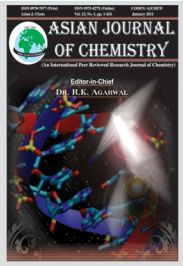
Asian Journal of Chemistry; Vol. 23, No. 3 (2011), 1044-1048

ASIAN JOURNAL OF CHEMISTRY

www.asianjournalofchemistry.co.in



Retention Indices for Programmed-Temperature Gas Chromatography of Polycyclic Aromatic Hydrocarbons: A QSRR Study

KHADIDJA BOUHARIS, MOHAMED LOTFI SOUICI and DJELLOUL MESSADI*

Pollution Laboratory, Badji Mokhtar University, B.P. 12, 23200 Annaba, Algeria

*Corresponding author: E-mail: d_messadi@yahoo.fr

(*Received*: 8 March 2010; *Accepted*: 30 October 2010) AJC-9231

The programmed-temperature gas chromatography retention index (I) of 60 compounds containing unsubstituted and substituted polycyclic aromatic hydrocarbons, is predicted by a statistically validated QSRR modeling approach. The applied multiple linear regression (ordinary least squares, OLS) is based on a variety of theoretical molecular descriptors selected by the genetic algorithms-variable subset selection procedure. The model was validated for predictivity by different internal and external validation approaches. For external validation the original data set was randomly split. The best two-dimensional model, developed on a reduced training set of 20 chemicals, has a predictivity of 99.67 % when applied on 40 validation chemicals (prediction set). Descriptors included in the QSRR model indicated the role of volume, dispersion and hydrophobic interactions in retention mechanism.

Key Words: Theoretical molecular descriptors, Polycyclic aromatic hydrocarbons, Hybrid QSRR model.

INTRODUCTION

Polycyclic aromatic hydrocarbons (PAHs) are an important class of organic compounds, which usually have two to six fused benzene rings, with occasional incorporation of cyclopentene rings. A wide variation of alkyl substituents gives rise to thousands of different PAHs and many have been identified in environmental samples. Polycyclic aromatic hydrocarbons are generally highly toxic and carcinogenic compounds[1] and ubiquitous conta-minants of aquatic and atmospheric ecosystems, where they are present as a result of natural processes such as forest fires, volcanic emissions, but the predominant PAH sources in the environment are related to human activities such as oil spills, burning fossil fuels and domestic wastes, transport emissions.

Polycyclic aromatic hydrocarbons analysis can be well accomplished using gas chromatography[2-4], gas chromatography/mass spectrometry (GC-MS)[5], high performance liquid chromatography[6] or HPLC- MS[7]. There has been a lot of controversy over which of the chromatographic techniques, GC or HPLC, is more favourable. In general, there are some clear advantages of HPLC over GC and *vice versa*. Reversed phase and particularly liquid crystal HPLC columns[8] are capable of separating a number of PAHs that are difficult to separate by capillary GC. Also, the sample preparation procedure is less tedious.

However, many PAHs are thermally stable and exhibit low polarity and as such suitable for GC analysis. The advantages of GC separation of PAHs over the HPLC one are better peak resolution and, when coupled with MS, more reliable confirmation. Unlike separation of PAHs by isothermal GC that is somewhat obsolete nowadays, the temperature-programmed GC has the ability to separate both somewhat weakly and strongly retained PAHs in the same run.

Computers assisted approach to the optimization of chromatographic separations has been extensively used recently. Thus, a number of articles on HPLC of PAHs dealt with factorial design[9-11] and simplex[9,12-14] methodology, while GC separations of PAHs were mainly focused on quantitative structure-retention relationships (QSRRs) studies using multiple linear regression (MLR)[15-19] or artificial neural networks (ANNs)[20,21].

However, there has been a general lack of chromatographic data on isomeric PAHs due to the unavailability of reference compounds. Thus methods that can predict chromatographic retention data of PAHs from its structure are important.

In present work, a hybrid genetic algorithm (GA)/MLR approach was used to model the GC retention index on SE-52 capillary columns of 60 compounds containing unsubstituted and substituted PAHs, when calculating (DRAGON software[22]) theoretical descriptors (independent variables) from the chemical structure alone.

