EXCLI Journal 2015;14:484-495 – ISSN 1611-2156
Received: February 02, 2015, accepted: March 24, 2015, published: March 30, 2015

Original article:

2D-QSAR STUDY OF SOME 2,5-DIAMINOBENZOPHENONE FARNESYLTRANSFERASE INHIBITORS BY DIFFERENT CHEMOMETRIC METHODS

Saeed Ghanbarzadeh[1], Saeed Ghasemi[2*], Ali Shayanfar[3], Heshmatollah Ebrahimi-Najafabadi[2]

[1] Drug Applied Research center and Faculty of Pharmacy, Tabriz University of Medical Sciences, Tabriz, Iran
[2] Department of Medicinal Chemistry, School of Pharmacy, Guilan University of Medical Sciences, Rasht, Iran
[3] Department of Medicinal Chemistry, Faculty of Pharmacy, Tabriz University of Medical Sciences, Tabriz, Iran

* Corresponding author: Saeed Ghasemi, Department of Medicinal Chemistry, School of Pharmacy, Guilan University of Medical Sciences complex, Fouman-Saravan highway, Rasht, Iran, Tel: 00981333486470, Fax: 00981333486475, Postal code: 73774-41941. E-mail: ghasemi_saeed@yahoo.com

http://dx.doi.org/10.17179/excli2015-177

ABSTRACT

Quantitative structure activity relationship (QSAR) models can be used to predict the activity of new drug candidates in early stages of drug discovery. In the present study, the information of the ninety two 2,5-diaminobenzophenone-containing farnesyltranaferase inhibitors (FTIs) were taken from the literature. Subsequently, the structures of the molecules were optimized using Hyperchem software and molecular descriptors were obtained using Dragon software. The most suitable descriptors were selected using genetic algorithms-partial least squares and stepwise regression, where exhibited that the volume, shape and polarity of the FTIs are important for their activities. The two-dimensional QSAR models (2D-QSAR) were obtained using both linear methods (multiple linear regression) and non-linear methods (artificial neural networks and support vector machines). The proposed QSAR models were validated using internal validation method. The results showed that the proposed 2D-QSAR models were valid and they can be used for prediction of the activities of the 2,5-diaminobenzophenone-containing FTIs. In conclusion, the 2D-QSAR models (both linear and non-linear) showed good prediction capability and the non-linear models were exhibited more accuracy than the linear models.

Keywords: QSAR, multiple linear regression, artificial neural network, support vector machine

INTRODUCTION

Malaria is a deadly disease which is cause of more than 2-3 million deaths every year in the world and is estimated to be endemic in over 100 different countries. There are over 200 different *Plasmodium* species, but only 4 known types actually cause human malaria. *Plasmodium falciparum* is more dangerous and deadly than other species of *Plasmodium*species that can cause malaria in human (Eastman et al., 2007; Olepu et al., 2008; Xie et al., 2006).

Because of problems with available drugs (Chloroquine), such as drug resistance,

EXCLI Journal 2015;14:484-495 – ISSN 1611-2156
Received: February 02, 2015, accepted: March 24, 2015, published: March 30, 2015

finding new drugs with new mechanisms for treatment of malaria is required (Gupta and Prabhakar, 2008; Xie et al., 2006).

The RAS proteins belong to a family of related polypeptides that are present in all eukaryotic organisms from yeast to human. The RAS proteins are critical in signal transduction pathway and in cell growth. Several studies on RAS proteins have showed that some post-translational modifications are essential for its biological activity (Ghasemi et al., 2013b; Lu et al., 2007; Puntambekar et al., 2008). The first step of these modifications is farnesylation by farnesyltransferase enzyme (FTase). FTase is a heterodimeric metalloenzyme that contain a zinc ion (Gilleron et al., 2007; Puntambekar et al., 2008; Xie et al., 2006). FTase adds a C-15 farnesyl group from farnesyl pyrophosphate (FPP) to the cysteine of the CAAX sequence (C=cys, A=an aliphatic amino acid, X is typically Met) in the carboxyl terminal of RAS proteins (Bolchi et al., 2007; Equbal et al., 2008; S Ghasemi et al., 2013a, b; Lu et al., 2007; Tanaka et al., 2007).

It has been showed that farnesyltranaferase inhibitors (FTIs) can inhibit the growth of *Plasmodium falciparum* in human red blood cells (Ohkanda et al., 2001). Therefore, these compounds can be used as antimalarial agents against *Plasmodium falciparum* (Shayanfar et al., 2013).

Several classes of antimalarial FTIs have been synthesized such as 2,5-diaminobenzophenone derivatives, biphenyl derivatives, tetrahydroquinoline and etc. (Ohkanda et al., 2001; S Olepu et al., 2008).

The drug development contributes to high cost and long time. Quantitative structure–activity relationship (QSAR) approach as a computational methods can be used to predict drug biological activity by finding a correlation between the structures and the activities of drugs, and therefore decreases the cost and time of the drug development (Shayanfar et al., 2013; Yee and Wei, 2012). This methods are based on correlation between molecular properties and differences in the features of the molecules (Jain et al., 2012).

Two-dimensional (2D) and three-dimensional (3D)-QSAR are the most common QSAR models. 2D-QSAR models investigate correlation between the activities of active molecules and structures without regarding the three-dimensional conformations of the molecules. However, 3D-QSAR models consider the 3D conformations of the molecules (Shayanfar et al., 2013).

