

# Application of Wavelet and Genetic Algorithms for QSAR Study on 5-Lipoxygenase Inhibitors and Design New Compounds

Fatemeh Bagheban Shahri\*, Ali Niazi and Ahmad Akrami

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

Corresponding author. Email: f-bagheban@iau-arak.ac.ir, f.baghebanshahri@gmail.com

Fax: +98 86 34057336

Received January 1<sup>st</sup>, 2015; Accepted September 4<sup>th</sup>, 2015

**Abstract.** A quantitative structure-activity relationship (QSAR) modeling was carried out for the prediction of inhibitory activity of 1-phenyl[2H]-tetrahydro-triazine-3-one analogues as inhibitors of 5-lipoxygenase. Partial least squares (PLS) algorithm was employed to model the relationships between molecular descriptors and inhibitory activity of molecules using the genetic algorithm (GA) method as variable selection tool. Pre-processing methods such as wavelet transform (WT) were also used to enhance the predictive power of multivariate calibration methods. To evaluate the models applied in this study (PLS, GA-PLS and WT-GA-PLS), the inhibitory activities of several compounds, not included in the modeling procedure, were predicted. The results of models showed high prediction ability with root mean square error of prediction 0.194, 0.161 and 0.140 for PLS, GA-PLS and WT-GA-PLS, respectively. The WT-GA-PLS method was employed to predict the inhibitory activity of the new inhibitor derivatives.

**Key words:** 1-phenyl[2H]-tetrahydro-triazine-3-one analogues, genetic algorithms, wavelet transform, QSAR, PLS.

**Resumen.** Se realizó la modelación de una relación estructura-actividad (QSAR, por sus siglas en inglés) para la predicción de la actividad inhibitoria de análogos del 1-fenil[2H]-tetrahydro-triazina-3-ona como inhibidores de la 5-lipoxigenasa. Se utilizó el algoritmo de mínimos cuadrados parciales (PLS, por sus siglas en inglés) para simular las relaciones entre los descriptores moleculares y la actividad inhibitoria de moléculas, usando el método de algoritmos genéticos (GA, por sus siglas en inglés) como herramienta de selección de variables. También se utilizaron métodos de preprocesamiento como transformada de ondeleta (WT, por sus siglas en inglés) para incrementar el poder predictivo de los métodos de calibración multivariable. Para evaluar los modelos aplicados en este estudio (PLS, GA-PLS y WT-GA-PLS), se evaluaron las actividades inhibitorias de varios compuestos no incluidos en el proceso de modelación. Los resultados de los modelos mostraron una alta capacidad predictiva con un error cuadrático medio en la predicción de 0.194, 0.161 y 0.140 para PLS, GA-PLS y WT-GA-PLS, respectivamente. Se utilizó el método WT-GA-PLS para predecir la actividad inhibitoria de nuevos derivados inhibidores.

**Palabras clave:** Análogos del 1-fenil[2H]-tetrahydro-triazina-3-ona, algoritmos genéticos, transformada de ondeleta, QSAR, PLS.

## 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, the inhibition of 5-lipoxygenase offers a potential approach in the therapy of such diseases [2]. In the present study, the inhibitory activity of 1-phenyl[2H]-tetrahydro-triazine-3-one analogues as inhibitors of 5-lipoxygenase were used to construct a mathematical model with structural information, so called a quantitative structure-activity relationship (QSAR).

QSAR is an important tool in Agrochemistry, Pharmaceutical Chemistry and Toxicology [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 the 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]. Since a large number of molecular descriptors are available for QSAR analysis, the most relevant descriptors should be selected. Many variable selection methods such as stepwise regression [6], simulated annealing [7] and genetic algorithms [8-10] are available. It has been shown that genetic algorithms (GAs) can be successfully used as a feature selection technique [11-13].

