

# 3D-QSAR Modeling of Substituted Thiophene-Anthranilamides as Potent Inhibitors of Human Factor Xa Using Quantum Chemical Descriptors

K. Dguigui<sup>1</sup>, M. Elhallaoui<sup>2</sup>

<sup>1,2</sup>Équipe de Chimie Informatique et modélisation, laboratoire d'ingénierie des matériaux, de modélisation et d'environnement, LIMME, Faculté des sciences Dhar El Mehraz, Université Sidi Mohammed Ben Abdullah, USMBA. BP 1796, 30000, Atlas, Fes, Morocco.

**Abstract:** *The potent inhibitors of human factor Xa of 54 thiophene derivatives were modeled by quantitative structure-activity relationship (QSAR) using density functional theory (DFT) to generate quantum descriptors. From the pool of descriptors chosen to generate the QSAR model, four descriptors are selected by multiple linear regression MLR method with a correlation coefficient  $R_{RLM}=0,95$ . The predictive ability of the proposed model was assessed using neural network  $R_{NN}=0,98$  and validated by internal leave one out cross validation  $R_{CV}=0,90$ .*

**Keywords:** factors Xa inhibitor, anticoagulant, QSAR, MLR, Neural Network, Cross Validation.

## 1. Introduction

Anticoagulant therapy is the primary strategy for treatment and prevention of thromboembolic disease. Warfarin (Coumadin). A vitamine K antagonist, is the only oral anticoagulant in clinical use. Coumadin is highly efficacious, but suffers from a delayed onset of action, and the need for careful clinical monitoring. Alternate anticoagulants such as the low molecular weight heparins and the direct thrombin inhibitor hirudin must be administered parenterally. The unmet clinical need for a safe and orally active anticoagulant has resulted in a widespread drug discovery effort [1].

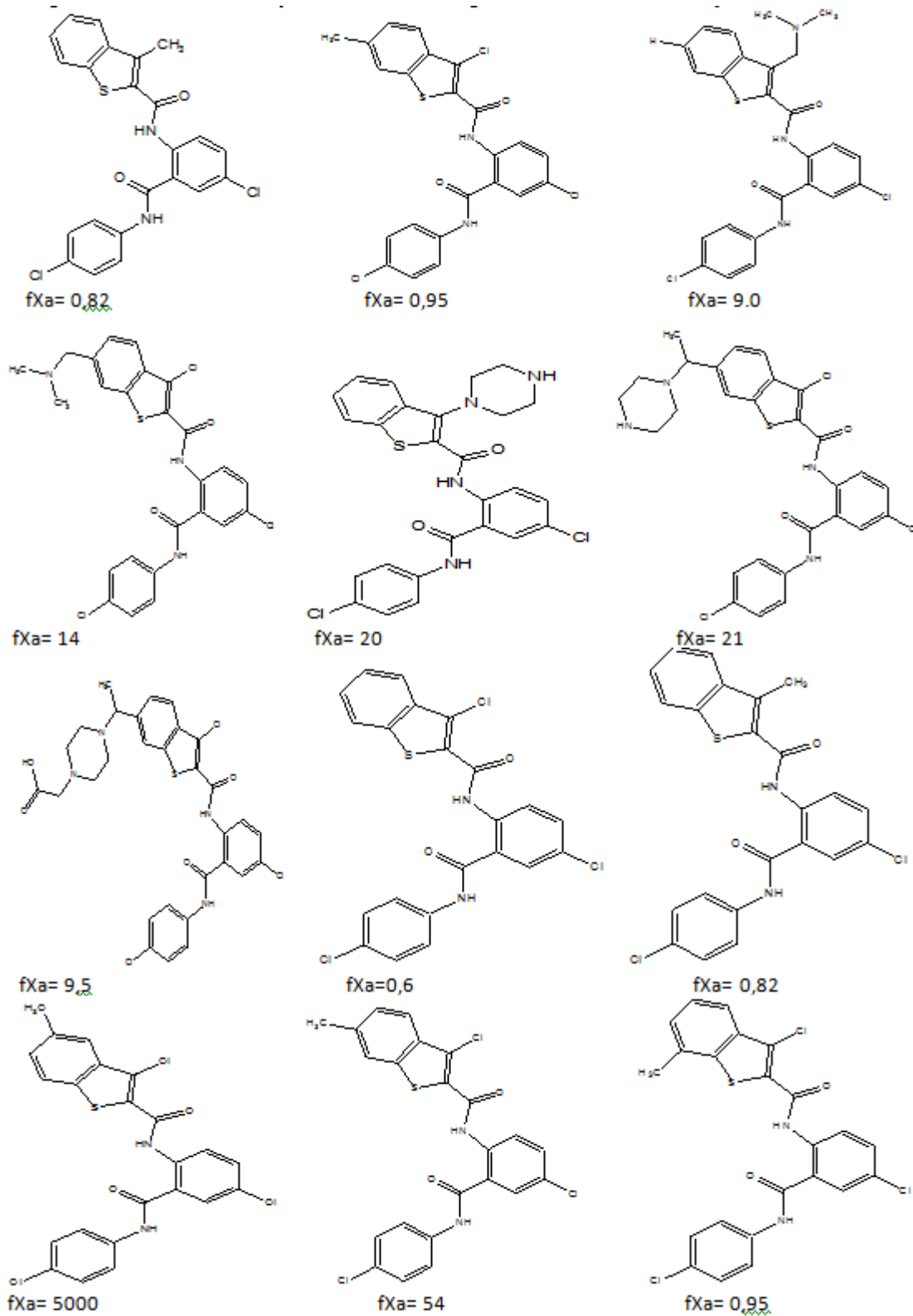
Factor Xa (fXa) is a trypsin like serine protease that forms a prothrombinase complex with ion  $Ca^{2+}$ , and phospholipids to produce thrombin. The key enzyme functions at the convergence of the intrinsic and extrinsic coagulation pathways in a process that involves signal amplification. One molecule of fXa activates many molecules of prothrombin to thrombin. As inhibition of fXa prevent thrombin formation without affecting pre-existing thrombin, fXa inhibitors are predicted to cause less impairment of hemostasis than direct thrombin inhibitors, leading to wider therapeutic window [2].

