



# **Octanol-water Partition Coefficients Determination and QSPR Study of Some 3-hydroxy Pyridine-4-one Derivatives**

**R. Sabet<sup>1\*</sup>, A. Fassih<sup>1</sup> and L. Saghaie<sup>2</sup>**

<sup>1</sup>*Department of Medicinal Chemistry, School of Pharmacy, Shiraz University of Medical Sciences, Shiraz, Iran.*

<sup>2</sup>*Department of Medicinal Chemistry, School of Pharmacy and Pharmaceutical Sciences, Isfahan University of Medical Sciences, Isfahan, Iran.*

## **Authors' contributions**

*This work was carried out in collaboration between all authors. Author RS designed the study, performed the statistical analysis, wrote the protocol, wrote the first draft of the manuscript and managed the literature searches. Authors AF and LS managed the analysis of the study. All authors read and approved the final manuscript.*

## **Article Information**

DOI: 10.9734/JPRI/2018/41142

### Editor(s):

(1) Mohamed Fathy, Professor, Faculty of Pharmacy, Assiut University, Assiut, Egypt.

(2) Sanjay Nilapwar, Senior Research Investigator (Purification Development), CMC Large Molecule, Wilmington, USA.

### Reviewers:

(1) Amel Bendjeddou, Souk Ahras University, Algeria.

(2) Ravi Bansal, Jiwaji University Gwalior, India.

(3) Geetha Ramakrishnan, Sathyabama University, India.

Complete Peer review History: <http://www.sciencedomain.org/review-history/24968>

**Original Research Article**

**Received 24<sup>th</sup> February 2018**

**Accepted 7<sup>th</sup> May 2018**

**Published 4<sup>th</sup> June 2018**

## **ABSTRACT**

The partition coefficients ( $K_{part}$ , in octanol/water system) of a range of bidentate ligands containing the 3-hydroxy pyridine-4-one moiety were determined using shake flask. These derivatives were subjected to quantitative structure-property relationships (QSPR) analysis. A collection of chemometrics methods, including partial least squares combined with the genetic algorithm as variable selection method (GA-PLS), factor analysis-based multiple linear regression (FA-MLR) and principal component regression (PCR) were employed to make connections between structural parameters and  $\log p_{o/w}$ . The results revealed the significant role of constitutional parameters in the partition coefficient of the studied compounds. The most significant QSAR model, obtained by GA-PLS, could explain and predict 96% and 91% of variances in the  $\log p_{o/w}$  data.

\*Corresponding author: E-mail: Sabet\_r@sums.ac.ir;

**Keywords:** Partition coefficient; 3-Hydroxypyridine-4-one; QSPR; Log P.

## 1. INTRODUCTION

Drug disposition processes, i.e. absorption, distribution, metabolism and excretion (ADME) seriously depend on the ability or inability of the molecules to cross the biological membranes or attach to the hydrophobic binding sites of the proteins involved in drug disposition. Thus, a high correlation is expectable between the ADME processes and measures of lipophilicity [1]. Not only pharmacokinetics of the drug but also its pharmacodynamics is affected by this molecular property. Hydrophobic drug-receptor interactions sometimes determine the potency of the drug molecule [2]. For an organic compound, its partition between *n*-octanol and water is generally accepted as a physicochemical parameter for characterization of lipophilicity [2]. Experimentally, this partition is defined as the ratio of concentrations of the compound at equilibrium between the organic and aqueous phases. The logarithm of this ratio,  $\log p_{o/w}$ , has been widely used as a measure of hydrophobicity and lipophilicity and it is a determinant physicochemical property considered in Lipinski's rule of 5 for potential drug molecules [3].

The procedure of obtaining a new drug is difficult and consumes much time to construct [4].

Currently, computational methods are used in order to simplify drug discovery, design, development and optimization. In particular, computer-aided drug discovery is being utilized to identify active drug candidates, select leads and optimize them, i.e. transform biologically active compounds into suitable drugs by improving their physicochemical, pharmaceutical and pharmacokinetic properties.

Pioneering work by Hansch has led to the use of  $\log p$  in Quantitative structure Activity relationship (QSAR). Computational methods for the investigation of  $\log p$  since 1964 when Fujita et al. correlate difference of benzene and substituted benzenes to experimental data of  $\log p$  and used these data for predicting  $\log p$  for another series [5].

Quantitative structure-property relationships models (QSPR) describe a mathematical relationship between the structure of a chemical compound and its physicochemical properties.

Different linear and nonlinear modelling methods are used in QSPR. There are two types of QSPR models: regression and classification models. Among regression models, multiple linear regression (MLR), principle component regression (PCR), and partial least squares (PLS) can be mentioned. MLR equations can describe the structure-property relationships well but some information will be discarded in MLR analysis. Due to the co-linearity problem in MLR analysis, variable selection methods including forward, backward, and stepwise selection. There are also some other methods which are inspired by nature, the most widely used is genetic algorithm [6-8]. Factor analysis identifies the important predictor variables contributing to the response variable and avoids collinearities among them. PLS analysis as a factor analysis-based method omits the multicollinearity problem in the descriptors. In this method, the descriptors data matrix is decomposed to orthogonal matrices with an inner relationship between the dependent and independent variables. Because a minimal number of latent variables are used for modelling in PLS; this modelling method coincides with noisy data better than MLR. MLR yields models that are simpler and easier to interpret than PCR and PLS, because these methods perform regression on latent variables that don't have a physical meaning. On the other hand, factor analysis-based methods can handle the collinear descriptors and therefore better predictive models will be obtained by PLS method [9].

$\log p_{o/w}$ , has also been shown to be one of the key parameters in QSPR studies. There are some reports about the application of MLR and ANN modelling to predict the *n*-octanol/water partition coefficient of organic compounds [10-15].

In the present paper, more than 600 topological, geometrical, constitutional, functional group, and chemical descriptors were used to develop different QSPR models for the  $\log p_{o/w}$  of the studied compounds. The methods used in this study for model construction were: (i) genetic algorithm-partial least squares (GA-PLS), (ii) factor analysis MLR (FA-MLR) and (iii) principal component regression analysis (PCRA). The key aim of this work was to investigate molecular descriptors important in determining *n*-octanol/water partition coefficient.

