Iranian Journal of Mathematical Chemistry, Vol. 7, No.1, March 2016, pp. 47– 59 **IJMC**

Quantitative Structure Activity Relationship Study of Inhibitory Activities of 5–Lipoxygenase and Design new Compounds by Different Chemometrics Methods

FATEMEH BAGHBAN SHAHRI, ALI NIAZI AND AHMAD AKRAMI

Department of Chemistry, Arak Branch, Islamic Azad University, Arak, Iran

Correspondence should be addressed to f-baghban@iau-arak.ac.ir (F. Bagheban Shahri)

Received 8 March 2015; Accepted 13 March 2015

ACADEMIC EDITOR: IVAN GUTMAN

ABSTRACT A quantitative structure-activity relationship (QSAR) study was conducted for the prediction of inhibitory activity of 1-phenyl[2H]-tetrahydro-triazine-3-one analogues as inhibitors of 5-Lipoxygenase. The inhibitory activities of the 1-phenyl[2H]-tetrahydro-triazine-3-one analogues modeled as a function of molecular structures using chemometrics methods such as multiple linear regression (MLR) and least squares support vector machines (LS-SVM). The obtained models were applied to predict the inhibitory activity of compounds which were not in the modeling procedure. The results of models showed high prediction ability with root mean square error of prediction of 0.167 and 0.061 for MLR and LS-SVM, respectively. The LS-SVM method was used for prediction of inhibitory activity of the new inhibitor derivatives.

KEYWORDS QSAR • 1–phenyl[2H]–tetrahydro–triazine–3–oneanalogues • MLR • LS–SVM.

1. INTRODUCTION

Lipoxygenases (LOs) are a class of widely occurring, non-heme iron-containing oxygenases that can be isolated from animals, higher plants, and fungi. Currently, three distinct mammalian LOs have been characterized, 5-LO, 12-LO, and 15-LO, which oxygenate arachidonic acid at specific carbon centers (C5, C12, and C15, respectively) [1].The 5-Lipoxygenase is the first dedicated enzyme in the biosynthetic pathway leading to the leukotrienes. Since leukotrienes have been implicated as important mediators in such diseases as asthma, psoriasis, ulcerative colitis, and rheumatoid arthritis, inhibition of 5-Lipoxygenase offers a potential approach for the therapy of these diseases [2].

In the present study, the inhibitory activity data of 1-phenyl[2H]-tetrahydro-triazine-3-one analogues as inhibitors of 5-Lipoxygenase were used to construct a mathematical model with structural information, a so-called QSAR (quantitative structure–activity relationship). Quantitative structure-activity relationships (QSAR) are an important tool in agrochemistry, pharmaceutical chemistry, toxicology, and eventually

most facts of chemistry [3, 4]. QSAR models are mathematical equations which relate chemical structure of a compound to its physical, chemical, biological and technological properties. The main goal of QSAR studies is to establish an empirical rule or function to relate the structural descriptors of compounds under investigation to bioactivities. This rule or function is then utilized to predict the same bioactivities of compounds which are not involved in the training set from their structural descriptors. Model development in QSAR studies comprises different critical steps as (1) descriptor generation, (2) data splitting to calibration (or training) and prediction (or validation) sets, (3) variable selection, (4) finding appropriate model between selected variables and activity and (5) model validation [5].

Among the investigation of QSAR, one of the most important factors affecting the quality of the model is the method to build the model. Many multivariate data analysis methods such as multiple linear regression (MLR) [6, 7], artificial neural network (ANN) [8] and partial least squares (PLS) [9] have been used in QSAR studies. MLR, as most commonly used chemometrics method, has been extensively applied to QSAR investigations. The artificial neural network (ANN) offers satisfactory accuracy in most cases but tends to over fit the training data. The PLS method is based on factor analysis that is originally suggested and chemically applied by Wold et al [10]. The support vector machine (SVM) is a popular algorithm developed from the machine learning community. Due to its advantages and remarkable generalization performance over other methods, SVM has attracted attention and gained extensive applications [11, 12]. As a simplification of traditional SVM, Suykens and Vandewalle [13, 14] have proposed the use of least-squares SVM (LS-SVM). LS-SVM encompasses similar advantages as SVM, but its additional advantage is that it requires solving a set of linear equations (linear programming), which is much easier computationally [15, 16]. In this study, the MLR and LS-SVM methods were proposed to model and predict the inhibitory activity of 1-phenyl[2H]-tetrahydro-triazine-3-one analogues as inhibitors of 5-Lipoxygenase.