The experimental values of retention index data (I) of PAHs, taken from the literature[23], were randomly split into a training set of 20 chemicals (a third of the total set), used to develop the QSRR model and a validation set of 40 chemicals

(two thirds of the total set), used only for statistical external validation.

The model was examined for robustness and predictive ability through both internal and external validation methods.

EXPERIMENTAL

Molecular descriptor calculation and selection: The structures of the molecules were drawn using Hyperchem 6.03 software[24]. The final geometries were obtained with the semi empirical method AM1. All calculations were carried out at the restricted Hartree-Fock (RHF) level with non-configuration interaction. The molecular structures were optimized using the algorithm Polak-Ribiere and a gradient norm limit of 0.001 kcal/Å. The resulted geometry was transferred into the software DRAGON version 5.4 22 to calculate 290 descriptors of the type Geometrical (74), Topological (120) and 2D Autocorrelations (96). Descriptors with constant or near constant values inside each group were discarded. For each pair of correlated descriptors (with a correlation coefficient $r \geq 0.95$), the one showing the highest pair correlation with the other descriptors was excluded.

The genetic algorithm (GA)[25] has been considered superior to other method of variable selection techniques. So, variable selection was performed on the training set, using GA in the Moby Digs version of Todeschini[26] by maximizing the cross-validated explained variance Q^2_{Loo}.

Model development and validation: Multiple linear regression analysis and variable selection were performed by package Moby Digs for windows/PC[26], using ordinary least squares regression (OLS) method and, as previously indicated, GA for variable subset selection (GA-VSS).

The goodness of fit of the calculated model was assessed by means of the multiple determination coefficient, R^2 and the standard deviation error in calculation (SDEC) defined as

$$SDEC = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)^2}{n}} \quad (1)$$

Cross-validation techniques allow the assessment of internal predictivity (Q^2_{LMO} cross-validation; bootstrap) in addition to the robustness of the model (Q^2_{Loo} cross-validation; Y- scrambling).

Cross-validation by the leave-one-out (Loo) procedure employs n training sets of n - 1 objects in and predicting each excluded object in the test set. The cross-validated explained variance Q^2_{Loo} is defined as:

$$Q^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_{i/i})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} = 1 - \frac{PRESS}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (2)$$

where y_i and $\bar{y}$ are, respectively, the measured and averaged (over the entire data set) values of the dependent variable; $\hat{y}_{i/i}$ denotes the response of the i-th object estimated by using a model obtained without using the i-th object; the summations run over all compounds in the training set.

The predictive residual sum of squares (PRESS) measures the dispersion of the predicted values. It is used to define Q^2 and the standard deviation error in prediction (SDEP).

$$SDEP = \sqrt{\frac{1}{n} PRESS} \quad (3)$$

A value $Q^2 > 0.5$ is generally regarded as a good result and a $Q^2 > 0.9$ as excellent[27,28].

However, studies have indicated that while Q^2 is a necessary condition for high predictive power in a model, its alone is not sufficient.

To avoid overestimating the predictive power of the model the leave-more-out (LMO up to 50 % of perturbation: LMO/50) procedure (repeated 5000 times in this study) was also performed. In a typical LMO validation, n objects of the data set are divided in G cancellation groups of equal size, mj (= n/G). Based on the value of n, G is generally selected between 2 and 10. A large number of models are developed with each of the n-mj objects in the training set and mj objects in the validation set. For each corresponding model, mj objects are predicted and computed (as average value of the number of validation runs).

In order to evidence the existence of fortuitous correlations, the randomization test (Y-scrambling)[29] was adopted. This test consists of building a property vector whose components are the components of the actual property vector, but randomly permuted in their position. This new property vector is used as if it was really an experimental one and a QSRR model is computed in the usual way. This process was repeated 300 times, in order to test the capacity factor of the model to extract actual structure/retention relationships.

By bootstrap validation technique, the original size of the data set (n) is preserved for the training set, by the selection of n objects with repetition; in this way the training set usually consists of repeated objects and the evaluation set of the objects left out[30]. The model is calculated on the training set and responses are predicted on the evaluation set. All the squared differences between the true response and the predictive response of the objects of evaluation set are collected in PRESS. This procedure of building training sets and evaluation sets is repeated 5000 times in this study, PRESS are summed and the average predictive power is calculated.