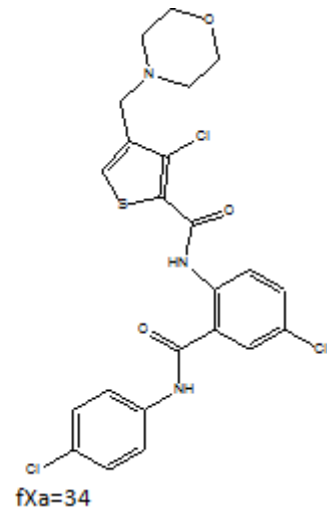
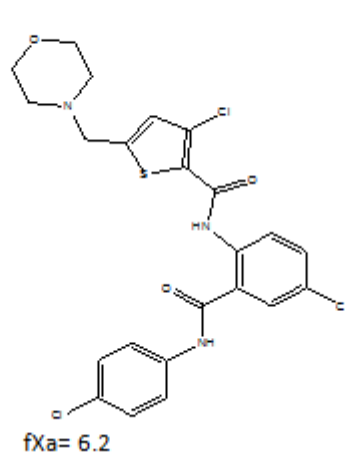
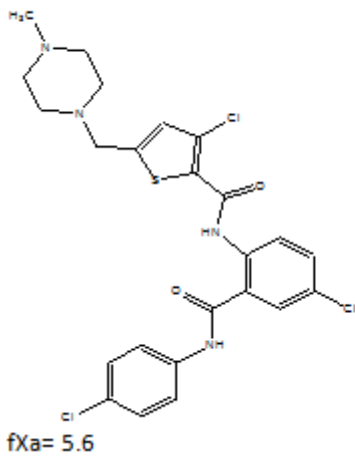
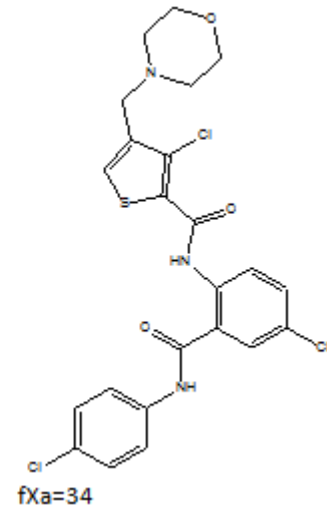
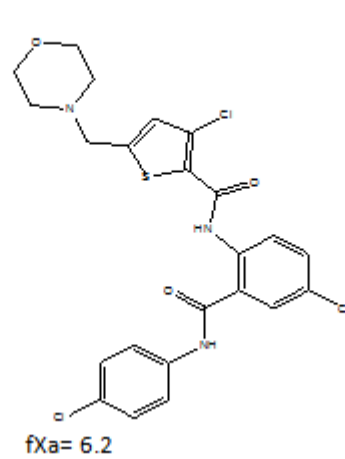
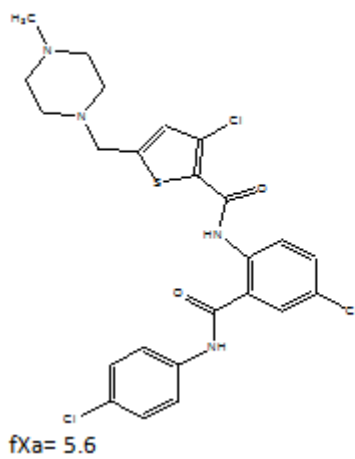
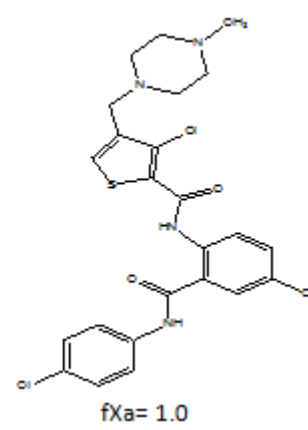
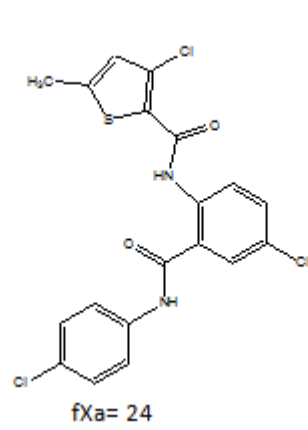
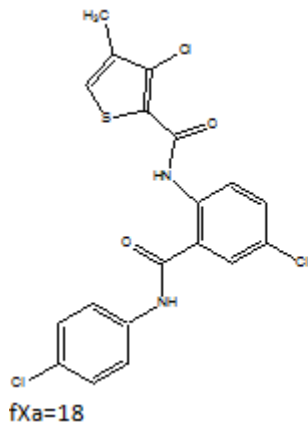
The discovery of orally active, small molecule competitive fXa inhibitors in preclinical animal thrombosis models in recent years has spurred the process for discovery of such molecules as antithrombotic drugs. Thiophene substituted anthranilamides have been reported as novel and potent inhibitors of human fXa. In this work, we used QSAR methods to study a series of thiophene derivatives tested for their inhibitory activity of human factors Xa. QSAR methods are among the most practical tools in computational physical chemistry. These methods are based on the axiom that the variance in the physicochemical properties and activities of chemical compounds is determined by the variance in their molecular structures [3]. Therefore, the aim of the present study is to develop QSAR model identifying the relevant molecular descriptors reliable with the activities of thiophene derivatives.

