ACADEMIA ROMÂNĂ
Revue Roumaine de Chimie
http://web.icf.ro/rrch/

Rev. Roum. Chim.,
2011, *56*(4), 381-389

Dedicated to Professor Alexandru T. Balaban on the occasion of his 80th anniversary

MODELING OF FOOD PRESERVATIVES CHROMATOGRAPHIC LIPOPHILICITY APPLYING GENETIC ALGORITHM AND MULTIPLE LINEAR REGRESSION

Dorina CASONI and Costel SÂRBU*

Babeş-Bolyai University, Faculty of Chemistry and Chemical Engineering, Arany Janos Str. No 11, RO-400028, Cluj-Napoca, Roumania

Received February 28, 2010

Quantitative structure-property relationships (QSPR) models were developed for the prediction of food preservatives lipophilicity based on a wide set of theoretical molecular descriptors and a set of published experimental chromatographic data. The lipophilicity data have been modeled using linear multivariate regression (MLR) methodology and genetic algorithm (GA) procedure for variable subset selection. The molecular descriptors employed in this study, computed using the DRAGON package, combine two (2D) and three-dimensional (3D) aspects of the molecular structure. The best models revealed some insights into 2D and 3D structural features, the results showing that the use of 3D descriptors did not significantly improve the predictability of preservatives lipophilicity. According to our findings, molecular descriptors including information indices, topological, 3D-MoRSE and WHIM descriptors provide a good prediction of preservatives lipophilicity. The most important descriptors, highly significant in the predictive lipophilicity models of preservatives, were related to the atomic polarizabilities, atomic Sanderson electronegativities and atomic van der Waals volume of the molecules.

INTRODUCTION

Understanding and predicting the effects of chemicals have become two major problems faced by chemists involved in the development of industrial chemicals, as well as by scientists studying the toxicology of natural and xenobiotic products. The development of efficient and inexpensive technologies for testing and predicting the physical, chemical and biological properties of new compounds, which would enable the estimation of the potential dangers of old compounds and allow effective risk assessment, is thus of major significance. Quantitative structure–activity and structure-property relationships (QSARs, QSPRs) have been used over the years to develop models to estimate and predict toxicity by relating it to chemical structures.[1,2] These models are particularly useful for screening chemical databases and virtual libraries before the synthesis of chemicals, for setting testing priorities, for reducing reliance on animal testing and, in conclusion, for the timely assessment of the health risks of chemicals.[3] Over the past decade, the quantitative structure-retention/property relationships (QSRR/QSPR) have also become a powerful theoretical tool for description and prediction of molecular systems in chromatographic research.[4-6] It is widely recognized that QSPR equations, derived in a purely empirical fashion from an arbitrary set of descriptors, can give considerable insight into the manner by which chemical structure controls physico-chemical and biological properties of compounds. Once a meaningful model is derived for a given chromatographic system using a representative training set of compounds, the retention and respectively property/activity for new candidates can be

* Corresponding author: csarbu@chem.ubbcluj.ro

predicted based on the properties included in the model and no additional experiments are needed. Nowadays, one of the major goals for the physico-chemical screening of chemicals is the prediction of human absorption, for example, the transport of a molecule through cellular membranes. There are several routes for a molecule to become absorbed, and the most frequent one is the passive transport through the gut wall, which strongly depends on the its solubility in lipids (lipophilicity).[7] This particular property referred to in this paper, defined as tendency of a chemical compound to distribute between an immiscible non-polar organic solvent and water, is one of the molecular parameters most frequently used in QSAR/QSPR studies.[8-12] It plays an important role in several ADME (absorption, distribution, metabolism and elimination) aspects, as well as in the pharmacodynamic and toxicological profile of drugs.[13] The role and importance of the concept of lipophilicity in the research of pharmacological and toxicological study of different compounds or degradation products and the possibility to explore a broader lipophilicity range has been recognized since many years.[14] The development of a model for predicting lipophilicity needs a test system able to provide reproducible and quantitative estimates values, the most widely studies using chromatographic method (RP-HPLC) for lipophilicity assessment of different classes of compounds.[15-19]