By using the selected model the values of the response for the test objects are calculated and the quality of these predictions is defined in terms, of Q^2_{ext}, which is defined as:

$$Q^2_{ext} = 1 - \frac{\sum_{i=1}^{n_{ext}}(\hat{y}_{i/i} - y_i)^2 / n_{ext}}{\sum_{i=1}^{n_{tr}}(y_i - \bar{y}_{tr})^2 / n_{tr}} \quad (4)$$

where n_{ext} and n_{tr} are the number of objects in the external set (or left out by bootstrap) and the number of training set objects, respectively.

Other useful parameters are R^2, calculated for the validation chemicals by applying the model developed on the training

set and external standard deviation error of prediction ($SDEP_{ext}$) defined as:

$$SDEP_{ext} = \sqrt{\frac{\sum_{i=1}^{n_{ext}} (y_i - \overline{y}_i)^2}{n_{ext}}} \quad (5)$$

where the sum runs over the test objects (n_{ext}).

Outlier and influential compounds for the developed QSRR model: The jacknified residuals (or Studentized residuals), are the standardized cross-validated residuals. Each residual is divided by its standard deviation, which is calculated without the i-th observation. Compounds, in the training or validation sets, with their standardized residuals greater than three standard deviation units (3σ) are outliers.

The leverage (h_i) value of a chemical in the original variable space is defined as:

$$h_i = x_i^T (X^T X)^{-1} x_i \quad (i = 1, \ldots, n) \quad (6)$$

where x_i is the descriptor row-vector of the query compound and X is the n × p matrix of p + 1 model parameter values for n training set compounds. The superscript T refers to the transpose of the matrix/vector.

The warning leverage value (h*) is defined as 3 (p + 1)/n. When h value of a compound is lower than h*, the probability of accordance between predicted and actual values is as high as that for the compounds in the training set. An influential chemical, with $h_i > h^*$, will reinforce the model if the chemical is in the training set. But such a chemical in the validation set and its predicted data may be unreliable. However, this chemical may not appear to be an outlier because its residuals may be low.

RESULTS AND DISCUSSION

Application of the GA-VSS for the 20 compounds of the training set (Table-1) led to several good models for the prediction of I. The best two dimensional model involved the gravitational index for the bonded atoms G_b and an 2D autocorrelation index, the Geary autocorrelation-lag 4/ weighted by atomic van der Waals volumes GATS 4v, these 2 parameters being not very correlated (r = - 0.299). The MLR analysis equation was obtained as follows:

$$I = -70.10 (\pm 18.82) + 36.97 (\pm 0.44)\ G_b + 36.16 (\pm 15.60)\ GATS4v \quad (7)$$

n = 20, s = 4.41, R^2 (%) = 99.80, F = 3738.18

It is found that gravitational index for the bonded atoms G_b itself yielded a one-variable equation I = - 29.05 (± 5.26) + 36.61 (± 0.47) G_b with a determination coefficient of R^2 (%) = 99.71 and standard deviation s = 4.88 for 20 PAHs. This means that gravitational index for the bonded atoms (G_b) is an important descriptor for the influence of molecular structure on retention behaviour for PAHs. However, a major drawback of gravitational index for the bonded atoms is its degeneracy, *i.e.*, isomers obtain identical numerical values. Hence, the model developed only by employing gravitational index for the bonded atoms is not accurate enough for PAHs. To improve the description a second regressor, GATS4v, was added as shown above in eqn. 7.

Gravitational index for the bonded atoms is an geometrical descriptor reflecting the mass distribution in a molecule, defined[31] as:

$$G_b = \sum_{b=1}^{B} \left(\frac{m_i m_j}{r_{ij}^2} \right)_b \quad (8)$$

where m_i and m_j are the atomic masses of the considered atoms, r_{ij} the corresponding interatomic distances and B the number of bonds of the molecule. This index is related to the bulk cohesiveness of the molecule, accounting, simultaneously, for both atomic masses (volumes) and their distribution within the molecular space; it is associated to the dispersion and hydrophobic interactions.

