Annals of West University of Timisoara

Series of Chemistry 20**(3)** (2011) 13-18

PRELIMINARY MLR STUDY OF PHOSPHORAMIDATE DERIVATIVES BASED ON DRAGON DESCRIPTOR

Mihaela Petric[a], Luminița Crișan[a], Andreea Micle[a], Manuela Crișan[a], Bianca Maranescu[a], G. Ilia[a,b]

[a]Institute of Chemistry of Romanian Academy, 24 Mihai Viteazul, Timisoara, 300223, ROMÂNIA

[b]West University of Timisoara, Faculty of Chemistry, Biology, Geography, 16 Pestalozzi, Timisoara, 300115, ROMÂNIA

Received: 17 September 2011 Modified: 22 September 2011 Accepted: 29 September 2011

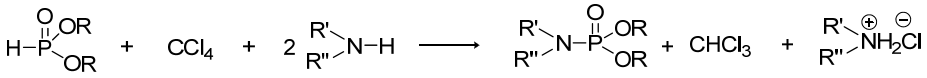
SUMMARY

Phosphoramidates belong to an amide class of phosphoric acid. It was found that these compounds have particular importance most likely in medicine [1], chemistry [2] and agriculture (pesticides, herbicides) [3]. Phosphoramidates are important biologically active compounds, often used as intermediates in organic synthesis [4]. Our goal is to identify theoretical descriptors that can be correlated with the capacity factor (k'). For this reason a large amount of 2D descriptors using the DRAGON software were calculated. The statistical qualities of MLR have been evaluated by several parameters such as: the squares of correlation coefficient (R^2), standard error of estimate (see), Fischer test (F).
Keywords: phosphoramidates; MLR; capacity factor.

INTRODUCTION

The molecular structure for phosphoramidate derivatives is shown in Table I.

Phosphoramidate derivatives are obtained from an amine, used as base, and the system phosphite/tetrahalometane. This method was developed by Atherton and Todd in 1945 [5].



A limited number of models QSPR (Quantitative Structure-Property Relationship) have been developed about this topic [6, 7]. Multiple Linear Regression (MLR) analyses were used in order to find the relationship between molecular descriptors and capacity factor (k') calculated from HPLC data [6].

QSPR models, once constructed and validated, can be used to predict different properties of compounds as yet unmeasured or even unknown [8].

In this paper, our aim is to find one or more equations that accurately predict the experimental property. The statistical qualities of MLR have been evaluated by several parameters such as: the squares of correlation coefficient (R^2>0.850), standard error of estimate (see<0.438), Fischer test (F>51.185).

MATERIALS AND METHODS

In QSPR, molecular descriptors (X-matrix) and response variable (Y-matrix) are correlated [8]. A set of 452 2D-molecular descriptors using the software Dragon 3.0 the web version [9], including parameters of all types such as: Constitutional descriptors, Topological descriptors, Walk and path counts, BCUT descriptors, Galvez topological charge index, 2D autocorrelations, Functional group counts, Atom-centered fragments, Empirical descriptors, Properties was calculated.

Table I. The structure of synthesized compounds

No.	$(R`)_2P(O)$-NH-R"	
	R`	R"
1	C_2H_5O-	-C_6H_5
2	C_6H_5O-	-$C_6H_4NO_2$
3	$C_6H_5CH_2O$-	-$C_6H_4NO_2$
4	C_4H_9O-	-C_6H_5
5	C_6H_5O-	-C_6H_5
6	$C_6H_5CH_2O$-	-C_6H_5
7	$C_8H_{17}O$-	-C_6H_5
8	$C_{10}H_{21}O$-	-C_6H_5
9	$C_8H_{17}O$-	-$C_6H_4NO_2$
10	C_2H_5O-	-$C_6H_4NO_2$
11	C_4H_9O-	-$C_6H_4NO_2$

In order to calculate all these descriptors, molecular structures for phosphoramidates derivatives were constructed with the Symyx Draw 3.3 [10] and were optimized using Molecular Mechanics MM+ in the first step and then AM1 algorithm included in HyperChem package [11].

All calculations were performed by using the statistical software package STATISTICA [9], and the best monoparametric equations are presented.

