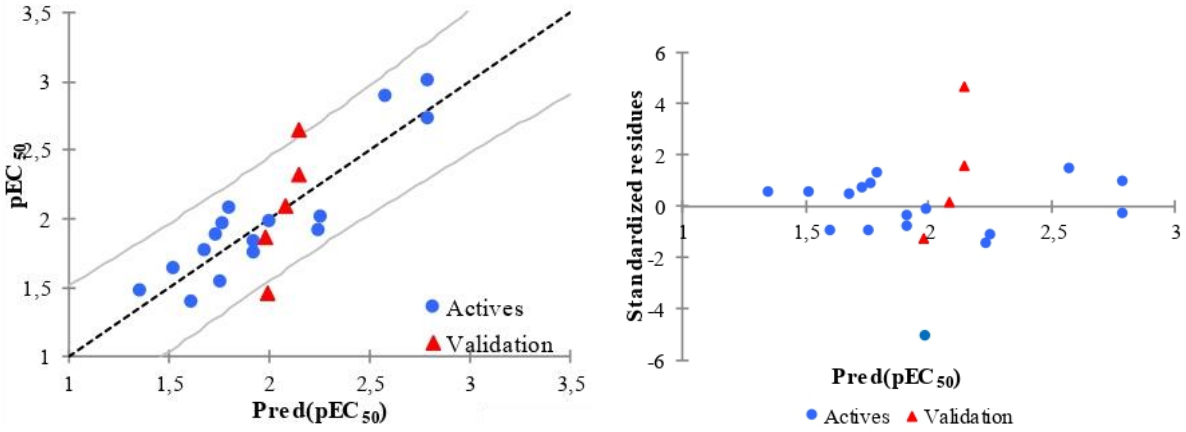


Fig. 4 Relationship between the observed pEC50 values and those predicted by the PLS model

3.3 3D-QSAR study

3.3.1 Molecular alignment

One of the most important steps in creating a 3D-QSAR model is molecular alignment. The set of molecules was aligned around compound 10, which was selected as the model. The alignment of every three-dimensional chemical structure in the shared nucleus is depicted in Fig. 5.

