QSAR Study of (5-Nitroheteroaryl-1,3,4-thiadiazole-2-yl)piperazinyl Derivatives to Predict New Similar Compounds as Antileishmanial Agents

ABDELLAH OUSAA[1*], BOUHYA ELIDRISSI[1],
MOUNIR GHAMALI[1], SAMIR CHTITA[1], ADNANE AOUIDATE[1],
MOHAMMED BOUACHRINE[2] and TAHAR LAKHLIFI[1]

[1]Molecular Chemistry and Natural Substances Laboratory, Faculty of Science, Moulay Ismail University, Meknes, Morocco

[2]MEM, ESTM, Moulay Ismail University, Meknes, Morocco

abdellahousaa@gmail.com

Abstract: To search of newer and potent antileishmanial drugs, a series of 36 compounds of 5-(5-nitroheteroaryl-2-yl)-1,3,4-thiadiazole derivatives were subjected to a quantitative structure-activity relationship (QSAR) analysis for studying, interpreting and predicting activities and designing new compounds using several statistical tools, The multiple linear regression (MLR), non-linear regression (RNLM) and artificial neural network (ANN) models were developed using 30 molecules having pIC_{50} ranging from 3.155 to 5.046. The best generated MLR, RNLM and ANN models show conventional correlation coefficients R of 0.750, 0.782 and 0.967 as well as their leave-one-out cross-validation correlation coefficients R_{CV} of 0.722, 0.744 and 0.720, respectively. The predictive ability of those models was evaluated by the external validation using a test set of 6 molecules with predicted correlation coefficients R_{test} of 0.840, 0.850 and 0.802, respectively. The applicability domains of MLR and MNLR transparent models were investigated using William's plot to detect outliers and outsides compounds. We expect that this study would be of great help in lead optimization for early drug discovery of new similar compounds.

Introduction

Leishmaniasis is a parasitic disease caused by protozoan parasites of the genus Leishmania, is generally recognized as an important public health problem, which touching millions of people living mainly in large areas of tropical and subtropical regions. Currently, there are only a limited number of drugs that are available for the treatment and control of this Leishmaniasis disease, all of which are associated with limiting factors such as high toxicity,

variable efficacy, long dosing schedules and/or parenteral administration[1,2]. To date, no vaccine against any clinical form of Leishmaniasis has been commercialized and its treatment relies solely on chemotherapy that has been based on the use of pentavalent antimonial drugs. Other drugs, such as pentamidine, miltefosine and amphotericin B, have been used as alternative medications towards resistant parasites. With the emergence of some resistant strains, all those drug therapies cause serious side effects. Furthermore, general treatment is unaffordable for many afflicted countries. Therefore, there is always a need for designing newly potent, safer and cheaper drugs[3-5].

A promising strategy for discovering new therapeutic leads is to study classes of compounds potentially bioactive or old active compounds for alternative uses. Nitrogen heterocycles such as quinolines, pyrimidines, acridines, phenothiazines, indoles quinones, in general and particularly thiadiazole derivatives as well as their reduced derivatives have been tested in the last years in antileishmanial tests.

Thiadiazole derivatives since their discovery in twentieth century have demonstrated a broad spectrum of pharmacological properties[6]. They used initially as antibacterial agents[7], and rapidly revealed an interesting antiproliferative activity against both protozoa and tumor cells [8,9]. Consequently, these derivatives have been extensively used in antiparasitic chemotherapy and a large variety of new thiadiazole derivatives were synthesized and evaluated for their *in vitro* antileishmanial activity[10,11]. Today, traditional methods for drug discovery and development have been gradually replaced by modern approaches, wherein computational techniques have become inevitable in drug development pipeline by reducing the amount of synthetic works and biological evaluations needed to achieve the required results.

In the present study, QSAR studies based on principal components analysis (PCA), multiple linear regression (MLR), non-linear regression (RNLM) and artificial neural network (ANN) calculations were performed on a series of 36 of (5-nitroheteroaryl-1,3,4-thiadiazole-2-yl) piperazinyl derivatives[12], in order to identify the key structural features required to design new potent lead candidates of this class. The results extracted from this study might be helpful to design highly potent antileishmanial drugs.
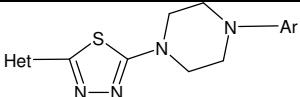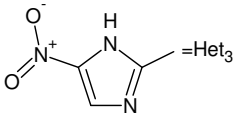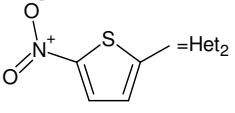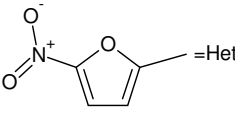
Experimental
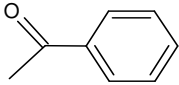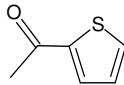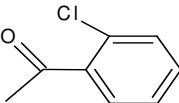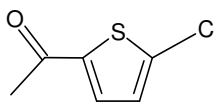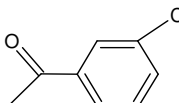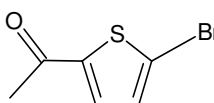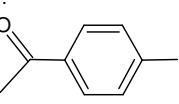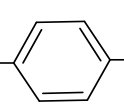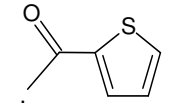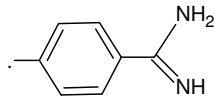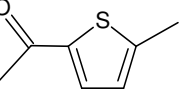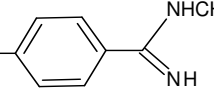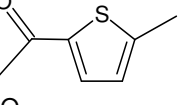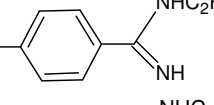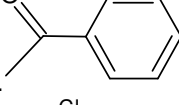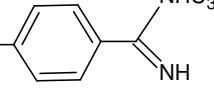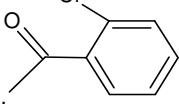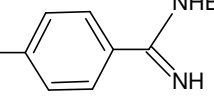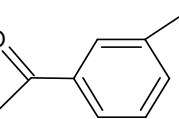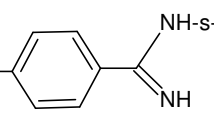
The experimental antileishmanial activity IC_{50} (μM) of 36 thiadiazole derivatives were collected in previously study[12]. The dataset was split randomly into a training set (thirty molecules) to build the quantitative model and the remaining molecules were used to test the performance of the proposed model (Test set). The molecular structures of the studied molecules with their antileishmanial activity converted into pIC_{50} ($-logIC_{50}$) are presented in Table 1.

