

# Alkylphosphate Chromatographic Index Estimation by Computational Methods

S. Funar-Timofei<sup>\*</sup>, L. Kurunczi<sup>\*\*</sup>, G. Iliu<sup>\*</sup>

<sup>\*</sup> Institute of Chemistry of the Romanian Academy, Bd. Mihai Viteazul 24, 300223 Timisoara, Romania

<sup>\*\*</sup> „Victor Babes” University of Medicine and Pharmacy Timisoara, Faculty of Pharmacy, 300034 Timisoara, Romania

Author for correspondence: Simona Funar-Timofei, e-mail: timofei@acad.icht.tm.edu.ro

**Abstract:** The chromatographic retention is based on the interaction between the solute and the stationary phase, which may include dispersion forces, hydrogen bonds, etc. In this study structure-retention relationships were applied to the logarithm of the Kovats chromatographic index of 14 alkylphosphates. 1D and 2D descriptors of phosphate structures were calculated. The influence of the structural parameters on their chromatographic behaviour was studied by MLR (multiple linear regression). Good correlations with descriptors which describe the phosphate dimension and hydrophobicity and, also, predictable models were obtained. The predictability was checked by the crossvalidation and external validation approaches. The proposed MLR equations can be useful for the estimation of chromatographic indexes of new synthesized phosphonates, as a rapid method of structure analysis.

**Keywords:** alkylphosphate, Kovats chromatographic index, MLR, genetic algorithm

## 1. Introduction

Dispersion forces, hydrogen bonds are included in chromatographic interaction forces between the analyzed solute and the stationary phase. It is considered that some intermolecular basic interactions determine the behaviour of chemical compounds both in biological and chromatographic environment.

The n-alkyl esters of the phosphoric acid have been studied mostly on non-polar stationary phases by gas-phase chromatography [1].

Such compounds having the general formula:  $(R'O)_2(RO)PO$ , with  $R'$  = ethyl, propyl, butyl;  $R$  = methyl, ethyl, propyl, butyl, pentyl have been studied on siliconic polar stationary phases [2]. Structural features of the stationary phases and of solutes have been discussed. Influence of the substituents attached to the benzene carbonyl group of some n-alkyl esters of the phosphonic acid on their Kovats chromatographic index have been discussed qualitatively [3].

In this study structure-retention relationships were applied to the logarithm of the Kovats chromatographic index of 14 alkylphosphates (see table 1) in order to estimate the retention indices of new synthesized compounds.

## 2. Methods

### A. Definition of Target Property and Molecular Structures

The logarithm of the Kovats chromatographic index was used as dependent parameter, measured on different stationary phases: OV-225, OV-275, OV-1 [2]. The

computer DRAGON software [4] was used to prepare a set of molecular descriptors ranging from 1D to 2D descriptors. Constitutional and functional group descriptors were used. Hydrophobicity (ClogP) parameter was calculated by the ClogP software [5] and by the ALOGPS 2.1 program [6a, 6b]. Molecular volumes were computed by the Molinspiration ( $V_{molins}$ ) [7] and SPARC ( $V_{SPARC}$ ) [8] softwares, based on group contribution. In addition, the number of  $CH_2$  groups ( $n_{CH_2}$ ) included in the alkyl moieties was considered, too.

### B. Multiple Linear Regression (MLR)

Multiple Linear Regression [9] relates one experimental variable  $y_k$  to one or several structural variables  $x_i$  by the equation:

$$y_k = b_0 + \sum_i b_i \cdot x_{ik} + e_k \quad (1)$$

where  $b$  represents regression coefficients and  $e$  the deviations and residuals. The leave-one-out cross validation procedure [10] was used to test the reliability of our results. MLR calculations were performed by the STATISTICA package [11]. Structural descriptors used in the final MLR analysis are presented in Table 1.

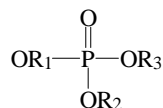
## 3. Results and discussion

Starting from the total set of potential descriptors the MLR analysis has been applied to model the retention of the n-alkyl phosphates. Variable selection was carried out by the genetic algorithm included in the MobyDigs program [12], using the RQK function [13], as fitness function. Good and predictable MLR equations were

obtained (see Table 2). Results of external validation for the OV-1 stationary phase are presented in Table 3.