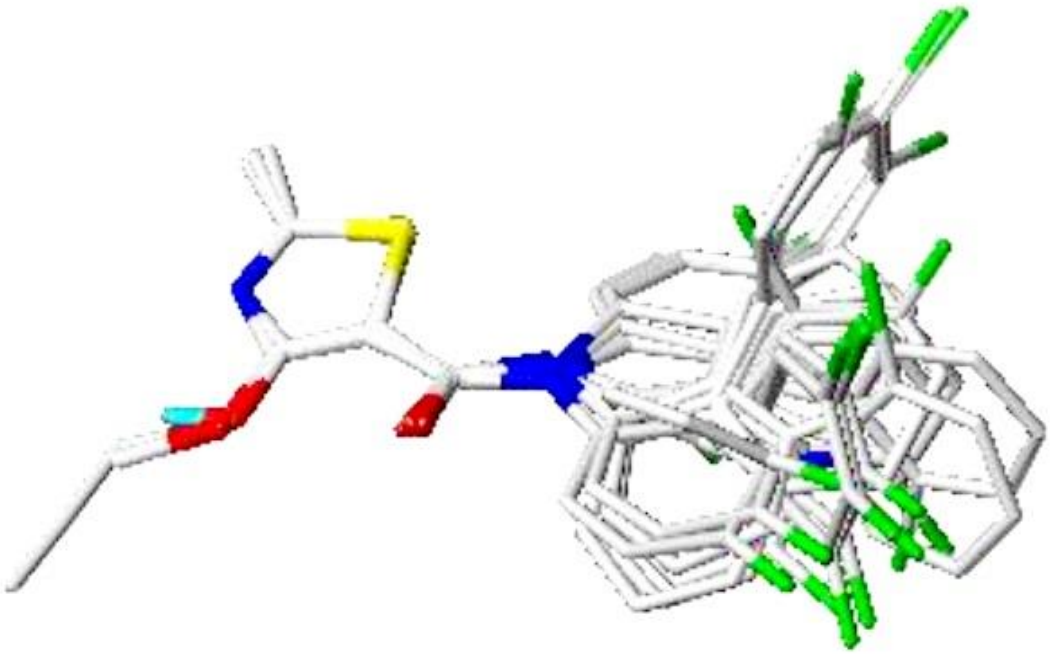


Fig. 5 3D-QSAR structure aligned compounds of the training set

3.3.2 CoMSIA results

We employed various combinations of the five fields steric (S), electrostatic (E), hydrophobic (H) and hydrogen bond donor (D) and acceptor (A) to create a CoMSIA model. Table 7 shows the results of the statistical parameters of the CoMSIA models, with the CoMSIA model/SEHD having the best biological activity predictions and providing the best statistical parameters R^2 = 0.957 with four as the ideal number of main components and a reliable SEE of 0.108 and a F test value of 60.61.

Table 7 Statistical results of CoMSIA models with different combinations of molecular fields

Models	R^2	Q^2	SEE	F	NOC	R^2test	Fractions				
							Ster.	Elec.	Hyd.	Don.	Acc.
CoMSIA/SE	0.865	0.343	0.190	17.603	2	0.780	0.538	0.462			
CoMSIA/SH	0.886	0.208	0.175	21.337	2	0.079	0.318		0.682		
CoMSIA/SD	0.883	0.448	0.177	20.761	3	0.846	0.689			0.311	
CoMSIA/SA	0.869	0.321	0.187	18.263	1	0.847	0.764				0.236
CoMSIA/EH	0.951	0.432	0.115	52.951	3	0.422		0.322	0.678		
CoMSIA/ED	0.774	0.389	0.246	9.397	2	0.228		0.480		0.520	
CoMSIA/EA	0.799	0.293	0.232	10.941	1	0.019		0.725			0.275
CoMSIA/HD	0.939	0.499	0.128	42.050	3	0.442			0.654	0.346	
CoMSIA/HA	0.934	0.360	0.133	38.970	4	0.056			0.790		0.210
CoMSIA/DA	0.643	0.360	0.309	4.950	2	0.053				0.509	0.491
CoMSIA/SEH	0.946	0.444	0.120	48.607	3	0.740	0.211	0.255	0.534		
CoMSIA/SED	0.882	0.538	0.178	20.477	4	0.716	0.367	0.255		0.378	
CoMSIA/SEA	0.879	0.341	0.180	19.961	1	0.811	0.493	0.377			0.129
CoMSIA/SHD	0.946	0.555	0.120	48.321	4	0.758	0.203		0.497	0.300	
CoMSIA/SHA	0.924	0.340	0.142	33.607	2	0.418	0.247		0.584		0.169
CoMSIA/SDA	0.899	0.413	0.165	24.388	3	0.733	0.525			0.305	0.171
CoMSIA/EHD	0.960	0.597	0.103	66.747	4	0.513		0.195	0.485	0.319	
CoMSIA/EDA	0.799	0.387	0.232	10.930	2	0.224		0.430		0.436	0.135
CoMSIA/EHA	0.955	0.356	0.109	58.617	4	0.171		0.240	0.639		0.121
CoMSIA/HDA	0.940	0.466	0.127	42.986	3	0.328			0.612	0.309	0.079
CoMSIA/SEHA	0.956	0.384	0.108	59.993	2	0.807	0.157	0.260	0.505		0.078
CoMSIA/SEHD	0.957	0.614	0.108	60.612	4	0.80	0.15	0.15	0.415	0.285	
CoMSIA/SHDA	0.945	0.500	0.121	47.659	4	0.710	0.190		0.470	0.281	0.058
CoMSIA/SEDA	0.896	0.523	0.167	23.670	4	0.679	0.327	0.229		0.340	0.105
CoMSIA/EHDA	0.959	0.564	0.105	64.395	4	0.497		0.181	0.467	0.296	0.056
CoMSIA/ ALL	0.978	0.587	0.081	88.046	5	0.667	0.15	0.136	0.378	0.294	0.042

R^2 = is the square of the non CV coefficient; Q^2 = the square of the LOO cross-validation (CV) coefficient; R^2test = the square of prediction coefficient; N = the optimum number of components; SEE is the standard error of estimation of non CV analysis; F is the F-test value.

S, E, H, D and A are the steric, electrostatic, hydrophobic, hydrogen-bond donor and hydrogen-bond acceptor contributions, respectively.