## 2. METHODOLOGY

### 2.1 Determination of Partition Coefficients Using the Shake Flask Method

Partition coefficients ( $K_{part}$ ) of the molecules used in the present study were determined using the shake flask method. The two phases used in determination were tris buffer (50 mM, pH 7.4, prepared using distilled water) and 1-octanol, each of which was pre-equilibrated with the other phase before use (the solubility of water in 1-octanol is 2.3 M [16]). A solution of compounds with a concentration of  $10^{-4}$  M was prepared in tris buffer and the absorbance of the solution was measured in the ultraviolet region of a wavelength of approximately 290 nm using the buffer as a blank. A known volume (normally 10-50 ml) sample of the solution was stirred vigorously with a suitable volume of 1-octanol in a glass vessel for 1 h. The two layers were separated by centrifugation for 5 minutes. An aliquot of the aqueous layers was then carefully removed using a glass Pasteur pipette ensuring that the sample was not contaminated with 1-octanol. The absorbance of the sample was measured as above and the partition coefficient was then calculated using the following formula:

$$K_{part} = \frac{A_1 - A_2}{A_2} \times \frac{V_w}{V_o} \quad (1)$$

Where

$A_1$  = Absorbance reading in the aqueous layer before partitioning

$A_2$  = Absorbance reading in the aqueous layer after partitioning

$V_w$  = Volume of the aqueous layer used in partitioning

$V_o$  = Volume of 1-octanol layer used in partitioning

For each compound, the experiment was repeated four times which led to the calculation of a mean  $K_{part}$  value and standard deviation (Table 1).

The results for the mean  $K_{part}$  value and standard deviation of the compounds **16-19** are described elsewhere [17].

### 2.2 Software

Two-dimensional structures of molecules were drawn using Hyperchem 7.0 software [18]. The molecules were optimized with the same software, in order to find generically stable conformation. The structures were pre-optimized with the Force-Field Molecular Mechanics (MM<sup>+</sup>) procedure. Then, the resulting geometries were further optimized by means of the Semi-Empirical Molecular Orbital Method AM1 (Austin Model 1) by using the Polak-Ribiere's algorithm until the root mean square gradient of 0.01 kcal/(Å mol). The resulted geometry was transferred into Dragon program package, developed by Milano Chemometrics and QSAR Group [19]. MATLAB R2017 software was used for the PLS regression method.

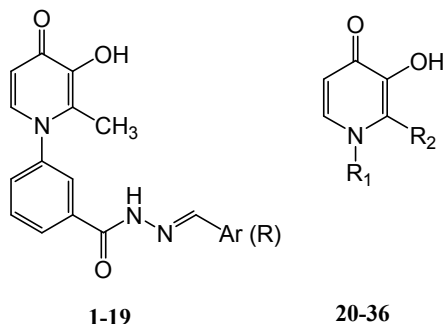
### 2.3 Partition Coefficient Data and Descriptor Generation

The data used in this study were  $\log p_{o/w}$  of some 3-hydroxy pyridine-4-one derivatives [20]. The structural features of these compounds are listed in Table 3. They were used for subsequent QSPR analysis as independent variables. The  $\log p_{o/w}$  values are shown in Table 1.

A large number of molecular descriptors was calculated using Dragon package. The Dragon software calculated several classes of descriptors such as topological, functional groups, geometrical and constitutional descriptors for each molecule. The calculated descriptors for each molecule are summarized in Table 2. Finally, constant descriptors (i.e., variables that take the same value for all compounds in the dataset) and near constants (i.e., variables that take the same value, but a small number of samples to take other values) were omitted.

### 2.4 Data Screening and Model Building

The selected descriptors from each class and the experimental data were analyzed by SPSS (version 22.0) software. The calculated descriptors were collected in a data matrix whose number of rows and columns were the number of molecules and descriptors, respectively. Partial least squares combined with genetic algorithm (GA-PLS), MLR with factor analysis (FA-MLR) and principal component regression analysis (PCRA) methods were used to derive the QSPR equations.

**Table 1. Chemical structure of the compounds used in QSAR is 3-hydroxy pyridine-4-one**

Compound	Ar (R)	Experimental logp <sub>o/w</sub>
1	4-Methoxyphenyl	1.41
2	2-Methoxyphenyl	1.37
3	4-Methylphenyl	1.67
4	2-Chlorophenyl	1.69
5**	4-Chlorophenyl	1.73
6	1-Methyl-5-nitro imidazole-2-yl	1.13
7	5-Nitrofuranyl	1.11
8**	1-Methyl-2-methylthio imidazol-5-yl	1.85
9	2-Hydroxyphenyl	1.14
10	N, N-Dimethyl aminophenyl	1.83
11	Phenylvinyl	1.85
12	CH <sub>3</sub>	1.29
13*	3-Bromophenyl	1.91
14	4-Bromophenyl	1.93
15**	4-Nitrophenyl	1.45
16*	Thiophen-2-yl	1.32
17	Furan-2-yl	1.29
18**	Phenyl	1.92
19	3-Nitrophenyl	1.41

Compound	R <sub>1</sub>	R <sub>2</sub>	Experimental logp <sub>o/w</sub>
20	CH(CH <sub>2</sub> OH) <sub>2</sub>	Me	0.02
21	CH <sub>2</sub> CH <sub>2</sub> NH <sub>2</sub>	Me	0.03
22	CH <sub>2</sub> CH <sub>2</sub> CH <sub>2</sub> COOH	Et	0.04
23*	CH <sub>2</sub> CH <sub>2</sub> OH	Me	0.08
24	CH <sub>2</sub> CH <sub>2</sub> CH <sub>2</sub> OH	Me	0.13
25	CH <sub>3</sub>	Me	0.17
26	(CH <sub>2</sub> ) <sub>4</sub> OH	Me	0.18
27*	CH <sub>2</sub> CH <sub>2</sub> OH	Et	0.22
28	H	Me	0.32
29	CH <sub>2</sub> CH <sub>2</sub> OCH <sub>3</sub>	Me	0.39
30	CH <sub>2</sub> CH <sub>3</sub>	Me	0.49
31	(CH <sub>2</sub> ) <sub>4</sub> OH	Et	0.52
32	CH <sub>3</sub>	Et	0.62
33*	CH <sub>3</sub> CH=CH <sub>2</sub>	Me	1.07
34	H	Et	1.11
35	CH <sub>2</sub> CH <sub>2</sub> CH <sub>3</sub>	Me	1.51
36	CH <sub>2</sub> CH <sub>3</sub>	Et	1.70