Several studies by 2D-QSAR modeling were performed for prediction of FTIs biological activities. Freitas and Castilho (2008) investigated the activities of tetrahydroquinoline FTIs using multiple linear regression (MLR) models. Gupta and his coworker also correlated FTI activities to tetrahydroquinoline analogues structures with 2D-QSAR model with the Combinatorial Protocol in Multiple Linear Regression (CP–MLR), a filter based variable selection procedure (Gupta and Prabhakar, 2008). Modeling studies were performed for some thiol and non-thiolpeptidomimetic inhibitors using artificial neural networks (ANN) and radial distribution function (RDF) approaches by Gonzalez et al. (2006). Recently Gaurav et al. (2011) and Shayanfar et al. (2013) also studied QSAR of imidazole containing FTIs.

Despite of the many benefits of 3D-QSAR models, 2D-QSAR models have some beneficial advantages. In 2D-QSAR models it is not necessary to align the structures that can create some limitation in 3D-QSAR. Furthermore, development of 2D-QSAR models is very faster and easier than 3D-QSAR models (Shayanfar et al., 2013).

Literature review indicated that, no 2D-QSAR study has been reported for 2,5-diaminobenzophenone-containing FTIs. Therefore in the present work, 92 FTIs with 2,5-diaminobenzophenone scaffold were used to develop 2D-QSAR models by various chemometric methods. Multiple linear regression (MLR), ANN and support vector machine (SVM) methods were used to predict the IC_{50} of the 2,5-diaminobenzophenone-containing FTIs. Genetic algorithms-partial

EXCLI Journal 2015;14:484-495 – ISSN 1611-2156
Received: February 02, 2015, accepted: March 24, 2015, published: March 30, 2015

least squares (GA-PLS) and stepwise-regression methods were used to select molecular descriptors. Internal validation method was used for confirmation of the validities of the developed models.

MATERIAL AND METHODS

Data Set

The pIC_{50}, negative logarithm of the IC_{50} (half maximal enzyme inhibitory concentration), values of the ninety two 2,5-diaminobenzophenone-containing FTIs were collected from the literature (Xie et al., 2006). This data set is formed of the five different groups of 2,5-diaminobenzophenone-containing FTIs. Chemical structures of these compounds are shown in Figure 1. In order to compare the results of the present study (2D-QSAR) with previous 3D-QSAR study, the same carefully-selected training and test sets were employed in the model development (Xie et al., 2006).

Molecular descriptors

The structures of the all studied compounds were drawn using Hyperchem 8.0 software and pre-optimized with the molecular mechanics force field (MM+) method to calculate molecular descriptors. Subsequently, AM1 semiempirical calculations were performed for optimization of the 3D geometries of the molecules with the Polak-Ribière (conjugate gradient) algorithm.

Finally, Hyperchem 8.0 software was fed into the Dragon 3.0 software and the molecular descriptors of these compounds were calculated.