The statistical indicators shown in Table 7 demonstrate that the CoMSIA model of the combination of steric, hydrophobic, electrostatic and hydrogen bond donor (SEHD) has significant statistical indicators and is capable. The various combinations of the fields mentioned were carried out to establish the different 3D CoMSIA models.

The contributions of the steric, electrostatic, hydrophobic and hydrogen bond donor fields account for 15%, 15%, 41.5% and 28.5% of the variance in a high-performance pEC_{50} prediction. Table 8 compiles the various observed and predicted values for the 2D and 3D models developed with its residuals. The coefficient of determination of the model not validated R^2 CoMSIA/SEHD was of the order of 0.957 with an estimate of the

standard error SEE reliably of 0.105. The value of test F is 60.612 and the optimal number of principal components is to take the value 4.

Table 8 Values predicted by the 2D-QSAR / MLR / PLS and 3D-QSAR / CoMSIA models

pEC_{50}	MLR		PLS		CoMSIA	
	Pred.(pEC_{50})	Residual	Pred.(pIC_{50})	Residual	Pred.(pIC_{50})	Residual
1.833	1.825	0.007	1.826	0.007	1.861	-0.028
1.910	2.322	-0.412	2.331	-0.421	1.961	-0.051
1.472	1.367	0.106	1.452	0.020	1.464	0.008
1.631	1.547	0.084	1.576	0.055	1.556	0.075
1.742	1.916	-0.174	1.788	-0.045	1.834	-0.092
1.398	1.503	-0.105	1.520	-0.122	1.334	0.064
3.000	2.687	0.313	2.715	0.285	3.052	-0.052
2.010	2.174	-0.164	2.196	-0.186	2.207	-0.197
1.975	2.058	-0.084	2.015	-0.041	1.973	0.002
2.638	2.259	0.379	2.153	0.485	2.452	0.186
1.854	2.003	-0.149	2.000	-0.146	1.915	-0.061
1.886	1.784	0.103	1.917	-0.031	1.958	-0.072
2.097	2.159	-0.063	2.189	-0.092	2.030	0.067
2.721	2.829	-0.108	2.845	-0.123	2.667	0.054
2.319	2.139	0.180	2.144	0.174	2.205	0.114
1.959	1.872	0.087	1.778	0.181	1.977	-0.018
1.456	1.830	-0.374	1.992	-0.536	1.781	-0.325
1.770	1.614	0.156	1.654	0.116	1.737	0.033
1.538	1.794	-0.257	1.937	-0.399	1.900	-0.362
2.886	2.511	0.375	2.460	0.426	2.208	0.678

3.4 Validation of QSAR models developed by 2D / 3D-QSAR

3.4.1 Internal validation

The reliability and static relevance of the QSAR suggested by the 2D/3D-QSAR algorithms (MLR, PLS and CoMSIA) were assessed internally using cross-validation with no return (LOOCV). As a result, all three of the models developed have square cross-validation coefficients Q^2 larger than 0.6: $Q^2 = 0.63$ for the MLR model and $Q^2 = 0.64$ for the PLS model. Q2 = 0.64 for the 3D-QSAR CoMSIA model. These cross-validation findings demonstrate the relevance of the descriptors and the ability of the models suggested by 2D/3D QSAR to predict activity with excellent performance.

3.4.2 External validation

Four randomly selected compounds of thiazole derivatives that were not included in the training phase were subjected to external validation in order to confirm the predictive power of the best QSAR models created by the 2D and 3D-QSAR approaches (MLR, PLS and CoMSIA). The test results for the set of four thiazole derivative chemicals reveal that both the MLR and PLS models have strong external predictive power, with the MLR model having an advantage (R^2test = 0.82 for MLR and R^2test = 0.62 for PLS).

Similarly, the CoMSIA model recorded an acceptable coefficient of determination (R2) with a value of R^2test = 0.80 in the test set to verify and ensure that the models provided by 2D and 3D-QSAR are externally validated. In order to meet Tropsha and his collaborators criterion, we calculated the measures Δrm^2 and $\hat{r}^2 m$. The findings are summarized in Table 8.