A GA is a stochastic method to solve optimization problems defined by a fitness criterion applying evolution hypothesis of Darwin and different genetic functions, i.e. crossover and

mutation. Leardi [14] demonstrated that GA, after suitable modifications, produces more interpretable results, since the selected variables are less dispersed compared to other methods. 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) [15, 16] artificial neural network (ANN) [17] and partial least squares (PLS) [18] have been used in QSAR studies. MLR, as most commonly used chemometrics method, has been extensively applied to QSAR investigations. However, due to the collinearity between descriptors, MLR is unable to extract useful information from the data, and the over-fitting problem will occur as a consequence. 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 the factor analysis which is originally suggested and chemically applied by Wold et al [19]. In order to enhance the predictive power of multivariate calibration methods, molecular descriptors are often corrected prior to the data analysis. One of the data preprocessing techniques is wavelet transform (WT) [20], which used to eliminate undesirable background effects and enhance the subsequent PLS regression model. In this study, the PLS, GA-PLS and WT-GA-PLS methods were proposed to model and predict the inhibitory activity of 1-phenyl[2H]-tetrahydro-triazine-3-one analogues as inhibitors of 5-Lipoxygenase.

## Material and Computational Methods

### Hardware and Software

The computations were made with the ASUS Personal Computer that was equipped with the Windows 7 operating system and MATLAB (Version 11.0, Math Work Inc.). All the required evaluations for GA variable selection and PLS modeling were carried out using the PLS program from PLS-Toolbox Version 4.0 and MATLAB from Eigenvector Research Inc. The calculations for data processing implemented in software using MATLAB Wavelet Toolbox and DWT technique. Kennard-Stones program was written in MATLAB according to the algorithm [21, 22]. ChemOffice package (Version 2013) 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.

### Data Set

The inhibitory activity values of 1-phenyl[2H]-tetrahydro-triazine-3-one analogues were taken from the literature [2]. The chemical structures of 1-phenyl[2H]-tetrahydro-triazine-3-one analogues (Fig. 1) and their corresponding inhibitory activity values have been listed in Table 1. In order to assure that train-

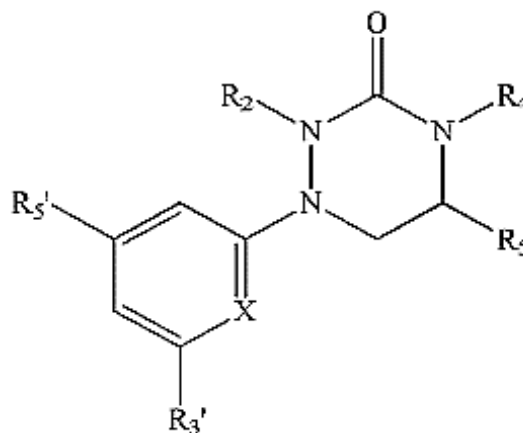
ing and prediction sets cover the total space occupied by the original data set, it was divided into parts of training and prediction set according to the Kennard-Stones algorithm [21, 22]. The Kennard-Stones algorithm is known as one of the best ways of building training and prediction sets, and it has been used in many QSAR studies. The Kennard-Stone algorithm selects a set of molecules in studied set of data, which are uniformly distributed over the space defined by the candidates. This is a classic technique to extract a representative set of molecules from a given data set. In this technique the molecules are selected consecutively. The first two objects are chosen by selecting the two farthest apart from each other. The third sample chosen is the one farthest from the first two objects, etc. Supposing that  $m$  objects have already been selected ( $m < n$ ), the  $(m + 1)^{\text{th}}$  sample in the calibration set is chosen using the following criterion:

$$\max(\min(d_{1r}, d_{2r}, \dots, d_{mr})) \{ \max(m < r \leq n) \} \quad (1)$$

Where,  $n$  stands for the number of samples in the training set,  $d_{jr}$ ,  $j=1, \dots, m$  are the squared euclidean distances from a candidate sample  $r$ , not yet included in the representative set, to the  $m$  samples already included in the representative set [23].