2. THEORY

The LS-SVM [13] is capable of dealing with linear and nonlinear multivariate calibration. In LS-SVM a linear estimation is made in kernel-induced feature space ($y = w^T\phi(x) + b$). As in SVM, it is necessary to minimize a cost function (C) containing a penalized regression error, as follow:

$$C = \frac{1}{2} w^T w + \frac{1}{2} \gamma \sum_{i=1}^{N} e_i^2 \tag{1}$$

such that:

$$y_i = w^T \phi(x_i) + b + e_i \tag{2}$$

for all $i = 1,...,N$, where ϕ denotes the feature map.

The first part of this cost function is a weight decay which is used to regularize weight sizes and penalize large weights. Due to this regularization, the weights converge to similar value. Large weights deteriorate the generalization ability of the LS-SVM

because they can cause excessive variance. The second part of Eq. (1) is the regression error for all training data. The parameter γ, which has to be optimized by the user, gives the relative weight of this part as compared to the first part. The restriction supplied by Eq. (2) gives the definition of the regression error. Eq. (1) and its restriction given by Eq. (2), could be concluded to be a typical problem of convex optimization [14] which might be solved by the Lagrange multipliers method, as follow:

$$L = \frac{1}{2}\|w\|^2 + \gamma \sum_{i=1}^{N} e_i^2 - \sum_{i=1}^{N} \alpha \{w^T \phi(x_i) + b + e_i - y_i\} \tag{3}$$

where

$$y_i = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}, \quad e_i = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_N \end{bmatrix} \text{ and } \quad \alpha_i = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_N \end{bmatrix}.$$

Obtaining the optimum, that is, carrying out $\partial L(w,b,e_i,\alpha_i)/\partial w$, $\partial L(w,b,e_i,\alpha_i)/\partial b$, $\partial L(w,b,e_i,\alpha_i)/\partial e_i$, $\partial L(w,b,e_i,\alpha_i)/\partial \alpha_i$ and setting all partial first derivatives to zero, generates the weights that are the linear combinations of the training data:

$$\frac{\partial L(w,b,e,\alpha)}{\partial w} = w - \sum_{i=1}^{N} \alpha_i \phi(x_i) = 0 \therefore w = \sum_{i=1}^{N} \alpha_i \phi(x_i) \tag{4}$$

$$\frac{\partial L(w,b,e,\alpha)}{\partial e} = \sum_{i=1}^{N} \gamma e - \alpha = 0 \tag{5}$$

and then

$$w = \sum_{i=1}^{N} \alpha_i \phi(x_i) = \sum_{i=1}^{N} \gamma e_i \phi(x_i) \tag{6}$$

where a positive definite kernel is used as follows:

$$K(x_i, x_j) = \phi(x_i)^T \phi(x_j) \tag{7}$$

An important result of this approach is that the weights (w) can be written as linear combinations of the Lagrange multipliers with corresponding data training (x_i). Substituting the result of Eq. (6) into the original regression line $(y = w^T \phi(x) + b)$, the following result is obtained:

$$y = \sum_{i=1}^{N} \alpha_i \phi(x_i)^T \phi(x) + b = \sum_{i=1}^{N} \alpha_i \langle (\phi(x_i)^T, \phi(x) \rangle + b \tag{8}$$

for a point y_i to be evaluated it is:

$$y_i = \sum_{i=1}^{N} \alpha_i \phi(x_i)^T \phi(x_j) + b = \sum_{i=1}^{N} \alpha_i \langle (\phi(x_i), \phi(x_j) \rangle + b \tag{9}$$

The α vector follows from solving a set of linear equation:

$$M \begin{bmatrix} \alpha \\ b \end{bmatrix} = \begin{bmatrix} y \\ 0 \end{bmatrix} \tag{10}$$

where M is a square matrix given by:

$$M = \begin{bmatrix} K + \dfrac{I}{\gamma} & 1_N \\ 1_N^T & 0 \end{bmatrix} \tag{11}$$

Where K denotes the kernel matrix with $ijth$ element $K = (x_i, x_j) = \phi(x_i)^T \phi(x_j)$ and I denotes the identity matrix $N \times N$, $1_N = [1 \quad 1 \quad \cdots \quad 1]^T$. Hence, the solution is given by:

$$\begin{bmatrix} \alpha \\ b \end{bmatrix} = M^{-1} \begin{bmatrix} y \\ 0 \end{bmatrix} \tag{12}$$

