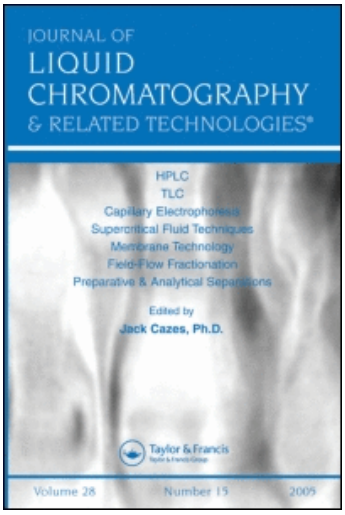

Journal of Liquid Chromatography & Related Technologies

Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713597273

Rapid HPLC Method Development of Polynuclear Aromatic Hydrocarbons Separation Based on Quantitative Structure Retention Relationships

Tomasz Bączek[ab]; Katarzyna Macur[a]; Leszek Bober[c]

[a] Medical University of Gdańsk, Department of Biopharmaceutics and Pharmacodynamics, Gdańsk, Poland [b] Department of Medicinal Chemistry, Faculty of Pharmacy, Collegium Medicum, Nicolaus Copernicus University, Bydgoszcz, Poland [c] POLPHARMA SA Pharmaceutical Works, Starogard Gdański, Poland

To cite this Article Bączek, Tomasz , Macur, Katarzyna and Bober, Leszek(2009) 'Rapid HPLC Method Development of Polynuclear Aromatic Hydrocarbons Separation Based on Quantitative Structure Retention Relationships', Journal of Liquid Chromatography & Related Technologies, 32: 5, 668 — 679

To link to this Article: DOI: 10.1080/10826070802711147

URL: http://dx.doi.org/10.1080/10826070802711147

Journal of Liquid Chromatography & Related Technologies®, 32: 668–679, 2009

ISSN: 1082-6076 print/1520-572X online
DOI: 10.1080/10826070802711147

Rapid HPLC Method Development of Polynuclear Aromatic Hydrocarbons Separation Based on Quantitative Structure Retention Relationships

Tomasz Bączek,[1,2] Katarzyna Macur,[1] and Leszek Bober[3]

[1]Medical University of Gdańsk, Department of Biopharmaceutics and Pharmacodynamics, Gdańsk, Poland
[2]Department of Medicinal Chemistry, Faculty of Pharmacy, Collegium Medicum, Nicolaus Copernicus University, Bydgoszcz, Poland
[3]POLPHARMA SA Pharmaceutical Works, Starogard Gdański, Poland

Abstract: Previous studies demonstrated that quantitative structure retention relationships (QSRR) combined with the linear solvent strength (LSS) model allow for prediction of gradient reversed phase high performance liquid chromatography (HPLC) retention time for any analyte of a known molecular structure under defined HPLC conditions. The QSRR model derived at the selected gradient time was tested at the same gradient time. Presently, in the first step, experimental retention data for model sets of just 5 analytes were used to derive appropriate QSRR models at two gradient times. Additionally, a new molecular modeling approach based on a more accurate *ab initio* method was here proposed. Those QSRR models were used to further predict gradient retention times for sets of 16 test analytes belonging to polynuclear aromatic hydrocarbons (PAHs), at two selected gradient times. Then, applying linear solvent strength (LSS) theory, only predicted retention times for PAHs were used to find the optimal gradient conditions to separate them. Satisfactory predictions of gradient retention times for PAHs were obtained. Contrary to the previous achievements, the proposed QSRR provides the chance to predict retention of PAHs with the appropriate selectivity achieving the same sequence of analytes eluted in the experiment and during the simulation performed on the computer screen.