The 2D Auto class descriptors represent the topological structure of the compounds. The 2D Auto descriptor considered in this study has its origin in autocorrelation of topological structure of Geary (GAST 4v)[32]. The computation of this descriptor involves the summation of an autocorrelation function corresponding to a topological distance (*i.e.*, the lag in the autocorrelation terms) d = 4. At the same time, this descriptor indicates the role of volume in retention mechanism.

All relevant statistical parameters of the proposed model are reported hereafter:

n_{tr}	n valid	Q_2	R^2	$Q^2_{LMO/50}$	Q^2_{boot}	R^2_{adj}
20	40	99.73	99.77	99.66	99.67	99.75
Q^2_{ext}	SDEC	SDEP	$SDEP_{ext}$	F	s	
99.67	4.07	4.47	4.85	3738.19	4.41	

Values of R^2 and R^2_{adj} attest the good fitting performances of the model which, moreover, is very highly significant (great value of the Fisher parameter F).

The model is robust, the difference between R^2 and Q^2 is negligible (0.04 %). Fig. 1 shows a plot contrasting experimental and cross-validated I. As can be seen the point dispersion is small.

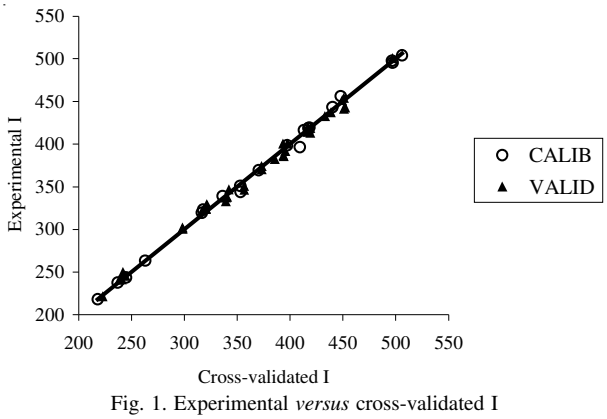


Fig. 1. Experimental *versus* cross-validated I

Standard deviation error of prediction is similar to SDEC, so the model has internal predictivity not too dissimilar from fitting power. The model demonstrates an excellent stability in internal validation (difference between Q^2_{Loo} and $Q^2_{LMO/50}$ is 0.07 %), while bootstrapping confirms the internal predictivity and stability of the model. $SDEP_{ext}$ is a little bit different from SDEP. The model works slightly worse in external prediction than in internal prediction.