As demonstrated in Eqs. (11) and (12), all Lagrange multipliers (the support vectors) are usually nonzero, which means that all training objects contribute to the solution. In contrast, with standard SVM the LS-SVM solution is usually not sparse. However, as described by Suykens and J. Vandewalle [13] a sparse solution can be easily achieved via pruning or reduction techniques. Depending on the number of training data set either direct solvers or iterative solvers such as conjugate gradients methods (for large data sets) can be used in both cases with numerically reliable methods.

In applications involving nonlinear regression, it is enough to change the inner product $\langle \phi(x_i), \phi(x_j) \rangle$ of Eq. (9) by a kernel function and the $ijth$ element of matrix K equals $K_{ij} = \phi(x_i)^T \phi(x_j)$. If this kernel function meets Mercer's condition, the kernel implicitly determines both a nonlinear mapping, $x \to \phi(x)$ and the corresponding inner product $\phi(x_i)^T \phi(x_j)$. This leads to the following nonlinear regression function:

$$y = \sum_{i=1}^{N} \alpha_i K(x_i, x) + b \tag{13}$$

for a point x_j to be evaluated it is:

$$y_j = \sum_{i}^{N} \alpha_i K(x_i, x_j) + b \tag{14}$$

The attainment of the kernel function is cumbersome and it will depend on each case. However, the kernel function is more used as the radial basis function (RBF), $\exp(-(\|x_i - x_j\|^2)/2\sigma^2)$, a simple Gaussian function, and polynomial functions $\langle x_i, x_j \rangle^d$, where σ^2 is the width of the Gaussian function and *d* is the polynomial degree, which should be optimized by the user, to obtain the support vector. For α of the RBF kernel and *d* of the polynomial kernel it is of significant importance to do a careful model selection of the tuning parameters, in combination with the regularization constant γ, in order to achieve a good generalization model.

3. MATERIALS AND COMPUTATIONAL METHODS

3.1. HARDWARE AND SOFTWARE

The computations were made with the ASUS Personal Computer (1 GB RAM) that was equipped with the Windows 7 operating system and MATLAB (Version 9.0, Mathwork Inc.). The LS-SVM optimization and model results were obtained using the LS-SVM lab toolbox (Matlab/C Toolbox for Least-Squares Support Vector Machines). The MLR analysis with a stepwise forward selection method was carried out by using the SPSS 21 software. Kennard-Stones program was written in MATLAB according to the algorithm [17, 18]. ChemOffice package (Version 2010) was used to draw the molecular structure and optimization by the AM1. Descriptors were calculated using Dragon software (Milano Chemometrics and QSAR research group, http://www.disat.unimib.it/chm/). These descriptors are calculated using two-dimensional representation of the molecules and therefore geometry optimization is not essential for calculating these types of descriptors.

3.2. DATA SET

The inhibitory activity values of 1-phenyl[2*H*]-tetrahydro-triazine-3-one analogues were taken from literature [2]. The chemical structures of 1-phenyl[2*H*]-tetrahydro-triazine-3-one analogues (Figure 1) and their corresponding inhibitory activity values have been listed in Table 1. In order to assure that training and prediction sets cover the total space occupied by the original data set, it was divided into parts of training and prediction set according to the Kennard-Stones algorithm [17, 18]. The Kennard-Stones algorithm is known as one of the best ways of building training and prediction sets and it has been used in many QSAR studies.

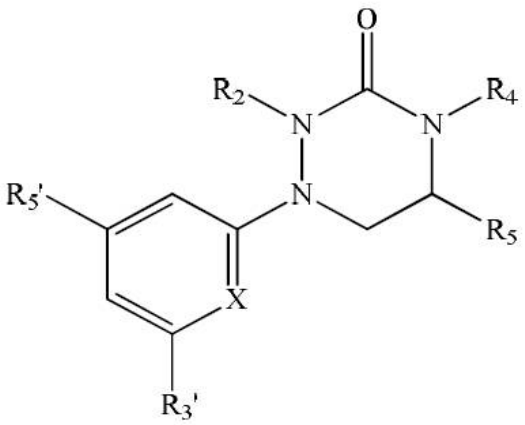

Figure 1. Chemical structure of 1-phenyl[2H]-tetrahydro-triazine-3-one analogues.

