Available online at www.pelagiaresearchlibrary.com

Pelagia Research Library

Der Chemica Sinica, 2011, 2(6):288-293

Pelagia Research Library
ISSN: 0976-8505
CODEN (USA) CSHIA5

Prediction of 1 Octanol/Water Partition Coefficient of Adamantane derivatives using QSAR method

M. Zanoozi and Z. Bayat

Department of Chemistry, Islamic Azad University-Quchan Branch, Iran

---

ABSTRACT

In this study, a quantitative structure– Activity relationship technique has been used for the simultaneous prediction of 1-Octanol/Water Partition Coefficient of Adamantane derivatives, using a Multivariable Linear Regression (MLR). The best-selected descriptors that appear in the models are the Molecule surface area (SA), Mullikenl Chargeg (MC), Mass(M), solvation Free Energyin Octanol (ΔG_{oct}) .After optimization of the network parameters, the network was trained using a training set. For the evaluation of the predictive power of the generated (MLR), an optimized network was used to predict the 1-Octanol/Water Partition Coefficient of the prediction set. Quantitative structure– Activity relationships (QSARs) have been used to obtain simple models to explain and predict the 1-Octanol/Water Partition Coefficient of Adamantane derivatives .In this report, a MLR was employed to generate a QSAR model between the molecular based structural parameters and observed 1-Octanol/Water Partition Coefficient of Adamantane derivatives.

Keywords: partition coefficient octanol- water (LogPo/w), Adamantane derivatives , QSAR, HF, MLR.

INTRODUCTION

The prediction of physicochemical and biological data of substances through the application of Quantitative Structure Property-Activity Relationships Theory (QSPR-QSAR) has acquired an increasing importance in the last decades. This is specially so when the experimental values of an endpoint can not be determined in the laboratory due to several circumstances, such as economical reasons or simply because the measurements demand too much time. The QSPR-QSAR studies are considered to be the most effective computational approaches for the

estimation of different type of properties [1-3].Although there is a great number of definitions for molecular descriptors available in the literature, it is well known that a single variable is unable to carry all the information on molecular structure, and this leads to the employment of more parameters in the QSPR-QSAR relationship. Nowadays, different standard statistical methods constitute a common practice for the model design, such as Multivariable Linear Regression (MLR) [4], Principal Component Analysis (PCA) [5], Partial Least Squares (PLS) [6], Genetics Algorithms [7–9] or Artificial Neural Networks (ANN) [10]. However, all of these elaborated techniques require the knowledge of a specific functional form of the model (linear or non-linear) and also optimized regression parameters to be present in the equation which, however, may not lead to the best results. QSPR-QSAR studies are usually based on such complex statistical analyzes and sophisticated local and global descriptor definitions.

MATERIALS AND METHODS

The data set was randomly divided into two groups; a training set and a prediction set consisting of 31 and 8 molecules, respectively. The training set was used for the model generation and the prediction set was used for the evaluation of the generated model. The molecules were drawn into the Hyper Chem. (Version 7.0 Hypercube, Alberta, Canada) software. The Gaussian 03 was used for calculating the molecular descriptors. Some of the descriptors are obtained directly from the chemical structure, e. g. constitutional, geometrical, and topological descriptors. Other chemical and physicochemical properties were determined by the chemical structure (lipophilicity, hydrophilicity descriptors, electronic descriptors, energies of interaction). In this work, we used Gaussian 03 for ab initio calculations.The log*P* values calculated in this approach are closer to the experimental values compared to other ab initio methods. 1 Octanol/Water Partition Coefficient are estimated at the Hartree–Fock level with 6-31+G** basis set.

Table 1. The calculated descriptors used in this study

Descriptors	Symbol	Abbreviation	Descriptors	Symbol	Abbreviation
Quantum	Molecular Dipole	MDP	Quantum	difference between	E_{GAP}
	Molecular Polarizability	MP		Hardness	H
	Natural Population	NPA		Softness (S=1/ η)	S
	Electrostatic Potentialc	EP		Electro negativity	X
	Highest Occupied	HOMO		El Electro philicity ($\omega=\chi^2/2$	Ω
	Lowest Unoccupied	LUMO		MullikenlChargeg	MC
chemical descriptors	Partition Coefficient	Log P	chemical descriptors	Molecule surface area	SA
	Mass	M		Hydration Energy	HE
	Molecule volume	V		Refractivity	REF

Due to the diversity of the molecules studied in this work, 90different descriptors were calculated. These parameters encoded different aspects of the molecular structure and consist of electronic, geometric and topological descriptors. Geometric descriptors were calculated using optimized Cartesian coordinates [11,12].Topological descriptors were calculated using two dimensional representation of the molecules. Some of the descriptors generated for each compound encoded similar information about the molecule of interest. Therefore it was desirable

to test each descriptor and eliminate those that show high correlation (R =0.90) with each other. A total of 33 out of 90 descriptors showed high correlation and were removed from the next consideration. Subsequently, the method of stepwise multiple linear regression was used for selection of important descriptors. The descriptors that appear in the best MLR equations for Partition Coefficient of Adamantane derivatives are identical and are shown in Table 1.