Table 8 Golbraikh and Tropsha test results

Parameter	Validation Criteria	2D-QSAR		3D-QSAR
		MLR	PLS	CoMSIA/SEHA
R^2	R^2>0.6	0.82	0.62	0.80
k	0.85 <k< 1.15	1.003	0.964	1.020
$(R^2 - \mathbf{R_0^2}) /R^2$	$(R^2 - R_0^2) /R^2$ <0.1	-0.226	-0.594	-0.250
k’	0.85 <k’< 1.15	0.974	0.997	0.938
$(R^2 - \mathbf{R'^2_0}) /R^2$	$(R^2 - \mathbf{R'^2_0}) /R^2$ <0.1	-0.224	-0.602	-0.237
$\mathbf{r_m^2}$	$\mathbf{r_m^2}$ > 0.5	0.693	0.530	0.679
$\mathbf{r'^2_m}$	$\mathbf{r'^2_m}$ > 0.5	0.709	0.542	0.695
$\hat{\mathbf{r}}_\mathbf{m}^\mathbf{2}$	$\hat{\mathbf{r}}_\mathbf{m}^\mathbf{2}$> 0.5	0.701	0.536	0.687
$\Delta\mathbf{r_m^2}$	$\Delta\mathbf{r_m^2}$<0.2	0.016	0.013	0.016
$\Delta\mathbf{r_0^2}$	$\Delta\mathbf{r_0^2}$ < 0.3	0.002	-0.530	0.011

Table 8 unequivocally demonstrates that the constructed 2D and 3D-QSAR models fully adhere to the Golbraikh and Tropsha criteria; this outcome validates the models' strong external prediction.

❖ Randomization test

To evaluate the resilience of the models produced by 2D/3D-QSAR, our data set was randomized. Ten times, this procedure was carried out.

Table 9 Randomization test for MLR, PLS and CoMSIA models

Model	2D-QSAR (MLR)				2D-QSAR (PLS)				3D-QSAR (CoMSIA)			
	R^2	MSE	Q^2	R^2test	R^2	MSE	Q^2	R^2test	R^2	MSE	Q^2	R^2test
Random1	0.225	0.136	-0.310	-0.230	0.230	0.240	0.160	0.021	0.230	0.112	0.205	0.130
Random2	0.420	0.142	0.270	-0.290	0.290	0.160	0.220	0.350	0.132	0.152	0.211	0.240
Random3	0.240	0.140	0.170	0.025	0.450	0.190	-0.290	0.023	0.260	0.310	0.341	0.360
Random4	0. 31	0.126	0.340	0.150	0.150	0.030	0.190	-0.270	0.207	0.241	-0.240	-0.420
Random5	0.120	0.340	-0.260	0.041	0.410	0.030	0.33	-0.315	0.321	0.187	0.401	0.210
Random6	0.410	0.013	0.210	0.390	0.190	0.067	-0.240	0.270	0.120	0.191	0.134	0.370
Random7	0.030	0.163	0.150	0.137	0.170	0.017	0.280	-0.130	0.320	0.201	0.256	-0.440
Random8	0.360	0.210	-0.241	-0.230	0.230	0.020	0.220	0.360	0.147	0.302	-0.319	0.310
Random9	0.191	0.312	0.040	0.420	0.232	0.100	0.340	-0.210	0.201	0.140	0.177	0.290
Random10	0.290	0.192	0.160	0.310	0.137	0.109	0.110	-0.164	0.124	0.192	0.314	0.430
Average	**0.254**	**0.177**	**0.052**	**0.072**	**0.249**	**0.096**	**0.11**	**-0.006**	**0.206**	**0.203**	**0.148**	**0.148**
Original	**0.80**	**0.054**	**0.63**	**0.82**	**0.78**	**0.041**	**0.64**	**0.62**	**0.86**	**0.043**	**0.35**	**0.86**

The results of the model randomization test obtained by 2D/3D-QSAR are compiled in Table 9. The results show that all of the new models have values of R^2 and Q^2 lower than those of the original models, indicating that the original models are no longer robust and are not the result of chance.

3.5 Applicability domain (DA)

Leverage analysis was used to evaluate outliers in the domain of application, as illustrated in Fig. 6, which shows each compound's leverage as a function of the normalized residues. The values of the levers for the dataset are all less than the critical value h * = 0.75 (hi < h *), which indicates that there are no influential points and outliers during the training and test sets. Fig. 6 shows that all of the standardized residues are less than 2 standard deviation units 2σ. Consequently, this outcome demonstrates that their action anticipated the values' dependability.