\*Compounds used as prediction set

\*\*Compounds were an outlier and deleted in QSPR study

**Table 2. A brief description of some descriptors used in this study**

Descriptor Type	Molecular Description
Constitutional	Mean atomic van der Waals volume (Mv) (scaled on Carbon atom), no. of heteroatoms, no. of multiple bonds (nBM), no. of rings, no. of circuits, no of H-bond donors, no of H-bond acceptors, no. of Nitrogen atoms (nN), chemical composition, sum of Kier-Hall electrotopological states (Ss), mean atomic polarizability (Mp), number of rotatable bonds (RBN), mean atomic Sanderson electronegativity (Me), etc.
Topological	Narumi harmonic topological index (HNar), Total structure connectivity index (Xt), information content index (IC), mean information content on the distance degree equality (IDDE), total walk count, path/walk-Randic shape indices (PW3, PW4, PW5, Zagreb indices, Schultz indices, Balaban J index (such as MSD) Wiener indices, Information content index (neighborhood symmetry of 2-order) (IC2), Ratio of multiple path count to path counts (PCR), Lovasz-Pelikan index (leading eigenvalue) (LP1), total information content index (neighborhood symmetry of 1-order) (TIC1), reciprocal hyper-detour index (Rww), Average connectivity index chi-5 (X5A), piLD (conventional bond-order ID number), etc.
Geometrical	3D Petijean shape index (PJI3), Asphericity (ASP), Gravitational index, Balaban index, Wiener index, Length-to-breadth ratio by WHIM (L/Bw), etc.
Functional group	Number of total secondary C(sp <sup>3</sup> ) (nC <sub>s</sub> ), Number of total tertiary carbons (nC <sub>t</sub> ), Number of H-bond acceptor atoms (nHAcc), Number of secondary amides (aliphatic) (nCONHR), Number of unsubstituted aromatic C (nCaH), Number of ethers (aromatic) (nRORPh), Number of ketones (aliphatic) (nCO), Number of tertiary amines (aliphatic) (nNR2), Number of phenols (nOHPh), Number of total primary C(sp <sup>3</sup> ) (nC <sub>p</sub> ), etc.

A genetic algorithm was applied as a feature selection method for partial least squares regression (GA-PLS) to model the structure- $\log_{o/w}$  relationships [21-23]. Partial least squares (PLS) linear regression is a recent technique that generalizes and combines features from principal component analysis and multiple linear regressions. PLS is a method suitable for overcoming the problems in MLR related to multicollinear or over-abundant descriptors [9]. This method is normally used in combination with cross-validation to obtain the optimum number of components [24-25]. The PLS regression method used was the NIPALS-based algorithm existed in the chemometrics toolbox of MATLAB software (version 2017b Math Work Inc.). In order to obtain the optimum number of factors based on the Haaland and Thomas F-ratio criterion, leave-one-out cross-validation procedure was used [26]. Factor analysis (FA) was used to reduce the number of variables and to detect structure in the relationships between them [27]. Principle component regression analysis, PCRA, was also exploited for the dataset along with FA-MLR. With PCRA, collinearities among X variables are not a disturbing factor and the number of variables included in the analysis may exceed the number of observations [28]. In this method,

factor scores, as obtained from FA, are used as the predictor variables [27]. In PCRA, all descriptors are assumed to be important while the aim of factor analysis is to identify relevant descriptors.

## 2.5 Variable Importance in the Projection (VIP)

In order to investigate the relative importance of the variables appeared in the final model obtained by GA-PLS method, variable importance in projection (VIP) was employed [29]. VIP values reflect the importance of terms in PLS model. According to Erikson et al. [30] X-variables (predictor variables) could be classified according to their relevance in explaining y (predicted variable), so that  $VIP > 1.0$  and  $VIP < 0.8$  mean highly or less influential, respectively, and  $0.8 < VIP < 1.0$  mean moderately influential.

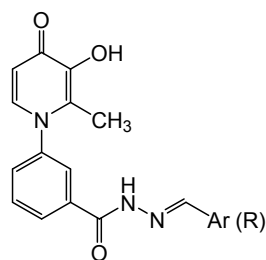
## 3. RESULTS AND DISCUSSION

### 3.1 $K_{part}$ Values of 3-hydroxy Pyridine-4-One Derivatives

The method used for  $k_{part}$  determination involves mixing of an aqueous solution of a known

concentration of the studied compound with a known volume of the organic phase and to allow the system to attain equilibrium. When equilibrium is established, either the concentration remaining in the aqueous phase or the concentration in the organic phase is measured (usually spectrophotometrically) and the  $K_{part}$  values are calculated. The studied derivatives covered a range of  $K_{part}$  values of 1.11-1.93. Among the studied compounds **14** possesses the highest partition coefficient (Table 3).

**Table 3.  $K_{part}$  values of the compounds between 1-octanol and tris buffer at pH 7.4. A number of determination = 4.**



1-15

Compound	$K_{part}$
1	1.41 ± 0.14
2	1.37 ± 0.01
3	1.67 ± 0.08
4	1.69 ± 0.12
5	1.73 ± 0.09
6	1.13 ± 0.12
7	1.11 ± 0.15
8	1.85 ± 0.19
9	1.14 ± 0.13
10	1.83 ± 0.09
11	1.85 ± 0.18
12	1.29 ± 0.16
13	1.91 ± 0.17
14	1.93 ± 0.15
15	1.45 ± 0.11