Correspondence: Tomasz Bączek, Hallera 107, 80-416 Gdańsk, Poland. E-mail: tbaczek@amg.gda.pl

Keywords: High performance liquid chromatography (HPLC), Linear solvent strength (LSS) model, Polynuclear aromatic hydrocarbons (PAHs), Quantitative structure retention relationships (QSRR)

INTRODUCTION

Polyaromatic hydrocarbons (PAHs) are usually formed during the burning of coal, oil, gas, wood, tobacco, rubbish, and other organic substances. They are present in coal tars, crude oil and petroleum products such as creosote and asphalt. Generally, PAHs are ubiquitous environmental contaminants. Benzo(a)pyrene (BaP) is the most toxic substance of all occurring in mixtures of PAHs, therefore it is considered as an indicator of such a contamination. There are some natural sources of PAHs, such as forest fires and volcanoes, but they mainly arise from combustion made sources (related to oil or just man). Concern about PAHs initially focused on their ability to cause cancer. However, more recently concern has turned to their interference with hormone systems and their potential effects on reproduction and their ability to depress immune function. A concern is also associated with the effects of PAHs on egg production in fish, and their potential effects on the numerous early life stages that reside in the surface microlayer of the oceans, where PAHs can become concentrated.[1]

Previous work[2] has shown that gradient reversed phase liquid chromatography separation can be predicted at the given gradient retention time as a function of molecular structural descriptors (total dipole moment, electron excess charge of the most negatively charged atom, and water accessible molecular surface area) for any structurally defined small molecular weight analyte. It was proven that quantitative structure retention relationships (QSRR) combined with the linear solvent strength (LSS) model, allow for reliable, however approximate, prediction of gradient reversed phase HPLC retention time of any structurally defined analyte on a once characterized column. Information, which can guide further optimization of separation procedures can thus be obtained, offering a rational alternative to simple guessing.

Consecutive studies[3] demonstrated that the predictability of the individual QSRR equation for the specific HPLC system was different, depending on the physicochemical properties of the stationary phase, mobile phase, and analytes (neutral, acids, bases) used in the experiment. Later on it was confirmed[4] that one gradient experiment, carried out at the given gradient time for an appropriately designed series of 15 model analytes, provides retention data sufficient to derive a general QSRR equation' characterizing a given column/eluent system. That equation, once established, can next be used to evaluate gradient retention times for any analyte of a known molecular structure, which is going to be

chromatographed in the given HPLC system. It must be emphasized that the QSRR equation derived is applicable for the specific HPLC system under consideration.

It was also suggested[2–4] that two gradient experiments, carried out at two different gradient times for an appropriately designed series of 15 model analytes, provides retention data which can serve to derive two general QSRR equations. These equations can next be used to evaluate gradient retention times for any other analyte of a known molecular structure, which is going to be chromatographed in the given HPLC systems.[9,10]

However, although predictions were accurate in those studies, in fact, they did not allow obtaining the appropriate selectivity achieving the same sequence of analytes eluted in experiment and during the predictions performed on the screen of the computer. Due to the fact, that obtaining the chromatographic conditions, potentially predetermined *a priori* for a new analyte and to optimize its separation is still limited, there is a need to develop new QSRR strategy, which would manage the selectivity criterion. In the current study concerning the specific application, the individual QSRR solution is proposed. As a test set of analytes, a mixture 16 polynuclear aromatic hydrocarbons (PAHs) was chosen. The reason for that choice is, first, PAHs are common combustion products (e.g., in ingredients of smoke) and many of them are known to be harmful. Therefore, PAHs are vigorously monitored in the environment. Secondly, results from previous works[3] suggested that QSRR based predictions performed with the use of the older strategy possess a high prediction accuracy for that group of analytes. Nevertheless, considering the aims of the application of QSRR during the optimization of HPLC separations, the required selectivity criterion was not obtained then.

The aim of the present study was to evaluate QSRR models used to predict gradient retention times for a mixture of 16 polynuclear aromatic hydrocarbons (PAHs). These predictions should possess not only satisfactory accuracy but also should be correct in view of selectivity of the analytes considered. That, in turn, would provide a chance for real optimization of the test analytes separation applying the proposed QSRR model.

EXPERIMENTAL

Equipment

Chromatographic measurements were made with an HPLC apparatus (Waters Corporation, Milford, MA, USA) equipped with a pump, variable wavelength UV/VIS detector, autosampler, thermostat, and the Waters Millennium 2.15 software for data collection and instrument control. Chromatographic measurements on that equipment were performed

with a standard HPLC column, XBridge C18, 150.0 × 4.6 mm (Waters Corporation, Milford, MA, USA) packed with octadecyl-bonded silica.