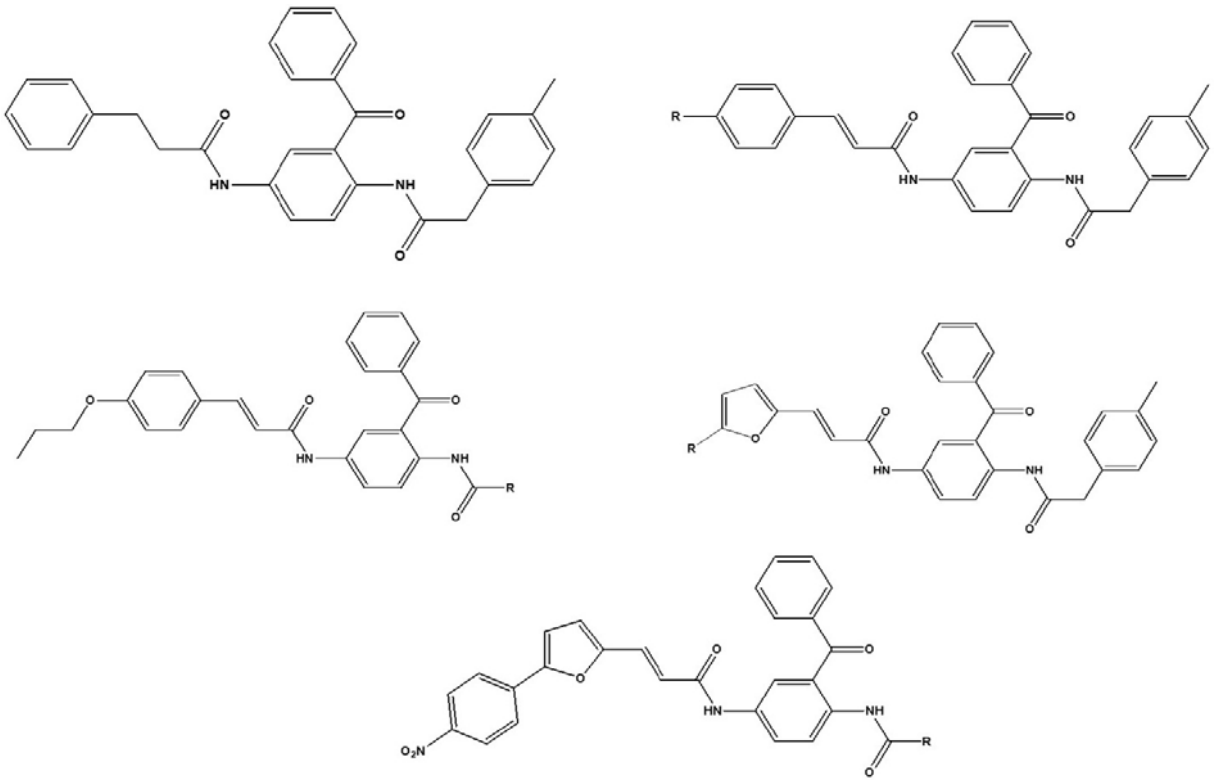

Figure 1: Structures of the studied 2,5-diaminobenzophenone-containing farnesyltranaferase inhibitors

EXCLI Journal 2015;14:484-495 – ISSN 1611-2156
Received: February 02, 2015, accepted: March 24, 2015, published: March 30, 2015

Descriptors selection

With the aim of reduction in the number of descriptors, the descriptors belonging to 74 compounds in the training set with higher than 50 % repeated values or collinear descriptors ($R > 0.9$) were excluded and further reduction in the number of descriptors was performed with Genetic algorithm and partial least square, a valuable tool for data reduction, (GA-PLS). GA simulates the process of natural evolution and has been used commonly as an acceptable method for reducing the number of descriptors (Dastmalchi et al., 2008; Habibi-Yangjeh, 2009; Soltani et al., 2010). The MATLAB 7.8 software was used to run the GA-PLS method developed by Leardi et al. (2002). Population size is one of the major factors which affect the performance of the algorithm and it is necessary to have good population to produce optimal result in quick time. The population size of genetic algorithms in this study was considered as 100. Ten percent of the descriptors with top scores were selected and the descriptor selection was performed using stepwise regression. High correlations with response and low inter-correlation between descriptors (using Pearson correlation) were considered as selection criteria before stepwise regression.

MODEL BUILDING

MLR Model

The selected descriptors were employed to develop a MLR equation using SPSS 16 software. Statistical properties of the proposed equation including correlation coefficient (R), adjusted correlation coefficient (Radj), standard error of estimate (SEE), probability values (p-value) of each descriptor, and Fischer statistic or variance ratio (F), recommended by Dearden et al., (2009) were obtained. The proposed model was validated using the leave one-out (LOO) method to evaluate prediction capability of the model.

ANN with the Levenberg-Marquardt Algorithm

ANN, which mimics human brain process information, is useful in detecting complex non-linear relationship between a set of inputs and outputs. Briefly, the general structure of ANN has one input layer, one or more hidden layers and one output layer. Each layer has some units corresponding to neurons. The units in neighboring layers are fully interconnected with links corresponding to synapses. The strengths of connections between two units are called ‘weights’. Selected descriptors are neurons of the input layer, pIC_{50} values of compounds are the output neurons and a three layer networks with three neurons in the hidden layer was designed. ANN learns an approximate non-linear relationship by a training procedure, which involves varying weight values. Training means a search process for the optimized set of weight values, which can minimize the squared error between the estimated and experimental data of units in the output layer. The number of training cycles was selected on the basis of the Mean Squared Error (MSE) of the validation subset, which prevents the network from over-training (Jalali-Heravi et al., 2008). Neural networks for modeling in conjunction with genetic algorithms have proved very powerful for optimization. There are different algorithms for weight update functions in the literature. In the recent QSAR studies, the Levenberg–Marquardt algorithm was considered as one of the most effective algorithm (Arab Chamjangali, 2009, 2007; Jalali-Heravi et al., 2008) . In this study, we used the nftool (network-fitting tool) toolbox of MATLAB 7.8 software for training of the network. This toolbox is user-friendly and uses Levenberg–Marquardt back propagation algorithms (Trainlim) for ANN training. For training a valid network and preventing over fitting the 56 data points of the training set, described for MLR, were randomly classified into training (70 %), validation (15 %) and test (15 %) sets.

Support Vector Machine

SVM is a new and very promising classification and regression method developed by Vapnik (2000) SVM have been successfully used to solve classification and correlation problems, such as cancer diagnosis, identification of HIV protease cleavage sites, protein class prediction, etc. SVMs have also been applied in chemistry and QSAR studies (Cheng et al., 2010; Darnag et al., 2010; Shahlaei et al., 2010; Vapnik, 2000). In this method a hyperplane is constructed in a multidimensional space which provides the minimum error by employing a non-linear kernel function for classification or regression tasks. Some parameters should be optimized in SVM analysis include the capacity parameter (C) that is a regularization parameter that adjusts maximizing the distance from the hyperplane to any training set data points and minimizing the error. ε is another parameter which is related to noise in the data. A common type of kernel function is a radial basis function (RBF) (Asadpour-Zeynali and Soheili-Azad, 2010; Katritzky et al., 2010; Louis et al., 2010; van de Waterbeemd and Testa, 2008). This function has a parameter (γ) which should be optimized and controls the generalization ability of the SVM. The C and ε parameters were optimized using the leave-many-out cross-validation method. SVM was performed using STATISTICA 7 software.