### Molecular Descriptors

A major step in constructing QSAR model is the 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 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



**Fig. 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 <sub>50</sub> ) obs <sup>a</sup>
	X	R <sub>3</sub> '	R <sub>5</sub> '	R <sub>2</sub>	R <sub>4</sub>	R <sub>5</sub>	
1	CH	H	H	H	H	CH <sub>2</sub> OCH <sub>2</sub> Ph	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 <sup>b</sup>	CH	H	H	H	H	Me <sub>2</sub>	5.17
6	CH	H	H	H	H	Et	5.16
7 <sup>b</sup>	CH	H	H	H	H	Me	4.94
8	CH	H	H	H	H	CH <sub>2</sub> OC <sub>2</sub> H <sub>4</sub> OMe	4.85
9 <sup>b</sup>	CH	H	H	H	H	Me(S)	4.85
10	CH	H	H	H	H	CO <sub>2</sub> Me	4.70
11 <sup>b</sup>	CH	H	H	H	H	H	4.68
12	CH	H	OCH <sub>2</sub> Ph	H	H	H	5.96
13	CH	H	Br	H	H	H	5.31
14	CH	H	Cl	H	H	H	5.20
15 <sup>b</sup>	CH	H	Et	H	H	H	4.89
16	CH	H	SMe	H	H	H	4.85
17 <sup>b</sup>	CH	H	Me	H	H	H	4.82
18	CH	H	CF <sub>3</sub>	H	H	H	4.77
19	CH	H	F	H	H	H	4.72
20 <sup>b</sup>	CH	H	CN	H	H	H	4.43
21 <sup>b</sup>	CH	H	OMe	H	H	H	4.33
22	CH	H	NO <sub>2</sub>	H	H	H	4.31
23	CH	H	NH <sub>2</sub>	H	H	H	3.75
24	CH	H	Br	H	H	Me	5.59
25	CH	H	Cl	H	H	Me	5.57
26 <sup>b</sup>	CH	H	F	H	H	Me	5.20
27 <sup>b</sup>	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 <sub>2</sub> Ph	Me	5.37
33 <sup>b</sup>	CH	H	H	H	OH	Me	5.22
34	CH	H	H	H	OEt	Me	5.13
35	CH	H	H	H	OCH <sub>2</sub> Ph	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 <sup>b</sup>	N	Me	H	H	H	Me	5.42
43	N	Me	H	H	H	H	5.26
44 <sup>b</sup>	N	OMe	H	H	H	Me	5.26

No.	Substitution						log(1/IC <sub>50</sub> )
	X	R <sub>3</sub> '	R <sub>5</sub> '	R <sub>2</sub>	R <sub>4</sub>	R <sub>5</sub>	
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 <sup>b</sup>	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 <sup>b</sup>	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 <sup>b</sup>	CH	Me	Me	H	H	H	4.92
57 <sup>b</sup>	N	Cl	H	H	H	H	5.48
58 <sup>b</sup>	CH	H	Cl	H	H	H	5.35
59 <sup>b</sup>	CH	H	H	H	H	H	4.77
60 <sup>b</sup>	CH	Cl	Me	H	H	H	5.48

<sup>a</sup> Observed inhibitory activity.

<sup>b</sup> The compounds selected as the test set.

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. A total number of 293 descriptors were calculated by Dragon for each molecule and 60×293 data matrix X was obtained. The rows and columns of this matrix are the number of molecules and molecular descriptors respectively.

### Discrete Wavelet Transform

Transform-based methods are of fundamental importance in signal and image processing. Wavelet transform decomposes a signal into a set of basic functions. These basic functions are obtained from a single prototype wavelet called mother wavelet by dilations and shifting [20]. The discrete wavelet transform (DWT) is a linear transformation that operates on a data vector whose length is an integer power of two, transforming it into a numerically different vector of the same length. It is a tool which separates the data into different frequency components, and then studies each component with resolution matched to its scale. DWT [24] is computed with a cascade of filtering followed by a factor 2 sub-sampling (Fig. 2).

*H* and *L* denote high and low-pass filters respectively, ↓2 denotes sub-sampling. Outputs of these filters are given by equations (2) and (3):

$$a_{j+1}[p] = \sum_{n=-\infty}^{+\infty} l[n-2p] a_j[n] \quad (2)$$

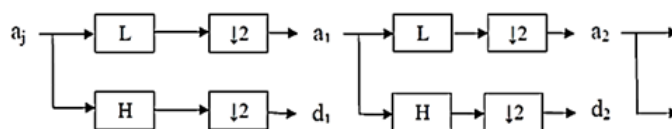
$$d_{j+1}[p] = \sum_{n=-\infty}^{+\infty} h[n-2p] d_j[n] \quad (3)$$

The elements  $a_j$  are used for next step (scale) of the transform and the elements  $d_j$ , the wavelet coefficients, determine the output of the transform.  $l[n]$  and  $h[n]$  are coefficients of low and high-pass filters respectively. It can be assumed that on scale  $j+1$  there is only half from number of  $a$  and  $d$  elements on scale  $j$ . These elements are called scaling function coefficients.