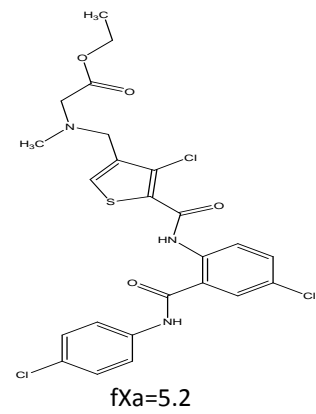
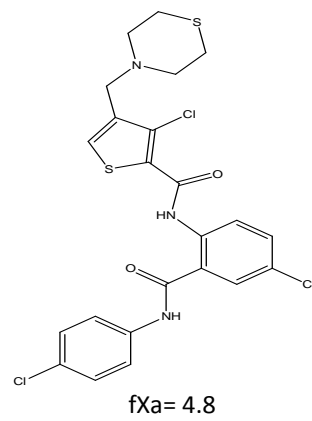
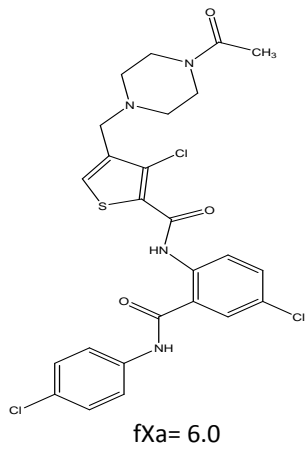
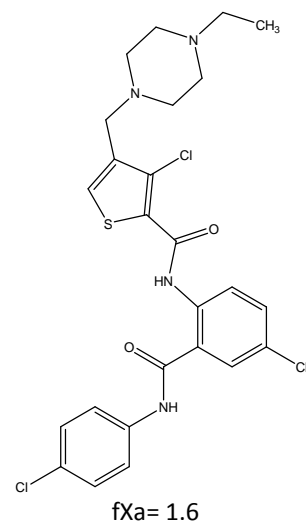
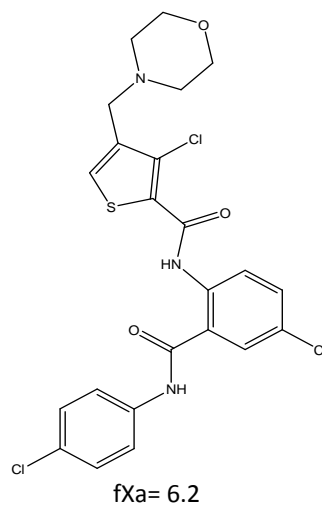
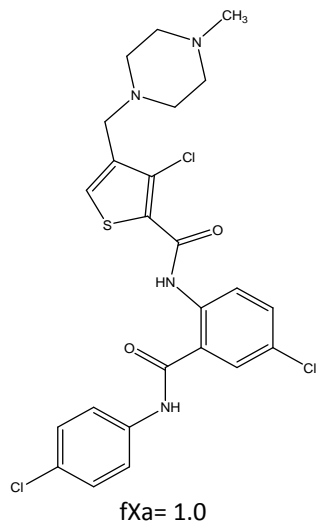
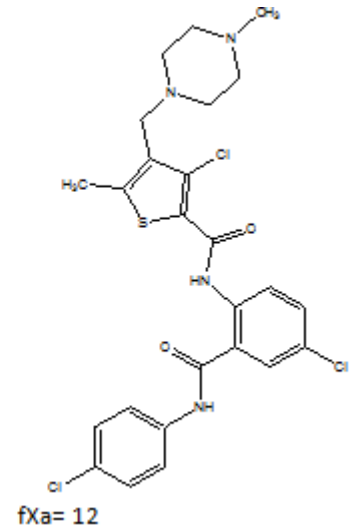
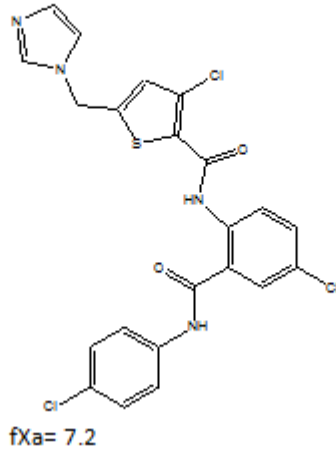
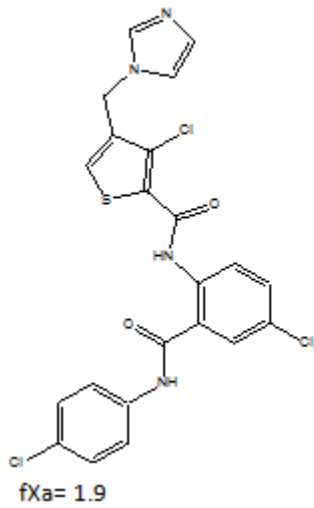
## 2. Materials and Methods

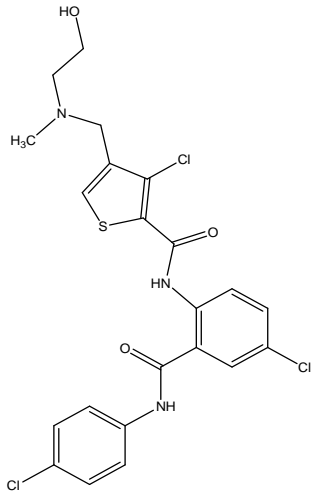
### 2.1 Data Set

A total of 54 compounds of thiophene-anthranilamides are sketched in the below figure with their experimental activities. All original activities fXa values are converted into logarithm scale in QSAR study.