The injected sample volume was 20 μL. All the chromatographic measurements were done at 35°C with eluent flow rate of 2 mL/min. The experiments were performed at a detection wavelength of 254 nm. The dead time (equaled 1.27 min) was determined by injection of solvent B (acetonitrile with the addition of 0.10% trifluoroacetic acid). Dwell volume was 4.3 mL. All samples were prepared by dissolving the analytes in methanol.

Chemicals

Acetonitrile (HPLC grade) from Merck (Darmstadt, Germany), methanol from P.C. Odczynniki (Gliwice, Poland), and trifluoroacetic acid from Fluka (Buchs, Switzerland) were used. Water used during analyses was prepared with a Milli-Q Water Purification System (Millipore Corporation, Bedford, MA, USA). The following 5 analytes (previously carefully selected from the previously proposed model set of analytes),[2] were used to derive the model QSRR: benzene, naphthalene, biphenyl, phenanthrene, and pyrene, all from all from Sigma-Aldrich (St. Louis, MO, USA). The following 16 analytes were used to test the retention prediction potency of the approaches studied: azulene, acenaphtene, fluorene, triphenylene, 11H-benzo(b)-fluorene, benz(a)anthracene, benz-(b)anthracene, benz(e)acephenanthrylene, perylene, benzo(e)pyrene, benzo(k)fluoranthene, 1,2,3,4-dibenzoanthracene, dibenzo(a,h)anthracene, benzo(g,h,i)perylene, naphto(2,3-a)pyrene, and coronene, all from Sigma-Aldrich (St. Louis, MO, USA).

Determination of Retention Parameters for QSRR Studies

Gradient HPLC elution was carried out with solvent A (water with the addition of 0.10% trifluoroacetic acid) and solvent B (acetonitrile with the addition of 0.10% trifluoroacetic acid). The mobile phase used was filtered through a GF/F glass microfiber filter (Whatman, Maidstone, UK) and degassed with helium during the analysis.

Gradient experimental retention times, $t_{R\ exp}$, of the model series of analytes were measured on the XBridge C18 column washed with linear gradient of 5–100% of acetonitrile. First, the analyses were performed for the model series of analytes within gradient times, $t_G = 5$ min and 15 min (Table 1). Retention data from those two gradient experiments were used to derive model QSRR for 5 preselected analytes. The analyses necessary to confirm the predictability of the model proposed were performed for

Table 1. Molecular descriptors along with experimental gradient retention times (min) for a subseries of model analytes determined at gradient times $t_G = 5$ min and $t_G = 15$ min

Analyte	*X2v*	$t_{R\ exp}$ ($t_G = 5$ min)	$t_{R\ exp}$ ($t_G = 15$ min)
Benzene	1.155	6.88	10.90
Biphenyl	2.732	8.02	15.13
Naphthalene	2.347	7.75	14.10
Phenantrene	3.508	8.25	15.80
Pyrene	4.290	8.57	16.62

testing analytes within gradient times, $t_G = 5$ and 15 (the same gradient time as for model analytes) and additionally $t_G = 45$ min (arbitrary chosen gradient time other than those ones used for model analytes), all with gradient of 5–100% of acetonitrile.

Structural Descriptors of Analytes

The calculations were done by the use of HyperChem program for personal computers with the extension ChemPlus (HyperCube Inc., Gainesville, FL, USA) using the molecular mechanics force field method (MM+) with the Polak-Ribière conjugate gradient algorithm with an RMS gradient of 0.01 kcal/(A mol) as a stopping criterion, followed by quantum chemical calculations according to *ab initio* method with 6–31G** function mode. Thirteen descriptors were obtained directly with the use of HyperChem software. Twelve hundred and twenty nine molecular descriptors were additionally calculated with Dragon professional 5.0 software (Milano Chemometrics and QSAR Research Group – Talete, Milano, Italy) using previously optimized molecules in HyperChem software.