Table 1. Structures and observed inhibitory activity of 5-Lipoxygenase of 1-phenyl[2H]-tetrahydro-triazine-3-one analogues.

No.	Substitution						log(1/IC_{50})
	X	R_3'	R_5'	R_2	R_4	R_5	obs[a]
1	CH	H	H	H	H	CH_2OCH_2Ph	6.00
2	CH	H	H	H	H	Bu	5.82
3	CH	H	H	H	H	i-Pr	5.17
4	CH	H	H	H	H	Me(R)	5.17
5[b]	CH	H	H	H	H	Me_2	5.17

6	CH	H	H	H	H	Et	5.16
7 [b]	CH	H	H	H	H	Me	4.94
8	CH	H	H	H	H	$CH_2OC_2H_4OMe$	4.85
9	CH	H	H	H	H	Me(S)	4.85
10	CH	H	H	H	H	CO_2Me	4.70
11 [b]	CH	H	H	H	H	H	4.68
12	CH	H	OCH_2Ph	H	H	H	5.96
13	CH	H	Br	H	H	H	5.31
14	CH	H	Cl	H	H	H	5.20
15	CH	H	Et	H	H	H	4.89
16	CH	H	SMe	H	H	H	4.85
17 [b]	CH	H	Me	H	H	H	4.82
18	CH	H	CF_3	H	H	H	4.77
19	CH	H	F	H	H	H	4.72
20	CH	H	CN	H	H	H	4.43
21	CH	H	OMe	H	H	H	4.33
22	CH	H	NO_2	H	H	H	4.31
23	CH	H	NH_2	H	H	H	3.75
24	CH	H	Br	H	H	Me	5.59
25	CH	H	Cl	H	H	Me	5.57
26	CH	H	F	H	H	Me	5.20
27 [b]	CH	H	Me	H	H	Me	4.72
28	CH	H	H	H	C(=O)-i-Pr	H	5.89
29	CH	H	H	H	C(=O)Et	H	5.59
30	CH	H	H	H	C(=O)Me	Me	5.48
31	CH	H	H	H	C(=O)Me	H	5.47
32	CH	H	H	H	OCH_2Ph	Me	5.37
33	CH	H	H	H	OH	Me	5.22
34	CH	H	H	H	OEt	Me	5.13
35	CH	H	H	H	OCH_2Ph	H	5.08
36	CH	H	H	C(=O)Et	C(=O)Et	H	4.90
37	CH	H	H	H	OMe	Me	4.65
38	CH	H	H	C(=O)Me	C(=O)Me	H	4.40
39	N	Br	H	H	H	Me	5.62
40	N	Br	H	H	H	H	5.46
41	N	Cl	H	H	H	Me	5.46
42 [b]	N	Me	H	H	H	Me	5.42
43	N	Me	H	H	H	H	5.26
44	N	OMe	H	H	H	Me	5.26
45	N	Cl	H	H	H	H	5.25
46	N	F	H	H	H	Me	5.18
47	N	F	H	H	H	H	5.04
48	N	OMe	H	H	H	H	5.02
49	N	H	H	H	H	Me	4.66
50	N	H	H	H	H	H	4.59
51	CH	H	Cl	H	C(=O)Me	H	5.89
52	CH	H	Cl	H	OH	Me	5.41
53	CH	H	F	H	OH	Me	5.16
54	CH	Me	Me	H	OH	H	5.08
55	CH	F	F	H	H	H	5.05
56	CH	Me	Me	H	H	H	4.92
57 [b]	N	Cl	H	H	H	H	5.48
58 [b]	CH	H	Cl	H	H	H	5.35
59 [b]	CH	H	H	H	H	H	4.77
60	CH	Cl	Me	H	H	H	5.48

[a] Observed inhibitory activity .

[b] The compounds selected as the test set.