RESULTS AND DISCUSSION

Selection of descriptors

The details of the selected descriptors by using GA-PLS and stepwise regression (where less than 1 descriptor per 9 compounds was selected) are shown in Table 1. Results indicated that a mixture of 2D and 3D descriptors showed the best predictability for the pIC_{50} value of the studied structures. Three of the used descriptors are topological descriptors, which is a connectivity index is a type of a molecular descriptor that is calculated based on the molecular graph of a chemical compound. In addition, BCUT as another 2D descriptors are selected (Todeschini and V Consonni, 2008). On the other hand, four 3D descriptors including RDF, WHIM and GETAWAY were in the selected descriptors. According to the selected descriptors, it was found that both of the volume, shape and polarity of the molecules were important for the activity of the studied compounds.

A correlation matrix indicated that there was no intercorrelation ($R < 0.6$) between the selected descriptors (Table 2) which showed that the selected descriptors were linearly independent and as a result could be used simultaneously in the QSAR models development.

Model building using different methods

The selected descriptors were used for the QSAR models development by using MLR, ANN as well as SVM. Based on the obtained results, a linear model, as the simplest and most straightforward model was proposed. The standard error of estimate, coefficient and the p-value of the selected descriptors of the most accurate MLR model were presented in Table 3. Furthermore, statistical information which is necessary to validate QSAR models are presented in Table 4 for the proposed models in the present study. The results indicated that there was no significant difference between Rand Radj and the correlation coefficient was acceptable ($P < 0.05$).The influence of the number of descriptors on Rand Radj for the developed model are presented in Figure 2. The increase in the descriptor number resulted in increase in Radj value confirmed the influence of all the selected descriptors (Dearden et al., 2009)

The obtained data was used to develop an ANN model with three optimal hidden neurons. Table 4 shows the statistical parameters of the developed ANN model for the data set including training, validation and test sets. There are no significant changes between statistical properties of these sets.

Selected descriptors were used to develop SVM models. The STATISTICA 7 software was employed for optimization of the

EXCLI Journal 2015;14:484-495 – ISSN 1611-2156
Received: February 02, 2015, accepted: March 24, 2015, published: March 30, 2015

Table 1: Selected descriptors by GA-PLS and stepwise regression from DRAGON software

Number	Symbol	Definition	Class
1	IC3	Information content index (neighborhood symmetry of 3-order)	Topological descriptors
2	VRA1	Randic-type eigenvector-based index from adjacency matrix	Topological descriptors
3	RDF060e	Radial Distribution Function - 6.0 / weighted by atomic Sanderson electronegativities	RDF descriptors
4	E2s	2nd component accessibility directional WHIM index / weighted by atomic electrotopological states	WHIM descriptors
5	HATS2e	Leverage-weighted autocorrelation of lag 2 / weighted by atomic Sanderson electronegativities	GETAWAY descriptors
6	BELm7	Lowest eigenvalue n. 7 of Burden matrix / weighted by atomic masses	BCUT descriptors
7	BELm6	Lowest eigenvalue n. 6 of Burden matrix / weighted by atomic masses	BCUT descriptors
8	GNar	Narumi geometric topological index	Topological descriptors
9	R7m+	R maximal autocorrelation of lag 7 / weighted by atomic masses	GETAWAY

Table 2: Correlation matrix between selected descriptors

	IC3	VRA1	RDF060e	E2s	HATS2e	BELm7	BELm6	GNar	R7m+
IC3	1								
VRA1	0.091	1							
RDF060e	0.220	0.068	1						
E2s	0.173	0.244	0.203	1					
HATS2e	0.135	0.194	0.454	0.055	1				
BELm7	0.502	0.252	0.458	0.026	0.533	1			
BELm6	0.156	0.257	0.136	0.012	0.413	0.557	1		
GNar	0.062	0.015	0.277	0.063	0.411	0.095	0.209	1	
R7m+	0.048	0.039	0.160	0.027	0.062	0.232	0.283	0.067	1

SVM parameters (C, ε and γ) with 10-fold cross-validation. A robust model can develop by selecting parameters which give the lowest error. The optimized values of C, ε and γ were 7, 0.001 and 0.1, respectively. The statistical properties of the proposed SVM model for the training set are listed in Table 4.

Experimental and predicted pIC_{50} as well as absolute error values using MLR, ANN and SVM models are summarized in Table 5.

Figure 3 shows the experimental versus predicted values for training (74 data points) and test sets (18 data points) using MLR, ANN and SVM models. The AAE values of training and test compounds are listed in Table 6. These data indicated that the developed models have good predictability. The AAE's of the ANN and SVM models were better than those of the MLR model and accordingly, as shown in Tables 4, R values of the ANN and SVM models (as non-linear models) were also greater than that of the MLR model (as linear model) indicating that the SVM and ANN models are more accurate than the MLR model.

EXCLI Journal 2015;14:484-495 – ISSN 1611-2156
Received: February 02, 2015, accepted: March 24, 2015, published: March 30, 2015

CONCLUSION

Different chemometric methods were used to developed QSAR models to predict the activities of 2,5-diaminobenzophe–containing FTIs employing a collection of 2D and 3D descriptors to display the FTI structures. The obtained results demonstrated that the volume, shape and polarity are important parameters for the activity of the studied compounds. Furthermore, developed 2D-QSARmodels by using linear (MLR) and especially nonlinear(ANN and SVM) methods can be used to predict the activities of FTIs with high accuracy. In conclusion, the proposed models could be used in drug design to evaluate novel 2,5-diaminobenzophe-containing FTIs.