QSRR Analysis

Multiple linear regression (MLR) analysis was performed employing Statistica v. 8.0 software (StatSoft, Tulsa, OK, USA) run on a personal computer. From the total number of molecular descriptors (1242) taken into account, finally only *X2v* – valence connectivity index chi-2 was chosen. It is because this descriptor characterizes the chromatographic system in the best manner, considering statistical significance and the possibility to predict further the retention of PAHs, keeping their

selectivity simultaneously. That last criterion caused the extensive reduction of the available descriptors to be applied. Without that assumption, a number of QSRR models, including two and three parameters MLR equations, could be proposed. However, for the first time, selectivity and the correct sequence of the following analytes in both experimental and predicted retention times of test set of analytes was considered as the most significant criterion for the construction of the final QSRR model. The values of the chosen descriptor for the model set of analytes are collected in Table 1.

Connectivity indices belong to the most popular topological indices and are calculated form vertex degree of the atoms in the H-depleted molecular graph.[5] Randic connectivity index was the first connectivity index proposed.[6] By replacing the vertex degree by the valence vertex degree, valence connectivity indices were previously proposed.[7] They are able to account for the presence of heteroatoms as well as double and triple bonds in the molecule.

Linear regression equations for model set of analytes based on the experimental retention times were derived for $t_G = 5$ min and $t_G = 15$ min (Table 2). Regression coefficients (±standard deviations), correlation coefficients, R, standard errors of estimate, s, significance levels the whole equations, p, and values of the F-test of significance, F, were calculated. Regression equations derived were treated as the basis of the structure retention relationships (QSRR)[8,9] analysis performed.

Moreover, in the present study, a computer simulation software DryLab 2000 plus (LC Resources, Walnut Creek, CA, USA) was employed during the prediction of reversed phase HPLC retention of analytes, using retention times predicted by the individual QSRR models derived for the appropriate gradient times ($t_G = 5$ min and $t_G = 15$ min). The linear solvent strength (LSS) model,[10] upon which DryLab computer simulation software is based, allows the prediction of gradient separation from two gradient runs where only gradient time is varied.

Table 2. Coefficients $k_1 - k_4$ (±standard deviations) with their significance levels, p, and statistical parameters: R, s, F and p, of regression equations of the form: $t_R = k_1 + k_2\ X2v$ for the series of model analytes designed to derive general QSRR equations characterizing individual stationary/mobile phase HPLC systems

Gradient retention time [min]	k_1	k_2	R	s	F	p	Eq. No.
5 min	6.4085 (±0.1916)	0.5294 (±0.0638)	0.9789	0.1516	69	0.0037	1
15 min	9.422 (±0.8827)	1.7987 (±0.2942)	0.9621	0.6984	37	0.0088	2

RESULTS AND DISCUSSION

The scheme of the QSRR based experimental design in optimization of HPLC separations was depicted in Figure 1. To examine prediction ability of a given model one should apply a set of analytes, which was not used previously to derive the QSRR equation. Therefore, the comparison between the experimental and the predicted retention times was done for a test set of analytes.