A very simple descriptor, like the number of CH<sub>2</sub> groups included in the alkylphosphate structure, was considered at first glance as important for the prediction of new chromatographic retention indices.

TABLE 1. The logarithm of the Kovats chromatographic index (*logIOV-1*, *logIOV-275*, *logIOV-225*), the hydrophobicity *IAllogP* parameter [6a] and the SPARC molecular volume (*V<sub>SPARC</sub>*)



No	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	logIOV-1	logIOV-275	logIOV-225	IAllogP	V <sub>SPARC</sub>
1	C <sub>2</sub> H <sub>5</sub>	C <sub>2</sub> H <sub>5</sub>	CH <sub>3</sub>	3.019116	3.304059	3.1917	0.66	158.4
2	C <sub>2</sub> H <sub>5</sub>	C <sub>2</sub> H <sub>5</sub>	C <sub>2</sub> H <sub>5</sub>	3.041393	3.303412	3.203	0.96	170.9
3	C <sub>2</sub> H <sub>5</sub>	C <sub>2</sub> H <sub>5</sub>	C <sub>3</sub> H <sub>7</sub>	3.075547	3.312389	3.2247	1.44	189.5
4	C <sub>2</sub> H <sub>5</sub>	C <sub>2</sub> H <sub>5</sub>	C <sub>4</sub> H <sub>9</sub>	3.106531	3.324899	3.2479	1.91	205.3
5	C <sub>2</sub> H <sub>5</sub>	C <sub>2</sub> H <sub>5</sub>	C <sub>5</sub> H <sub>11</sub>	3.136086	3.337459	3.2718	2.37	221.8
6	C <sub>3</sub> H <sub>7</sub>	C <sub>3</sub> H <sub>7</sub>	CH <sub>3</sub>	3.081707	3.323252	3.234	1.51	190
7	C <sub>3</sub> H <sub>7</sub>	C <sub>3</sub> H <sub>7</sub>	C <sub>2</sub> H <sub>5</sub>	3.102777	3.321184	3.2417	1.82	205.1
8	C <sub>3</sub> H <sub>7</sub>	C <sub>3</sub> H <sub>7</sub>	C <sub>3</sub> H <sub>7</sub>	3.130977	3.329805	3.2629	2.23	220.7
9	C <sub>3</sub> H <sub>7</sub>	C <sub>3</sub> H <sub>7</sub>	C <sub>4</sub> H <sub>9</sub>	3.158965	3.341435	3.2828	2.69	236.3
10	C <sub>3</sub> H <sub>7</sub>	C <sub>3</sub> H <sub>7</sub>	C <sub>5</sub> H <sub>11</sub>	3.185259	3.352954	3.304	3.24	252.8
11	C <sub>4</sub> H <sub>9</sub>	C <sub>4</sub> H <sub>9</sub>	CH <sub>3</sub>	3.144885	3.345766	3.281	2.41	221.7
12	C <sub>4</sub> H <sub>9</sub>	C <sub>4</sub> H <sub>9</sub>	C <sub>2</sub> H <sub>5</sub>	3.160769	3.344589	3.288	2.76	236.6
13	C <sub>4</sub> H <sub>9</sub>	C <sub>4</sub> H <sub>9</sub>	C <sub>3</sub> H <sub>7</sub>	3.184975	3.350636	3.3047	3.17	252
14	C <sub>4</sub> H <sub>9</sub>	C <sub>4</sub> H <sub>9</sub>	C <sub>4</sub> H <sub>9</sub>	3.209247	3.359835	3.322	3.58	267.6

MLR models chosen by genetic algorithm included hydrophobicity and molecular volume descriptors. The statistical results of these models were better in comparison to the ones with nCH<sub>2</sub> groups, as descriptor. Better statistical results and predictable models were noticed in MLR models including hydrophobicity and molecular volume descriptors, in comparison to those obtained in models including the number of CH<sub>2</sub> groups. These predicted values of chromatographic retention indices, when compared to the experimental ones, can give rapid valuable information on the structure of new synthesized compounds, as a rapid method of structure analysis.