Despite the benefits attributed to food additives, for several years there have also been a number of concerns regarding the potential short- and long-term risks of consuming these substances. Critics of additives are concerned with both indirect and direct impacts of using such kind of compounds. As for many of the benefits, there is not always adequate scientific proof of whether or not a particular additive is safe. Toxicological problems resulting from the long-term consumption of additives are not well documented. Cancer and reproductive problems are of primary concern, although there is no direct evidence linking additive consumption with their occurrence in humans.[20] Little or no data are available concerning the health risks or joint effects of the additive. Among many additives, preservatives have been widely used in food, drug and cosmetic products to prevent their aging and decay.[21] Some of them act as antimicrobial agents, some act as anti-oxidants and some of them can perform both functions. While sorbates and benzoates are specific inhibitors of bacteria, parabens have both antimicrobial and antioxidant activity.[22] Owing to their potency to induce allergic contact dermatitis[23] and potential direct toxicological effects, preservatives have attracted the attention of food safety chemists, toxicologists and regulatory agencies for some decades.[24] So, the QSPR and QSAR studies of the food preservatives have become a necessary tool for food safety and toxicological risk assessment. One of the most relevant objectives of the development of QSPR/QSAR models for the prediction of toxicological property is to obtain knowledge of the lipophilicity of substances that still have not been tested, or for which reliable experimental data are not available. In the above considerations, the objective of the current study was to develop multilinear regression models (MLR) to predict reliable lipophilicity of food preservatives using theoretical descriptors calculated directly from their molecular structures. Such models should allow us to accurately predict the lipophilicity of new compounds that have not yet been used in the model training set but that still belong to the same chemical domain as the training set.

RESULTS AND DISCUSSION

An extensive investigation was carried out for quantitative structure-property (lipophilicity) relationships of some food preservatives by using multiple linear regression analysis (MLR) method. The experimental retention parameters converted into the lipophilicity indices (log k_w, S and φ_0) (Table 1) were used as dependent variables for development of preservatives lipophilicity prediction models.

From a variety of potential models with various combinations of descriptors calculated in Dragon software, the statistically significant two and tri-dimensional MLR models, internally validated by the *leave-one-out* procedure, were generated by using genetic algorithms (GA). The best predictive models (for log k_w, S and φ_0) were chosen by examining the regression statistical parameters Q^2_{loo} (leave-one-out crossvalidation coefficient), R^2_{fit} (determination coefficient) PRESS (predictive error sum of squares) and s (standard error of estimate). The three-dimensional model with the highest predictive ability and related statistical parameters, obtained for the prediction of preservatives lipophilicity, are reported in Table 2.

Table 1

The lipophilicity indices obtained on different chromatographic columns[18]

Nr.	Compound	log k_{ow}	C18			C8			CN		
			log k_w	S	φ_0	log k_w	S	φ_0	log k_w	S	φ_0
1	Sorbic acid	1.33	1.879	-0.030	-62.22	1.935	-0.033	-57.95	0.657	-0.023	-28.57
2	Benzoic acid	1.87	1.633	-0.025	-65.06	2.003	-0.034	-58.41	0.743	0.022	33.76
3	Salicylic acid	2.26	0.942	-0.010	-98.14	2.044	-0.033	-62.33	-	-	-
4	3-Hydroxybenzoic acid	1.20	0.532	-0.008	-70.00	1.461	-0.034	-43.48	0.567	-0.022	-25.77
5	4-Hydroxybenzoic acid	1.58	0.693	-0.013	-52.11	1.237	-0.032	-38.41	0.560	-0.023	-24.35
6	4-Aminobenzoic acid	0.83	0.095	-0.005	-18.55	0.727	-0.030	-24.16	0.393	-0.017	-23.12
7	Methylparaben	1.96	2.106	-0.034	-62.13	1.893	-0.034	-56.18	0.875	-0.025	-34.98
8	Ethylparaben	2.47	2.520	-0.037	-68.11	2.349	-0.037	-62.81	1.102	-0.028	-39.35
9	Propylparaben	3.04	3.228	-0.046	-69.87	2.917	-0.042	-69.46	1.349	-0.031	-43.51
10	Butylparaben	3.57	3.023	-0.036	-83.97	3.494	-0.048	-72.80	1.616	-0.034	-47.54
11	Propyl gallate	1.80	2.049	-0.036	-56.45	2.161	-0.041	-52.45	0.716	-0.023	-31.14
12	Tert-butylhydroquinone	-	2.415	-0.037	-64.57	2.574	-0.042	-62.01	1.100	-0.028	-39.29

Table 2

Multiple regression models and their statistics based on molecular descriptors available in Dragon software