Molecular descriptors generation
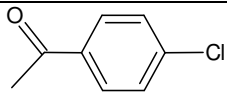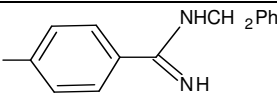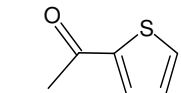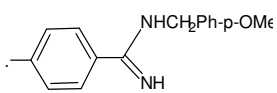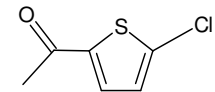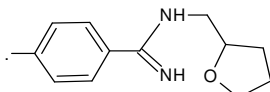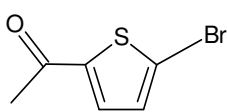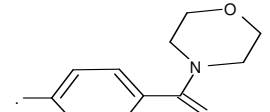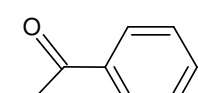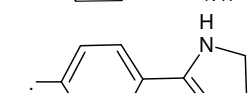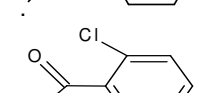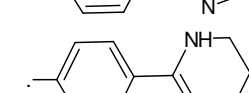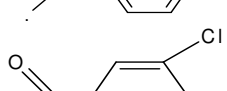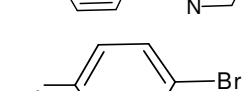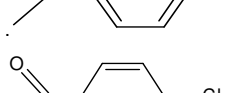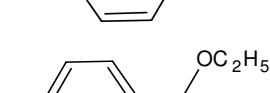
The calculation of electronic descriptors was carried out by using the Gaussian03W package[13]. The geometries of 36 thiadiazole derivatives were optimized with DFT method with the B3LYP functional and 6-31G (d) base set. Then, several related structural features were opted from the obtained results of calculations as follows: highest occupied molecular orbital energy (E_{HOMO}), lowest unoccupied molecular orbital energy (E_{LUMO}), dipole moment (μ), energy gap (ΔE) and total energy (E_T).

ChemSketch program[14] was employed to calculate the others molecular descriptors such as: the molar volume MV (cm^3), the molecular weight MW (g/mol), the molar refractivity MR (cm^3), the parachor Pc (cm^3), the density D (g/cm^3), the refractive index n and the octanol/water partition coefficient log *P*.

Table 1. Chemical structures of thiadiazole derivatives used in this study

N°	Het	Ar	pIC$_{50}$	N°	Het	Ar	pIC$_{50}$
1	Het$_1$		4.932	19	Het$_3$		4.592
2	Het$_1$		4.969	20	Het$_3$		5.029
3*	Het$_1$		4.880	21	Het$_3$		4.797
4*	Het$_1$		4.581	22	Het$_1$		3.983
5	Het$_1$		4.717	23	Het$_1$		3.955
6	Het$_1$		4.921	24	Het$_1$		4.638
7	Het$_1$		4.882	25	Het$_1$		4.482
8	Het$_2$		4.602	26	Het$_1$		5.000
9	Het$_2$		4.463	27	Het$_1$		4.959
10*	Het$_2$		4.222	28	Het$_1$		4.482

Contd….

No	Het	Structure	Value	No	Het	Structure	Value
11*	Het₂		4.618	29	Het₁		5.046
12	Het₂		4.549	30	Het₁		4.097
13	Het₂		4.880	31	Het₁		4.022
14	Het₂		4.581	32	Het₁		4.032
15	Het₃		4.983	33	Het₁		3.155
16	Het₃		4.693	34	Het₁		3.222
17*	Het₃		4.748	35	Het₁		4.143
18*	Het₃		4.367	36	Het₁		4.319

Test set

Methods of data analysis

Principal component analysis (PCA)

The principal component analysis is a data analysis method used to transform a set of variables correlated to a new set of variables, called principal components. They are fewer but independent. Using these new variables, the dimensionality of the system is reduced with a minimum loss of information[15]. The obtained matrix of coordinates allows us to analyze the dispersion of individuals in the new defined space[16-18]. After that, the principal component analysis (PCA) was used to determine the non-linearity and non-multicollinearity among variables and to select descriptors that correlate with the activity.

Multiple linear and non-linear regressions

Multiple linear regression is used to study the relation between one dependent and several independent variables. The aim of this method is to minimize the difference between the actual and predicted values of any experimental effect and was used to select the descriptors to be used as input in the multiple non-linear regression and as well as in the artificial neural network. Multiple linear and non-linear regressions are generated using the software XLSTAT version 2013 that were used to predict effects on the antileishmanial activity.