Hydrophobic n-alkyl-phosphates - chromatographic substrate interactions would be expected. They can be explained by the presence of alkyl moieties included in both solute and substrate molecules. Intermolecular induction and dispersion interactions and hydrogen bonding between the injected compound and the medium of the gas chromatographic column were mentioned in the literature [14]. Increase of molecular dimension of solutes increases the retention index.

TABLE 2. MLR results (*r*<sup>2</sup> - squared correlation coefficient, *SE* - standard error of estimates, *q*<sup>2</sup> - leave-one out cross validated correlation coefficients, *q*<sub>boot</sub><sup>2</sup> - bootstrap parameter, *r*<sub>adj</sub><sup>2</sup> - adjusted *r*<sup>2</sup>, *F* - Fischer test

No	Equation	r <sup>2</sup>	SE	q <sup>2</sup>	q <sub>boot</sub> <sup>2</sup>	r <sub>adj</sub> <sup>2</sup>	F
1	logIOV-1 = 2.98 (± 0.003) + 0.06(± 0.002)IAllogP	0.995	0.004	0.993	0.993	0.995	2477.84
2	logIOV-1 = 2.75 (± 0.008) + 0.002(±0.00004)V <sub>SPARC</sub>	0.965	0.004	0.993	0.993	0.994	2277.30
3	logIOV-1 = 2.97 (± 0.004) + 0.03(±0.0007) n <sub>CH2</sub>	0.992	0.005	0.989	0.989	0.991	1489.30
4	logIOV-225 = 3.16 (± 0.003) + 0.05(±0.001) IAllogP	0.992	0.004	0.989	0.990	0.991	1432.24
5	logIOV-225 = 3.15 (± 0.005) + 0.02(±0.0007) n <sub>CH2</sub>	0.981	0.006	0.976	0.977	0.979	610.92
6	logIOV-275 = 3.29 (± 0.003) + 0.02(±0.001) IAllogP	0.957	0.004	0.944	0.949	0.954	268.30
7	logIOV-275 = 3.28 (± 0.004) + 0.09(±0.0007) n <sub>CH2</sub>	0.932	0.005	0.911	0.920	0.926	163.50

TABLE 3. MLR results (*r*<sup>2</sup> - squared correlation coefficient, *SE* - standard error of estimates, *q*<sup>2</sup> - leave-one out cross validated correlation coefficients, *q*<sub>boot</sub><sup>2</sup> - bootstrap parameter, *q*<sub>ext</sub><sup>2</sup> - external validation coefficient obtained from models having *logIOV-1* as dependent parameter

No	Descriptor	Test comp.	r <sup>2</sup>	SE	q <sup>2</sup>	q <sub>boot</sub> <sup>2</sup>	q <sub>ext</sub> <sup>2</sup>	Obs.
1	IAllogP	4, 6	0.995	0.004	0.992	0.992	0.998	
2	V <sub>SPARC</sub>	4, 6	0.995	0.004	0.993	0.993	0.991	Compound no.11-outlier
3	IAllogP	2, 10	0.995	0.003	0.991	0.992	0.994	Compound no.1-outlier
4	V <sub>SPARC</sub>	2, 10	0.994	0.004	0.990	0.991	0.998	
5	IAllogP	7, 8	0.995	0.005	0.993	0.993	0.991	
6	V <sub>SPARC</sub>	7, 8	0.996	0.004	0.994	0.994	0.925	
7	IAllogP	1, 10	0.996	0.003	0.993	0.993	0.993	
8	V <sub>SPARC</sub>	1, 10	0.993	0.004	0.989	0.989	0.998	
9	IAllogP	5, 13	0.995	0.004	0.991	0.991	0.999	
10	V <sub>SPARC</sub>	5, 13	0.994	0.005	0.992	0.992	0.998	
11	IAllogP	5, 9	0.995	0.004	0.993	0.993	0.993	
12	V <sub>SPARC</sub>	5, 9	0.995	0.005	0.993	0.993	0.996	

## 4. Conclusions

Chromatographic behaviour of n-alkyl esters of the phosphoric acid of some n-alkyl esters of the phosphonic acid were studied by multiple linear regression. Models obtained by genetic algorithm included hydrophobic and molecular volume descriptors.