Stationary phase	The multiple regression equations based on 2D descriptors	Q^2_{loo}	$R^2_{fit.}$	PRESS	s
C18	$\log k_w$ = -23.074 + 62.088RBF + 2.247Espm14u – 1.311VRD1	95.74	97.54	0.472	0.184
	$\log k_w$ = -10.155 + 2.242MAXDP – 28.699BICO + 44.270Vev2	93.57	97.42	0.713	0.189
	S = 0.031 – 0.084MATS4v – 0.011ESpm14u + 0.020GGI1	94.03	96.99	0.001	0.003
	S = -2.238 + 0.055J + 0.015ICR – 0.096MATS4p	92.89	96.39	0.001	0.003
C8	$\log k_w$ = 9.527 – 0.262SPI – 5.338ICO + 0.039AEige	97.58	98.97	0.146	0.088
	$\log k_w$ = -0.165 – 4.545ICO + 2.007MATS2p + 3.032BEHm3	97.23	98.55	0.167	0.104
	φ_0 = -98190.940 + 123.678ICO + 15.538ESpm03u – 36.164BEHv4	94.92	97.30	105.523	2.645
	φ_0 = -268.814 – 333.595RBF + 34.923WA + 451.882BICO	93.89	96.90	126.806	2.838
CN	$\log k_w$ = -2.713 -0.159Har2 + 0.143TPC + 1.172CICO	98.69	99.24	0.018	0.039
	$\log k_w$ = 7.419 – 0.975Jhete – 0.257X4 – 8.607SICO	98.11	99.08	0.026	0.043
	φ_0 = -1144.007 – 19.511ATS5m – 169.262EEig03d + 191.445ESpm05x	95.72	97.74	204.962	3.293
	φ_0 = -1143.408 – 19.510ATS4e – 169.179EEig03d + 191.340ESpm05x	95.71	98.42	205.439	4.789
	$\log k_{ow}$ = 4.255 + 0.588PCD – 1.175x4sol – 3.864AAC	97.71	98.84	0.149	0.104
	$\log k_{ow}$ = 5.542 + 0.452PCD – 0.048RDSQ – 15.997 SICO	97.34	98.51	0.173	0.118
	The multiple regression equations based on 3D descriptors				
C18	$\log k_w$ = 1.629 – 0.473RDF010u – 3.068Mor06v + 10.888Mor25p	94.91	97.50	0.564	0.186
	$\log k_w$ = 1.534 – 0.392RDF010e – 2.870Mor06v + 10.139Mor25p	94.48	97.38	0.612	0.190
	S = -0.030 + 0.019RDF015v – 0.009RDF040e + 0.017Mor06u	90.35	95.51	0.001	0.003
	S = -0.029 + 0.064Mor06v – 0.154Mor25v + 0.003Tv	90.34	95.50	0.001	0.003
	φ_0 = -94.206 – 4.512DP18 + 174.067Mor11v + 70.869Mor09e	92.56	96.29	291.864	4.267
	φ_0 = -118.150 + 71.292Mor09e + 211.879Mor11v+ 38.092HATS2u	91.96	95.84	315.223	4.518
C8	$\log k_w$ = 2.928 – 0.399DP09 – 3.103Mor23u – 6.834Mor11p	94.73	97.44	0.317	0.139
	$\log k_w$ = 5.443 + 0.319RGyr – 4.626Mor06p – 6.449Ku	94.63	97.37	0.323	0.144
	φ_0 = -97.427 + 12.494SP08 + 214.560Mor11p + 91.364Mor18p	91.57	95.50	175.165	3.416
	φ_0 = -177.095 + 29.871SP04 + 245.894Mor11p + 114.606Mor18p	90.53	94.98	196.605	3.609
CN	$\log k_w$ = -2.129 – 1.346Mor06u + 3.721HATS1u – 1.917H4p	97.77	99.04	0.054	0.054
	$\log k_w$ = -2.181 – 1.279Mor06u – 1.765E2m + 4.619HATS1u	97.41	98.47	0.054	0.055
	φ_0 = -472.500 + 1.537SEig + 1929.279G1e + 337.121R4m$^+$	98.44	99.27	98.635	2.409
	φ_0 = -431.668 + 0.141Mor01e + 1891.143G1e + 322.395R4m$^+$	97.48	99.29	159.519	2.367
	$\log k_{ow}$ = 21.973 + 0.705Mor7u – 2.905HATS7u – 6.535H0e	96.30	98.11	0.241	0.133
	$\log k_{ow}$ = 23.707 + 0.746Mor07u – 2.883HATS7u – 2.414HATSe	96.09	98.01	0.255	0.136