RESULTS AND DISSCUSIONS

In order to correlate the capacity factor (k') with molecular descriptors, all 452 monoliniar correlations were performed. The monoparametric models with the squares of correlation coefficient, R^2>0.850, indicated the importance of Atom-centered fragments, 2D Autocorrelations, Functional Group Counts descriptor in correlation with the capacity factor (Table II).

Table II. The capacity factor and significant descriptors included in the final equations values

	k'_HPLC	C-002	ATS5e	nH	S3K	H-046	GATS8m	ATS3e
1	1.46	0	66.434	16	4.145	0	0.004	63.172
2	1.72	0	88.591	15	5.172	0	0.003	91.845
3	1.03	0	99.175	19	6.196	0	0.003	103.775
4	1.83	4	90.381	24	6.886	10	0.006	96.567
5	1.86	0	80.65	16	4.622	0	0.004	85.67
6	1.37	0	91.234	20	5.666	0	0.004	97.6
7	3.14	12	157.135	40	13.197	26	0.011	163.284
8	4.2	14	195.945	50	14.885	30	0.011	207.546
9	3.73	12	165.075	39	13.055	26	0.009	169.459
10	1.38	0	74.374	15	4.606	0	0.002	69.347
11	1.36	4	98.321	23	7.158	10	0.004	102.742

The best 7 monoparametric equations developed by MLR method were shown below in descending order of squares of correlation coefficient. The standard errors of regression coefficients are given within parenthesis. The significant molecular descriptors for these seven equations are presented in Table III. The casewise plot of outliers was analyzed, but according to standard deviation values (± 2SD) no compounds should be removed from the data set as outliers.

Equation 1. Dependence of capacity factor from C-002

k' = 1.355(± 0.177) + 0.148 (±0.022) **C-002**

N=11; $\mathbf{R^2 = 0.882}$; see = 0.389; F = 67.239; t (9) = 9.1362

Equation 2. Dependence of capacity factor from ATS5e

k' = -0.489(± 0.023) + 0.355 (±0.03) **ATS5e**

N=11; $\mathbf{R^2 = 0.870}$; see = 0.408; F = 60.281; t (9) = -1.379

Equation 3. Dependence of capacity factor from nH

k' = 0.0192(± 0.083) + 0.297 (±0.011) **nH**

N=11; $\mathbf{R^2 = 0.868}$; see = 0.412; F = 59.001; t (9) = 0.064

Equation 4. Dependence of capacity factor from S3K

k' = 0.129(± 0.253) + 0.291 (±0.034) **S3K**

N=11; $\mathbf{R^2 = 0.863}$; F = 56.481; see = 0.420; t (9) = 0.442

Equation 5. Dependence of capacity factor from H-046

k' = 1.344(± 0.081) + 0.163 (±0.011) **H-046**

N=11; $\mathbf{R^2 = 0.861}$; see = 0.423; F = 55.604; t (9) = 8.253

Equation 6. Dependence of capacity factor from GATS8m

k' = 0.409(± 304.61) + 0.268 (±41.739) **GATS8m**

N=11; $\mathbf{R^2 = 0.855}$; see = 0.432; F = 53.263; t (9) = 1.5404

Equation 7. Dependence of capacity factor from ATS3e

k' = -0.368(± 0.022) + 0.369 (±0.003) **ATS3e**

N=11; $\mathbf{R^2 = 0.850}$; see = 0.438; F = 51.185; t (9) = -0.9978

Table III. Descriptor ID, Descriptor Classes and Descriptor significance

No.	Descriptor ID	Descriptor Classes	Descriptor significance
1	C-002	Atom-centred fragments	CH_2R_2
2	ATS5e	2D autocorrelations	Broto-Moreau autocorrelation of lag 5 (log function) weighted by Sanderson electronegativity
3	nH	Functional group counts	Number of Hydrogen Atoms
4	S3K	Topological indices	3-path Kier alpha-modified shape index
5	H-046	Atom-centred fragments	H attached to C0(sp3) no X attached to next C
6	GATS8m	2D autocorrelations	Geary autocorrelation of lag 8 weighted by mass
7	ATS3e	2D autocorrelations	Broto-Moreau autocorrelation of lag 3 (log function) weighted by Sanderson electronegativity

From the data analyzed Equation 1 shows the best statistical results with the highest value for the squares of correlation coefficient (R^2) and the smallest value for standard error of estimate (see).