Equations of models were mainly justified by the correlation coefficient ®, the mean squared error (MSE), the Fisher's F statistic and the significance level (p-value)[19].

Artificial neural network (ANN)

This ANN method is performed in order to rise the chance of characterizing the studied compounds and to generate a predictive QSAR model between the molecular descriptors selected in the MLR equation and the observed activity values. The ANN analysis was applied by using the Matlab (2014) software, neural mounting tool (nntool) toolbox. Artificial neural network is a non-linear empirical method[20] that is used in the prediction of an experimental effect, while its application is booming in many disciplines. It is, among others, a very interesting alternative method to traditional statistics for the data processing.

Internal and external validation

To determine the stability of the predictive model and to test the influence of each sample on the final model, two basic principles, internal validation and external validation, were carried out in this study. Cross-validation technique is one of the most popular methods for internal validation. In fact, there are at least three cross-validation techniques: "holdout method", "k-fold cross-validation" and "leave-one-out cross validation". In this study, the internal predictive capability of the model was evaluated using leave-one-out cross-validation (Rcv). A good Rcv of ten indicates a good robustness and high internal predictive power of a QSAR model. However, recent studies[21] indicate that there is no evident correlation between the value of Rcv and actual predictive power of a QSAR model, suggesting that the Rcv remains inadequate as a reliable estimate of the model's predictive power for all new chemicals. To test that reliably predictive power, the use of a set of external validation, not used for the development of the model is required. As long as the original data set is large enough, the latter can be easily divided into two: a learning set in which the model is developed and a set of validation used to characterize its predictive power.

Results and Discussion

Data set for analysis

QSAR study was performed using the activity values of 36 thiadiazole derivatives, as previously reported[12] to determine the quantitative relationship between the structures of the studied compounds and biological activity. The values of 12 chemical descriptors are shown in Table 2.

Table 2. Values of the obtained parameters of the studied thiadiazoles

N	pIC_{50}	MW	MR	MV	Pc	n	D	E_T	E_{HOMO}	E_{LUMO}	ΔE	μ	log P
1	4.932	385.40	97.67	265.60	36.597	1.66	1.45	-44404.76	-6.59	-2.84	3.75	3.04	1.844
2	4.969	419.84	102.56	277.50	34.399	1.66	1.51	-56919.44	-6.70	-2.85	3.85	3.77	2.402
3*	4.880	419.84	102.56	277.50	34.399	1.66	1.51	-56919.28	-6.41	-2.92	3.49	5.40	2.402
4*	4.581	419.84	102.56	277.50	34.399	1.66	1.51	-56919.54	-6.69	-2.88	3.82	2.20	2.402
5	4.717	391.42	96.05	255.10	42.830	1.68	1.53	-53138.85	-6.38	-3.08	3.30	7.55	1.825
6	4.921	425.87	100.95	267.00	40.057	1.68	1.59	-65653.52	-6.70	-2.88	3.82	5.74	2.199
7	4.882	470.32	103.74	271.20	46.026	1.69	1.73	-123149.92	-6.65	-2.87	3.78	6.74	2.534
8	4.602	401.46	103.75	272.30	36.730	1.69	1.47	-53199.45	-6.37	-2.48	3.89	3.56	3.210
9	4.463	435.91	108.65	284.20	34.520	1.69	1.53	-65714.15	-6.69	-2.93	3.76	6.05	3.768