3.3. MOLECULAR DESCRIPTORS

A major step in constructing QSAR model is generation of the corresponding numerical descriptors of the molecular structures. Molecular descriptors define the molecular structure and physicochemical properties of molecules by a single number. To calculate the different kinds of theoretical descriptors for each molecule, the Dragon (Milano Chemometrics and QSAR research group, http://www.disat.unimib.it/chm/) software was utilized. The Dragon is able to calculate different molecular descriptors such as constitutional, topological, molecular walk counts, BCUT, Galvez topol. Charge indices, 2D autocorrelations, charge, aromaticity indices, Randic molecular profiles, geometrical, RDF, 3D-MoRSE, WHIM, GETAWAY, functional groups, atom-centered fragments, properties and empirical. In this study, just GETAWAY (geometry, topology, and atom-weights assembly) and WHIM (weighted holistic invariant molecular) descriptors were used. Here, 293 descriptors were calculated by Dragon for each molecule, therefore we have 60×293 data matrix X. The rows and columns of this matrix are the number of molecules and molecular descriptors respectively.

4. RESULTS AND DISCUSSION

4.1. PRINCIPAL COMPONENT ANALYSIS OF THE DATA SET

Principal components analysis (PCA) was performed on the calculated structural descriptors to the whole data set (Table1), for investigation the distribution in the chemical space, which shows the spatial location of samples to assist the separation of the data into training and prediction sets. The PCA results show that three PCs (PC1, PC2 and PC3) describe 81.35% of the overall variances (Figure 2). Since almost all variables can be accounted for the first three PCs, their score plot is a reliable presentation of the spatial distribution of the points for the data set. As can be seen in Figure 2, there is not a clear clustering between compounds. The data separation is very important in the development of reliable and robust QSAR models. The quality of the prediction depends on the data set used to develop the model. For regression analysis, data set was separated into two groups, a training set (51 data) and a prediction set (9 data) according to Kennard-Stones algorithm. As shown in Figure 2, the distribution of the compounds in each subset seems to be relatively well-balanced over the space of the principal components.

4.2. MLR ANALYSIS

The multivariate calibration is a powerful tool for modeling, because it extracts more information from the data and allows building more robust models. Among the descriptors calculated, the most significant molecular descriptors were identified using multiple linear regression analysis with a stepwise forward selection method. According to inhibitory activity data (Table 1), the data classified to training and prediction sets according to Kennard-Stones algorithm and the MLR model was run.

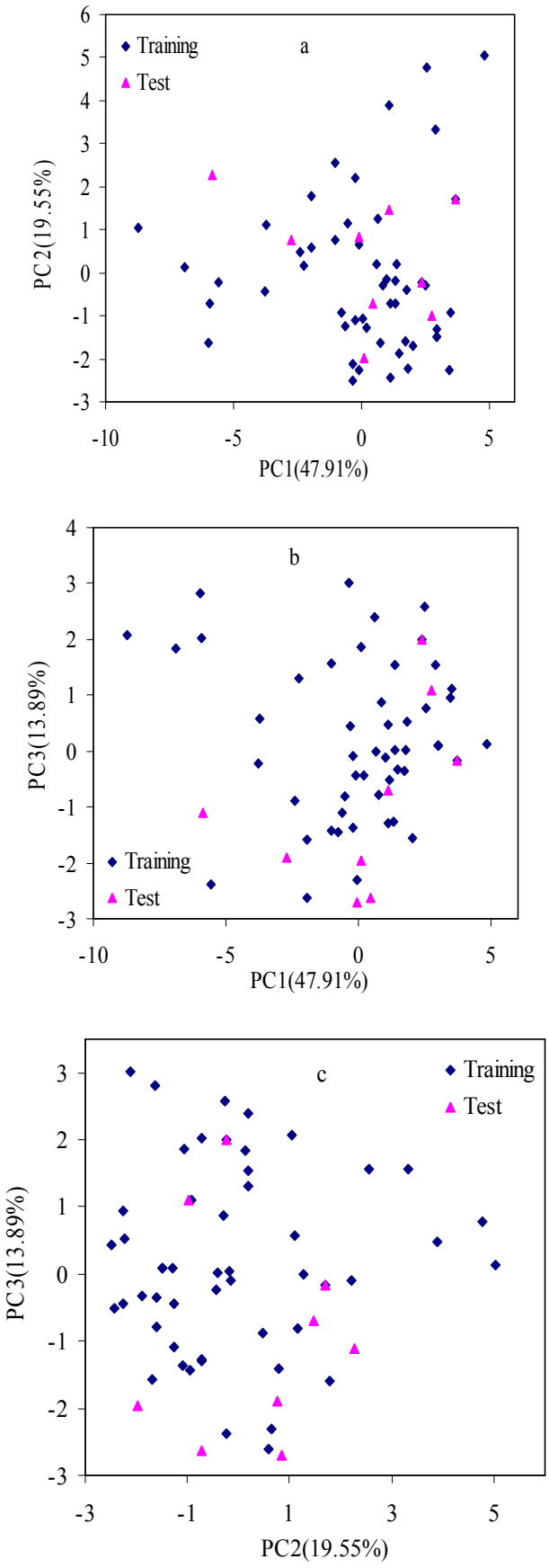

Figure 2. Principal components analysis of the descriptors for the data set, (a) PC2 versus PC1, (b) PC3 versus PC1 and (c) PC3 versus PC2.