Most of the regression coefficients are statistically significant and all equations obtained can be acceptable from statistical point of view (goodness of fit and prediction). The use of 3D descriptors did not improve significantly the predictability of preservatives lipophilicity. The best models were examined in order to gain some insights into structural features that are important for preservatives chromatographic behavior. The most important selected descriptors indicate that the following descriptors are highly significant in the predictive lipophilicity models developed in this study: (a) molecular descriptors calculated as information content of molecules, based on the calculation of equivalence classes from the molecular graph, the indices of neighborhood symmetry taking into account also neighbor degree and edge multiplicity (information indices – BICO, CICO, TICO, SICO); (b) molecular descriptors, conformationally independents, obtained from molecular graph (usually H-depleted) (topological descriptors – Har2, Jhete, Wap); (c) molecular descriptors calculated by summing atoms weights viewed by a different angular scattering function (3D-MoRSE descriptors – Mor06p, Mor06u); (d) molecular descriptors obtained as statistical indices of the atoms projected onto the 3 principal components obtained from weighted covariance matrices of the atomic coordinates (WHIM descriptors – Ku, Km, G1e); (e) molecular descriptors obtained by radial basis functions centred on different interatomic distances (RDF descriptors – RDF015v, RDF035v, RDF010e, RDF070e).

The most important descriptors in these models, accounting for two (2D) and three-dimensional (3D) aspects of the molecular structure, can be classified as RDF (Radial Distribution Function), GETAWAY (autocorrelation), 3D-MoRSE signal, Burden Eigenvalues and edge adjacency descriptors. The selected RDF descriptors are related to the atomic van der Waals volumes (v) and atomic Sanderson electronegativities (e). The GETAWAY descriptors are related to the atomic Sanderson electronegativies (e) and atomic van der Waals volumes (v). Also, the use of 2D descriptors (information indices, Burden eigenvalues, 2D autocorrelation and edge adjacency) show that dipole moments (d) and atomic polarizabilities (p) are the most important properties responsible for preservatives lipophilicity estimated by HPLC. The symbol of these descriptors and their meanings can be found in Table 3.

The best models found for log k_w prediction yield a determination coefficient over than 0.8807 in case of 2D descriptors and over than 0.9676 in case of 3D descriptors. The best QSPR equations were obtained in the case of C8 and respectively CN stationary phases (Fig. 1).

The model with the highest predictive ability (Fig. 1e), obtained for the prediction of preservatives lipophilicity, is a three-dimensional model with 2D descriptors (for CN column), the corresponding correlation showing a determination coefficient higher than 0.99 between predicted and experimental log k_w values.

In order to compare the efficiency and applicability of the best founded predictive lipophilicity models, it is imperative that a set of data, not used in the construction of the models, should be used. Unfortunately, our set includes only 12 solutes and it proves very difficult to identify a new set of HPLC data included enough of data set lipophilicity values to be statistically significant (to have the some HPLC conditions). We therefore resorted to the experimental n-octanol-water partition coefficient values founded in Human Metabolome online database (http://www.metabolomics.ca/) for a test set of other 5 compounds (ascorbic acid – log k_{ow} = 1.85; gallic acid – log k_{ow} = 0.70; 4-ethylbenzoic acid – log k_{ow} = 2.89; m-chlorobenzoic acid – log k_{ow} = 2.68; 2-aminobenzoic acid – log k_{ow} = 1.21) with related structures with the training set compounds. The performance of the best models for partition coefficient octanol-water values prediction (Table 2) was initially tested for the training set of compounds. Based on prediction criteria, the best equations for the log k_{ow} values, could predict 98.84 % of variance in case of 2D descriptors and 98.11 % in case of 3D descriptors (Fig. 2, a and b). These models were finally subjected to a externally validation strategy (using the test set compounds) in order to check the reliability for their possible application on a new set of data. External predicted variance is 99.91% in case of 2D descriptors and 79.67 % in case of 3D descriptors (Fig. 2, c and d). The differences between predicted and experimental were less than ± 0.03 in all cases. These results show that the developed models should allow to accurately prediction the octanol-water partition coefficients of new compounds that have not yet been used in the model training set but that still belong to the same chemical domain as the training set.