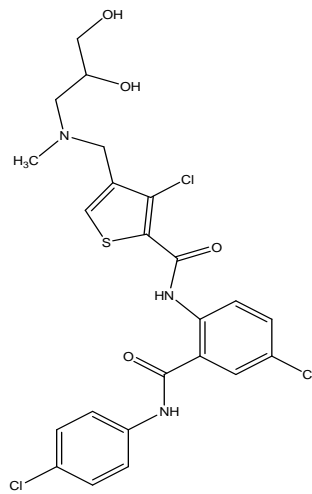




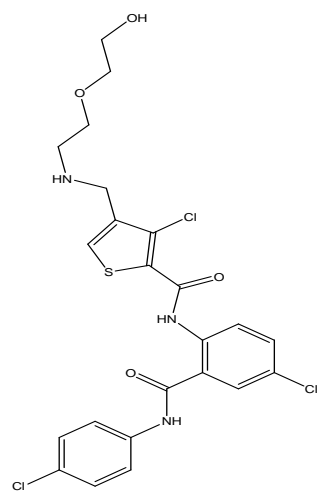




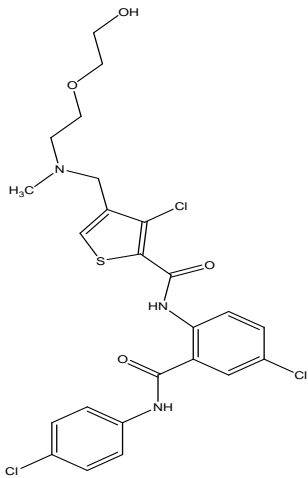
fXa=0.8



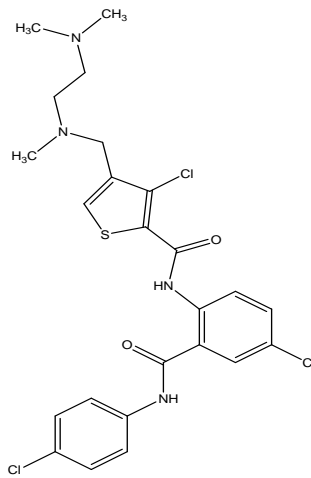
fXa= 1.4



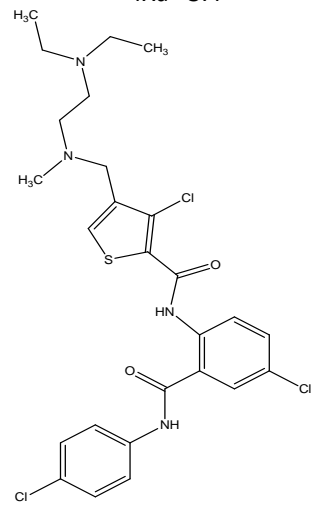
fXa= 3.4



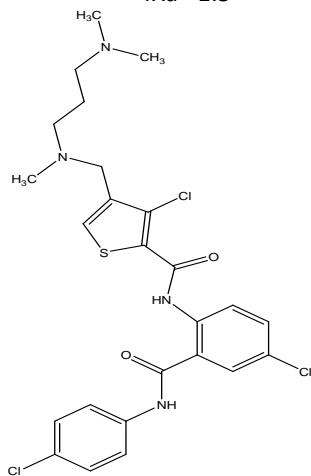
fXa= 1.5



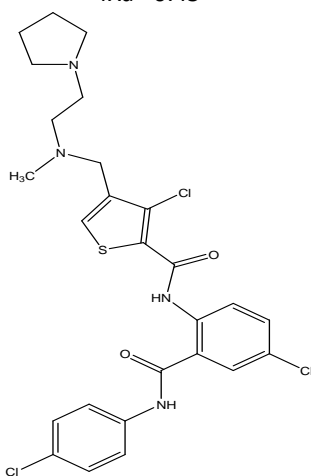
fXa= 0.45



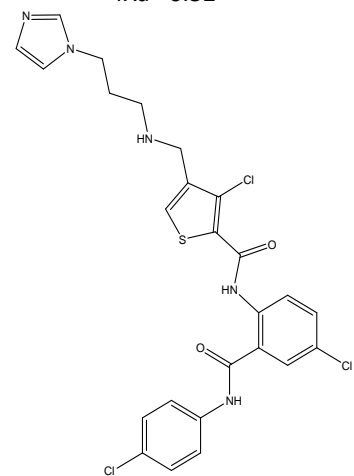
fXa= 0.52



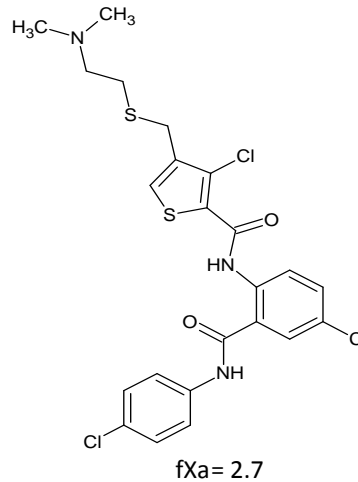
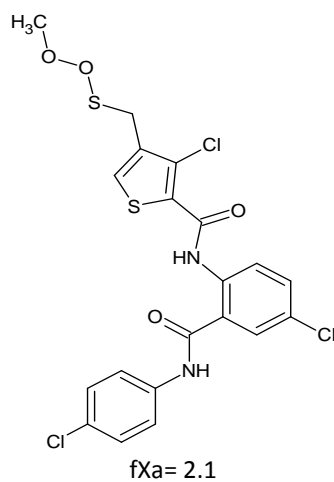
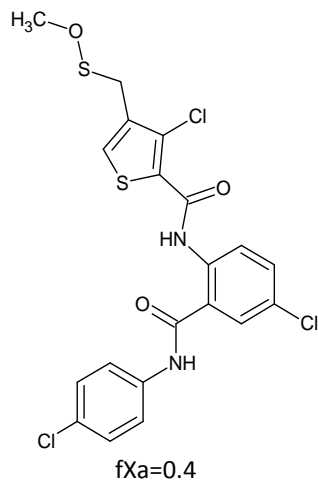
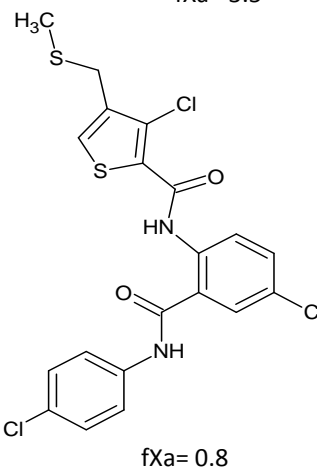
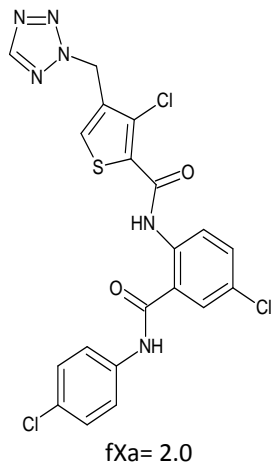
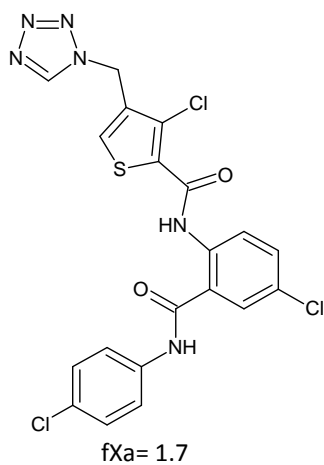
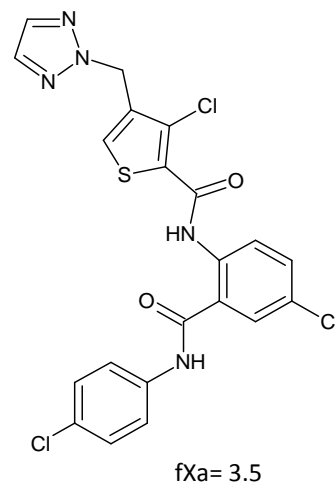
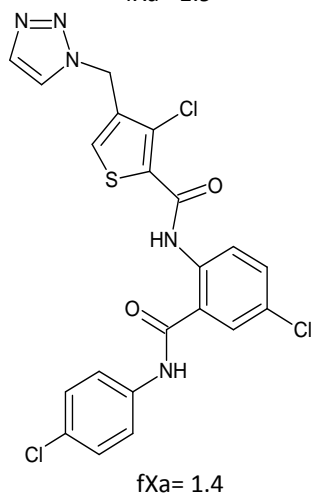
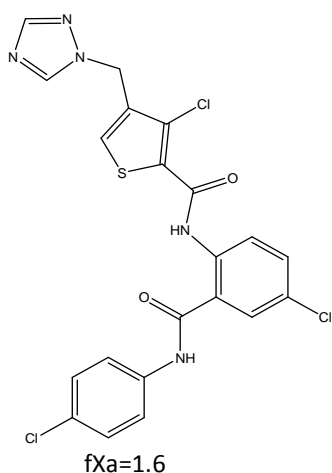
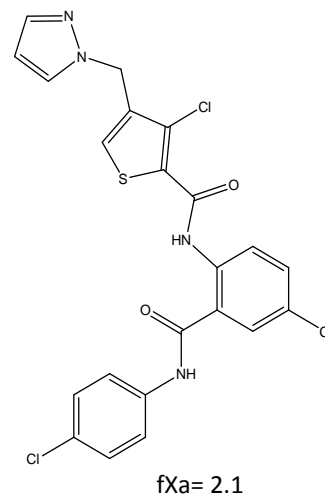
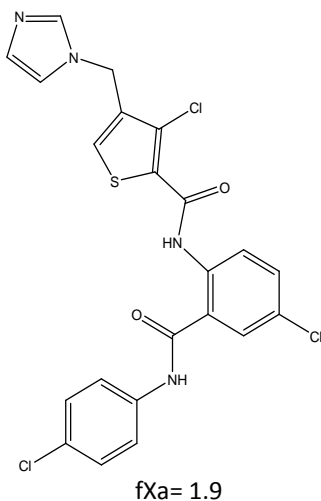
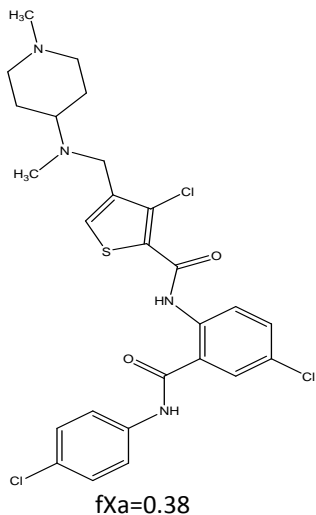
fXa= 0.6