The best equation obtained for the inhibitory activity of 1-phenyl[2H]-tetrahydro-triazine-3-one analogues derivatives was:

$\log(1/IC_{50}) = 3.095 - 1.131R7e - 38.269R6U^{+} - 4.942R1u + 4.897R2u - 5.247P2u + 3.787H2v + 18.425G3e - 0.008Vu$

As seen, the resulting model has eight significant descriptors. Table 2 shows the calculated descriptors for each molecule, the descriptors coefficients, the standard error of coefficients, the t values for null hypothesis, and their related P values.

Table 2. Results of multiple linear regression analysis.

Notation	Descriptors	Coefficient	S.E.[a] of coefficient	t value	P value
Intercept	-	3.095	2.358	1.312	0.197
R7e	R autocorrelation of lag 7 / weighted by Sanderson electronegativity	-1.131	0.282	-4.018	0.000
$R6U^{+}$	R maximal autocorrelation of lag 6 / unweighted	-38.269	8.865	-4.317	0.000
R1u	R autocorrelation of lag 1 / unweighted	-4.942	0.747	-6.613	0.000
R2u	R autocorrelation of lag 2 / unweighted	4.897	0.794	6.169	0.000
P2u	2nd component shape directional WHIM index / unweighted	-5.247	1.147	-4.575	0.000
H2v	H autocorrelation of lag 2 / weighted by van der Waals volume	3.787	0.455	8.330	0.000
G3e	3rd component symmetry directional WHIM index / weighted by Sanderson electronegativity	18.425	5.821	3.165	0.003
Vu	V total size index / unweighted	-0.008	0.003	-2.209	0.033

[a] Standard error.

4.3. LS–SVM Analysis

LS-SVM was performed with radial basis function (RBF) as a kernel functions. In the model development using LS-SVM and RBF kernel, γ and σ^2 parameters were a manageable task. To determine the optimal parameters, a grid search was performed

based on leave-one-out cross-validation on the original training set for all parameter combinations of γ and σ^2 from 0.1 to 10 and 1 to 100, respectively. In Table 3 is shown the optimum γ and σ^2 parameters for the LS-SVM and RBF kernel, using the training sets for 51 inhibitory activity data.

Table 3. Observation and calculation values of $\log(1/IC_{50})$ using MLR and LS-SVM models.

No. of compounds (Table 1)	Observation $\log(1/IC_{50})$	MLR		LS-SVM	
		Predicted	Error (%)	Predicted	Error (%)
5	5.170	4.980	-3.675	5.081	-1.721
7	4.940	4.792	-2.996	4.856	-1.700
11	4.680	4.836	3.333	4.716	0.769
17	4.820	4.851	0.643	4.832	0.249
27	4.720	4.916	4.152	4.771	1.080
42	5.420	5.189	-4.262	5.356	-1.181
57	5.480	5.327	-2.792	5.389	-1.660
58	5.350	5.227	-2.299	5.309	-0.766
59	4.770	4.966	4.109	4.746	-0.503
γ				0.500	
σ^2				10.000	
RMSEP		0.167		0.061	
RSEP (%)		3.315		1.212	

4.4. MODEL VALIDATION AND PREDICTION OF INHIBITORY ACTIVITY

The predictive ability of these methods (MLR and LS-SVM) were investigated by prediction of inhibitory of 9 molecules (their structures are given in Table 1). Validation of predictive ability is another key step in QSAR studies. Several statistical parameters have been used for the evaluation of the suitability of the developed QSAR models for prediction of the property of the studied compounds this include the root mean square error of prediction (RMSEP) and relative standard error of prediction (RSEP), validation through an external prediction set.

$$RMSEP = \sqrt{\frac{\sum_{i=1}^{n}(y_{i,pred} - y_{i,obs})^2}{n}} \tag{15}$$

$$RSEP(\%) = 100 \times \sqrt{\frac{\sum_{i=1}^{n}(y_{i,pred} - y_{i,obs})^2}{\sum(y_{i,obs})^2}} \tag{16}$$

where $y_{i,pred}$ is the predicted of the inhibitory activity using different model, $y_{i,obs}$ is the observed value of the inhibitory activity, and n is the number of compounds in the prediction set. The statistical parameters obtained by these methods are listed in Table 3.