Table 3

Information about the descriptors in the best model found

Type		Symbol	Meaning
3D descriptors	RDF	RDF015v RDF010e RDF010u	Radial Distribution Function – 1.5 and 3.5/weighted by atomic van der Waals volumes; 1.0; 4.0; 7.0 /weighted by atomic Sanderson electronegativities and 10.0/unweighted
	GETAWAY	R4m^{+} H4p H0e HATS7u HATSe	R maximal autocorrelation of lag 4 /weighted by atomic mass H autocorrelation of lag 4/weighted by atomic polarizabilities H autocorrelation of lag 0 / weighted by atomic Sanderson electronegativities Leverage-weighted autocorrelation of lag 7 /unweighted Leverage-weighted total index / weighted by atomic Sanderson electronegativities
	WHIM	Ku Tv G1e	K global shape index/ unweighted Total size index/ weighted by atomic van der Waals volumes 1st component symmetry directional WHIM index/ weighted by atomic Sanderson electronegativities
	3D – MoRSE	Mor07u Mor06m Mor25p Mor06v	3D-MoRSE signal / unweighted (u); weighted by atomic masses (m)and weighted by atomic van der Waals volumes (v)
2D descriptors	Edge adjacency	ESpm14u ESPm03u ESPm05x	Spectral moment from edge adj. matrix; unweighted (u) Spectral moment 05 from edge adj. matrix weighted by edge degrees
		EEig03d	Eigenvalue from edge adj. matrix weighted by dipole moments (d)
	Burden eigenvalues	BEHv4 BEHm3	Highest eigenvalue n. 4 of Burden matrix weighted by atomic van der Waals volumes Highest eigenvalue n. 3 of Burden matrix / weighted by atomic masses
	2D autocorrelations	MATS4p MATS4v ATS5m ATS4e	Moran autocorrelation lag 4/weighted by atomic polarizabilities Moran autocorrelation - lag 4 / weighted by atomic van der Waals volumes Broto-Moreau autocorrelation lag 5/weighted by atomic mass Broto-Moreau autocorrelation of a topological structure - lag 4 / weighted by atomic Sanderson electronegativities
	Information indices	BICO CICO TICO SICO AAC	Bond information content (neighborhood symmetry of 0-order) Complementary information content (neighborhood symmetry of 0-order) Total information content index (neighborhood symmetry of 0-order) Structural information content (neighborhood symmetry of 0-order) Mean information index on atomic composition
	Walk and path counts	PCD TPC	Difference between multiple path count and path count Total number of paths of any length (from 0 to the maximum path length) in the graph.
	Eigenvalue-based indice	VRD1 Vev2	Randic-type eigenvector-based index from distance matrix average eigenvector coefficient sum from van der Waals weighted distance matrix
	Topological charge indices	GGI1	topological charge index of order 1
Others descriptors	Constitutional descriptors	RBF	Rotable band fraction
	Topological descriptors	Har2 Jhete ICR SPI MAXDP	Square reciprocal distance sum index Balaban-type index from electronegativity weighted distance matrix Radial centric information index Superpendentic index Maximal electrotopological positive variation
	Connectivity indices	X4 RDSQ	Connectivity index chi-4 Reciprocal distance squared Randic-type index