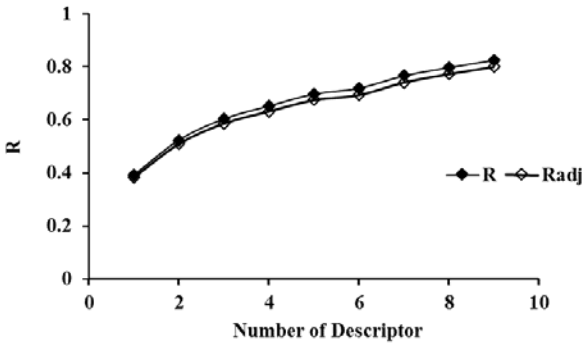

Figure 2: Effects of the number of descriptors on R and Radj values.

Table 3: Coefficients and standard error of estimate and the p-value of the selected descriptors of the most accurate MLR model

Descriptors	Coefficient	SEE	p-value
Constant	-20.294	4.005	<0.001
IC3	1.511	0.343	<0.001
VRA1	0.000	0.000	<0.001
RDF060e	0.019	0.004	<0.001
E2s	-1.596	0.407	<0.001
HATS2e	19.844	3.656	<0.001
BELm7	5.029	0.881	<0.001
BELm6	-2.938	0.744	<0.001
GNar	5.581	1.577	0.001
R7m+	7.280	2.302	0.002

Table 4: Statistical information for the proposed models for the Training Set

	N	R^2	$Radj^2$	SEE[a]	F[a]
MLR					
Training set	74	0.826	0.801	0.313	33.66
ANN					
Training Set	52	0.849	0.846	0.277	280.56
Test	11	0.897	0.886	0.254	78.46
Validation	11	0.850	0.834	0.279	51.18
overall	74	0.849	0.847	0.274	405.83
SVM					
Training set	74	0.841	0.839	0.282	380.96

EXCLI Journal 2015;14:484-495 – ISSN 1611-2156
Received: February 02, 2015, accepted: March 24, 2015, published: March 30, 2015

Table 5: Experimental (exp) pIC_{50}, predicted (pred) IC_{50} and absolute error (AE) values of 74 training and 18 test set compounds

No.	pIC_{50}exp	MLR pIC_{50}pred	AE	ANN pIC_{50}pred	AE	SVM pIC_{50}pred	AE
Training set							
1	5.57	5.366	0.204	5.578	0.008	5.417	0.153
2	5.19	5.237	0.047	5.154	0.036	5.156	0.034
3	4.4	5.020	0.620	4.812	0.412	5.074	0.674
4	6	6.302	0.302	6.218	0.218	6.171	0.171
5	5.24	5.044	0.196	5.056	0.184	5.087	0.153
6	5.26	5.292	0.032	5.260	0.000	5.295	0.035
7	5.49	5.605	0.115	5.509	0.019	5.579	0.089
8	4.37	4.761	0.391	4.467	0.097	4.740	0.370
9	5.85	5.403	0.447	5.625	0.225	5.443	0.407
10	5.89	6.126	0.236	6.304	0.414	6.091	0.201
11	5.92	6.033	0.113	6.100	0.180	5.961	0.041
12	5.92	5.827	0.093	5.821	0.099	5.757	0.163
13	5.52	5.339	0.181	5.188	0.332	5.448	0.072
14	6.07	6.271	0.201	6.301	0.231	6.240	0.170
15	5.96	5.510	0.450	5.312	0.648	5.807	0.153
16	6.47	6.298	0.172	6.258	0.212	6.267	0.203
17	5.05	5.210	0.160	5.517	0.467	5.202	0.152
18	5.6	5.502	0.098	5.145	0.455	5.549	0.051
19	5.62	5.952	0.332	5.829	0.209	6.006	0.386
20	6.46	6.240	0.220	6.426	0.034	6.270	0.190
21	6.51	6.410	0.100	6.703	0.193	6.358	0.152
22	5.55	5.836	0.286	5.874	0.324	5.870	0.320
23	6.92	6.529	0.391	6.489	0.431	6.520	0.400
24	4.62	4.745	0.125	4.644	0.024	4.773	0.153
25	4.64	4.396	0.244	4.654	0.014	4.487	0.153
26	6	6.327	0.327	5.983	0.017	6.152	0.152
27	6.7	7.011	0.311	6.876	0.176	6.854	0.154
28	6.92	6.889	0.031	6.895	0.025	6.766	0.154
29	7.06	6.987	0.073	7.014	0.046	6.908	0.152
30	7.07	6.919	0.151	6.824	0.246	6.917	0.153
31	6.89	6.935	0.045	7.102	0.212	6.927	0.037
32	6.52	6.827	0.307	6.952	0.432	6.812	0.292
33	6.12	6.305	0.185	6.130	0.010	6.235	0.115
34	6.68	6.846	0.166	6.793	0.113	6.733	0.053
35	6.49	6.645	0.155	6.834	0.344	6.666	0.176
36	6.84	6.661	0.179	6.912	0.072	6.683	0.157
37	6.9	6.744	0.156	6.857	0.043	6.747	0.153
38	7.12	6.717	0.403	6.721	0.399	6.680	0.440
39	6.17	6.677	0.507	6.698	0.528	6.736	0.566
40	7.11	7.123	0.013	7.112	0.002	6.969	0.141
41	6.59	6.804	0.214	6.606	0.016	6.777	0.187