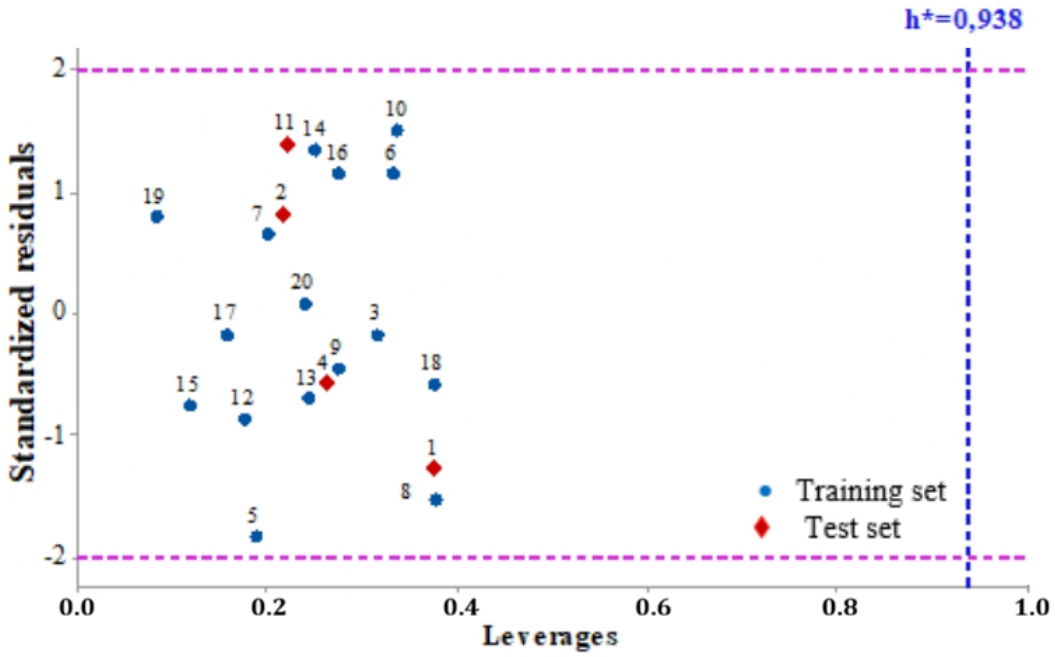


Fig. 6 Williams plot of standardized residual versus leverage for the model 2D-QSAR

3.6 Graphical interpretation of the CoMSIA model

Using the most active chemical in the data set (compound 10) as a reference, the CoMSIA contour map lets you see the data in the chosen 3D-QSAR model. The chosen model contour map (SEHA) of steric, electrostatic, hydrophobic, and hydrogen bond donor fields is shown in Fig. 7.