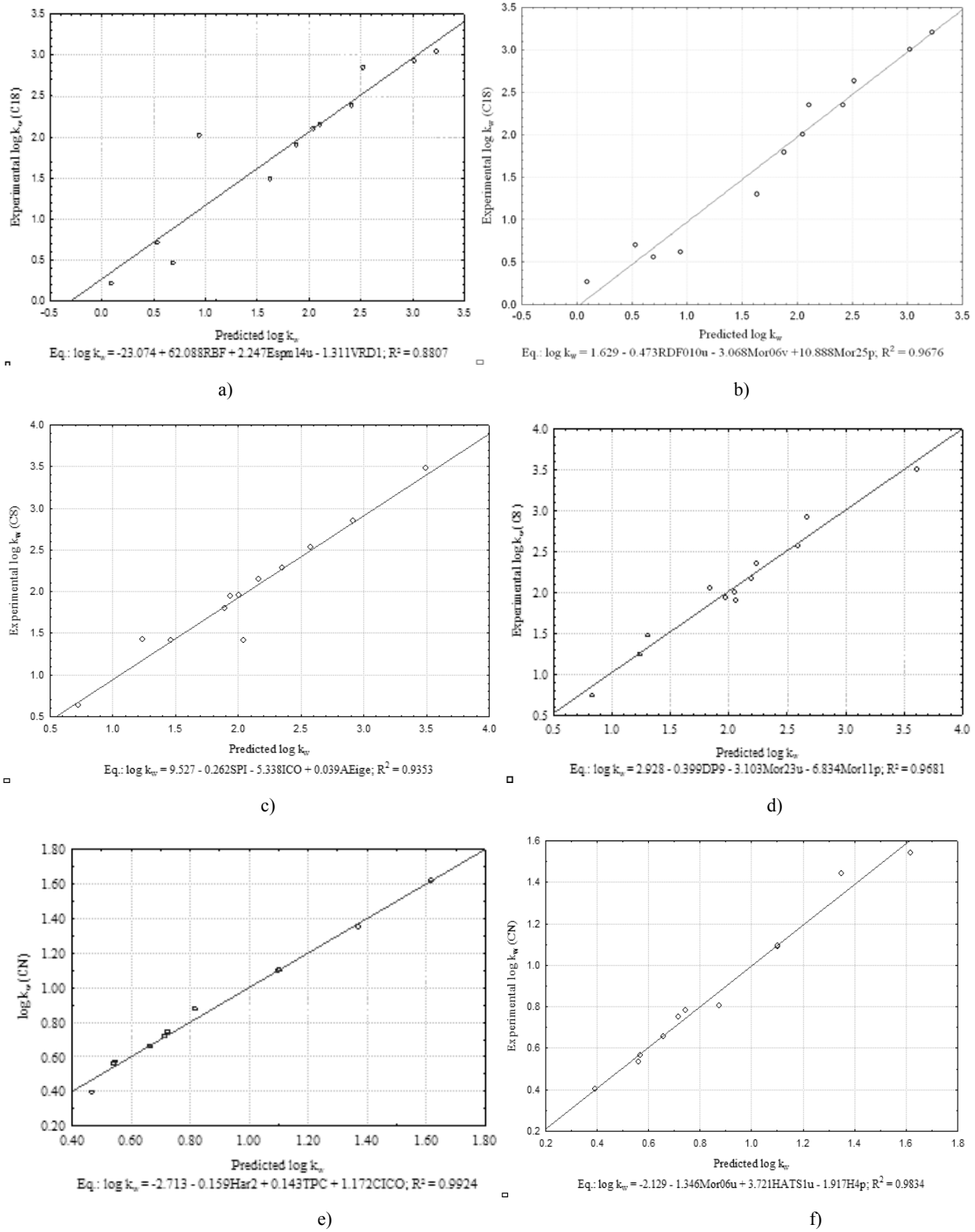

Fig. 1 – The predicted versus experimental lipophilicity indices in the case of 2D and 3D descriptors respectively: C18 (a,b); C8 (c,d); CN (e,f).