EXCLI Journal 2015;14:484-495 – ISSN 1611-2156
Received: February 02, 2015, accepted: March 24, 2015, published: March 30, 2015

No.	pIC_{50}exp	MLR pIC_{50}pred	AE	ANN pIC_{50}pred	AE	SVM pIC_{50}pred	AE
42	6.66	6.541	0.119	6.469	0.191	6.508	0.152
43	7.17	6.984	0.186	7.057	0.113	6.969	0.201
44	6.77	6.645	0.125	6.736	0.034	6.641	0.129
45	6.25	6.487	0.237	6.452	0.202	6.528	0.278
46	6.55	6.536	0.014	6.470	0.080	6.564	0.014
47	7.43	6.825	0.605	6.990	0.440	6.812	0.618
48	7.22	7.153	0.067	7.333	0.113	7.068	0.152
49	6.7	6.628	0.072	6.796	0.096	6.642	0.058
50	5.52	5.360	0.160	5.534	0.014	5.439	0.081
51	5.6	5.942	0.342	5.857	0.257	5.955	0.355
52	6.11	5.714	0.396	5.828	0.282	5.666	0.444
53	6.57	6.531	0.039	6.688	0.118	6.549	0.021
54	6.82	6.418	0.402	6.532	0.288	6.516	0.304
55	6.19	6.233	0.043	6.305	0.115	6.260	0.070
56	6.64	6.621	0.019	6.533	0.107	6.593	0.047
57	7.15	7.319	0.169	7.331	0.181	7.164	0.014
58	7.33	6.807	0.523	7.104	0.226	6.574	0.756
59	6	6.477	0.477	6.281	0.281	6.343	0.343
60	5.85	6.292	0.442	6.107	0.257	6.209	0.359
61	6.6	6.420	0.180	6.332	0.268	6.446	0.154
62	6.68	6.146	0.534	6.117	0.563	6.208	0.472
63	6	5.938	0.062	5.937	0.063	5.946	0.054
64	5.26	5.435	0.175	5.372	0.112	5.421	0.161
65	6.51	5.828	0.682	5.929	0.581	5.955	0.555
66	5.48	6.334	0.854	6.429	0.949	6.407	0.927
67	6.36	6.122	0.238	6.269	0.091	6.207	0.153
68	6.2	6.460	0.260	6.337	0.137	6.469	0.269
69	5.96	6.087	0.127	5.862	0.098	6.113	0.153
70	6.36	6.084	0.276	6.062	0.298	6.168	0.192
71	6.77	6.604	0.166	6.565	0.205	6.641	0.129
72	6.05	6.240	0.190	6.008	0.042	6.203	0.153
73	6.38	6.420	0.040	6.439	0.059	6.360	0.020
74	6.24	6.349	0.109	6.249	0.009	6.354	0.114
Test set							
75	5.24	4.948	0.292	4.859	0.381	5.017	0.223
76	5.26	5.047	0.213	5.233	0.028	5.170	0.090
77	5.47	5.773	0.303	5.979	0.509	5.821	0.351
78	5.89	5.546	0.344	5.329	0.561	5.612	0.278
79	6.38	6.297	0.083	6.374	0.006	6.327	0.053
80	6.52	6.360	0.160	5.869	0.652	6.215	0.305
81	7.08	6.752	0.328	6.914	0.167	6.817	0.263
82	6.23	6.527	0.297	6.450	0.220	6.563	0.333
83	6.25	6.338	0.088	6.257	0.007	6.441	0.191
84	5.6	5.408	0.192	5.564	0.036	5.513	0.087

EXCLI Journal 2015;14:484-495 – ISSN 1611-2156
Received: February 02, 2015, accepted: March 24, 2015, published: March 30, 2015

No.	pIC_{50}exp	MLR pIC_{50}pred	AE	ANN pIC_{50}pred	AE	SVM pIC_{50}pred	AE
85	6.49	6.780	0.290	6.714	0.224	6.809	0.319
86	7.19	7.330	0.140	7.546	0.356	7.239	0.049
87	6	6.864	0.864	6.640	0.640	6.851	0.851
88	5.89	6.054	0.164	6.052	0.162	6.176	0.286
89	7.21	6.854	0.356	7.109	0.101	6.668	0.542
90	6.89	6.729	0.161	6.965	0.075	6.738	0.152
91	5.85	5.705	0.145	5.458	0.392	5.696	0.154
92	5.92	5.730	0.190	5.820	0.100	5.885	0.035

Table 6: AAE's of the proposed models using different chemometrics methods

	AAE of Training Set (N=74)	AAE of Test Set (N=18)
MLR	0.2339±0.174	0.2561±0.175
ANN	0.2028±0.182	0.2564±0.220
SVM	0.2166±0.181	0.2534±0.199