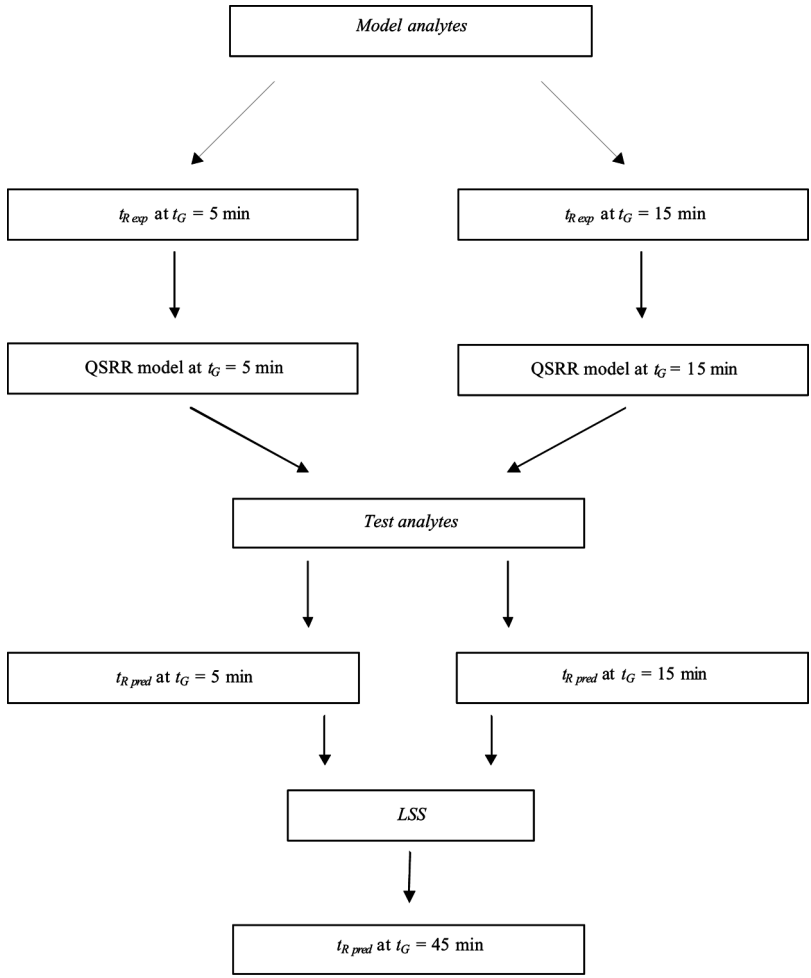

Figure 1. Scheme of the QSRR based experimental design in optimization of HPLC separations.

First, employing gradient retention time data determined experimentally for a series of model analytes (compounds Nos. 1–5 in Table 1), the corresponding regression QSRR equations were derived, characterizing XBridge column (Equations 1 and 2 in Table 2). For that column, the description of t_R by the set of applied structural parameters is good at both gradient times. Correlation coefficients, R, standard errors of estimate, s, significance levels the whole equations, p, and the values of the F-test of significance, F, all are also good.

Now, employing QSRR equations derived at $t_G = 5$ min and $t_G = 15$ min for the model series of analytes (Equations 1 and 2 in Table 2, respectively), retention times were predicted for testing set the

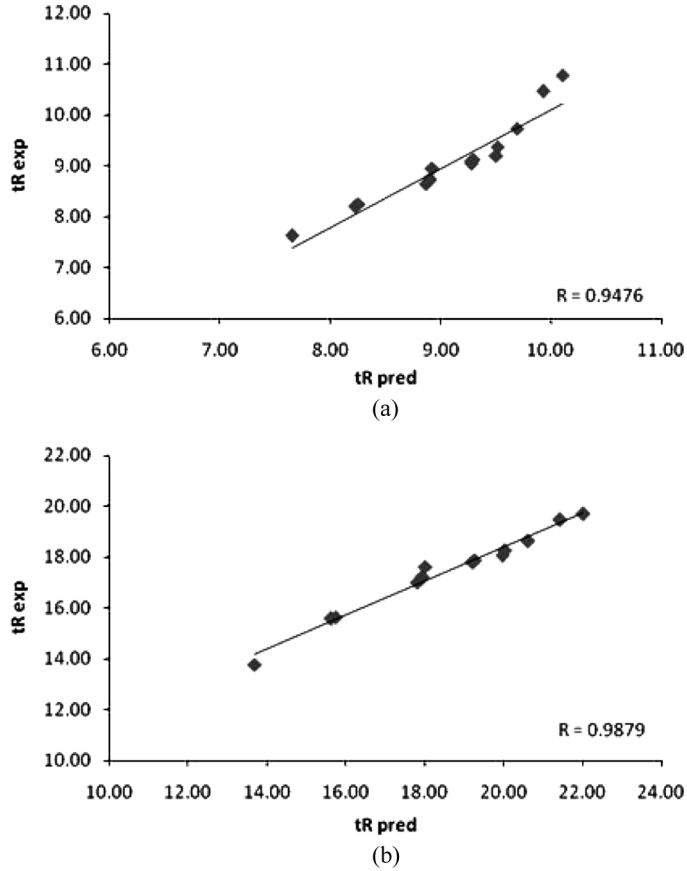

Figure 2. Correlation plots for test series of analytes. Experimental vs. QSRR-predicted gradient retention times for gradient time, t_G : (a) 5 min, and (b) 15 min.