Contd…

10[*]	4.222	435.91	108.65	284.20	34.520	1.69	1.53	-65714.24	-6.70	-2.96	3.74	5.57	3.768
11[*]	4.618	435.91	108.65	284.20	34.520	1.69	1.53	-65714.25	-6.70	-2.97	3.73	4.39	3.768
12	4.549	407.49	102.14	261.70	42.999	1.71	1.56	-61933.62	-6.61	-2.93	3.69	6.03	3.191
13	4.624	441.94	107.03	273.70	40.210	1.71	1.61	-74448.23	-6.68	-2.96	3.72	5.14	3.565
14	4.608	486.39	109.83	277.90	46.214	1.72	1.75	-131944.63	-6.65	-2.96	3.69	5.21	3.900
15	4.983	399.43	105.86	254.80	37.730	1.77	1.56	-45371.46	-6.25	-2.81	3.44	5.62	3.856
16	4.693	433.87	110.46	264.10	35.430	1.78	1.64	-57886.14	-6.32	-2.82	3.50	5.92	3.039
17[*]	4.748	433.87	110.46	264.10	35.430	1.78	1.64	-57886.23	-6.35	-2.86	3.49	4.92	2.825
18[*]	4.367	433.87	110.46	264.10	35.430	1.78	1.64	-57886.24	-6.40	-2.72	3.68	3.96	3.956
19	4.592	405.45	104.45	237.20	44.267	1.83	1.70	-54105.54	-6.35	-2.87	3.49	5.42	3.756
20	5.029	439.90	109.05	246.50	41.356	1.84	1.78	-66620.16	-5.81	-1.95	3.86	6.30	2.956
21	4.797	484.35	112.01	249.70	47.628	1.85	1.93	-124116.56	-6.42	-2.90	3.52	4.48	2.984
22	3.983	382.40	97.23	247.90	32.430	1.71	1.54	-43830.08	-6.25	-2.94	3.31	2.15	2.784
23	3.955	399.42	103.40	243.00	32.209	1.80	1.64	-45370.34	-5.75	-2.84	2.91	8.11	1.894
24	4.638	413.45	108.96	266.10	27.642	1.76	1.55	-46440.57	-5.75	-2.82	2.94	7.46	2.414
25	4.482	427.48	113.57	282.10	25.050	1.74	1.51	-47511.32	-5.73	-2.80	2.93	8.00	2.753
26	5.000	441.51	118.18	298.20	22.805	1.72	1.48	-48581.84	-5.73	-2.12	3.60	7.38	3.239
27	4.959	455.53	122.79	314.20	20.850	1.71	1.44	-49652.19	-5.71	-2.82	2.89	7.61	3.656
28	4.482	455.53	122.95	317.90	21.003	1.70	1.43	-49652.25	-5.20	-1.92	3.27	9.94	3.557
29	5.046	489.55	134.26	334.70	21.693	1.73	1.46	-52732.28	-5.73	-2.82	2.91	7.91	4.288
30	4.100	519.58	140.07	356.30	19.321	1.72	1.45	-55850.74	-5.72	-2.82	2.90	7.84	4.162
31	4.022	483.54	126.43	309.10	22.355	1.75	1.56	-52737.79	-5.76	-2.83	2.92	7.31	2.555
32	4.032	469.52	122.72	299.10	24.975	1.76	1.56	-51666.81	-5.77	-2.83	2.94	7.35	2.392
33	3.155	425.46	112.35	262.40	34.888	1.80	1.62	-47478.41	-5.42	-3.00	2.42	5.80	0.821
34	3.222	439.49	116.96	278.50	32.136	1.78	1.57	-48548.98	-5.57	-2.81	2.76	8.18	0.927
35	4.143	436.28	100.16	268.90	40.777	1.67	1.62	-111329.43	-5.92	-2.86	3.06	5.18	3.579
36	4.319	428.46	111.91	283.90	24.314	1.72	1.50	-48052.33	-5.73	-2.82	2.91	8.65	3.130

[*]*Test set*

Principal component analysis

The total of 12 descriptors coding the 36 molecules was submitted to principal components analysis (PCA). The first three axes F1, F2 and F3 represent respectively (37.69%; 18.69% and 17.81%) of the total variance and they estimate 74.19% of the total information.

The PCA was conducted to identify the correlation between the different descriptors. It is also helpful for understanding the distribution of the compounds[20]. The correlation's matrix of the 12 descriptors is shown in Table 3.

The correlation coefficients in the obtained matrix provide the information about the high or low interrelationship between the descriptors. Generally good co-linearity ($r > 0.5$) was present between the majority of the variables. In this study, no descriptor strongly correlated with the others.

Multiple linear regressions

The obtained descriptors were used for the development of a mathematical linear model to predict quantitatively the physicochemical effects of substituents on the antileishmanial activity of 30 thiadiazole derivatives using backward selection in the multiple regression analysis.

The study of the descendant MLR multiple linear regression based on the elimination of descriptors until a valid model was obtained methods were employed to determine the best regression models.

**Table 3.** Correlation matrix between different obtained descriptors

Variables	pIC_{50}	E_T	E_{HOMO}	E_{LUMO}	μ	ΔE	$\log P$	MW	MV	MR	n	DPc
pIC_{50}	1											
E_T	-0.167	1										
E_{HOMO}	-0.424	0.381	1									
E_{LUMO}	0.222	0.160	0.456	1								
μ	-0.159	0.181	0.707	0.292	1							
ΔE	0.616	-0.320	-0.817	0.142	-0.597	1						
$\log P$	0.416	-0.227	-0.163	0.155	-0.011	0.282	1					
MW	-0.033	-0.446	0.249	0.099	0.379	-0.213	0.364	1				
MV	-0.022	0.135	0.384	0.167	0.455	-0.318	0.354	0.695	1			
MR	-0.168	0.149	0.612	0.256	0.582	-0.515	0.313	0.796	0.837	1		
n	-0.229	0.006	0.372	0.169	0.194	-0.304	-0.066	0.153	-0.323	0.247	1	
D	0.014	-0.680	-0.233	-0.086	-0.193	0.203	-0.057	0.176	-0.584	-0.240	0.629	1
Pc	0.185	-0.608	-0.677	-0.281	-0.523	0.571	-0.098	-0.290	-0.756	-0.722	0.099	0.707	1

Many attempts have been made to develop a relationship with the indicator variable of activity pIC_{50}, but the best relationship obtained by this method is only one corresponding to the linear combination of three descriptors selected, the energy E_{LUMO}, the energy E_{HOMO} and the octanol/water partition coefficient $\log P$.