## Results and discussion

### Principal Component Analysis

Principal component analysis (PCA) was performed on the calculated structural descriptors to the whole data set (Table 1), 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. PCA is a useful multi-



**Fig. 2.** DWT tree. (a) *H* and *L* denote high and low-pass filters respectively, (b) ↓2 denotes sub-sampling (c)  $a_1$  and  $a_2$  denote the approximation of the data and (d)  $d_1$  and  $d_2$  denote the detail of the data.

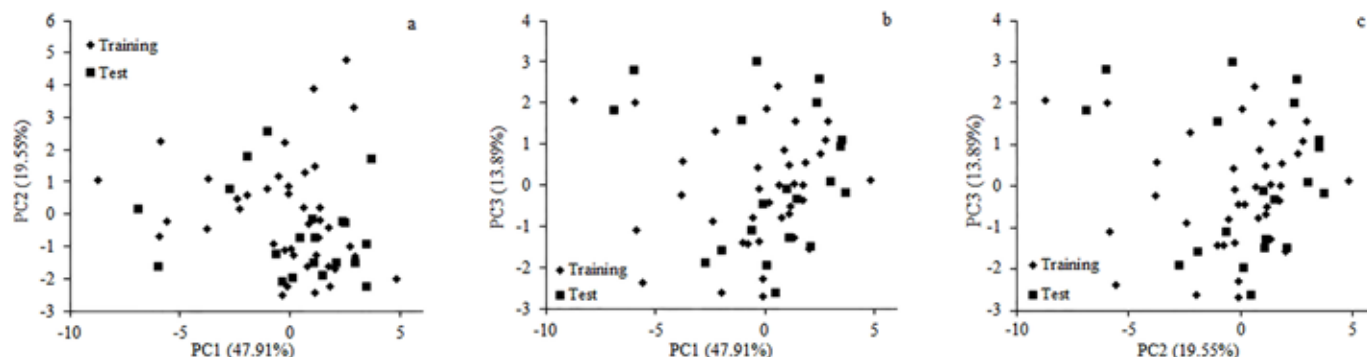


Fig. 3. Principal components analysis of the descriptors for the data set, (a) PC2 versus PC1, (b) PC3 versus PC1 and (c) PC3 versus PC2.

variate statistical technique in which new variables (called principal components, PCs) are calculated as linear combinations of the old ones. These PCs are sorted by decreasing information content so that most of the information is preserved in the first few PCs. An important feature is that the obtained PCs are uncorrelated, and they can be used to derive scores which can be used to display most of the original variations in a smaller number of dimensions. These scores can also allow us to recognize groups of samples with similar behavior. A total of 293 descriptors were initially calculated by PCA for the entire data set of 60 compounds. The total number of descriptors was reduced to 35 descriptors by eliminating the descriptors that were deemed insignificant (i.e. where the one-parameter correlation confinement with the activity is less than 0.1). The PCA results indicate that three PCs (PC1, PC2 and PC3) described 81.35% of the overall variances: 47.91%, 19.55% and 13.89% for PC1, PC2 and PC3, respectively (Fig. 3). As seen in Fig. 3, 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, the data set was separated into two groups, a training set (40 data) and a prediction set (20 data) according to the Kennard-Stones algorithm. As shown in Fig. 3, the distribution of the compounds in each subset seems to be relatively well-balanced over the space of the principal components.