### 3.2 GA-PLS

In order to find the more convenient set of descriptors in PLS modeling, a genetic algorithm was used. To do so, many different GA-PLS runs were conducted using different initial sets of populations. The data set was divided into two groups: calibration set ( $n = 27$ ) and prediction set ( $n = 5$ ). Given 27 calibration samples; leave-one-out cross-validation procedure was used to find the optimum number of latent variables for each PLS model. The most

convenient GA-PLS model that resulted in the best fitness contained 15 indices. The PLS estimate of coefficients for these descriptors are given in Fig. 1. The greater the absolute value of a coefficient, the greater the weight of the variable in the model. As it is observed, a combination of topological, geometrical, constitutional, and functional group descriptors have been selected by GA-PLS to account the partition coefficient of the studied compounds. The majority of these descriptors are constitutional indices. The resulted GA-PLS model possessed very high statistical quality  $R^2 = 0.96$  and  $Q^2 = 0.91$ . To measure the significance of the 15 selected PLS descriptors, VIP was calculated for each descriptor. The VIP analysis of PLS equation is shown in Fig. 2. VIP shows that Ms, RBN and nH, which are constitutional parameters, are the most important indices in the QSPR equation derived by PLS analysis. In addition, ASP as a geometrical parameter and  $n=CH_2$ , nRSR as functional group descriptors have been found to be moderately influential parameters.

### 3.3 FA-MLR and PCRA

Table 2 shows the five-factor loadings of the variables (after VARIMAX rotation). As it is observed, about 73% of variances in the original data matrix could be explained by the selected four factors. Based on the procedure explained in the experimental section the following three-parametric equation was derived.

$$\text{Logp}_{o/w} = 5.378 (\pm 1.70) - 2.80 (\pm 0.43) \text{Ms} + 3.308 (\pm 1.02) \text{SPH} + 0.42 (\pm 0.20) \text{nR05}$$

$$R^2 = 0.77 \quad S.E = 0.35 \quad F = 25.78 \quad Q^2 = 0.71 \\ \text{RMScv} = 0.34 \quad N = 32 \quad (E_1)$$

Equation 1 could explain 77% of the variance and predict 71% of the variance in  $\text{logp}_{o/w}$  data. This equation describes the effect of constitutional (Ms and nR05) and geometrical (SPH) indices on partition coefficient.

When factor scores were used as the predictor parameters in a multiple regression equation using forward selection method (PCRA), the following equation was obtained:

$$\text{Logp}_{o/w} = 0.967 (\pm 0.07) + 0.36 (\pm 0.07) f_2 + 0.27 (\pm 0.07) f_5 - 0.24 (\pm 0.07) f_3 + 0.19 (\pm 0.07) f_1$$

$$R^2 = 0.83 \quad S.E = 0.19 \quad F = 14.74 \quad Q^2 = 0.76 \\ \text{RMScv} = 0.20 \quad N = 32 \quad (E_2)$$

Equation 2 also shows high equation statistics (83% explained variance and 76% predict variance in  $\log p_{ow}$  data). Since factor scores are used instead of selected descriptors, and any factor-score contains information from different descriptors, loss of information is thus avoided and the quality of PCRA equation is better than those derived from FA-MLR.

As it is observed from Table 4, in the case of each factor, the loading values for some descriptors are much higher than those of the others. These high values for each factor indicate that this factor contains higher information about some descriptors. It should be noted that all factors have information from all descriptors but the contributions of descriptors in different factors

are not equal. For example, factors 1 and 2 have higher loadings for constitutional and geometrical indices, whereas information about topological, functional group and constitutional descriptors is highly incorporated in factors 3, 4 and 5. Therefore, from the factor scores used by equation  $E_2$ , the significance of the original variables for modelling the activity can be obtained. Factor score 1 indicates the importance of SPH and FDI (geometrical). Factor score 2 indicates the importance of Ms, RBN and nH (constitutional descriptors). Factor score 3 signify the importance of PW2, PW3 and SIC2 (topological descriptors). Factor scores 4 and 5 signify the importance of nR05, nO, nX and nRSR (constitutional and functional descriptors).