The resulting equation is:

$$pIC_{50} = 2.0453 - 0.6673\, E_{HOMO} + 0.7821\, E_{LUMO} + 0.1898 \log P \qquad (1)$$

The positive correlation of these factors E_{LUMO} and $\log P$ with the value of the pIC_{50} in eq. 1 shows that an increase in the values of these factors implies an increase in the value of the pIC_{50}, whereas a negative correlation of the E_{HOMO} shows that an increase in the value of this factor indicates a decrease in the value of the pIC_{50}. The correlations of the predicted and observed activities are illustrated in Figure 1.
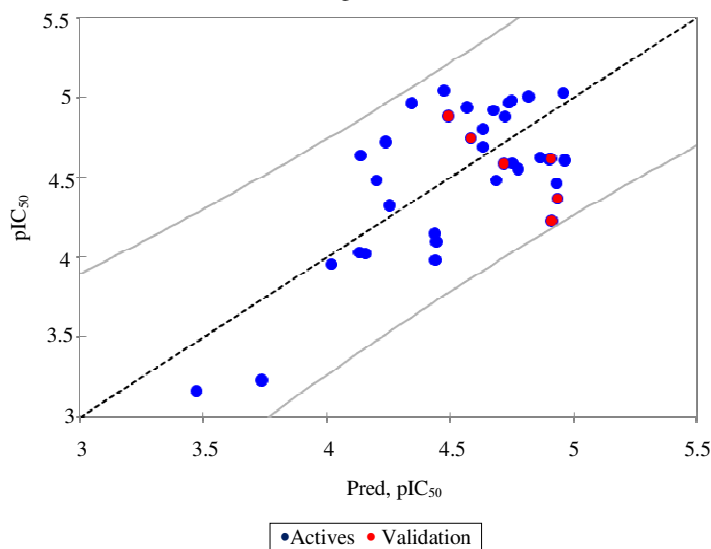
Figure 1. Graphical representation of calculated and observed activity by MLR

N = 30, R = 0.75, R_{CV} = 0.722, MSE = 0.119, F = 11.43, P-value <0.0001
N_{test} = 6, R_{test} = 0.840, MSE_{test} = 0.124

Where N is the number of compounds, R is the coefficient of correlation, Rcv is the coefficient of correlation for cross-validation, MSE is the mean squared error, F is the Fisher's criterion and P is the significance level.

High coefficient of correlation (R=0.75) and lower mean squared error indicate that the model is more reliable. P is smaller than 0.05 means that the obtained equation is statistically significant at the 95% level. The obtained model was cross-validated by its applicable Rcv value (Rcv= 0.722) using the leave-one-out (LOO) method. The value of Rcv greater than 0.5 is the basic criteria to qualify a model as valid[22]. This reported model was also used to predict the pIC_{50} values of the remaining 6 compounds in the test set. Such predicted values are also recorded in Table 5. R_{test} is 0.84 and mean squared error MSEtest is 0.119, which confirms that the proposed model has a better predictive ability. The correlation coefficients among descriptors in the model were calculated using the variance inflation factor (VIF), as shown in Table 4.

Table 4. The variance inflation factors (VIF) of descriptors in QSAR model

Statistique	E_{HOMO}	E_{LUMO}	$\log P$
Tolerance	0.773	0.758	0.915
VIF	1.294	1.319	1.093

The VIF was defined as $1/(1-R^2)$, where R is the multiple correlation coefficient for an independent variable against all other descriptors in the model. If VIF is greater than 5, the models are unstable and must be eliminated; models with a VIF value of 1 to 4 can be accepted. Table 4 shows that all of the VIF values of the three descriptors are smaller than 5.0. Thus, there is no collinearity among the selected descriptors, and the obtained model is stable.

With the MLR models, the values of predicted pIC_{50} calculated from equation 1 the observed values are given in Table 5.

Multiple non-linear regressions

We also used a non-linear regression model to improve the structure-activity relationship and to evaluate the effect of the substituent. We applied the proposed descriptors by multiple linear regressions for 30 molecules in the whole formation and we used the correlation coefficient (R) and the Mean squared error (MSE) to select the best performance of regression. We used a pre-programmed function of XLSTAT as follows:

$$Y = a + (b\ X1 + c\ X2 + d\ X3 + e\ X4\ldots) \tag{2}$$

Where a, b, c, d...: represent the parameters and X1, X2, X3, X4….: represent the variables.

The resulting equation:

$$pIC_{50} = -7.544 - 3.591\ E_{HOMO} + 0.5711\ E_{LUMO} + 0.7071\ \log P - 0.247\ E_{HOMO}^2 - 0.025\ E_{LUMO}^2 - 9.724\ 10^{-2} \log P^2 \tag{3}$$

Where N = 30, R = 0.782, R_{CV} = 0.744, MSE = 0.121, N_{test} = 6, R_{test} = 0.850, MSE_{test} = 0.126

The activity values pIC_{50} predicted by this model are almost similar to that observed. The Figure 2 shows a very regular distribution of activity values based on the observed values. The obtained coefficient of correlation in equation 3 is quite very interesting (0.782). The QSAR model expressed by equation 3 is cross-validated by its appreciable Rcv values (Rcv = 0.744) obtained using the leave-one-out (LOO) method. A value of Rcv is greater than 0.5 is the important criterion for qualifying a QSAR model as valid[22].
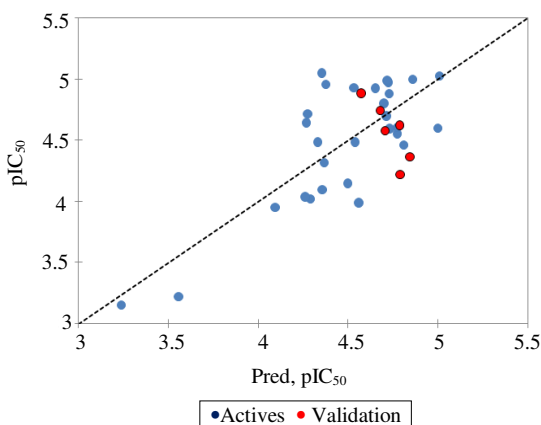
Figure 2. Graphical representation of calculated and observed activity by MNLR

The robustness and predictive power of the model were further supported by the significant R_{test} value (0.850) of the test set data. The observed and calculated pIC_{50} values are given in Table 5.