TABLE-1
QSRR RESULTS OBTAINED BY MULTI LINEAR REGRESSION FOR RETENTION INDEX, USING THE VARIABLES G_b AND GATS4v

ID	Object	Status	G_b	GATS4v	Y Exp.	Y-Pred	Hat	Std. Err. Pred.
1	2,3-Dimethylnaphthalene	Training	7.335	1.232	243.55	244.9462	0.373	0.3998
2	3-Methylnaphthalene	Training	9.457	1.066	319.46	316.8499	0.076	-0.6156
3	Benz[a]anthracene	Training	11.580	1.119	398.50	397.3075	0.085	-0.2826
4	5-Methylbenz[a]anthracene	Training	12.153	1.087	418.72	417.3765	0.070	-0.3158
5	12-Methylbenz[a]anthracene	Training	12.153	1.111	419.39	418.2065	0.091	-0.2814
6	7,12-Dimethylbenz[a]anthracene	Training	12.726	1.148	443.38	440.2425	0.178	-0.7846
7	Dibenzo[a,h]anthracene	Training	14.275	1.112	495.45	497.1638	0.207	0.4362
8	1-Methyltriphenylene	Training	12.141	0.992	416.32	413.3013	0.128	-0.7327
9	Fluoranthene	Training	10.421	1.066	344.01	353.1941	0.055	2.1414
10	Pyrene	Training	10.415	1.079	351.22	353.0225	0.053	0.4198
11	4-Methylpyrene	Training	10.988	0.969	369.54	370.4094	0.187	0.2186
12	Cyclopenta[cd]pyrene	Training	11.959	1.036	396.54	409.4075	0.070	3.0252
13	Perylene	Training	13.110	0.984	456.22	447.931	0.157	-2.0469
14	Dibenzo[def,mno]chrysene	Training	14.641	0.974	503.89	505.9805	0.235	0.5417
15	2-Methylnaphthalene	Training	6.762	1.085	218.14	218.051	0.220	-0.0229
16	2,6-Dimethylnaphthalene	Training	7.335	1.027	237.58	237.0989	0.247	-0.1257
17	2,3,6-Trimethylnaphthalene	Training	7.908	1.163	263.31	263.1573	0.186	-0.0383
18	9-Methylphenanthrene	Training	9.457	1.097	323.06	317.7565	0.072	-1.2481
19	2,7-Dimethylphenanthrene	Training	10.030	1.022	339.23	336.3969	0.103	-0.6781
20	Benzo[b]chrysene	Training	14.275	1.112	497.66	496.5875	0.207	-0.2730
21	1-Methylnaphthalene	Test	6.762	1.215	221.04	222.6411	0.336	0.4454
22	1,3-Dimethylnaphthalene	Test	7.335	1.061	240.25	238.4132	0.188	-0.4622
23	1,6-Dimethylnaphthalene	Test	7.335	1.129	240.72	240.8039	0.172	0.0209
24	1,5-Dimethylnaphthalene	Test	7.335	1.232	244.98	244.4251	0.357	-0.1569
25	1,8-Dimethylnaphthalene	Test	7.335	1.164	249.52	242.0344	0.207	-1.9053
26	Phenanthrene	Test	8.884	1.137	300.00	298.3587	0.112	-0.3949
27	Anthracene	Test	8.884	1.137	301.69	298.3587	0.112	-0.8015
28	2-Methylanthracene	Test	9.457	1.066	321.57	317.0489	0.075	-1.0656
29	4-Methylphenanthrene	Test	9.457	1.097	323.17	318.1388	0.071	-1.1834
30	1-Methylphenanthrene	Test	9.457	1.158	323.90	320.2834	0.130	-0.8790
31	1-Methylanthracene	Test	9.457	1.158	323.33	320.2834	0.130	-0.7404
32	9-Methylanthracene	Test	9.457	1.188	329.13	321.3381	0.191	-1.9643
33	9-Ethylphenanthrene	Test	10.021	1.110	337.05	339.4494	0.067	0.5631
34	2-Ethylphenanthrene	Test	10.021	1.137	337.50	340.3987	0.094	0.6902
35	9-Isopropylphenanthrene	Test	10.584	1.005	345.78	356.5745	0.113	2.5986
36	1,8-Dimethylphenanthrene	Test	10.030	1.175	346.26	342.0674	0.160	-1.0373
37	9-*n*-Propylphenanthrene	Test	10.584	1.007	350.30	356.6449	0.110	1.5245
38	9,10-Dimethyl-3-ethylphenanthrene	Test	11.167	1.220	381.85	385.6896	0.310	1.0477
39	Benzo[c]phenanthrene	Test	11.580	1.063	391.39	395.4403	0.054	0.9440
40	Chrysene	Test	11.580	1.119	400.00	397.4091	0.083	-0.6135
41	11-Methylbenz[a]anthracene	Test	12.153	1.134	412.72	419.1229	0.121	1.5483
42	2-Methylbenz[a]anthracene	Test	12.153	1.063	413.78	416.6267	0.063	0.6665
43	1-Methylbenz [a] anthracene	Test	12.153	1.087	414.37	417.4705	0.069	0.7284
44	6-Methylbenz [a] anthracene	Test	12.153	1.087	417.57	417.4705	0.069	-0.0234
45	3-Methylchrysene	Test	12.153	1.134	418.10	419.1229	0.121	0.2473
46	2-Methylchrysene	Test	12.153	1.087	418.80	417.4705	0.069	-0.3124
47	5-Methylchrysene	Test	12.153	1.060	419.68	416.6267	0.063	-0.7149
48	1,12-Dimethylbenz[a]anthracene	Test	12.726	1.090	436.82	438.6569	0.087	0.4358
49	Picene	Test	14.275	1.110	500.00	496.8093	0.199	-0.8082
50	2-Phenylnaphthalene	Test	10.008	1.120	332.59	339.1797	0.072	1.5504
51	Triphenylene	Test	11.580	1.010	400.00	393.4715	0.099	-1.5589
52	1,3-Dimethyltriphenylene	Test	12.726	0.920	432.32	432.891	0.297	0.1544
53	2-Methylpyrene	Test	10.988	1.050	370.15	372.9889	0.059	0.6633
54	1-Methylpyrene	Test	10.988	1.050	373.55	372.9889	0.059	-0.1311
55	1-Ethylpyrene	Test	11.551	1.060	385.35	394.1219	0.055	2.0458
56	2,7-Dimetylpyrene	Test	11.561	1.030	386.34	393.5424	0.072	1.6951
57	Benzo[e]pyrene	Test	13.110	0.980	450.73	449.2339	0.152	-0.3683
58	Benzo[a]pyrene	Test	13.110	1.030	453.44	450.9566	0.097	-0.5924
59	Benzo[j]fluoranthene	Test	13.118	1.020	440.92	450.8305	0.105	2.3753
60	Benzo[k]fluoranthene	Test	13.112	1.070	442.56	452.4017	0.094	2.3440