RESULTS AND DISCUSSION

Table 2. The experimental Log P values of the Adamantane derivatives training set used in this study and their predicted values by MLR

Compound	logP[exp]	logP [Pred]	REF
Adamantane	2.69	2.51	[13]
1 3 dimethyl adamantane	3.56	3.51	[13]
1,3,5 -trim ethyl adamantane	3.99	3.92	[13]
1-adamantanol	2.66	1.58	[14]
1-buthyl adamantane	4.31	4.15	[13]
1-ethyl adamantane	3.52	2.96	[13]
1-isopropyl adamantane	3.85	3.36	[13]
1-propyl adamantane	3.92	3.66	[13]
2-buthyl adamantane	4.21	3.74	[13]
2-ethyl adamantane	3.42	3.91	[13]
2-isopropyl adamantane	3.75	3.49	[13]
2-methyl adamantane	3.02	3.96	[13]
2-propyl adamantane	3.81	3.69	[13]
1-bromo adamantane	2.66	3.29	[14]
methyl-(1-adamanthyl) ketone	2.9	3.66	[13]
propyl-(1-adamanthyl) ketone	3.93	3.25	[13]
2-adamantanon	2.31	3.35	[14]
ethyl-(1-adamanthyl)ketone	3.53	2.61	[13]
1-methyladamantane	3.13	3.72	[13]
1- sec butyl adamantane	4.25	2.59	[13]
1-tert-buthyl adamantane	4.29	3.29	[13]
1-amino adamantane	1.11	3.09	[13]
2-amino adamantane	2.44	4.06	[14]
1-carboxcylic acid adamantane	2.36	4.27	[14]
1-aceti acid adamantane	2.29	1.69	[14]
1.3-diacetic acid adamantane	1.89	2.79	[14]
1-adamantanol-3 carboxylic acid	1.12	1.93	[14]
1-adamantyl Ethan amine	3.28	2.78	[14]
3,5 -dimethyl-adamantane 1-amine	3.31	1.80	[13]
2-bromo ethyl adamantane	5.094	1.21	[14]
1-adamantane ethanol	3.227	3.51	[14]

Multiple linear regression analysis provided a useful equation that can be used to predict the log P_o/w of drug based upon these parameters. The best equation obtained for the solubility of the drug compounds

LogP= -0.027 (±0.008) SA2+1.122 (±0.194) MC_1+0.018 (±0.003) M
+0.140(±0.015) ΔG_{oct}+ 0 .759 (±0.515) (HF/6-31+G**)
R^2_{train}=0.880 , F_{train}=109.038 , R^2_{test}=0.848 , F_{test} = 4.35 R^2_{adj}=0.914 , Q^2_{LOO}=0.804

Q^2_{LGO}= 0.86 , N_{train}= 31, N_{test}=8

In the present study, the QSAR model was generated using a training set of 31 molecules (Table 2). The test set of 8 molecules (Table 3) with regularly distributed log P_o/w values was used to assess the predictive ability of the QSAR models produced in the regression.

Table 3. The experimental Log P values of the Adamantane derivatives test set used in this study and their predicted values by MLR

Compound	logP[exp]	logP [Pred]	REF
1-ethyl-3-methyl-adamantane	4.35	2.77	[13]
1.3.5.7.tetra methyl adamantane	4.42	4.86	[13]
1.3 diethyl adamantane	4.35	2.83	[13]
1n-methyl-amino adamantane	1.51	4.04	[13]
1-n-n dimethyl adamantane	1.88	4.88	[13]
2-chloro adamantane	3.865	4.36	[14]
1-chloroadamantane	2.6	2.09	[14]
2-isobuthyl adamantane	4.15	1.65	[13]

the predicted values for $LogP_{o/w}$ for the compounds in the training and test sets using equation $LogP_{o/w}$ were plotted against the experimental $LogP_{o/w}$ values in Figure 1.and the comparison between $LogP_{o/w}$ using prediction and the experimental .A plot of the residual for the predicted values of RI for both the training and test sets against the experimental $LogP_{o/w}$ values are shown in Figure 2. As can be seen the model did not show any proportional and systematic error, because the propagation of the residuals on both sides of zero are random. The real usefulness of QSAR models is not just their ability to reproduce known data, verified by their fitting power (R^2), but is mainly their potential for predictive application. For this reason the model calculations were performed by maximising the explained variance in prediction, verified by the cross-validated correlation coefficient,Q^2.This indicates that the obtained regression model has a good internal and external predictive power.

To derive QSAR models, an appropriate representation of the chemical structure is necessary. For this purpose, descriptors of the structure are commonly used. Also, in order to assess the robustness of the model, the chance correlation test was applied in this study. The dependent variable vector ($LogP_{o/w}$) was randomly shuffled and The new QSAR models (after several repetitions) would be expected to have low R^2 and R values (Table 4). If the opposite happens then an acceptable QSAR model cannot be obtained for the specific modeling method and data.