According to metric values, internal and external, the results obtained by MNLR are relatively better than those obtained by MLR, but the latter approach is more transparent and gives the most interpretable results and a good explanation of the descriptors associated with antileishmanial activity of 36 thiadiazole derivatives.

Artificial neural networks ANN

In order to increase the probability to characterize the compounds, artificial neural network can be used to generate predictive models of quantitative structure-activity relationship between a set of molecular descriptors obtained from the multiple linear regression and the observed activities. The ANN calculated activity model was developed using the parameters of the studied compounds. The correlation between ANN calculated and experimental activity values are very significant as illustrated in Figure 3.

N = 30, R = 0.967, R_{CV} = 0.720, MSE = 0.118, N_{test} = 6, R_{test} = 0.802, MSE_{test} = 0.110
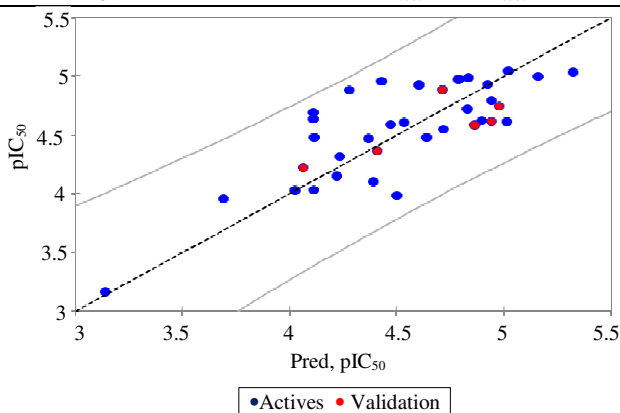
Figure 3. Correlations of observed and predicted activities calculated using ANN

The obtained correlation coefficient (R) value is 0.967 for this data set of the thiadiazole derivatives. This confirms that the artificial neural network (ANN) results are the best to predict the quantitative structure-activity relationship model. Furthermore, the good results obtained with test set shows that the ANN model is the high predictive power. The predicted activities calculated with the artificial neural network and the observed values are given in Table 5. A comparison of the quality of MLR, MNLR and ANN models shows that the ANN model have substantially better predictive capability because the ANN approach gives better results than MLR and MNLR. ANN was able to establish a satisfactory relationship between the selected descriptors and the activity of the studied compounds.

Table 5. Observed and predicted values of pIC_{50} according to different methods

	pIC50 (calcd.)					pIC50 (calcd.)		
N°	pIC_{50} (obs.)	MLR	MNLR	ANN	N°	pIC_{50} (obs.)	MLR	MNLR	ANN
1	4.932	4.573	4.535	4.922	19	4.592	4.755	4.734	4.940
2	4.969	4.741	4.724	4.793	20	5.029	4.959	5.007	4.504
3*	4.880	4.497	4.576	4.713	21	4.797	4.631	4.703	3.690
4*	4.581	4.718	4.704	4.865	22	3.983	4.444	4.563	4.115
5	4.717	4.242	4.278	4.828	23	3.955	4.019	4.098	4.120
6	4.921	4.679	4.651	4.607	24	4.638	4.140	4.268	5.163
7	4.882	4.722	4.729	4.276	25	4.482	4.199	4.329	4.432
8	4.602	4.969	4.999	4.534	26	5.000	4.821	4.859	4.645
9	4.463	4.930	4.809	4.366	27	4.959	4.342	4.377	5.024
10*	4.222	4.916	4.794	5.324	28	4.482	4.685	4.537	4.390
11*	4.618	4.908	4.786	4.471	29	5.046	4.479	4.354	4.030
12	4.549	4.775	4.773	4.718	30	4.097	4.448	4.359	4.114
13	4.624	4.864	4.788	4.894	31	4.022	4.156	4.290	3.138
14	4.608	4.906	4.772	5.012	32	4.031	4.136	4.265	4.940
15	4.983	4.749	4.721	4.833	33	3.155	3.471	3.236	4.504
16	4.693	4.632	4.717	4.116	34	3.222	3.739	3.560	2.409
17*	4.748	4.585	4.677	4.975	35	4.143	4.436	4.497	4.216
18*	4.367	4.936	4.847	4.408	36	4.319	4.254	4.365	4.231

* Test set

Domain of applicability

To estimate the reliability of any QSAR model and its ability to predict new compounds, the domain of applicability must be essentially defined[23]. The predicted compounds that fall within this domain may be considered as reliable. The applicability domains were discussed with the Williams graphs in Figures 4 and 5 of the MLR and MNLR models respectively, which the standardized residuals and the leverage values (h_i) are plotted.

It is based on the calculation of the leverage h_i for each molecule, for which QSAR model is used to predict its activity:

$$h_i = x_i \, (X^T \, X)^{-1} \, x_i{}^T \quad i = 1 \dots n \qquad (4)$$

Where x_i is the row vector of the descriptors of compound i and X is the variable matrix deduced from the training set variable values. The index T refers to the matrix/vector transposed. The critical leverage h^* is generally fixed at $(3k+1)/N$, where N is the number of training molecules, and k is the number of model descriptors. If the leverage value h of molecule is higher than the critical value (h^*) i.e., $h > h^*$, the prediction of the compound can be considered as not reliable.

From Figure 4, the leverage values (h_i) of any compound in the training and test sets are less than the critical value ($h^* = 0.4$). Also, the standardized residuals of all compounds in the training and test sets are less than three standard deviation units ($\pm3\sigma$). Therefore, the predicted activity by the developed MLR model is reliable.