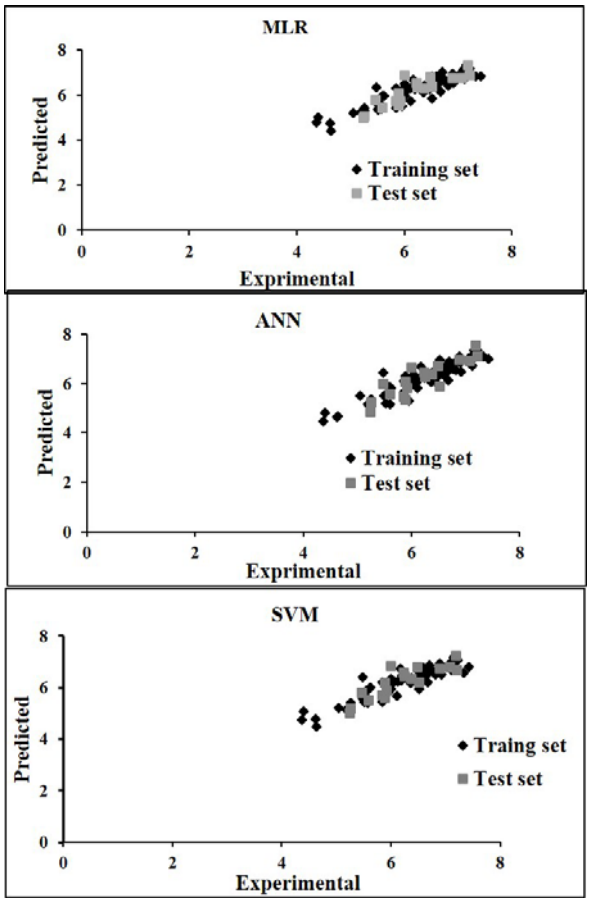

Figure 3: Experimental versus predicted pIC_{50} values using MLR, ANN and SVM models

Conflict of interest

The authors declare that they have no conflict of interest.

REFERENCES

Arab Chamjangali M. Modelling of cytotoxicity data (CC50) of anti-HIV 1-(5-chlorophenyl) sulfonyl)-1H-pyrrole derivatives using calculated molecular descriptors and Levenberg–Marquardt Artifical Neural Network. Chem Biol Drug Des. 2009;73:456-65.

Arab Chamjangali M, Beglari M, Bagherian G. Prediction of cytotoxicity data (CC50) of anti-HIV 5-pheny-l-phenylamino-1H-imidazole derivatives by artificial neural network trained with Levenberg–Marquardt algorithm. J Mol Graph Model. 2007;26: 360-7.

Asadpour-Zeynali K, Soheili-Azad P. Simultaneous polarographic determination of isoniazid and rifampicin by differential pulse polarography method and support vector regression. Electrochim Acta. 2010;55: 6570-6.

Bolchi C, Pallavicini M, Rusconi C, Diomede L, Ferri N, Corsini A, et al. Peptidomimetic inhibitors of farnesyltransferase with high in vitro activity and significant cellular potency. Bioorg Med Chem Lett. 2007; 17:6192-6.

Cheng Z, Zhang Y, Fu W. QSAR study of carboxylic acid derivatives as HIV-1 Integrase inhibitors. Eur J Med Chem. 2010;45:3970-80.

Darnag R, Mostapha Mazouz E, Schmitzer A, Villemin D, Jarid A, Cherqaoui D. Support vector machines: development of QSAR models for predicting anti-HIV-1 activity of TIBO derivatives. Eur J Med Chem. 2010;45:1590-7.

Dastmalchi S, Hamzeh-Mivehroud M, Ghafourian T, Hamzeiy H. Molecular modeling of histamine H3 receptor and QSAR studies on arylbenzofuran derived H3 antagonists. J Mol Graph Model. 2008;26:834-44.

Dearden J, Cronin M, Kaiser K. How not to develop a quantitative structure–activity or structure–property relationship (QSAR/QSPR). SAR QSAR Environ Res. 2009;20:241-66.

Eastman RT, White J, Hucke O, Yokoyama K, Verlinde CL, Hast MA, et al. Resistance mutations at the lipid substrate binding site of *Plasmodium falciparum* protein farnesyltransferase. Mol Biochem Parasitol. 2007;152:66-71.

Equbal T, Silakari O, Ravikumar M. Exploring three-dimensional quantitative structural activity relationship (3D-QSAR) analysis of SCH 66336 (Sarasar) analogues of farnesyltransferase inhibitors. Eur J Med Chem. 2008;43:204-9.

Freitas HF, Castilho MS. 2D QSAR studies of farnesyltransferase inhibitors againts Plasmodium falciparum. 4th Brazilian Symposium on Medicinal Chemistry - Braz Med Chem. 2008.

Gaurav A, Gautam V, Singh R. Exploring the structure activity relationships of imidazole containing tetrahydrobenzodiazepines as farnesyltransferase inhibitors: A QSAR study. Lett Drug Des Discov. 2011; 8:506-15.

Ghasemi S, Davaran S, Sharifi S, Asgari D, Abdollahi A, Mojarrad JS. Comparison of cytotoxic activity of L778123 as a farnesyltranferase inhibitor and doxorubicin against A549 and HT-29 cell lines. Adv Pharm Bull. 2013a;3:73-7.

Ghasemi S, Sharifi S, Davaran S, Danafar H, Asgari D, Mojarrad JS. Synthesis and cytotoxicity evaluation of some novel 1-(3-Chlorophenyl) piperazin-2-one derivatives bearing imidazole bioisosteres. Aust J Chem. 2013b;66:655-60.