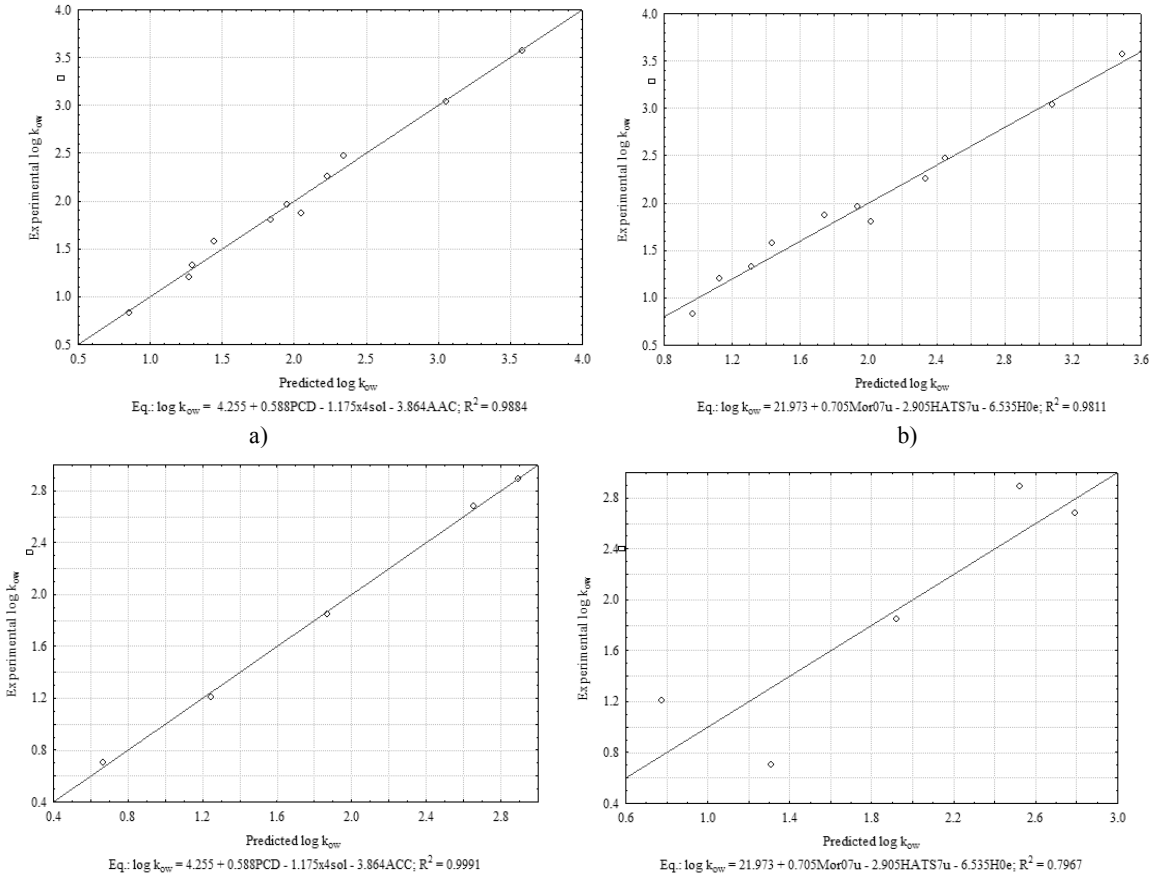

Fig. 2 – The predicted versus experimental octanol-water partition coefficient in the case of 2D and 3D descriptors respectively: for the training set compounds (a, b); for the test set compounds (c, d).

Although experimental properties are often used in the formulation of predictive models, however, theoretical descriptors can be obtained quickly and inexpensively compared to experimentally determined properties. In addition, they can be computed for compounds not yet synthesized and, as such, are useful in the design and hazard assessment of novel compounds.

EXPERIMENTAL

Materials and methods

In the study, we used a dataset of some food preservatives lipophilicity values, experimentally determined by us and already published.[18] Different indices of lipophilicity (log k_w – that refers to the isocratic retention factor values for pure water as mobile phase, obtained by extrapolation; S – related to the solvent strength of pure organic modifier as mobile phase and φ_0 -the isocratic chromatographic hydrophobicity index, φ_0 = log k_w/S) were determined by reversed-phase high-performance liquid chromatography on C18 (LiChroCART, Purosphere RP-18e), C8 (Zorbax, Eclipse XDB-C8) and respectively CN (Sälentechnik, Lichrosphere CN100) columns using methanol–water as mobile phase.

Computational methods - Calculation of molecular descriptors

Molecular descriptors, defined as numerical characteristics associated with chemical structures, are basic molecular properties of a compound. There are many types of molecular descriptors that each type is related to a specific type of interaction between chemical groups in the molecule. There are many software packages[25] dedicated to the computation of molecular descriptors of any desired chemical structure. One of the most widely used is Dragon software.[26] Dragon calculates more than 1600 molecular descriptors for any desired chemical structure. The descriptors employed in this study can be arranged in the following groups: *descriptors 2D*: 2D autocorrelations, edge adjacency, Burden eigenvalues, topological and connectivity indices; *descriptors 3D*: RDF, 3D-MORSE, GETAWAY, WHIM, geometrical properties and Randić molecular profiles; *other descriptors*: functional groups, atom-centered fragments, molecular properties, charge descriptors, and constitutional properties. Definitions and further information regarding all these molecular descriptors can be found in Todeschini and Consonni.[25] Three-dimensional (3D) descriptors require geometry optimization prior to the descriptor calculation. In all cases the structures of preservatives were preoptimized with the Molecular Mechanics Force Field (MM+) procedure included in Hyperchem version 7.5[27] and the resulting geometries were further refined by means of the semi empirical method PM3

(Parametric Method-3) using the Fletcher-Reeves algorithm and a gradient norm limit of 0.009 kcal/Å.

Chemometric Methods

Multiple linear regression analysis (MLR) was performed by the MobyDigs v.1.0 package.[28] As it is impossible to perform multilinear regression when descriptor variables are too many and correlation among them is too high, genetic algorithm procedure (GA)[29] was used to select the most significant variables. Models performance was described by means of statistical parameters related to model predictive capability (leave-one-out crossvalidation coefficient – Q^2_{loo}), fitting power (determination coefficient – R^2_{fit}), predictive error (sum of squares – PRESS) and standard error of estimate (s).

CONCLUSIONS

The chromatographic lipophilicity data for a dataset of food preservatives have been modeled by a wide set of theoretical molecular descriptors using linear multivariate regression and genetic algorithm methodologies. The best models, internally validated by the *leave-one-out* procedure, showed satisfactory predictive performance in all cases (for experimental data obtained on all three chromatographic columns). The best models revealed some insights into 2D and 3D structural features, the results showing that the use of 3D descriptors did not improve predictability of preservatives lipophilicity. The descriptors selected as the best combinations correlated to the different lipophilicity response, are not easily interpretable for an understanding of the complex underlying lipophilicity mechanism. However, the most important descriptors, highly significant in the predictive lipophilicity models of preservatives, were related to the atomic polarizabilities, atomic Sanderson electronegativities and atomic van der Waals volumes of the molecules. The models proposed here can be useful both in retention and octanol-water partition coefficient prediction of preservatives compounds for which experimental data are not available or the compounds are not yet synthesized.