The Williams plot for the MNLR model is shown in Figure 5. It is obvious that there is no outlier in training set and no outside in test set. Therefore, the predicted activity by the MNLR model is reliable.
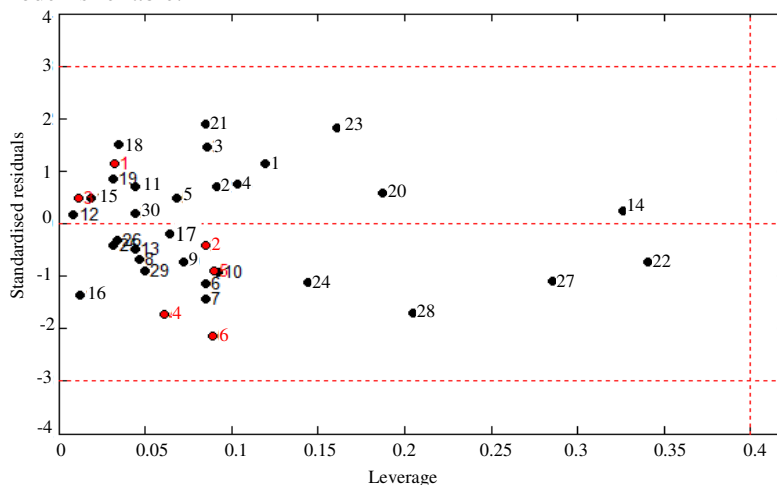
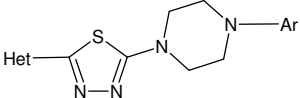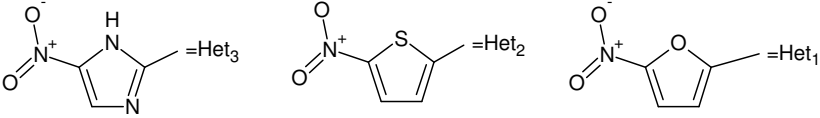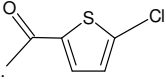**Figure 4.** Williams plot for the MLR model (with $h^* = 0.4$ and residual limits = $\pm3\sigma$)
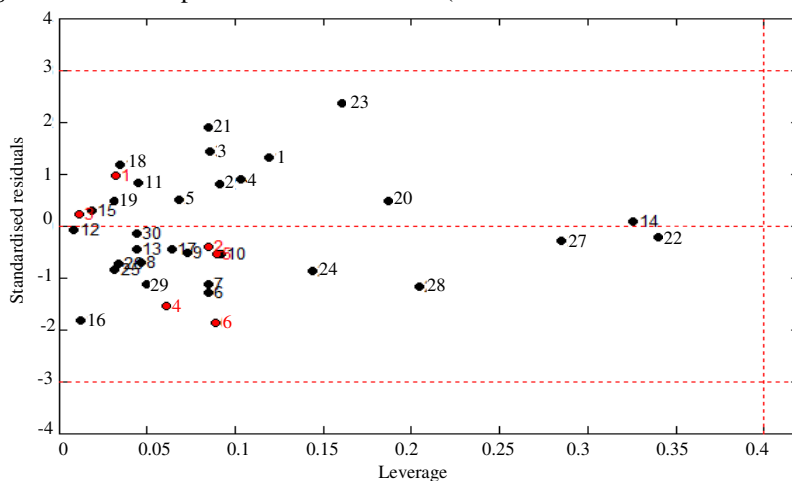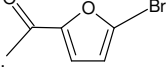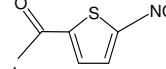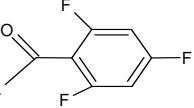
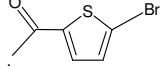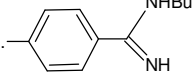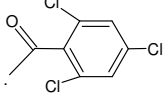
Figure 5. Williams plot for the MNLR model (with $h^* = 0.4$ and residual limits = $\pm3\sigma$)

Proposed novel compounds

Consequently, with MLR and MNLR approaches, we can design new compounds with different and improved values of antileishmanial activity than the studied compounds (Table 6). Taking into account the a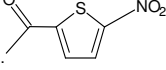bove results, we added suitable substitutions and then calculated the activities of the new compounds using the proposed model in equation1 and 2.

Table 6. Proposed compounds, values of calculated descriptors, and predicted values of pIC_{50} using MLR and MNLR model

Ar	Het	E_{HOMO}	E_{LUMO}	Log P	pIC_{50} RLM	pIC_{50} RNLM	h
X_1	$NHCH_2C_6H_4OCH_3$	-4.055	-1.801	4.16	4.16	3.09	0.272
X_2	$NHCH_2C_6H_4OCH_3$	-4.598	-1.456	4.622	4.85	4.05	0.398
X_3	$NHCH_2CHCH_2$	-4.931	-2.025	4.831	4.67	4.04	0.301
X_4	Het1	-5.232	-2.775	4.67	4.25	3.88	0.388
X_5	Het1	-5.732	-2.377	2.307	4.45	4.53	0.069
X_6	Het1	-5.555	-2.224	2.687	4.52	4.58	0.071
X_7	Het2	-5.556	-2.278	2.673	4.48	4.547	0.05
X_8	Het3	-6.312	-2.155	4.340	5.40	5.171	0.312

The leviers h of new compounds are less than the critical value ($h^* = 0.4$) Therefore are regarded reliable compounds for design new compounds with different and improved values of activity than the studied compounds. These results provided here can be used for drug design and development of new and safer drugs is warranted.