fXa= 0.3



fXa= 4.6



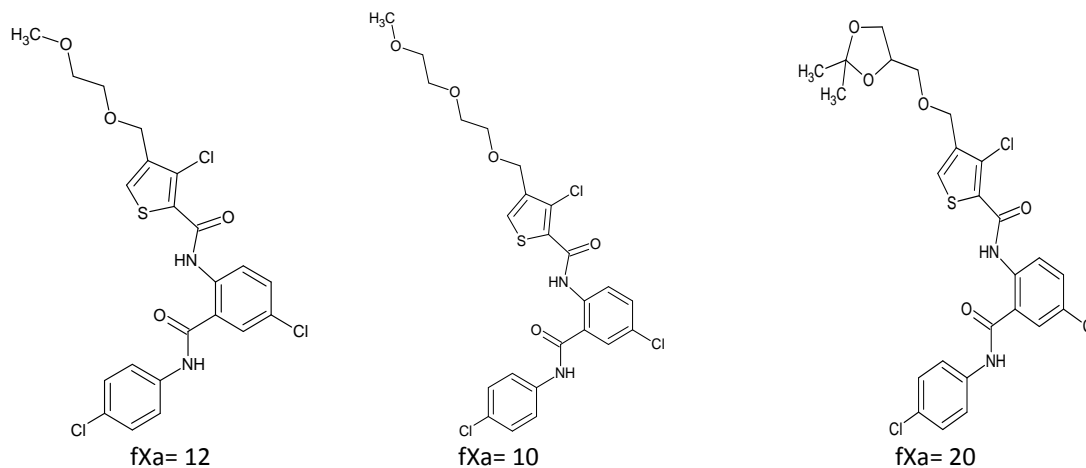


Figure 1: Structures Of Compounds with their observed fXa values

## 2.2 Descriptors generation

In order to obtain QSAR model, compounds are represented by the molecular descriptors [4]. The calculation process of the molecular descriptors is described as following: all molecules were drawn and pre-optimized in ACDLAB.12.0 program using MM2 force field. The resulting geometries formed the inputs for chemoffice 2012 software to calculate steric, thermodynamic and topological descriptors.

For electronic descriptors, G09 software [5] was used for quantum chemistry calculations. According to DFT/B3LYP/6-31G(d) basis set, the lowest unoccupied molecular orbital's energy ( $E_{LUMO}$ ), the highest occupied orbital's energy ( $E_{HOMO}$ ) and total energy  $E_T$  are defined (table 1).

Table 1: Molecular descriptors classified according to their categories

Name of descriptors	Category
Molecular weight Ovality Molecular volume Density logP molar refraction St	Steric
$E_{HOMO}$ $E_{LUMO}$ $E_s$ $E_b$ $\alpha$ TVC	Electronic
$E_{VDW}$ Bp $E_{gibbs}$	Thermodynamic
Shap coef Iw	Topological

## 2.3 Descriptors Selection

The MLR method is very useful tool in searching the best set of descriptors for multi-linear correlations. The model obtained is expected to yield maximum predictive ability. A step wise addition of further descriptor scales was performed to find the best multi-parameter regression models with the optimum values of statistical criteria, in this study, four descriptors are selected: St, d, logP and  $E_{LUMO}$ .

## 2.4 Neural Network

Artificial neural network has been successfully applied to pattern classification and many other fields. In general, NNs are suitable to represent the non linear relation between variables due to the complex connections among the neurons and the flexibility in using processing functions. It processes information in a similar way the human brain does. The network is composed of few layers of neurons which are interconnected between processing elements that operate in parallel. Neurons in adjusting layers are connected with relative quantitative weights [7]. In this work we use a neural network with three layers; the first (input) layer contains four neurons, representing the relevant descriptors obtained in MLR method. Although, there are neither theoretical nor empirical rules to determinate the number of hidden layers, one hidden layer seems to be sufficient in the most chemical application of NN [8], some authors have proposed a parameter  $\rho$  leading to determine the number of hidden neurons, which play a major role to determine the best NN architecture. It's defined as follows:

$$\rho = \frac{\text{nb of data points in the training set}}{\text{sum of the number of connections in the NN}}$$

The three hidden neurons were chosen to maintain  $\rho$  between 1 and 3 ( $1 < \rho < 3$ ). So, the NN architecture is as shown in figure 2 (4-3-1):