of analytes (PAHs). The comparison between the experimental and the predicted retention times for the testing series of analytes was done. It is seen that correlation between the experimental and the predicted retention times for the testing series of analytes is high in the case of both gradient times with $R = 0.9476$ and 0.9879, respectively (Figure 2). More importantly, predicted retention times are in agreement with the experimentally achieved elution of the individual analytes (Table 3). That is the most important consideration of QSRR based predictions as useful too during the optimization of analyte mixtures separations.

Moreover, an important issue is also the fact that similar values of correlation coefficients were obtained if the predictions were performed for another gradient time ($t_G = 45$ min). It must be emphasized here that the predictions of retention were done on the basis of the previously predicted retention times with the use appropriate QSRR model ($t_{R\ pred}$ from Table 3). Hence, if one predicts retention times of a test series of analytes for gradient time $t_G = 45$ min based on the QSRR calculated retention times for testing analytes at $t_G = 5$ min and $t_G = 15$ min, then correlation between the experimental and predicted retention times is high with

Table 3. Molecular descriptors along with experimental gradient retention times (min) and calculated gradient retention times (min) for a subseries of test analytes determined at gradient times $t_G = 5$ min and $t_G = 15$ min

		$t_G = 5$ min		$t_G = 15$ min	
Analyte	*X2v*	$t_{R\ pred}$	$t_{R\ exp}$	$t_{R\ pred}$	$t_{R\ exp}$
Azulene	2.350	7.65	7.63	13.69	13.77
Acenaphtene	3.430	8.22	8.20	15.63	15.57
Fluorene	3.490	8.26	8.25	15.74	15.63
Triphenylene	4.640	8.86	8.65	17.81	17.00
11H-benzo(b)-fluorene	4.690	8.89	8.70	17.90	17.13
Benz(a)anthracene	4.710	8.90	8.73	17.93	17.23
Benz(b)anthracene	4.750	8.92	8.95	18.01	17.62
Benz(e)acephenanthrylene	5.420	9.28	9.05	19.21	17.80
Benzo(e)pyrene	5.420	9.28	9.08	19.21	17.80
Perylene	5.420	9.28	9.10	19.21	17.80
Benzo(k)fluoranthene	5.450	9.29	9.12	19.27	17.88
1,2,3,4-dibenzoanthracene	5.840	9.50	9.20	19.97	18.05
Dibenzo(a,h)anthracene	5.870	9.52	9.37	20.02	18.25
Benzo(g,h,i)perylene	6.200	9.69	9.73	20.61	18.62
Naphto(2,3-a)pyrene	6.650	9.93	10.47	21.42	19.47
Coronene	6.980	10.10	10.78	22.02	19.70

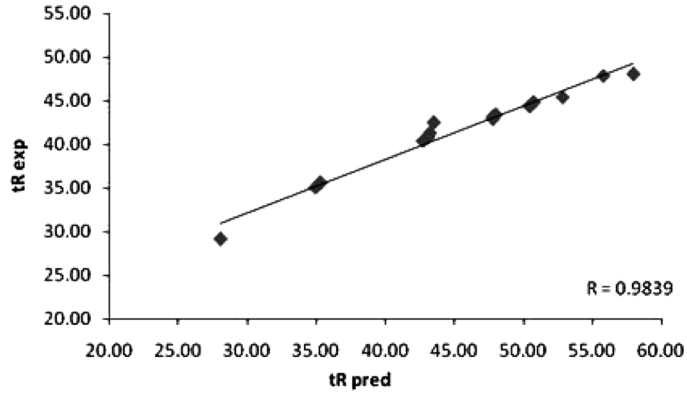

Figure 3. Correlation plot for test series of analytes. Experimental vs. QSRR-predicted gradient retention times for gradient time $t_G = 45$ min.