Gilleron P, Wlodarczyk N, Houssin R, Farce A, Laconde G, Goossens J-F, et al. Design, synthesis and biological evaluation of substituted dioxodibenzothiazepines and dibenzocycloheptanes as farnesyltransferase inhibitors. Bioorg Med Chem Lett. 2007;17: 5465-71.

González MP, Caballero J, Tundidor-Camba A, Helguera AM, Fernández M. Modeling of farnesyltransferase inhibition by some thiol and non-thiol peptidomimetic inhibitors using genetic neural networks and RDF approaches. Biorg Med Chem. 2006; 14:200-13.

Gupta MK, Prabhakar YS. QSAR study on tetrahydroquinoline analogues as plasmodium protein farnesyltransferase inhibitors: A comparison of rationales of malarial and mammalian enzyme inhibitory activities for selectivity. Eur J Med Chem. 2008;43: 2751-67.

Habibi-Yangjeh A. QSAR study of the 5-HT1A receptor affinities of arylpiperazines using a genetic algorithm–artificial neural network model. Monatsh Chem. 2009;140:523-30.

Jain SV, Ghate M, Bhadoriya KS, Bari SB, Chaudhari A, Borse JS. 2D, 3D-QSAR and docking studies of 1, 2, 3-thiadiazole thioacetanilides analogues as potent HIV-1 non-nucleoside reverse transcriptase inhibitors. Org Med Chem Lett. 2012;2(1):22.

Jalali-Heravi M, Asadollahi-Baboli M, Shahbazikhah P. QSAR study of heparanase inhibitors activity using artificial neural networks and Levenberg–Marquardt algorithm. Eur J Med Chem. 2008;43:548-56.

Katritzky AR, Kuanar M, Slavov S, Hall CD, Karelson M, Kahn I, et al. Quantitative correlation of physical and chemical properties with chemical structure: Utility for prediction. Chem Rev. 2010;110: 5714-89.

Leardi R, Seasholtz MB, Pell RJ. Variable selection for multivariate calibration using a genetic algorithm: prediction of additive concentrations in polymer films from Fourier transform-infrared spectral data. Anal Chim Acta. 2002;461:189-200.

Louis B, Agrawal VK, Khadikar PV. Prediction of intrinsic solubility of generic drugs using MLR, ANN and SVM analyses. Eur J Med Chem. 2010;45:4018-25.

Lu A, Zhang J, Yin X, Luo X, Jiang H. Farnesyltransferase pharmacophore model derived from diverse classes of inhibitors. Bioorg Med Chem Lett. 2007; 17:243-9.

Ohkanda J, Lockman JW, Yokoyama K, Gelb MH, Croft SL, Kendrick H, et al. Peptidomimetic inhibitors of protein farnesyltransferase show potent antimalarial activity. Bioorg Med Chem Lett. 2001;11: 761-4.

Olepu S, Suryadevara PK, Rivas K, Yokoyama K, Verlinde CL, Chakrabarti D, et al. 2-Oxo-tetrahydro-1,8-naphthyridines as selective inhibitors of malarial protein farnesyltransferase and as anti-malarials. Bioorg Med Chem Lett. 2008;18:494-7.

Puntambekar DS, Giridhar R, Yadav MR. Insights into the structural requirements of farnesyltransferase inhibitors as potential anti-tumor agents based on 3D-QSAR CoMFA and CoMSIA models. Eur J Med Chem. 2008;43:142-54.

Shahlaei M, Fassihi A, Saghaie L. Application of PC-ANN and PC-LS-SVM in QSAR of CCR1 antagonist compounds: a comparative study. Eur J Med Chem. 2010;45:1572-82.

Shayanfar A, Ghasemi S, Soltani S, Asadpour-Zeynali K, Doerksen RJ, Jouyban A. Quantitative structure-activity relationships of imidazole-containing farnesyltransferase inhibitors using different chemometric methods. Med Chem. 2013;9:434-48.

Soltani S, Abolhasani H, Zarghi A, Jouyban A. QSAR analysis of diaryl COX-2 inhibitors: comparison of feature selection and train-test data selection methods. Eur J Med Chem. 2010;45:2753-60.

Tanaka R, Rubio A, Harn NK, Gernert D, Grese TA, Eishima J, et al. Design and synthesis of piperidine farnesyltransferase inhibitors with reduced glucuronidation potential. Biorg Med Chem. 2007;15:1363-82.

Todeschini R, Consonni V. Handbook of molecular descriptors. New York: Wiley, 2008.

van de Waterbeemd H, Testa B (eds). Drug bioavailability: estimation of solubility, permeability, absorption and bioavailability. 2nd ed. New York: Wiley, 2008. (Methods and principles in medicinal chemistry, Vol. 40)

Vapnik V. The nature of statistical learning theory. Springer. 2000.

Xie A, Sivaprakasam P, Doerksen RJ. 3D-QSAR analysis of antimalarial farnesyltransferase inhibitors based on a 2,5-diaminobenzophenone scaffold. Biorg Med Chem. 2006;14:7311-23.

Yee LC, Wei YC. Current modeling methods used in QSAR/QSPR. In: Dehmer M, Varmuza K, Bonchev D (eds.): Statistical modelling of molecular descriptors in QSAR/QSPR (pp 1-32). New York: Wiley, 2012.