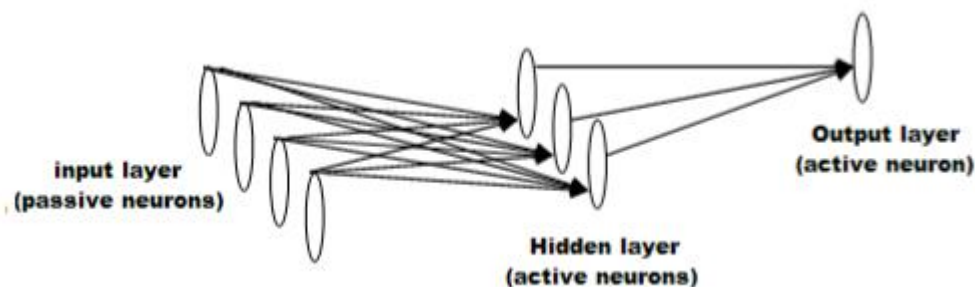


Figure 2: NN architecture used in this work

### 2.5 Cross Validation

CV used to measure a model's predictive ability and draw attention whether model is over-fitted. Over-fitting refers to the phenomenon in which a predictive model may well describe the relationship between predictors and response, but may subsequently fail to provide valid predictions for new compounds. The process of CV begins with the removal of one (leave one-out LOO) or a group (leave many-out LMO) of compounds, which becomes a temporary test set, from the training set. A CV model is created from the remaining data points using the descriptors from the original model, and tested on the removed molecules for its ability to correctly predict the activity. In this paper we used LOO method to test the QSAR model.

### MLR model:

Four –variables linear model obtained by applying step-wise multiple linear regression MLR is given in form regression equation and discussed below:

$$\text{Log}(fXa) = -6,26 - 0,027(\text{St}) + 3,37(\text{d}) + 0,339(\text{logP}) + 1,157(\text{E}_{\text{LUMO}})$$

$$N = 54, R = 0,95 ; R^2 = 0,9 ; ES = 0,24$$

Where N is the number of compounds in the training dataset, R is the correlation coefficient; SE is the standard error of estimate. The selected parameters in equation, as well as, the experimental activities and those predicted by multiple linear regressions, neural network and cross validation are listed in table 2.

### 3. Results and Discussions

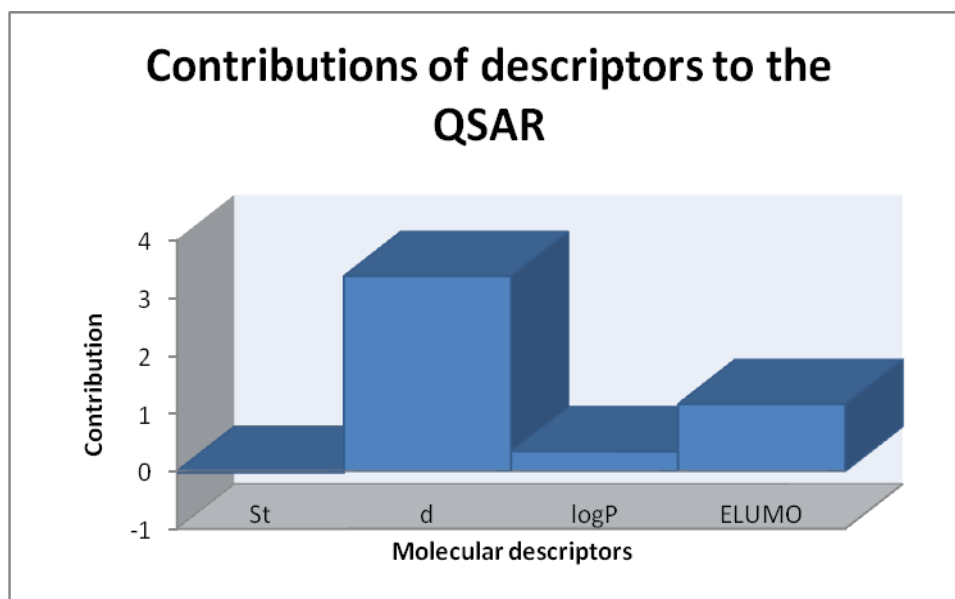
Table 2: selected descriptors, observed log(fXa) and predicted log(fXa), by MLR, NN and CV methods

molecules	St	D	logP	E <sub>LUMO</sub>	log(fXa) <sub>exp</sub>	Log(fXa) <sub>NN</sub>	Log(fXa) <sub>MLR</sub>	Log(fXa) <sub>CV</sub>
1	64,900	1,460	4,700	0,610	-0,086	-0,084	-0,850	-0,090
2	65,500	1,510	4,180	0,790	-0,022	-0,019	-0,021	-0,010
3	62,400	1,420	5,050	0,610	0,954	1,020	0,950	0,700
4	62,900	1,460	6,350	0,640	1,146	1,120	1,120	1,120
5	68,400	1,440	6,840	0,580	1,301	1,270	1,304	1,340
6	69,000	1,440	6,840	0,650	1,322	1,300	1,310	0,390
7	68,100	1,490	5,550	0,710	0,978	0,900	0,940	1,030
8	56,000	1,550	4,060	0,620	-0,222	-0,220	-0,222	-0,211
9	64,900	1,460	4,570	0,490	-0,086	-0,086	-0,087	-0,100
10	71,300	1,520	8,940	0,620	3,699	3,600	3,690	4,000
11	71,000	1,510	8,770	0,620	1,732	1,700	1,720	1,740
12	65,500	1,520	4,190	0,670	-0,022	-0,019	0,020	0,030
13	63,400	1,520	6,840	0,550	1,255	1,220	1,230	1,280
14	70,000	1,520	6,970	0,550	1,380	1,360	1,360	1,400
15	66,000	1,450	4,770	0,480	0,000	0,000	0,010	1,000
16	66,000	1,450	5,110	0,490	0,748	0,720	0,740	1,300
17	68,900	1,490	5,190	0,510	0,792	0,670	0,780	0,700
18	68,900	1,490	8,070	0,510	1,531	1,510	1,520	1,630
19	58,700	1,510	4,920	0,830	0,279	0,270	0,300	0,290
20	58,700	1,510	4,930	0,630	0,857	0,830	0,840	1,010
21	64,100	1,420	6,280	0,480	1,079	1,000	1,030	1,100
22	64,100	1,420	5,990	0,640	1,000	0,960	0,980	1,500
23	56,500	1,380	5,780	0,530	0,875	0,770	0,800	1,040
24	66,000	1,450	4,560	0,480	0,000	0,010	0,100	0,001
25	68,900	1,490	5,630	0,510	0,792	0,729	0,770	0,870
26	63,600	1,420	5,130	0,490	0,204	0,200	0,190	0,500
27	69,200	1,480	5,540	0,730	0,778	0,670	0,760	1,020
28	65,500	1,510	5,240	0,520	0,681	0,630	0,590	0,780
29	60,000	1,450	5,330	0,540	0,716	0,700	0,720	0,700
30	65,400	1,490	4,990	0,800	-0,097	0,060	-0,080	0,030
31	68,800	1,510	5,350	0,730	0,146	0,140	0,140	0,200
32	63,500	1,470	5,450	0,750	0,531	0,520	0,580	0,600
33	62,400	1,450	5,435	0,810	0,176	0,110	0,171	0,200