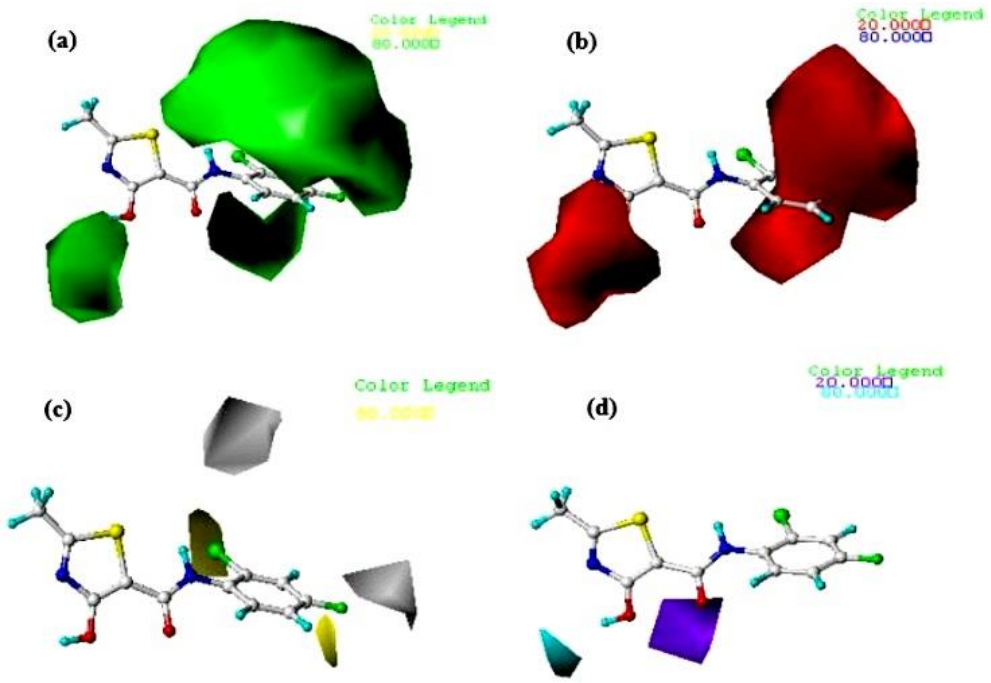


Fig. 7 Contour map of CoMSIA analysis (a) Steric, (b) electrostatic and (c) hydrophobic (d) hydrogen bonding donor with a grid spacing of 2.0 Å in combination with compound 10

The green contours (contribution 80%) in Fig. 7a steric contour maps show the areas where bulky substitutions boost the biological activity under investigation. However, the areas where the bulky substituents reduce the activity are indicated by the yellow-colored contours (20% contribution). The places where the positive electrostatic groups have been favored are indicated by the blue contours (80% contribution) in the CoMSIA electrostatic contour maps, Fig. 7b, whilst the negative electronegative groups are favored in the red regions (20% contribution). The favorable areas in the hydrophobic grouping (contribution of 80%) are indicated by the yellow contour in the CoMSIA hydrophobic contour maps Fig. 7c, while the favorable areas in the hydrophilic groups (contribution of 20%) are indicated by the white contour.

Additionally, in the CoMSIA hydrogen bond donor contour maps Fig. 7d, the purple contours (20% contribution) indicate the disadvantaged region, but the cyan contour maps (80% contribution) indicate that the hydrogen bond donor groups improve activity.
The various types of substitutions that affect the activity under study and can be used as a foundation for the creation of novel compounds with high pEC_{50} values are summarized in Fig. 8. This also enables us to describe how the action under study came to be.