Conclusion

The statistical analysis methods were used to develop quantitative structure–activity relation models of the antileishmanial activity of (5-nitroheteroaryl-1,3,4-thiadiazole-2-yl) piperazinyl derivatives. The artificial neural network had substantially better predictive capability than the multiple linear and non-linear regressions, with greater predictive power. We established satisfactory relations between several descriptors and antileishmanial activity. The results show that the proposed models in this paper can predict activity accurately and that the selected descriptors are pertinent to explain this activity.

The accuracy and predictability of the proposed models were illustrated by comparison of the key statistical indicators r or r2 for the different models (Table 5). The applicability domain of the proposed models was investigated using William's plot to detect the subspace of chemical structures that can be predicted reliably by the two regressions models. The proposed methods will reduce the time and cost of synthesis and determination of the antileishmanial activity of (5-nitroheteroaryl-1,3,4-thiadiazole-2-yl) piperazinyl derivatives. Furthermore, the descriptors are sufficiently rich in chemical, electronic and topological information to encode structural features that could be used with other descriptors in the development QSAR models.

The conflict of interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgment

References

1. Croft S L, Seifert K and Yardley V, *Indian J Med Res.*, 2006, 123(3), 399-410.
2. McConville M J and Handman E, *Int J Parasitology*, 2007, **37(10),** 1047-1051; DOI:10.1016/S0020-7519(07)00216-0
3. Rocha L G, Almeida J R, Macedo R O and Barbosa-Filho J M, *Phytomedicine*, 2005, **12(6-7),** 514-535; DOI:10.1016/j.phymed.2003.10.006
4. Kayser O, Kiderlen A F and Croft S L, *Stud Nat Prod Chem.*, 2002, **26,** 779-848; DOI:10.1016/S1572-5995(02)80019-9
5. Desjeux P, *Trans Roy Soc Trop Med Hyg.*, 2001, **95(3),** 239-243; DOI:10.1016/S0035-9203(01)90223-8
6. Denny W A, *Cur Med Chem.*, 2002, **9(18),** 1655-1665; DOI:10.2174/0929867023-369277
7. Wainwright M, *J Antimicrobial Chemotherapy*, 2001, **47(1),** 1-13; DOI:10.1093/jac/47.1.1
8. Antonini I, *Cur Med Chem.*, 2002, **9(18),** 1701-1716; DOI:10.2174/0929867023369268
9. Demeunynck M, Charmantray F and Martelli A, *Cur Pharm Des.*, 2001, **7(17),** 1703-1724; DOI:10.2174/1381612013397131
10. Marrapu V K, Mittal M, Shivahare R, Gupta S and Bhandari K, *European J Med Chem.*, 2011, **46(5)**, 1694-1700; DOI:10.1016/j.ejmech.2011.02.021
11. Marrapu V K, Srinivas N, Mittal M, Shakya N, Gupta S and Bhandari K, *Bioorg Med Chem Lett.*, 2011, **21(5)**, 1407-1410; DOI:10.1016/j.bmcl.2011.01.026
12. Tahghighia A, Hamzeh-Mivehroud M, Asadpour Zeynali K, Foroumadi A and Dastmalchi S, *J Chemometrics*, 2016, **30(5),** 284-293; DOI:10.1002/cem.2789
13. Frisch M J, Gaussian 03 Revision B.01, Gaussian, Inc., Pitts-burgh, PA, 2003.
14. ACDLABS 10, Advanced Chemistry Development, Inc., Toronto, Canada, 2015; http://www.acdlabs.com/.
15. XLSTAT 2015 software (XLSTAT Company), http://www.xlstat.com.
16. Chtita S, Ghamali M, Hmamouchi R, Bouhya E, Bourass M, Larif M, Bouachrine M and Lakhlifi T, *Adv Phys Chem.*, 2016, 2016, 1-16; DOI:10.1155/2016/5137289

17. Ghamali M, Chtita S, Ousaa A, Elidrissi B, Bouachrine M and Lakhlifi T, *J Taibah Uni Sci.*, 2017, **11(1),** 1-10; DOI:10.1016/j.jtusci.2016.03.002
18. Ghamali M, Chtita S, Aouidate A, Ghaleb A, Bouachrine M and Lakhlifi T, *J Taibah Uni Sci*, 2017, **11(3),** 422-433; DOI:10.1016/j.jtusci.2016.06.005
19. Larif M, Adad A, Hmammouchi R, Taghki A I, Soulaymani A, Elmidaoui A, Bouachrine M and Lakhlifi T, *Arabian J Chem.*, 2017, **10(1),** S946-S955; DOI:10.1016/j.arabjc.2012.12.033
20. Ousaa A, Elidrissi B, Ghamali M, Chtita S, Bouachrine M and Lakhlifi T, *J Comp Met Mol Des.*, 2014, **4(3),** 10-18.
21. Ousaa A, Elidrissi B, Ghamali M, Chtita S, Bouachrine M and Lakhlifi T, *J Mat Env Sci.*, 2017, **8(2)**, 476-484.
22. Golbraikh A and Tropsha A, *J Mol Grap Mod.*, 2002, **20(4)**, 269-276.
23. Eriksson L, Jaworska J, Worth A P, Cronin M T D, McDowell R M and Gramatica P, *Env Health Perspectives*, 2003, **111(10),** 1361-1375; DOI:10.1289/ehp.5758