34	55,600	1,390	4,360	0,730	-0,347	-0,334	-0,337	2,010
35	57,900	1,350	4,840	0,810	-0,284	-0,264	-0,270	-0,340
36	56,700	1,370	4,850	0,800	-0,222	-0,222	-0,221	-0,234
37	55,400	1,410	4,030	0,630	-0,523	-0,512	-0,580	-0,630
38	56,400	1,450	5,939	0,880	0,663	0,650	0,667	0,730
39	55,600	1,420	3,260	0,520	-0,420	-0,421	-0,420	-0,500
40	58,700	1,510	5,150	0,840	0,279	0,210	0,260	0,220
41	58,700	1,510	5,800	0,790	0,322	0,200	0,310	0,220
42	62,100	1,560	4,620	0,760	0,204	0,180	0,300	1,020
43	62,100	1,560	4,720	0,920	0,146	0,100	0,130	0,150
44	62,100	1,560	4,740	0,700	0,544	0,520	0,540	0,650
45	65,900	1,620	4,890	0,910	0,230	0,200	0,250	0,324
46	65,900	1,620	5,370	0,740	0,301	0,290	0,300	0,410
47	64,400	1,520	4,970	0,620	-0,097	0,100	-0,100	-0,089
48	64,800	1,540	4,290	0,610	-0,398	-0,392	-0,380	-0,400
49	65,300	1,550	5,560	0,930	0,322	0,321	0,330	0,560
50	60,600	1,440	5,630	0,590	0,431	0,380	0,410	0,440
51	58,900	1,460	5,910	0,560	1,079	1,080	0,990	1,060
52	56,800	1,420	5,810	0,560	1,000	0,990	1,070	0,960
53	58,600	1,430	6,290	0,540	1,301	1,300	1,320	0,990
54	66,000	1,550	5,340	0,510	0,756	0,730	0,760	0,850

We notice that steric, electronic and lipophilic properties that are considered as the principal factors engendering an interaction between a ligand and a binding site, constitute the relevant descriptors selected to establish the QSAR

model of potent inhibitors of human factor Xa. The contribution of each descriptor was illustrated in figure 3.



**Figure 3:** Contributions of descriptors to the QSAR model.

We notice that descriptors related to the density and LUMO energies are the most important factors in the establishment of the QSAR model of thiophene derivatives. This confirms that the activity is related to steric and electronic effects in

the same time the log P seems to have a non negligible impact on the QSAR model. The graphical correlation of observed and predicted  $\log(fX_a)$  by MLR is recorded in figure 4.

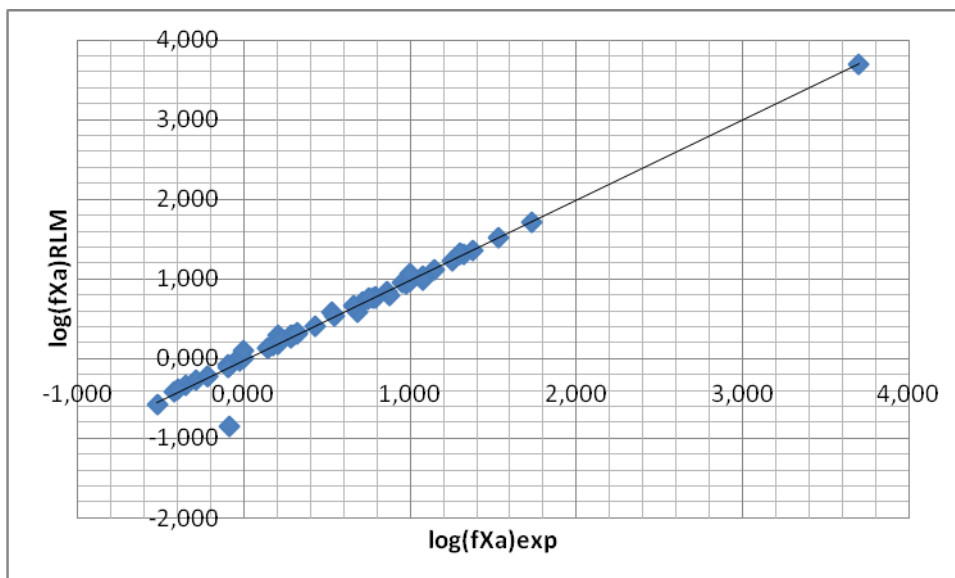


Figure 4: Correlation of observed and predicted activities by MLR

### Neural Network

The neural network model is generated using the four descriptors, appearing in the MLR model. One neuron, which encoded the activities, constituted the output layer, and we chose the hidden layer from the calculation of  $\rho$ . So, The best architecture is: 4-3-1 and the NN results are:  $R_{NN} = 0,98$  ES= 0,27 N=54

The correlation coefficient obtained by neural network implies that the selected descriptors by MLR are apposite and the model QSAR proposed to predict activities is relevant. The predicted power of 4-3-1 NN can be judged from the plot of observed activities versus predicted by NN shown in figure 5.

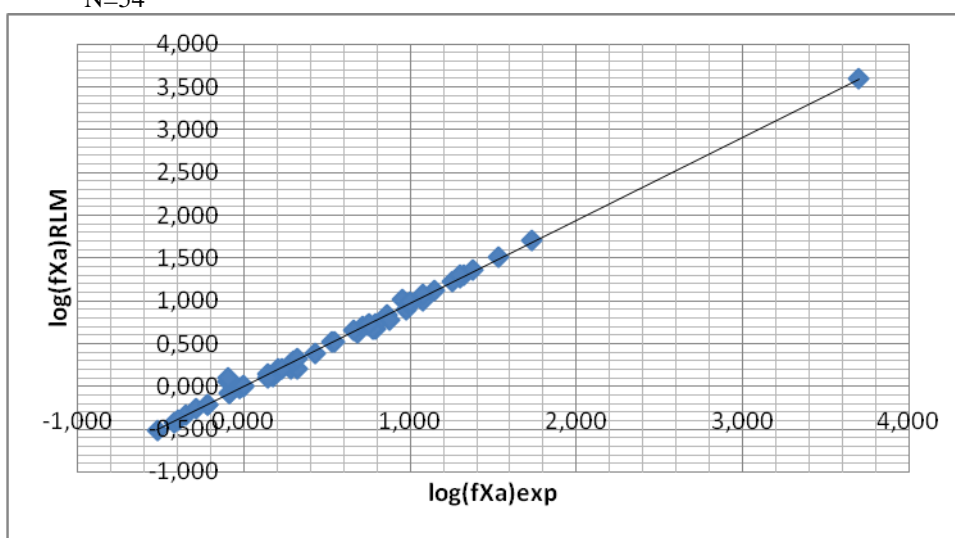


Figure 5: observed activities versus predicted ones by NN the training set.