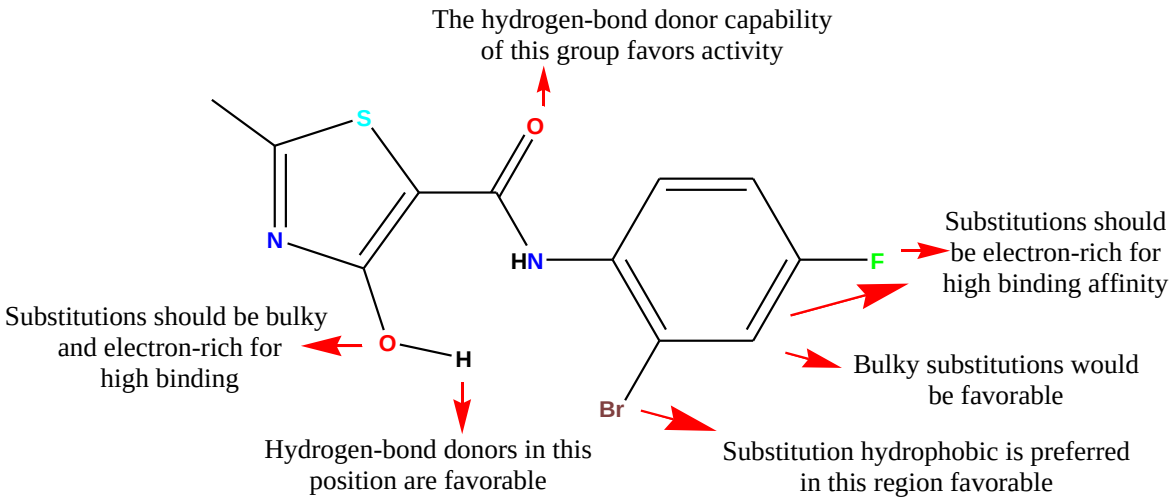


Fig. 8 Figure structure-activity relationship derived from CoMSIA- SEHD

4. Conclusion

With statistical values of $R^2 = 0.80$ and $Q^2 = 0.63$ for the MLR model and $R^2 = 0.78$ and $Q^2 = 0.64$ for the PLS model, the 2D-QSAR results showed that the two models created by the MLR method and the PLS method had significant statistical performance and good predictive capacity. An external prediction of the MLR model (R^2test = 0.82) was more powerful than the PLS model (R^2test = 0.62). In order to anticipate the efficacy of succinate dehydrogenase inhibitors against *S. sclerotiorum* pEC*50*, the most crucial parameters are the Balaban Index J, partition coefficient Log P and number of rotatable bonds NRB and the Dipole Moment MD, with J and Log P contributing negatively and NRB and MD contributing positively to the pEC_{50} values. The CoMSIA model exhibits an excellent predictive ability for pEC_{50} values with a mixture of steric, electrostatic, hydrophobic, and hydrogen bond donor fields (SEHD), with statistical values $R^2 = 0.957$, $Q^2 = 0.614$ and R^2test = 0.80, according to 3D-QSAR experiments. Additionally, the CoMSIA contour analysis results offered crucial information that can help clarify the structure-activity relationship and pinpoint the kind and positrons of substitutions that can enhance the biological activity under study. Consequently, novel molecules with strong inhibitory activity can be predicted.

Acknowledgements
We are grateful to the Department of Chemistry, PMCOE, Govt. Science College, Rewa (M.P.) for its pertinent help concerning the research.

References

[1] F. Sun. Cell,. 2005 121, no. 7. 1043–1057.
[2] C. E. Keohane. J. Am. Chem. Soc. 2018. 140. 5. 1774–1782.
[3] R. M. De Miccolis Angelini, M. Masiello, C. Rotolo, S. Pollastro, and F. Faretra, Pest Manag. Sci. 2014. 70. 12. 1884–1893.
[4] A. Ghaleb, A. Aouidate, M. Ghamali, A. Sbai, M. Bouachrine, and T. Lakhlifi, J. Mol. Struct., 2017. 1145. 278 284.
[5] S. Kumar and R. Aggarwal, Mini. Rev. Org. Chem. 2019 . 16. 1. 26–34. [6] [7] [8]
[6] D. J. White, W. Chen, and K. L. Schroeder, Crop Prot. 2019. 115. 7–12.
[7] M. Imran, Insect Pest Management, IntechOpen, 2020.

[8] X. Guo et al., J. Agric. Food Chem., 2019. 67. 6. 1647–1655.
[9] N. L. Haworth and L. L. Martin, J. Chem. Educ. 2018. 95. 12. 2256–2262.
[10] L. Petit, P. Maldivi, and C. Adamo, J. Chem. Theory Comput. 2005. 1. 5. 953 962.
[11] C. V. ACD, "12, Advanced Chemistry Development," Inc., Toronto, ON, Canada, 2009.
[12] K. TABTI, RHAZES Green Appl. Chem. 2020. 9. 70–91.
[13] L. Elmchichi, A. Belhassan, A. Aouidate, A. Ghaleb, T. Lakhlifi, and M. Bouachrine, Phys. Chem. Res.2020. 8. 1. 125–137.
[14] A. Aouidate, Struct. Chem., 2018. 29. 4. 1031–1043.
[15] B. L. Bush and R. B. Nachbar, J. Comput. Aided. Mol. Des., 1993. 7. 5. 587–619.
[16] M. D. M. AbdulHameed, A. Hamza, J. Liu, and C.-G. Zhan, J. Chem. Inf. Model. 2008. 48. 9. 1760–1772.
[17] M. Clark, R. D. Cramer III, and N. Van Opdenbosch, J. Comput. Chem. 1989. 10. 8. 982–1012.
[18] W. P. Purcell and J. A. Singer, J. Chem. Eng. Data, 1967. 12. 2. 235–246.
[19] G. Klebe, U. Abraham, and T. Mietzner, J. Med. Chem. 1994. 37. 24. 4130–4146.
[20] P. Cramer, J. Pers. 1991. 59. 1. 39–55.
[21] J. Zheng, Chem. Biol. Drug Des. 2011. 78. 2, pp. 314–321.
[22] K. Roy, Expert Opin. Drug Discov. 2007. 2. 12. 1567–1577.
[23] A. Golbraikh and A. Tropsha, J. Mol. Graph. Model. 2002. 20. 4. 269–276.
[24] S. Chtita, New J. Chem., 2020. 44. 5. 1747–1760.