The model was also verified by Y-scrambling. Fig. 2 gathers the randomization test for the optimal QSRR model. The statistics for the modified retention index vectors are clearly lower than the real QSRR model and for the major part a result of $Q^2 < 0$ is obtained. This ensures that a real structure-chromatographic retention relationship has been found out.

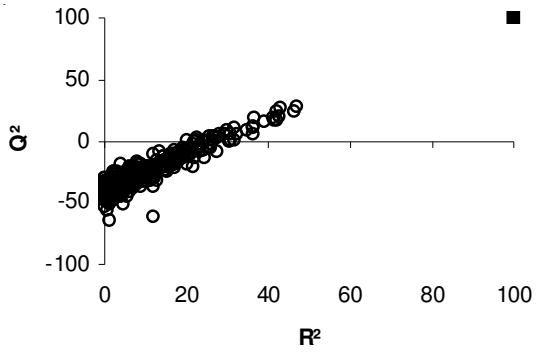


Fig. 2. Randomization test associated to the previous QSRR model. Circles represent the randomly ordered retention indices and the square corresponds to the real retention indices

As shown in Table-1, h_i values of all the compounds in the training and validation sets are lower than the warning value ($h^* = 0.45$). None of the compounds are particularly influential in the model space and the training set has great representativeness. For all the compounds in the training and validation sets, with the exception of the cyclopenta[cd]pyren, their standardized residuals (ei std) are smaller than three standard deviation units (3σ). Thus, in this model there is only one response outlier (cyclopenta[cd]pyren (12).

Conclusion

In summary, a multivariate linear QSRR model has been proposed to predict the programmed-temperature retention index of 60 unsubstituted and substituted polycyclic aromatic hydrocarbons. The ordinary least squares model is developed by a genetic algorithm selection of theoretical molecular descriptors from among a wide set of theoretical molecular descriptors. The proposed model is stable, robust, with good fitting and predictive performance. It is predictive for the chemicals used in the model development (internal validation on training chemicals) and also for chemicals not used in the model development (statistical external validation on validation set chemicals). To have "external" chemicals not used in the model development, the original data set is randomly split into a training set of 20 chemicals and a validation set of 40 chemicals. The factors governing chromatographic retention are the molecular size, as well as dispersion and hydrophobic interactions of molecule with the chromatographic system.