Hydrophobicity and molecular dimension of n-alkyl esters of the phosphoric acid increase the chromatographic retention index.

Dispersion and hydrophobic solute-stationary phase interactions are characteristic for the studied n-alkyl phosphates.

The proposed MLR models can be used for retention index prediction, as a rapid method of structure analysis.

#### ACKNOWLEDGEMENTS

This project was financially supported by Ministerul Educatiei, Cercetarii si Tineretului - Autoritatea Nationala pentru Cercetare Stiintifica (MEdC-ANCS), grant contract number: 71GR/2006.

#### REFERENCES

1. Crank, G., Haken, J.K., Gas chromatography of homologous esters XVII. Organophosphorus esters, *J. Chromatog.* **1982**, Vol. 245, pp. 346-349.
2. Soimu, P., Petric, M., Iliu G., Influenta continutului in grupe cian a fazelor stationare soliconice asupra comportarii gaz cromatografice a n-alkil esterilor acidului fosforic, *Rev. Chim. (Bucuresti)*, **2003**, Vol. 54(12), pp. 978-980.
3. Soimu, P., Rotaru, D., Macarie, L., Iliu, G., Consideratii privind comportarea gaz-cromatografica a esterilor n-alkil ai acizilor aroilfosfonici pe faze stationare siliconice, *Rev. Chim. (Bucuresti)*, **2002**, Vol. 53(2), pp. 179-181.
4. Dragon 2.1-**2002**, Talet SRL, Milano, Italy.
5. ClogP for windows software, v. 1.0.0, BioByte Corp., USA.
- 6a. Tetko I. V. and Tanchuk V. Y., Application of associative neural networks for prediction of lipophilicity in ALOGPS 2.1 program, *J. Chem. Inf. Comput. Sci.*, **2002**, Vol. 42, pp. 1136-1145; b. Tetko I. V., Tanchuk V. Y. and Villa A. E. P., Prediction of n-octanol/water partition coefficients from PHYSPROP database using artificial neural networks and E-state indices, *J. Chem. Inf. Comput. Sci.*, **2001**, Vol. 41, 1407-1421.
7. Molinspiration, <http://www.molinspiration.com>
8. SPARC v. 4.2, <http://ibmlc2.chem.uga.edu/sparc/>
9. Wold S. and Dunn III W.J., Multivariate quantitative structure-activity relationships (QSAR): conditions for their applicability, *J. Chem. Inf. Comput. Sci.* **1983**, Vol. 23, pp. 6-13.
10. Cramer, R.D., Bunce, J.D., Patterson, D.E. and Frank, I.E., Multivariate quantitative structure-activity relationships (QSAR): conditions for their applicability, *Quant. Struct.-Act. Relat.*, **1988**, Vol. 7, pp. 18-25.
11. STATISTICA 7.1, Tulsa, StatSoft Inc, OK, USA
12. Todeschini R., Consonni V., Mauri, A., Pavan, M., MobyDigs: software for regression and classification models by genetic algorithms in 'Nature-inspired Methods in Chemometrics: Genetic Algorithms and Artificial Neural Networks' (R. Leardi Ed.), Chapter 5, Elsevier, **2004**, pp. 141-167.
13. Todeschini, R., Consonni, V., Mauri, A., Pavan, M., Detecting 'bad' regression models: multicriteria fitness functions in regression analysis, *Anal. Chim. Acta*, **2004**, Vol. 515, pp. 199-208.
14. Katritzky, A.R., Mati Karelson, M., Lobanov, V.S., QSPR as a means of predicting and understanding chemical and physical properties in terms of structure, *Pure & Appl. Chem.*, **1997**, Vol. 69, pp. 245-248.