Table 3 shows RMSEP, RSEP and the percentage error for prediction of inhibitory activity of 1-phenyl[2H]-tetrahydro-triazine-3-one analogues. As can be seen, the good results were achieved in LS-SVM model with percentage error ranges from -1.721 to 1.080 for inhibitory activity of 1-phenyl[2H]-tetrahydro-triazine-3-one analogues. The plots of the predicted inhibitory activity versus actual values are shown in Figure 3 for each model (line equations and R^2 values are also shown). The correlation coefficients (R^2) for LS-SVM model were better than the MLR model. Also, it is possible to see that LS-SVM presents excellent prediction abilities when compared with MLR model.

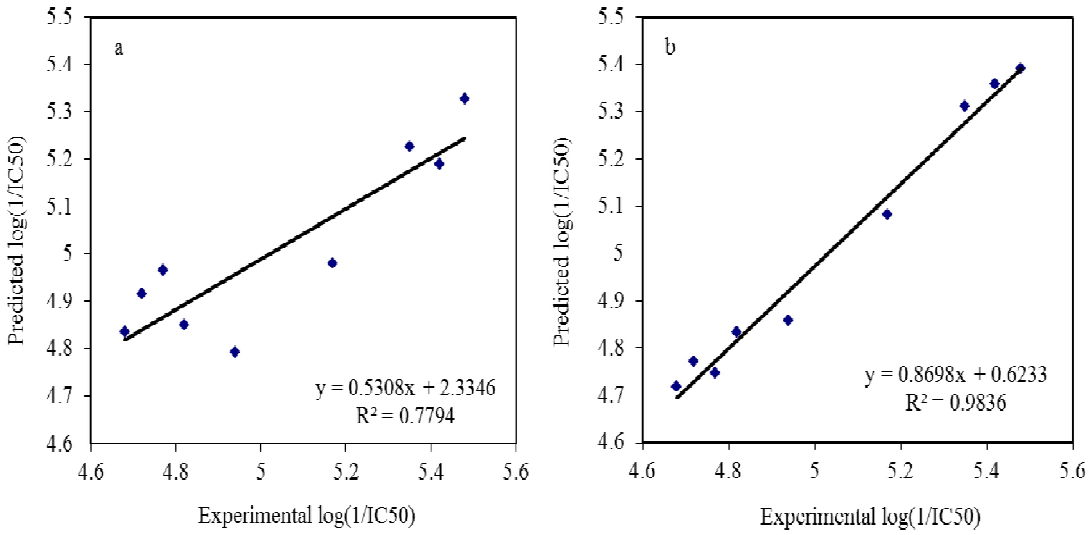

Figure 3. Plots of predicted versus actual log(1/IC_{50}), (a) MLR and (b) LS-SVM.

4.5. MOLECULAR DESIGN

As an application of proposed method, we investigated LS-SVM model to predict the inhibitory activity of four new 1-phenyl[2*H*]-tetrahydro-triazine-3-one analogues whose biological tests were not performed with them yet. Table 4 shows the chemical structure of four new compounds and their inhibitory activity calculated by this proposed method.

Table 4. New structures of 1-phenyl[2H]-tetrahydro-triazine-3-one analogues and predicted log(1/IC_{50}) by LS-SVM.

Number of Design	Substitution						log(1/IC_{50}) Calc.[a]
	X	R_3'	R_5'	R_2	R_4	R_5	
1	N	H	Me	H	H	Me	4.617
2	N	F	H	H	H	Br	6.138
3	CH	Br	H	H	H	Cl	5.856
5	CH	H	OH	H	Cl	H	5.149

[a] Calculated by LS-SVM model.

5. CONCLUSION

Using LS-SVM, a QSAR model has been successfully developed for the prediction of inhibitory activity for 60 compounds. The results well illustrate the power of descriptors in prediction of inhibitory activity of 1-phenyl[2*H*]-tetrahydro-triazine-3-one analogues. The model could predict the inhibitory activity of 1-phenyl[2*H*]-tetrahydro-triazine-3-one analogues derivatives not existed in the modeling procedure accurately. The work, shows that descriptors are capable to recognize the physicochemical information and be can useful to predict the inhibitory activity of the new compounds.