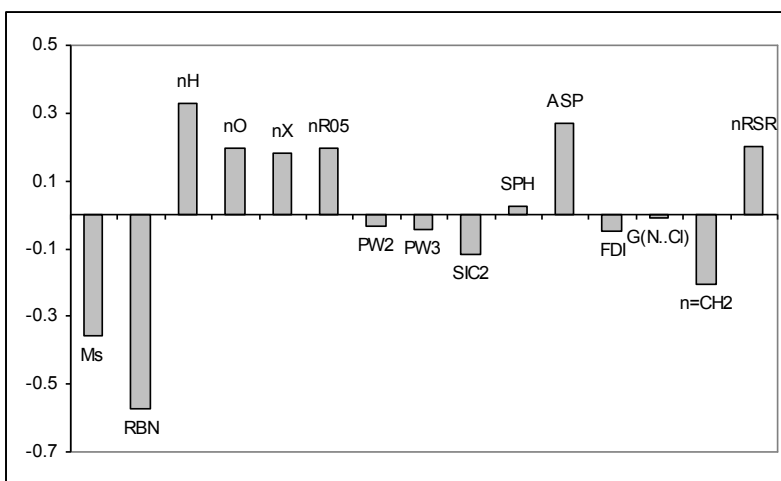


Fig. 1. PLS regression coefficients for the variables used in GA-PLS model

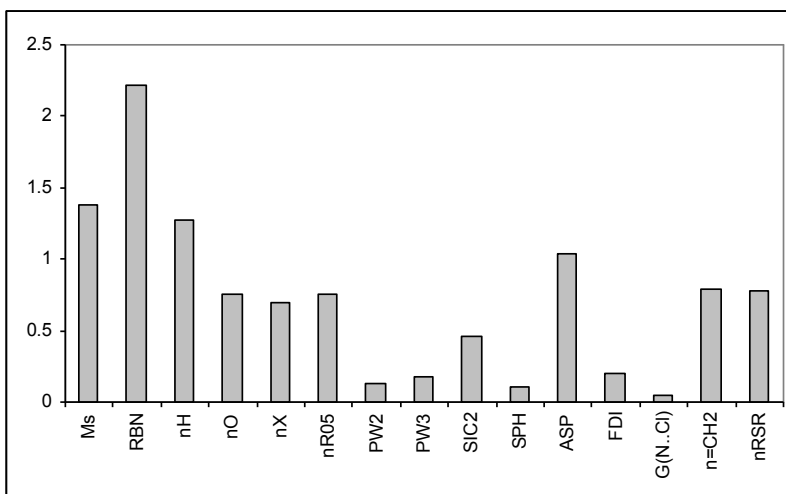
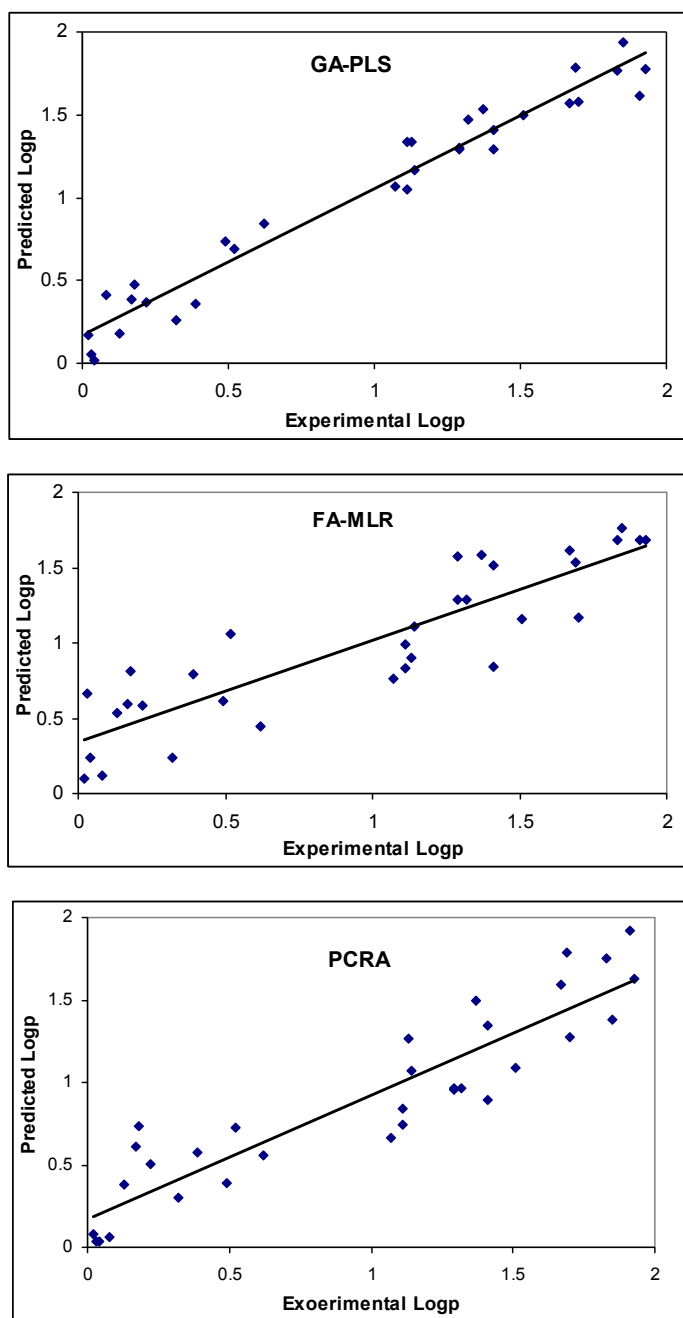


Fig. 2. Plot of variables importance in projection (VIP) for the descriptors used in GA-PLS model

Comparison between the results obtained by GA-PLS and the other employed regression methods indicates higher accuracy of this method in describing partition coefficient of the studied compounds. The difference in the accuracy of the different regression methods used in this study is visualized in Fig. 3 by plotting the predicted logp (by cross-validation) against the

experimental values. Obviously, all linear models represented scattering of data around a straight line with slope and intercept close to one and zero, respectively. As it is observed, the plot of data resulted by GA-PLS represents the lowest scattering and those obtained by FA-MLR and PCRA have lower accuracy.



**Fig. 3. Plots of the cross-validated predicted activity against the experimental activity for the QSAR models obtained by different chemometrics methods**



**Table 4. Numerical values of factor loading numbers 1–5 for some descriptors after VARIMAX rotation**

	1	2	3	4	5	Commonality
<b>Ms</b>	0.006	-0.804	0.219	0.269	-0.340	0.883
<b>RBN</b>	0.186	0.719	0.112	0.542	-0.067	0.863
<b>nH</b>	0.225	0.907	-0.026	0.289	0.056	0.961
<b>nO</b>	0.354	0.183	0.138	0.835	-0.070	0.881
<b>nX</b>	0.408	0.078	-0.034	-0.225	0.728	0.754
<b>nR05</b>	-0.171	-0.060	0.020	0.758	0.232	0.662
<b>PW2</b>	0.419	0.059	-0.688	0.127	0.060	0.672
<b>PW3</b>	0.292	-0.093	0.734	0.245	-0.070	0.698
<b>SIC2</b>	0.129	0.063	0.839	0.321	0.077	0.833
<b>SPH</b>	0.848	0.266	-0.102	-0.036	0.131	0.819
<b>ASP</b>	0.576	0.183	-0.481	0.043	0.421	0.776
<b>FDI</b>	0.871	0.136	0.226	0.198	0.106	0.879
<b>G(N..CI)</b>	0.233	0.507	-0.067	-0.105	-0.301	0.418
<b>n=CH2</b>	-0.112	-0.103	0.545	-0.366	0.086	0.462
<b>nRSR</b>	0.058	-0.023	0.059	0.372	0.743	0.697
<b>%Variance</b>	27.47	16.62	11.28	9.99	7.67	73.03

#### 4. CONCLUSION

The partition coefficients of the 3-hydroxy pyridine-4-one compounds were determined using shake flask. Among the studied compounds **14** (N'-(4-bromobenzylidene)-3-(3-hydroxy-2-methyl-4-oxopyridin-1(4H-yl)benzohydrazide) possesses the highest partition coefficient. Quantitative relationships between molecular structure and  $\log p_{o/w}$  data of derivatives were discovered by a collection of chemometrics methods including GA-PLS, FA-MLR and PCRA. The results revealed the significant role of constitutional parameters in the partition coefficient of the studied compounds. A comparison between the different statistical methods employed indicated that GA-PLS represented superior results and it could explain and predict 96% and 91% of variances in the  $\log p_{o/w}$  data. As it is observed, the plot of data resulted by GA-PLS represents the lowest scattering, and the impact of constitutional descriptors was the most.