In order to check the normality assumption for error terms, in the best equation (1), a normal probability plot of the residuals was constructed.

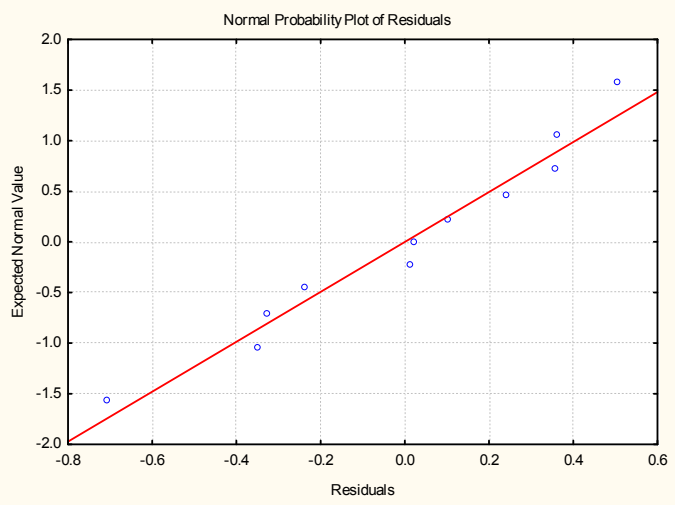


Figure 1. Normal Probability Plot of Residuals for eq.1

For equation 1 a strong linear trend in the plot (Figure 1) is indicative of normality. In addition, the predicted values for capacity factor show the robustness for this equation (Table IV).

Table IV. Predicted values for capacity factor

Compounds	1	2	3	4	5	6	7	8	9	10	11
K' predicted	1.36	1.36	1.36	1.95	1.36	1.36	3.13	3.43	3.13	1.36	1.95

The best prediction to capacity factor was resulted for compound 7 with the smallest difference (0.01).

CONCLUSION

In our QSPR studies, the goal is to find one or more equations that accurately predict the experimental properties. We have developed here seven equations derived from DRAGON descriptors associated with capacity factor.

The statistical performances for MLR analysis are satisfactory. The correlation coefficient (R^2) higher than 0.850, the standard error of estimate, see of approximately 0.4 for the capacity factor domain of 3.17 and Fischer test values higher than 51.185, indicates acceptable statistical performance

Linearity of the normal probability plot indicates the assumption of normal error for selected variable (C-002). The small difference between the predicted and experimental values for capacity factor shows the robustness for this equation.

ACKNOWLEDGMENTS

This project was financially supported by Project 1.2 and Project 2.2 of the Institute of Chemistry Timisoara of the Romanian Academy. We thank to Prof. Mircea Mracec for giving access to the Hyperchem software and to Dr. Simona Funar-Timofei for the access to the STATISTICA software.

REFERENCES

1. Mesmaeker A. D., Haner R., Martin P. , Moser H. E., *Acc. Vhem. Rev*., **28** (1995) 366.
2. Wuts P. G. M., Greene T. V., "*Greene's Protective Groups in Organic Synthesis",* 4th Ed., Wiley, New Jersey, 2006, p. 844.
3. Awad A. M., Sobkowski M., Seliger H., *Nucleotides & Nucleic Acids,* **23** (2004) 777.
4. Adams L.A., Cox R.J., Gibson J.S., Mazo-Martin M.B., Walter M., Whittingham W., *Chem. Commun*, (1998) 2004.
5. Atherton F.R., Openshaw H.T., Todd A. R., *J. Chem. Soc*. (1945) 660.
6. Maranescu B., Ilia G., Cozmiuc C., Glevitzky M., *Rev Chim*, **57** (10) (2006) 1055.
7. Maranescu B., Glevitzky G., Petric M., Cri□an L., *Annals of West University of Timisoara*, **14**(2) (2005) 239.
8. Ghasemi J., Saaidpour S., Brown S.D., *J Mol Struc-Theochem*, **805** (2007) 27.
9. DRAGON, web version, http://www.talete.mi.it
10. www.symyx.com
11. HYPERCHEM 7.52 RELEASE FOR WINDOWS; HYPERCUBE, INC., GAINESVILLE, FLORIDA, USA, http://www.hyper.com
12. STATISTICA 7.1, Tulsa, StatSoft Inc., OK, USA.