### PLS Modeling

The multivariate calibration is a powerful tool for modeling, as it extracts further information and allows building more robust models [25, 26]. The PLS method is used to establish relationships between the dependent variables of the activity matrix and the descriptors of the matrix as independent variables also that are called latent variables. Based on the inhibitory activity data (Table 1), the data was classified to training and prediction sets according to Kennard-Stones algorithm and the PLS model was run. The optimum number of factors to be included in the calibration model was determined by computing the root mean square error of calibration (*RMSEC*) from cross-validated models using a high number of factors (half the number of total

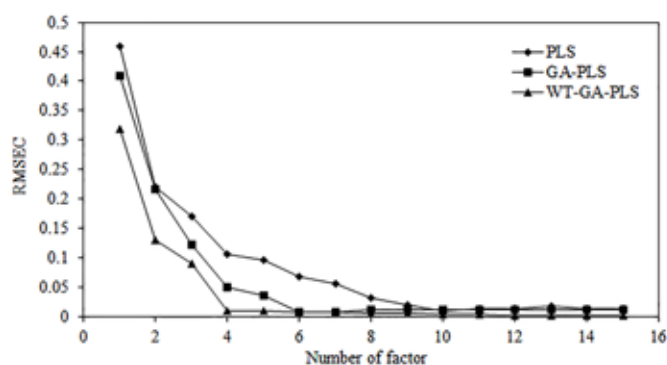


Fig. 4. Plot of *RMSEC* versus number of factors.

training set + 1). The cross-validation method was employed to eliminate only one compound at a time and then, PLS was employed to calibrate the remaining of the training set. The inhibitory activity of the left-out sample was predicted using this calibration. This process was repeated until each compound in the training set had been left out once. According to Haaland suggestion [26], the optimum number of factors selected. As it is shown in Fig. 4, the *RMSEC* is minimized when the number of factors is 10, thus, the optimum number of factors for the training set of PLS method was chosen to be 10. The data set was mean-centered, prior to the PLS analysis.

### GA-PLS Modeling

As mentioned before, one of the problems is choosing the set of molecular descriptors. GAs as intelligent selection techniques [27], was utilized to achieve this objective. The parameters of genetic algorithms used in this study are as below: the probability of mutation 1% and 90% for crossover, number of runs is 100 and window size for smoothing is 3. Prior to performing the GA-PLS, all the descriptors were mean-centered. After running of GAs for variables, the selected descriptors were used to run PLS. Among the descriptors calculated, the most significant molecular descriptors were identified. Table 2 shows the calculated descriptors for each molecule, the *t* values for null hypothesis and their related *P* values [28]. The number of factors reduced to 6 upon the application of GA-PLS (Fig. 4). As

**Table 2.** The most significant molecular descriptors used in this study.

Notation	Descriptors	t value	P value
H2v	H autocorrelation of lag 2 / weighted by van der Waals volume	8.330	0.000
R2u	R autocorrelation of lag 2 / unweighted	6.169	0.000
R1u	R autocorrelation of lag 1 / unweighted	-6.613	0.000
P2u	2nd component shape directional WHIM index / unweighted	-4.575	0.000
R6U <sup>+</sup>	R maximal autocorrelation of lag 6 / unweighted	-4.317	0.000
R7e	R autocorrelation of lag 7 / weighted by Sanderson electronegativity	-4.018	0.000

shown in Fig. 4, the *RMSEC* is minimized when the number of factors is 6, thus, the optimum number of factors for the training set of GA-PLS method was chosen to be 6. The present study shows that the GAs can be a good method for descriptor selection in analysis.

### WT-GA-PLS Modeling

In order to enhance the predictive power of multivariate calibration methods, molecular descriptors are often corrected prior to the data analysis. In this case, wavelet transform was used as the processing method. When WT-GA-PLS is used the number of factors reduced to 4 (Fig. 4). Table 2 shows the calculated descriptors for each molecule, the *t* values for null hypothesis and their related *P* values. In order to evaluate the models applied in this study (PLS, GA-PLS and WT-GA-PLS), the inhibitory activities were tested with set of tests. The WT-GA-PLS method combines WPT, which performs feature extraction and de-noising, GA which optimizes variation and selection of the fitness values, as well as PLS, which provides calibration model and reduces the dimension of the data. To be successful in obtaining a reliable result by the WT-GA-PLS method, different parameters such as wavelet functions (Coiflet 1, 2...5, and Daubechies 6, 7, 8), decomposition level (*L* = 1–2) and the number of PLS factors were tested. In these investigations, Coiflet 3, *L* = 1 and number of PLS factors = 4 were selected as the optimal parameters.