#### CONSENT

It is not applicable.

#### ETHICAL APPROVAL

It is not applicable.

#### COMPETING INTERESTS

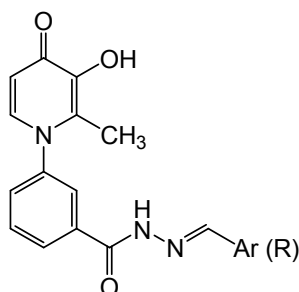
Authors have declared that no competing interests exist.

#### REFERENCES

1. Pliska V, Testa Priv.-Doz B, Van de Waterbeemd H. Lipophilicity in drug action and toxicology. VCH, Weinheim; 2008.
2. Saghaie L, Hider RC, Mostafavi SA. Comparison of automated continuous flow method with shake- flask method in determining partition coefficients of bidentate hydroxypyridinone ligands. Daru 2003;11:38.
3. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Delivery Rev. 1997;23:3
4. Fitzgerald GA. Opinion: Anticipating a change in drug development: The emerging era of translational medicine and therapeutics. Nat. Rev. Drug Discov. 2005; 4:815-818.
5. Fujita T, Iwasa J, Hansch C.  $\rho$ - $\sigma$ - $\pi$  Analysis. A Method for the Correlation of Biological Activity and Chemical Structure. J Am Chem Soc. 1964;86: 5175
6. Putta S, Eksterowicz J, Lemmen C, Stanton RJ. A Novel Subshape Molecular Descriptor. Chem Inf Comput Sci. 2003; 43:1623.
7. Gupta S, Singh M, Madan AK. Superpendentic index: A novel topological descriptor for predicting biological activity. J. Chem Inf Comput Sci. 1999;39:272.
8. Consonni V, Todeschini R, Pavan M. Structure/Response Correlations and

- Similarity/Diversity Analysis by GETAWAY Descriptors. 1. Theory of the Novel 3D Molecular Descriptors. *J Chem Inf Comput Sci.* 2002;42:693.
9. Sabet R, Fassihi A, Moeinifard B. QSAR study of PETT Derivatives as Potent HIV-Reverse Transcriptase Inhibitors. *J. Mol. Graph & Model.* 2009;28:146.
  10. Hemmateenejad B, Sabet R, Fassihi A. QSAR Studies on 2-amino-6-arylsulfonylbenzotriazoles as HIV-1 reverse transcriptase inhibitors using descriptors obtained from substituents and whole molecular structures. *Chem. Biol & Drug. Des.* 2009;74:405.
  11. Damián Palomba, María J. Martínez, Ignacio Ponzoni, Mónica F. Díaz, Gustavo E. Vazquez, Axel J. Soto. QSPR models for predicting log  $P_{\text{liver}}$  values for volatile organic compounds combining statistical methods and domain knowledge *Molecules.* 2012;17:14937.
  12. Gombar V, Enslein K. Assessment of noctanol/water partition coefficient: When is the assessment reliable?. *J Chem Inf Comput Sci.* 1996;36:1127.
  13. Tetko IV, Tanchuk VY, Villa AEP. Prediction of n-octanol/water partition coefficients from PHYSPROP database using artificial neural networks and E-state indices. *J Chem Inf Comput Sci.* 2001;41:1407.
  14. Duprat AF, Huynh T, Dreyfus G. Toward a Principled Methodology for Neural Network Design and Performance Evaluation in QSAR. Application to the Prediction of LogP. *J Chem Inf Comput Sci.* 1998;38:586.
  15. Sabet R, Fassihi A. QSAR study of Isatin analogues as *in vitro* anti-cancer agents. *Eur. J. Med. Chem.* 2010;45:1113.
  16. Sabet R, Behjati M, Vahhabpour R, Memarnejadian A, Rostami M, Fassihi A, Aghasadeghi MR, Saghale L, Miri R. Iron chelation afforded cardioprotection against  $H_2O_2$ -induced H9C2 cell injury: Application of novel 3-hydroxy pyridine-4-one derivatives. *Int. J. Cardiol.* 2012;15: 162(1):60-3.
  17. HyperChem (TM), Hypercube, Inc., 1115 NW 4th Street, Gainesville, Florida 32601, USA.
  18. Todeschini R, Consonni V, Pavan M. Dragon, version 2.1, Milano Chemometrics and QSPR Group., Milano, Italy; 2002.
  19. Sabet R, Fassihi A, Hemmateenejad B, Saghale L, Miri R, Gholami M. Computer-aided drug design of novel antibacterial 3-hydroxypyridine-4-ones: Application of QSAR methods based on the MOLMAP approach. *Journal of Computer-Aided Molecular Design.* 2012;26:349.
  20. Siedlecki W, Sklansky J. On Automatic Feature Selection. *Int J Pattern Recog Artif Intell.* 1988;2:197.
  21. Leardi R. Application of genetic algorithm-PLS for feature selection in spectral data sets. *J. Chemometr.* 2000;14:643.
  22. Leardi R, Gonzalez AL. Genetic algorithm applied to feature selection in PLS regression: how and when to use them. *Chemometr Intell Lab Syst.* 1998;41:195.
  23. Fassihi A, Sabet R. QSAR study of p56<sup>lck</sup> protein tyrosine kinase inhibitory activity of flavonoid derivatives using MLR and GA-PLS. *Int. J. Mol. Sci.* 2008;9:1876.
  24. Leardi R. Genetic Algorithms in Chemometrics and Chemistry: A Review. *J Chemometrics.* 2001;15:559.
  25. Karbakhsh R, Sabet R. Application of different chemometrics tools in QSAR Study of Azolo-adamantanes against influenza A virus. 2011;6:23.
  26. Franke R, Gruska A. Chemometrics methods in molecular design. In *Methods and Principles in Medicinal Chemistry*; van Waterbeemd, H., Ed.; VCH: Weinheim, Germany. 1995;2:113–119.
  27. Fassihi A, Abedi D, Saghale L, Sabet R, Fazeli H, Bostaki Gh, Deilami O, Sadinpour H. Synthesis, antimicrobial evaluation and QSAR study of some 3-hydroxypyridine-4-one and 3-hydroxypyran-4-one derivatives. *QSAR and Combinatorial Science.* 2009;751.
  28. Olah M, Bologa C, Oprea TIJ. *Comput Aided Mol Des.* 2004;18:437.
  29. Erikson L, Johansson E, Kettaneh-Wold N, Wold S. Multi-and megavariable data analysis-principle and application. *Umetrics Academy, Umea, Sweden; 2001.*
  30. Raziheh Sabet, Afshin Fassihi. QSAR Study of antimicrobial 3-Hydroxypyridine-4-one and 3-Hydroxypyran-4-one Derivatives Using Different Chemometric Tools, *Int. J. Mol. Sci.* 2008;9:2407-2423.