### Cross-validation

In order to check the reliability and the stability of the best elaborated QSAR model, we have used the leave-one-out validation, this step is very important because the high value of the square of the correlation coefficient indicates the best fit of the data but does not contain information about the ability to predict the dependent variable of no include data in

The LOO cross-validation procedure yielding high predictive R values “RCV= 0, 90 and SE = 0, 32” testifies on the high predictability of the proposed QSAR model.

Figure 6 represents observed activities versus LOO predicted  $\log(fXa)$ .

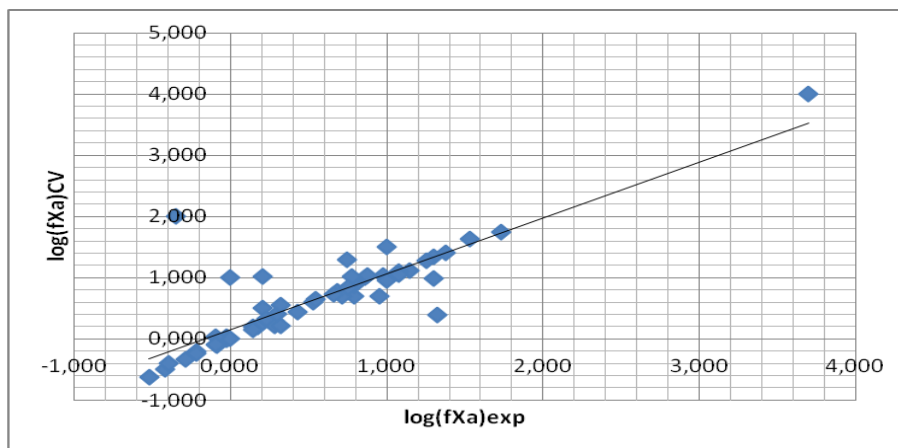


Figure 6: Observed activities versus predicted ones by LOO method

To estimate the predictive power of a QSAR model, Tropsha and Golbraikh [9] recommended use of the following statistical characteristic of the test set, (i) correlation coefficient  $R$  between the predicted and observed activities, (ii) coefficients determination ( $R^2$ ), (predicted vs observed activities  $R_0^2$ ), (iii) slopes  $K$  and  $K'$  of the regression lines through the origin. They consider that a QSAR model is predictive, if the following conditions are satisfied:

Tropsha and Golbraikh criteria's	Results
$R_{CV}^2 > 0,6$	$R_{CV}^2 = 0,81$
$R_{pred}^2 > 0,5$	$R_{pred}^2 = 0,90$
$\frac{R^2 - R_{pred}^2}{R^2} \leq 0,1$	$\frac{R^2 - R_{pred}^2}{R^2} = \frac{0,90 - 0,81}{0,90} = 0,1$

In addition, another criteria is proposed by Roy [10] to investigate the predictability of QSAR models,  $R_m^2 = R^2 (1 - (R^2 - R_{pred}^2)^{1/2})$

If  $R_m^2 > 0,5$ , it indicates a good predictability of the developed model, in this study we have:  
 $R_m^2 = 0,90 [1 - (0,90 - 0,81)^{1/2}] = 0,85$

#### 4. Conclusion

In this study, we have developed a QSAR model introducing quantum descriptors, regression quality indicates that these descriptors provide important information and have significant role in the rating of the activities. Neural network considering the relevant descriptors obtained by MLR, showed a good agreement between observed and predicted values; finally, we have validated our results by LOO-CV procedure.

#### References

- [1] Kochanny. M. K, Alder. M, Ewing. J, Griedel. B. D, Ho. E, Karanjawala. R, Lee. W, Lentz. D, Liang. A. M, Morrissey. M. M, Philips. G. B, Post. J, Sacchi. K. L, Sakata. S. T, Subramanyam. B, Vergona. R, Walters. J, White. K. A, Withlow. M, Ye. B, Zhao. Z, Shaw. K. J, substituted thiophene-anthranilamides as potent inhibitors of human factors Xa, *Bioorganic and Medicinal Chemistry* 15 (2007) 2127-2146.
- [2] Verma. A, Giridhar. R, Kanhed. A, Sinha. A, Modh. P, Yadav. M. R, Novel 2-aminobenzenamides as potential orally active antithrombotic agents, *ACS Medicinal Chemistry letters*, 4 (2013) 32-36.
- [3] Bellifa. K, Mekelleche. S. M, QSAR study of the toxicity of nitrobenzenes to tetrahymenapyriformis using quantum chemical descriptors, *Arabian journal of Chemistry*, (2012) 1-7.
- [4] Dguigui. K, El Hallaoui. M, Naceiri. N, QSAR Modeling of thiophene derivatives as potent agonist of S1P1, *IJARCSSE*, 4(8), (2014) 435-458.
- [5] Gaussian 09, Revision D.01, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, Gaussian, Inc., Wallingford CT, 2009.
- [6] Dguigui. K, El Hallaoui. M, Naceiri. N, QSAR Modeling of thiophene derivatives as potent agonist of S1P1, *IJARCSSE*, 4(8), (2014) 435-458.
- [7] Janlou. M. A. M, Abdolaleki. P., Tajbakhsh. M, Amanlou. M, Eidi. A, QSAR study of tyrosinase inhibitors using logistic regression and artificial neural network, *J Iran Chem SOC*, (2012) p 643-653.
- [8] Mbarki. S, El Hallaoui. M, 3D-QSAR models to predict antimoebic activities of the cyclised pyrazolines and 2-(quinolin-8-yloxy)acetohydrazones, *RJPBCS*, 5(1) (2014) 73-83.
- [9] Golbraikh. A, Tropsha. A, Beware of  $q^2$ , *Journal of molecular graph model*, vol 20, (2002) 269-276.
- [10] Ali. S, Shayanfar. S, is regression throuth origin useful in external validation of QSAR, *European Journal of pharmaceutical science*, vol 59, (2014) 31-35.