### Model Validation and Prediction of Inhibitory Activity

The predictive ability of these methods (PLS, GA-PLS and WT-GA-PLS) were investigated by prediction of inhibitory of 20 molecules (their structures are given in Table 1). The validation of predictive ability is another key step in the 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.

$$RMSEP = \sqrt{\frac{\sum_{i=1}^n (y_{i,pred} - y_{i,obs})^2}{n}} \quad (4)$$

$$RSEP(\%) = 100 \times \sqrt{\frac{\sum_{i=1}^n (y_{i,pred} - y_{i,obs})^2}{\sum (y_{i,obs})^2}} \quad (5)$$

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 WT-GA-PLS model with percentage error ranges from -5.844 to +4.512 % for inhibitory activity of 1-phenyl[2H]-tetrahydro-triazine-3-one analogues. Other statistical parameters have been used for the evaluation of the suitability of the developed models for prediction of the activity of the studied compounds this include cross validation coefficient ( $Q^2$  and  $R^2$ ) [29]. These parameters are defined as follows:

$$Q_{abs}^2 = 1 - \frac{\sum_Y (Y_{exp} - Y_{LOO})^2}{\sum_Y (Y_{exp} - \bar{Y}_{exp})^2} \quad (6)$$

$$R_{abs}^2 = 1 - \frac{\sum_Y (Y_{exp} - Y_{pred})^2}{\sum_Y (Y_{exp} - \bar{Y}_{exp})^2} \quad (7)$$

As seen in Table 3, these parameters show the good statistical qualities. The plots of the predicted inhibitory activity versus actual values are shown in Fig. 5 for each model. It is possible to see that WT-GA-PLS presents excellent prediction abilities when compared with other models.

### Y-randomization Test

This is a widely used technique to ensure the robustness of a QSAR model. In this test, the dependent-variable vector, Y-vector, is randomly shuffled and a new QSAR model is developed using the original independent-variable matrix. The process is repeated three times and the average of the three measurements showed low  $R^2$  values 0.226, 0.241 and 0.254 and  $Q^2$  values 0.137, 0.152 and 0.170 for the PLS, GA-PLS and WT-GA-PLS, respectively. If all QSAR models obtained in the Y-randomization test have relatively high  $R^2$  and  $Q^2$ , it implies that an acceptable QSAR model cannot be obtained for the given data set by the current modeling method [30, 31].

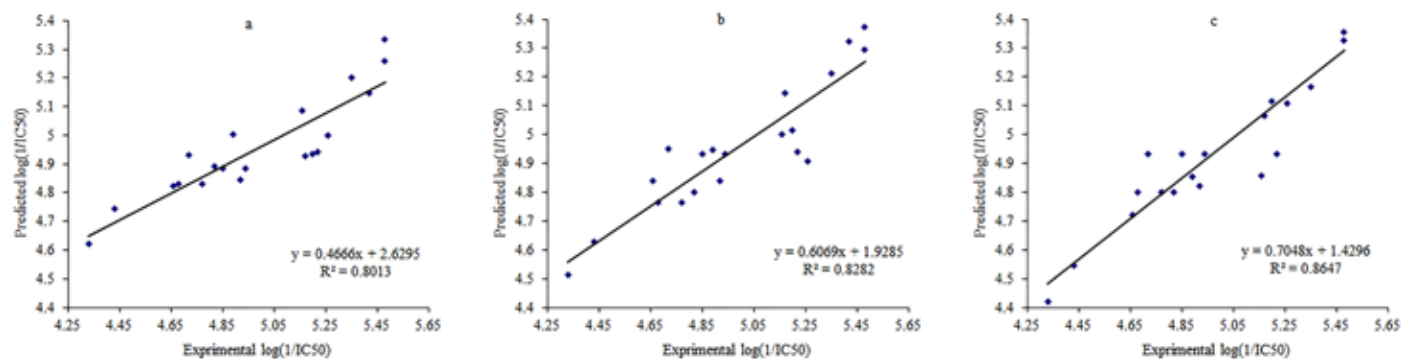


Fig. 5. Plots of predicted versus actual  $\log(1/IC_{50})$ , (a) PLS, (b) GA-PLS and (c) WT-GA-PLS.