Figure1. The predicted versus the experimental $LogP_{o/w}$ by MLR.

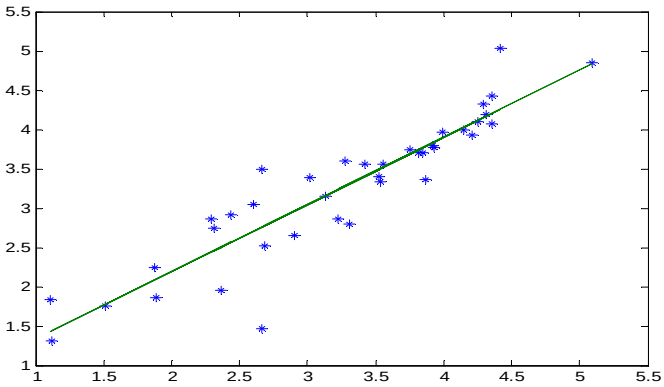
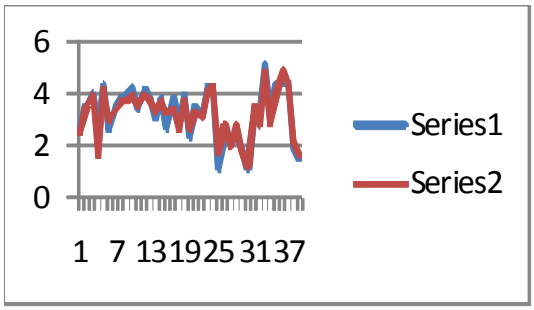


Figure 2. The comparison between properties ($LogP_{o/w}$) using experimental and prediction



Series 1: the values of log P were obtained by using prediction methods
Series 2: the values of log P were obtained by using Experimental methods

Table 4. The R^2 and R values after several chance correlation tests

N	R	R2
1	0.498	0.248
2	0.561	0.315
3	0.478	0.234
4	0.491	0.241
5	0.630	0.397
6	0.532	0.282
7	0.629	0.396
8	0.456	0.207
9	0.478	0.228
10	0.582	0.339

The MLR analysis was employed to derive the QSAR models for different Adamantane derivatives. MLR and correlation analyses were carried out by the statistics software SPSS (Table 5).

Table 5. The correlation coefficient existing between the variables used in different MLR and equations with HF/6-31+G method**

	MC_1	$SAAP_2$	LogP	ΔG_{oct}
MC_1	1	0	0	0
$SAAP_2$	0.239	1	0	0
LogP	0.895	0.163	1	0
ΔG_{oct}	0.203	0.341	0.378	1

CONCLUSION

The results of this study demonstrate that the QSAR method using the MLR techniques can generate a suitable model for the prediction of 1-Octanol/Water Partition Coefficient of Adamantane derivatives. The parameters of Molecule surface area , Mullikenl Chargeg , Mass, solvation Free Energyin Octanol can be considered as comprehensive descriptors for predicting the partition coefficient of Adamantane derivatives.

REFERENCES

[1] C. Hansch and A. Leo, Exploring QSAR. Fundamentals and Applications in Chemistry and Biology (American Chemical Society, Washington D. C., **1995**).
[2] Tropsha A, Gramatica P, Gombar VK. *QSAR Comb Sci* **2003**;22:69–77.
[3] Gramatica P, Pilutti P, Papa E. Validated QSAR prediction of OH tropospheric degradability: splitting into training-test set and consensus modeling. **2004**; 44: 1794–1802
[4] E.R. Malinowski, Factor Analysis in Chemistry (Wiley, New York, 1991).
[5] H. Hotelling, *J. Educ. Psychol*. 24 (**1933**) 417.
[6] S. Wold, M. Sjostrom and L. Eriksson, Encyclopedia of Computational Chemistry (Chichester, **1998**).
[7] D.E. Goldberg, Genetic Algorithms in Search, Optimization and Machine Learning (Addison- Wesley, Massachusetts, **1989**).
[8] R. Wehrens and L.M.C. Buydens, TrAC, *Trends Anal. Chem*. 17 (**1998**) 193.
636 P. R. Duchowicz et al. /Application of a novel ranking approach in QSPR-QSAR
[9] R. Leardi, J. Chemom. 8 (**1994**) 65.
[10] J. Zupan, Encyclopedia of Computational Chemistry (Chichester, **1998**).
[11] E.K. Whalden, P.C. Jurs, *Anal. Chem*. 53 (**1981**) 484.
[12] T.R. Stouch, P.C. Jurs, *J. Chem. Inf. Comput. Sci*. 26 (**1986**)26.
[13] S. V. Kurbatova and S. N. Yashkin, *Journal of Structural Chemistry*, Vol. 41, No. 4, **2000**
[14] K. K. Konstantinova and S. V. Kurbatova, *Russian Journal of Applied Chemistry*, **2006**, 79, 8, 1251-1259, **2006**.