Acknowledgements: Financial support by the Ministry of Education, Research Youth and Sport from Romania (CNCSIS, IDEI 560/2007) is gratefully acknowledged.

REFERENCES

1. J. C. Dearden, *J. Comput. Aid. Molecul. Des*., **2003**, *17*, 119-127.
2. J. Devillers and H. Devillers, *SAR QSAR Environ. Res.,* **2009**, *20*, 467-500.
3. P. Gramatica, V. Consonni and M. Pavan, *SAR QSAR Environ. Res.,* **2003**, *14*, 237-250.
4. R. Kaliszan, *Chem. Rev.,* **2007**, *107*, 3212-3246.
5. L. C. Porto, E. S. Souza, B. S. Junkes, R. A. Yunes and V. E. F. Heinzen, *Talanta,* **2008**, *76,* 407-412.
6. Y. -B. Liou, H. -O. Ho, C. -J. Yang, Y. -K. Lin and M.-T. Sheu, *J. Control. Release*, **2009**, *138*, 260-267.
7. A. Avdeef, *Curr. Topics Med. Chem*., **2001**, *1*, 277-351.
8. H. Chuman, S. Goto, M. Karasawa, M. Sasaki, U. Nagashima, K. Nishimura and T. Fujita, *Quant. Struct.-Act. Relat*., **2000**, *19*, 10-21.
9. K. Valkó, *J. Chromatogr. A,* **2004**, *1037*, 299-310.
10. M. P. Freitas and J. A. Martins, *Talanta*, **2005**, *67*, 182-186.
11. S. Kar and K. Roy, *J. Hazard. Mater.,* **2010**, *177*, 344-351.
12. C. Sârbu, D. Casoni, A. Kot-Wasik, A. Wasik and J. Namieśnik, *J. Sep. Sci.,* **2010**, *33*, 2219-2229.
13. B. Testa, P. Crivori and M. Reist, *Perspect. Drug Discov. Des.,* **2000**, *19*, 179-211.
14. A. Nasal, D. Siluk and R. Kaliszan, *Cur. Med. Chem.,* **2003**, *10*, 381-426.
15. K. Valkó, *J. Chromatogr. A,* **2004**, *1037*, 299-310.
16. X. Liu, H. Tanaka, A. Yamauchi, B. Testa and H. Chuman, *J. Chromatogr. A,* **2005**, *1091*, 51-59.
17. C. Giaginis, S. Theocharis and A. Tsantili-Kakoulidou, *J. Chromatogr. A,* **2007**, *1166*, 116-125.
18. D. Casoni, A. Kot-Wasik, J. Namieśnik and C. Sârbu, *J. Chromatogr. A,* **2009**, *1216,* 2456-2465.
19. R. D. Briciu, A. Kot-Wasik, J. Namieśnik and C. Sârbu, *J. Sep. Sci.,* **2009**, *32,* 2066-2074.
20. T. L. Ditschun and C. K. Winter, "Food Toxicology", W. Helferich and C. K. Winter, Eds., Boca Raton, CRC Press, London, **2001**, p. 187-198.
21. G. J. Nohynek, E. Antignac, T. Re and H. Toutain, *Toxicol. Appl. Pharmacol*., **2010**, *243*, 239-259.
22. M. G. Soni, G. A. Burdock, S. L. Taylor and N. A. Greenberg, *Food Chem. Toxicol*., **2001**, *39*, 513-532.
23. M. G. Soni, S. L. Taylor, N. A. Greenberg and G. A. Burdock, *Food Chem. Toxicol*., **2002**, *40*, 1335-1373.
24. D. V. Parke and D. F. Lewis, *Food. Addit. Contam,* **1992**, *9*, 561-577.
25. R. Todeschini and V. Consonni, "Handbook of Molecular Descriptors", M. Manhold, H. Kubinyi and H. Temmerman, Eds., Wiley-VCH, Weinheim, **2000**, p. 667-673.
26. R. Todeschini, V. Consonni, A. Mauri and D. Pavan, Dragon for Widows (Software for molecular Descriptor Calculation). Version 5.4, **2006**; Available at http://www.talete.mi.it/
27. Hyperchem Release 7.5 for Windows, Molecular Modeling System, Hypercube, Inc., **2002.**
28. R. Todeschini, Moby Digs Academic version software for variable subset selection by genetic algorithms, Rel. 1.0 for Windows, Talete, Milan, **2004**.
29. R. Leardi, R. Boggia and M. Terrile, *J. Chemom.,* **1992**, *6*, 267-281.