Table 3. Observation and calculation values of  $\log(1/IC_{50})$  using PLS, GA-PLS and WT-GA-PLS models.

No. of compounds (Table I)	Observation $\log(1/IC_{50})$	PLS		GA-PLS		WT-GA-PLS	
		Predicted	Error (%)	Predicted	Error (%)	Predicted	Error (%)
5	5.170	4.926	-4.702	5.144	-0.489	5.064	-2.050
7	4.940	4.883	-1.137	4.931	-0.182	4.930	-0.190
9	4.850	4.884	0.719	4.930	1.665	4.930	1.665
11	4.680	4.829	3.196	4.763	1.788	4.800	2.570
15	4.890	5.004	2.335	4.947	1.173	4.854	-0.736
17	4.820	4.892	1.510	4.80	-0.394	4.798	-0.452
20	4.430	4.745	7.112	4.627	4.467	4.545	2.598
21	4.330	4.620	6.713	4.511	4.196	4.418	2.036
26	5.200	4.935	-5.094	5.014	-3.559	5.115	-1.628
27	4.720	4.930	4.461	4.951	4.913	4.933	4.512
33	5.220	4.942	-5.312	4.937	-5.408	4.931	-5.536
42	5.420	5.149	-4.998	5.323	-1.787	5.424	0.086
44	5.260	5.001	-4.920	4.906	-6.726	5.106	-2.912
49	4.660	4.824	3.527	4.837	3.804	4.720	1.300
53	5.160	5.087	-1.399	5.001	-3.069	4.858	-5.844
56	4.920	4.846	-1.500	4.838	-1.666	4.820	-2.028
57	5.480	5.259	-4.025	5.295	-3.361	5.355	-2.279
58	5.350	5.202	-2.749	5.210	-2.614	5.164	-3.471
59	4.770	4.829	1.249	4.763	-0.132	4.800	0.635
60	5.480	5.333	-2.682	5.371	-1.978	5.327	-2.788
N. F. <sup>a</sup>		10		6		4	
R <sup>2</sup>		0.652		0.759		0.818	
Q <sup>2</sup>		0.536		0.691		0.744	
RMSEP		0.194		0.161		0.140	
RSEP (%)		3.882		3.229		2.807	

<sup>a</sup> Number of factors.

### Molecular Design

The role of computation in molecular design has grown steadily since the late 1960s [32, 33]. In the early days emphasis was on statistical and computational approaches aimed at quantify-

ing the relationship of chemical structure to biological properties. In addition, recent modeling by computational approaches has become a critical tool in the drug discovery process. As an application of proposed method, we investigated WT-GA-PLS model to predict the inhibitory activity of five new

**Table 4.** New structures of 1-phenyl[2H]-tetrahydro-triazine-3-one analogues and predicted  $\log(1/IC_{50})$  by WT-GA-PLS.

Number of Design	Substitution						$\log(1/IC_{50})$ Calc. <sup>a</sup>
	X	R <sub>3</sub> '	R <sub>5</sub> '	R <sub>2</sub>	R <sub>4</sub>	R <sub>5</sub>	
1	N	H	Me	H	H	Me	4.931
2	N	Cl	H	H	H	F	5.450
3	CH	Br	Me	H	H	H	5.778
4	CH	H	Cl	H	OH	H	5.191
5	CH	Br	H	H	H	Me	5.860

<sup>a</sup> Calculated by WT-GA-PLS model.

1-phenyl[2H]-tetrahydro-triazine-3-one analogues whose biological tests were not performed with them yet. Table 4 shows the chemical structure of five new compounds and their inhibitory activity calculated by this proposed method.

## Conclusions

Using WT-GA-PLS, 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[2H]-tetrahydro-triazine-3-one analogues. The model could predict the inhibitory activity of 1-phenyl[2H]-tetrahydro-triazine-3-one analogues derivatives not existed in the modeling procedure accurately. The work is the first application of WT-GA-PLS for QSAR study and shows that descriptors are capable to recognize the physicochemical information and be can useful to predict the inhibitory activity of the new compounds.