## APPENDIX

Table 1.  $K_{part}$  values of the compounds between 1-octanol and tris buffer at pH 7.4. Number of determination = 4.

1-15

Compound	$K_{part}$
1	1.41 ± 0.14
2	1.37 ± 0.01
3	1.67 ± 0.08
4	1.69 ± 0.12
5	1.73 ± 0.09
6	1.13 ± 0.12
7	1.11 ± 0.15
8	1.85 ± 0.19
9	1.14 ± 0.13
10	1.83 ± 0.09
11	1.85 ± 0.18
12	1.29 ± 0.16
13	1.91 ± 0.17
14	1.93 ± 0.15
15	1.45 ± 0.11

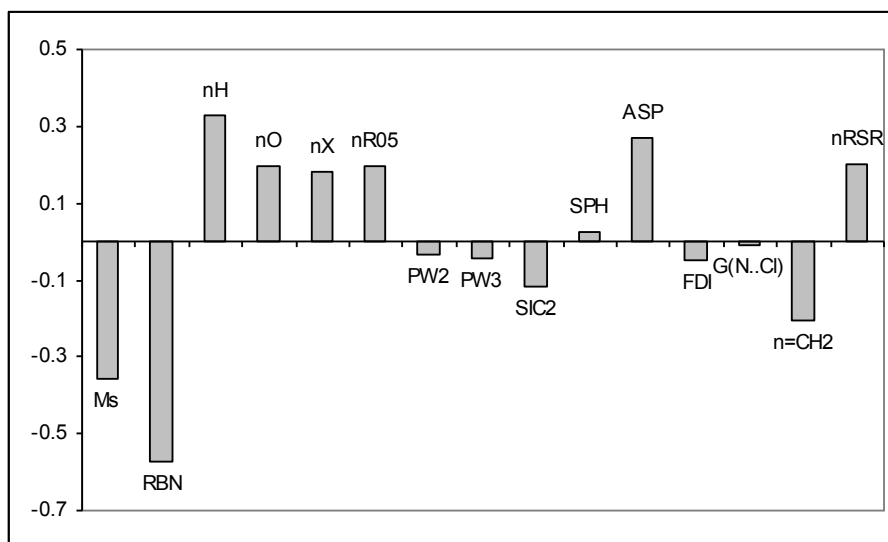
Table 2. Numerical values of factor loading numbers 1–5 for some descriptors after VARIMAX rotation

	1	2	3	4	5	Commonality
Ms	0.006	-0.804	0.219	0.269	-0.340	0.883
RBN	0.186	0.719	0.112	0.542	-0.067	0.863
nH	0.225	0.907	-0.026	0.289	0.056	0.961
nO	0.354	0.183	0.138	0.835	-0.070	0.881
nX	0.408	0.078	-0.034	-0.225	0.728	0.754
nR05	-0.171	-0.060	0.020	0.758	0.232	0.662
PW2	0.419	0.059	-0.688	0.127	0.060	0.672
PW3	0.292	-0.093	0.734	0.245	-0.070	0.698
SIC2	0.129	0.063	0.839	0.321	0.077	0.833
SPH	0.848	0.266	-0.102	-0.036	0.131	0.819
ASP	0.576	0.183	-0.481	0.043	0.421	0.776
FDI	0.871	0.136	0.226	0.198	0.106	0.879
G(N..CI)	0.233	0.507	-0.067	-0.105	-0.301	0.418
n=CH2	-0.112	-0.103	0.545	-0.366	0.086	0.462
nRSR	0.058	-0.023	0.059	0.372	0.743	0.697
%Variance	27.47	16.62	11.28	9.99	7.67	73.03



**Table 4. Brief description of some descriptors used in this study.**

Descriptor Type	Molecular Description
Constitutional	Mean atomic van der Waals volume (Mv) (scaled on Carbon atom), no. of heteroatoms, no. of multiple bonds (nBM), no. of rings, no. of circuits, no of H-bond donors, no of H-bond acceptors, no. of Nitrogen atoms (nN), chemical composition, sum of Kier-Hall electrotopological states (Ss), mean atomic polarizability (Mp), number of rotatable bonds (RBN), mean atomic Sanderson electronegativity (Me), etc.
Topological	Narumi harmonic topological index (HNar), Total structure connectivity index (Xt), information content index (IC), mean information content on the distance degree equality (IDDE), total walk count, path/walk-Randic shape indices (PW3, PW4, PW5, Zagreb indices, Schultz indices, Balaban J index (such as MSD) Wiener indices, Information content index (neighborhood symmetry of 2-order) (IC2), Ratio of multiple path count to path counts (PCR), Lovasz-Pelikan index (leading eigenvalue) (LP1), total information content index (neighborhood symmetry of 1-order) (TIC1), reciprocal hyper-detour index (Rww), Average connectivity index chi-5 (X5A), piID (conventional bond-order ID number), etc.
Geometrical	3D Petijean shape index (PJI3), Asphericity (ASP), Gravitational index, Balaban index, Wiener index, Length-to-breadth ratio by WHIM (L/Bw), etc.
Functional group	Number of total secondary C(sp3) (nC <sub>s</sub> ), Number of total tertiary carbons (nC <sub>t</sub> ), Number of H-bond acceptor atoms (nHAcc), Number of secondary amides (aliphatic) (nCONHR), Number of unsubstituted aromatic C (nCaH), Number of ethers (aromatic) (nRORPh), Number of ketones (aliphatic) (nCO), Number of tertiary amines (aliphatic) (nNR2), Number of phenols (nOHPh), Number of total primary C(sp3) (nC <sub>p</sub> ), etc.