ACKNOWLEDGMENT

The authors gratefully acknowledge the support to this work from, Islamic Azad University, Arak Branch, and research council.

REFERENCES

1. A. B. Camargo, E. Marchevsky and J. M. Luco, QSAR study for the soybean 15-lipoxygenase inhibitory activity of organosulfur compounds derived from the essential oil of garlic, *J. Agric. Food. Chem*. **55** (2007) 3096–3103.
2. Y. P. Zhou, J. H. Jiang, W. Q. Lin, H. Y. Zou, H. L. Wu, G. L. Shen and R. Q. Yu, Boosting support vector regression in QSAR studies of bioactivities of chemical compounds, *Eur. J. Pharm. Sci.* **28** (2006) 344–353.
3. G. Ioele, M. D. Luca, F. Oliverio and G. Ragno, Prediction of photosensitivity of 1,4-dihydropyridine antihypertensives by quantitative structure-property relationship, *Talanta* **79** (2009) 1418–1424.
4. W. Cui and X. Yan, Adaptive weighted least square support vector machine regression integrated with outlier detection and its application in QSAR, *Chemomtr. Intell. Lab. Syst.* **98** (2009) 130–135.
5. J. B. Ghasemi, S. H. Ahmadi and S. D. Brown, A quantitative structure–retention relationship study for prediction of chromatographic relative retention time of chlorinated monoterpenes, *Environ. Chem. Lett.* **9** (2011) 87–96.
6. M. Kompany-Zareh, A QSPR study of boiling point of saturated alcohols using genetic algorithm, Acta Chim. Slov. **50** (2003) 259-273.
7. B. Narasimhan, V. Judge, R. Narang, R. Ohlan and S. Ohlan, Quantitative structure activity relationship studies for prediction of antimicrobial activity of synthesized 2,4-hexadienoic acid derivatives, *Bioorg. Med. Chem. Lett.* **17** (2007) 5836–5845.
8. B. Hemmateenejad, M. A. Safarpour and F. Taghavi, Application of ab initio theory for the prediction of acidity constants of some 1-hydroxy-9,10-anthraquinone derivatives using genetic neural network, *J. Mol. Struc.* (*TheoChem*) **635** (2003) 183–190.

9. A. Niazi, S. Jameh-Bozorghi and D. Nori-Shargh, Prediction of acidity constants of thiazolidine-4-carbozylic acid derivatives using Ab initio and genetic algorithm-partial least squares, *Turk. J. Chem.* **30** (2006) 619–628.
10. K. G. Joreskog and H. Wold, "*System under indirect Observations*" North Holland, Amsterdam, 1982.
11. A. I. Belousov, S. A. Verzakov and J. Von-Frese, A flexible classification approach with optimal generalization performance: support vector machines, *J. Chemometr. Intell. Syst*. **64** (2002) 15–25.
12. R. Burbidge, M. Trotter, B. Buxton and S. Holden, Drug design by machine learning: support vector machines for pharmaceutical data analysis, *Comput. Chem.* **26** (2001) 5–14.
13. J. A. K. Suykens and J. Vandewalle, Least squares support vector machine classifiers, *Neural Process. Lett.* **9** (1999) 293–300.
14. J. A. K. Suykens, T. Van-Gestel, J. De-Brabanter, B. De-Moor and J. Vandewalle, "Least-Squares Support Vector Machines" World Scientifics, Singapore, 2002.
15. A. Niazi, J. B. Ghasemi and M. Zendehdel, Simultaneous voltammetric determination of morphine and noscapine by adsorptive differential pulse stripping method and least-squares support vector machines, *Talanta* **74** (2007) 247–254.
16. A. Niazi, J. B. Ghasemi and A. Yazdanipour, Simultaneous spectrophotometric determination of nitroaniline isomers after cloud point extraction by using least-squares support vector machines, *Spectrochim. Acta Part A*. **68** (2007) 523–530.
17. R. W. Kennard and L. A. Stones, Computer ided Design of Experiments, *Technometrics* **11** (1969) 137–148.
18. M. Daszykowski, B. Walczak and D. L. Massart, Representative subset selection, *Anal. Chim. Acta* **468** (2002) 91–103.