## Acknowledgments

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

## References

1. Camargo, A. B.; Marchevsky, E.; Luco, J. M. *J. Agric. Food Chem.* **2007**, 55, 3096-3103.

2. Zhou, Y. P.; Jiang, J. H.; Lin, W. Q.; Zou, H. Y.; Wu, H. L.; Shen, G. L.; Yu, R. Q. *Eur. J. Pharm. Sci.* **2006**, 28, 344-353.

3. Ioele, G.; Luca, M. D.; Oliverio, F.; Ragno, G. *Talanta* **2009**, 79, 1418-1424.

4. Cui, W.; Yan, X. *Chemometr. Intell. Lab. Syst.* **2009**, 98, 130-135.

5. Ghasemi, J. B.; Ahmadi, S. H.; Brown, S. D. *Environ. Chem. Lett.* **2011**, 9, 87-96.

6. Hocking, R. R. *Biometrics.* **1976**, 32, 1-49.

7. Shen, Q.; Lu, Q. Z.; Jiang, J. H.; Shen, G. L.; Yu, R. Q. *Eur. J. Pharm. Sci.* **2003**, 20, 63-71.

8. Leardi, R. *J. Chemom.* **1994**, 8, 65-79.

9. Depczynski, U.; Frost, V. J.; Molt, K. *Anal. Chim. Acta.* **2000**, 420, 217-227.

10. Hibbert, D. B. *Chemom. Intell. Lab. Syst.* **1993**, 19, 277-293.

11. Leardi, R. *J. Chemom.* **2000**, 14, 643-655.

12. Leardi, R. *J. Chemom.* **2001**, 15, 559-569.

13. Ghasemi, J. B.; Niazi, A.; Leardi, R. *Talanta* **2003**, 59, 311-317.

14. Leardi, R.; Boggia, R.; Terrile, M. *J. Chemom.* **1992**, 6, 267-281.

15. Kompany-Zareh, M. *Acta Chim Slov.* **2003**, 50, 259-273.

16. Narasimhan, B.; Judge, V.; Narang, R.; Ohlan, R.; Ohlan, S. *Bio-org. Med. Chem. Lett.* **2007**, 17, 5836-5845.

17. Hemmateenejad, B.; Safarpour, M. A.; Taghavi, F. *J. Mol. Struct.* **2003**, 635, 183-190.

18. Niazi, A.; Jameh-Bozorghi, S.; Nori-Shargh, D. *Turk. J. Chem.* **2006**, 30, 619-628.

19. Joreskog, K. G.; Wold, H. *System Under Indirect Observations*, Amsterdam, North Holland, **1982**.

20. Blessie, A. A.; Nalini, J.; Ramesh, S. C. *Int. J. Comput. Sci. Issues.* **2011**, 8, 449-453.

21. Kennard, R. W.; Stones, L. A. *Technometrics.* **1969**, 11, 137-148.

22. Daszykowski, M.; Walczak, B.; Massart, D. L. *Anal. Chim. Acta.* **2002**, 468, 91-103.

23. Ahmadi, M.; Shahlaei, M. *Res. Pharm. Sci.* **2015**, 10, 307-325.

24. Mallat, S. A. *Wavelet Tour of Signal Processing*, Academic Press, San Diego, **1998**.

25. Ghasemi, J. B.; Niazi, A. *Talanta* **2005**, 65, 1168-1173.

26. Haaland, D. M.; Thomas, E. V. *Anal. Chem.* **1988**, 60, 1193-1202.

27. Niazi, A.; Leardi, R. *J. Chemom.* **2012**, 26, 345-351.

28. Cherkasov, A.; et al. *J. Med. Chem.* **2014**, 57, 4977-5010.

29. Tropsha, A. *Mol. Inf.* **2010**, 29, 476-488.

30. Tropsha, A.; Gramatica, P.; Vijay, K.G. *QSAR. Com. Sci.* **2003**, 22, 69-77.

31. Golbraikh, A.; Tropsha, A. *J. Mol. Graph. Model.* **2002**, 20, 269-276.

32. Richon, A. B. *Drug Discov. Today* **2008**, 13, 659-664.

33. Clark, D. E. *Exp. Opin. Drug Discov.* **2006**, 1, 103-110.