REFERENCES

1. R.C. Harvey, Polycyclic Aromatic Hydrocarbons: Chemistry and Carcinogenecity, Cambridge University Press, Cambridge (1991).
2. S. Sremac, A. Popovic, Z. Todorovic, D. Cokeša and A. Onjia, *Talanta*, **76**, 66 (2008).
3. J.C. Fetzer, in ed.: J.D. Winefordner, Polycyclic Aromatic Hydrocarbons Chemistry and Analysis, Chemical Analysis: A Series of Monographs on Analytical Chemistry and its Applications, Wiley Interscience, Vol. 158, pp. 163-204 (2000).
4. C.D. Simpson, W.R. Cullen, K.B. Quinlan and K.J. Reimer, *Chemosphere*, **31**, 4143 (1995).
5. M. Angels Olivella, *Talanta*, **69**, 267 (2006).
6. J.L. Capelo, M.M. Galesio, G.M. Felisberto, C. Vaz and J.C. Pessoa, *Talanta*, **66**, 1272 (2005).
7. O. Delhomme, M. Millet and P. Herckes, *Talanta*, **74**, 703 (2008).
8. K. Jinno, Y. Saito, R.M. née Chopa, J.J. Pesek, J.C. Fetzer and W.B. Briggs, *J. Chromatogr*., **557**, 459 (1991).
9. A.M. Siouffi and R. Phan-Tan-Luu, *J. Chromatogr. A*, **892**, 75 (2000).
10. A.M. Dorthe, J.L. Ramberti and A. Thienpont, *Analusis*, **28**, 587 (2000).
11. W.C. Lan, K.K. Chee, M.K. Wong, H.K. Lee and Y.M. Sin, *Analyst*, **120**, 281 (1995).
12. P. Araujo, *Trends Anal. Chem*., **19**, 524 (2000).
13. R.W. Fedeniuk, S. Ramamurthi and A.R. Mc Curdy, *J. Chromatogr. B*, **677**, 291 (1991).
14. N. Kuppithayanant, R. Raryanakorn, S. Wongpornchai, T. Prapamontol and R.L. Deming, *Talanta*, **61**, 879 (2003).
15. S. Liu, C. Yin, S. Cai and Z. Li, *Chem. Intell. Lab. Syst*., **61**, 3 (2002).
16. J. Kang, C. Cao and Z. Li, *J. Chromatogr. A*, **799**, 361 (1998).
17. E.R. Collanres, W. Tong, W.J. Welsh and W.L. Zielinski, *Anal. Chem*., **68**, 2038 (1996).
18. J.R. Wilmhust, *J. Chromatogr*., **17**, 50 (1965).
19. R. Kaliszan and H. Lamparczyk, *J. Chromatogr. Sci*., **16**, 246 (1978).
20. S. Srenac, B. Skribic and A. Onjia, *J. Serb. Chem. Soc*., **70**, 1301 (2005).
21. B. Škribic and A. Onjia, *J. Chromatogr. A*, **1108**, 279 (2006).
22. R. Todeschini, V. Consonni and M. Pavan, DRAGON, Software for the Calculation of Molecular Descriptors. Release 5.4 for Windows, Milano (2006).
23. M.L. Lee, D.L. Vassilaros, C.M. White and M. Novotony, *Anal. Chem*., **51**, 768 (1979).
24. Hyperchem TM Release 6.03 for Windows, Molecular Modeling System (2000).
25. R. Leardi, R. Boggia and M. Terrile, *J. Chem*., **6**, 267 (1992).
26. R. Todeschini, D. Ballabio, V. Consonni, A. Mauri and M. Pavan, MOBYDIGS Software for Multilinear Regression Analysis and Variable Subset Selection By Genetic Algorithm, Release 1. 0 for Windows, Milano (2004).
27. L. Eriksson, J. Jaworska, A. Worth, M. Geonin, R.M. Mc Dowell and P. Gramatica, *Environ. Health Persp*., **111**, 69 (2003).
28. A. Tropsha, P. Gramatica and V.K. Grombar, *QSAR Com. Sci*., **22**, 69 (2003).
29. S. Wold and L. Eriksson, in ed.: H. Van de Waterbeend, Statistical Validation of QSAR Results. Validation Tools, In Chemometrics Methods in Molecular Design, VCH Publishers, New York, Vol. 2, pp. 309-318 (1995).
30. B. Efron, The Jacknife, The Bootsrap and Other Resampling Planes, Society for Industrial and Applied Mathematics, Philadelphia (PA), p. 92 (1982).
31. A.R. Katritzky, L. Mu, V.S. Labanov and M. Karelson, *J. Phys. Chem*., **100**, 10400 (1996).
32. R.C. Geary, *Incorp. Statist*., **5**, 115 (1954).