**Fig. 1. PLS regression coefficients for the variables used in GA-PLS model**

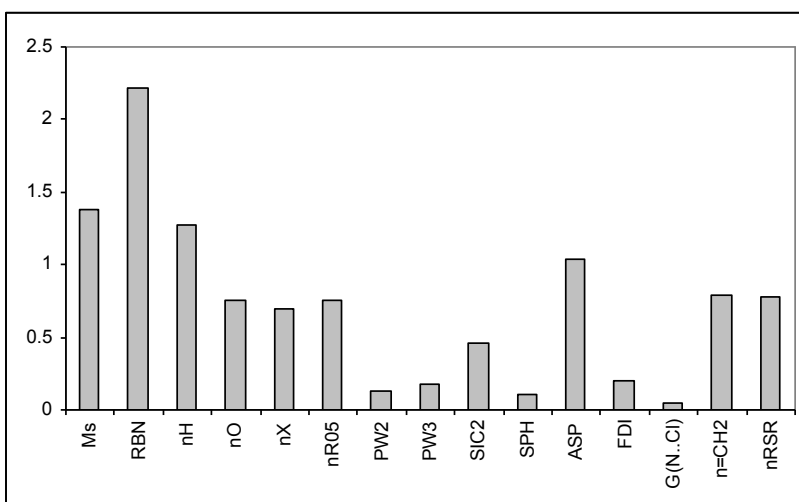
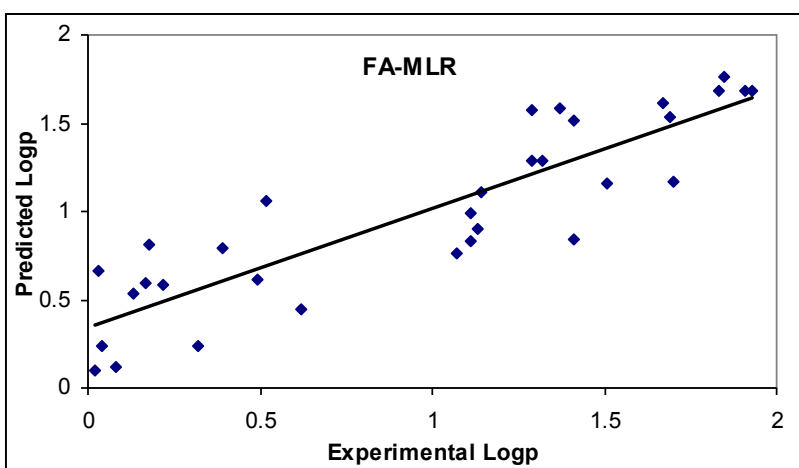
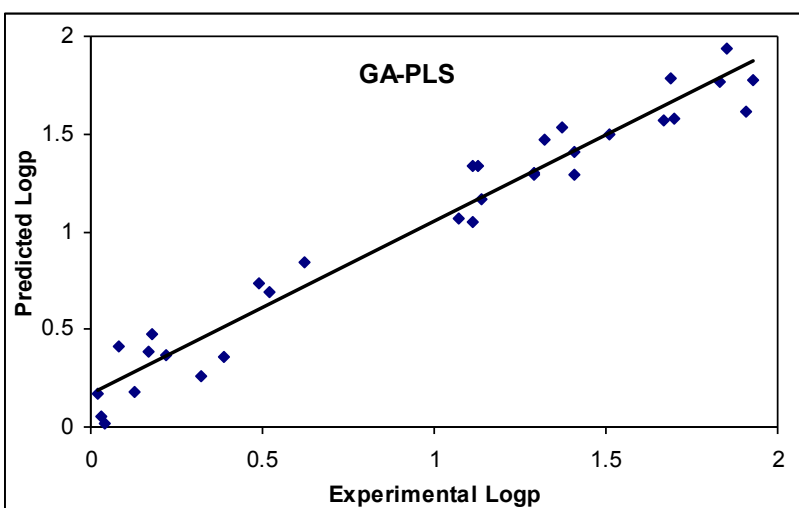
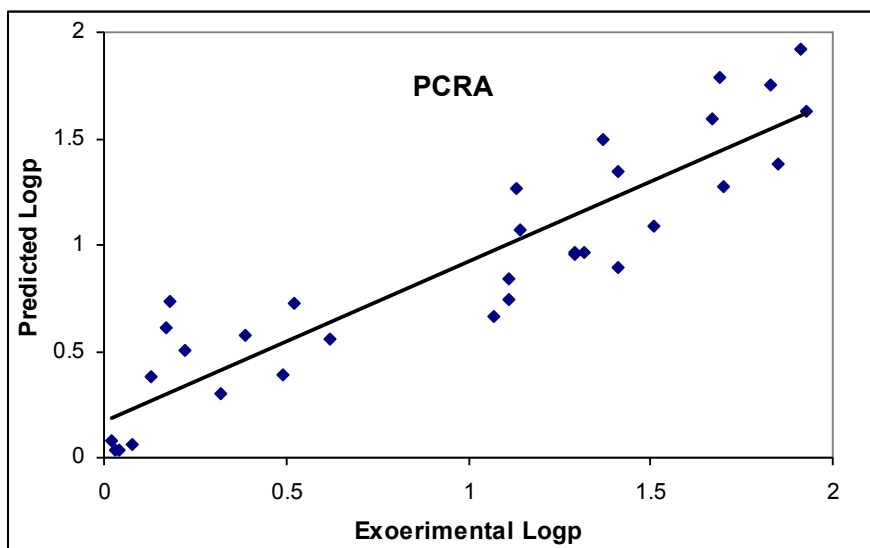


Fig. 2. Plot of variables importance in projection (VIP) for the descriptors used in GA-PLS model





**Fig. 3. Plots of the cross-validated predicted activity against the experimental activity for the QSAR models obtained by different chemometrics methods**

© 2018 Sabet et al.; This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

*Peer-review history:*  
The peer review history for this paper can be accessed here:  
<http://www.sciencedomain.org/review-history/24968>