$R = 0.9839$ (Figure 3). More importantly, again, predicted retention times are in agreement with the experimentally achieved elution of the individual analytes (Table 4).

Table 4. Experimental gradient retention times (min) and calculated gradient retention times (min) for a subseries of test analytes determined at gradient times $t_G = 45$ min

Analyte	$t_{R\ pred}$	$t_{R\ exp}$
Azulene	28.03	29.17
Acenaphtene	34.95	35.12
Fluorene	35.31	35.58
Triphenylene	42.76	40.45
11H-benzo(b)-fluorene	43.11	40.85
Benz(a)anthracene	43.21	41.28
Benz(b)anthracene	43.47	42.47
Benz(e)acephenanthrylene	47.78	42.93
Perylene	47.78	43.00
Benzo(e)pyrene	47.78	43.13
Benzo(k)fluoranthene	47.97	43.38
1,2,3,4-dibenzoanthracene	50.48	44.35
Dibenzo(a,h)anthracene	50.77	44.88
Benzo(g,h,i)perylene	52.83	45.43
Naphto(2,3-a)pyrene	55.79	47.88
Coronene	57.95	48.05

CONCLUSIONS

The prediction ability of general QSRR equations derived for the model series of analytes was tested for predictions of gradient retention data for the test series of PAHs. Important to notice from the study performed, is that QSRR based predictions may be considered as potentially useful in HPLC method development. Similar predictions were obtained employing QSRR equations derived for variable gradient times. Therefore, there is an option that the predictions of retention at different gradient times can be done on the basis of the previously predicted retention times with the use of appropriate QSRR models, without any significant increase of the prediction errors. The strategy proposed was designed to support the optimization of PAHs separations. The advantage of that QSRR based strategy is associated with the limited number of the initial experiments to be performed. Input retention data must be collected for just five model analytes (benzene, naphthalene, biphenyl, phenanthrene, and pyrene) at two variable gradient times. These retention data serve to derive the appropriate QSRR models comprising X2v descriptor (valence connectivity index chi-2), which can be easily calculated in Dragon software for any PAH molecular structure optimized previously in HyperChem software. Having QSRR equations, one is able to calculate retention times for any PAH at two gradient times. Furthermore, possessing those calculated retention times at two gradient times for these PAHs, and applying LSS theory (or just DryLab software), one can predict retention time for them at any other gradient time. This, in turn, could enable optimizing the final separation of the complicated mixture of PAHs.

Summarizing the novelty of the approach proposed, first, experimental retention data for a model set of just 5 analytes were used to derive QSRR models (previously, experimental effort was needed to be carried out for 15 model analytes). Moreover, a new molecular modeling approach based on more accurate *ab initio* method was proposed (previously less advanced quantum mechanic calculation method was used). Previously, Dragon software was also not used to provide several new molecular descriptors. Among them, the descriptor applied in the study was found. New testing sets of 16 analytes was used and, significant from an environmental and human health point of view, polynuclear aromatic hydrocarbons (PAHs) were applied. The approach is proposed for the separation of the defined group of analytes – PAHs. The current approach was extended to the predictions of retention performed with the use of LSS theory applying just predicted (not experimental) retention times at two gradient times. In contrary to the previous achievements, the proposed QSRR provides the chance to predict retention of PAHs with the appropriate selectivity, achieving

the same sequence of analytes eluted in the experiment and during the simulation performed on the screen of the computer.

ACKNOWLEDGMENTS

This work was supported by the Polish State Committee for Scientific Research Projects no. 2 P05F 041 30 and no. N N405 1040 33.

REFERENCES

1. A briefing to highlight WWF's work in conservation (www.wwf-uk.org), 1997.
2. Bączek, T.; Kaliszan, R. Combination of linear solvent strength model and quantitative structure-retention relationships as a comprehensive procedure of approximate prediction of retention in gradient liquid chromatography. J. Chromatogr. A. **2002**, *962*, 41–55.
3. Bączek, T.; Kaliszan, R. Predictive approaches to gradient retention based on analyte structural descriptors from calculation chemistry. J. Chromatogr. A. **2003**, *987*, 29–37.
4. Kaliszan, R.; Bączek, T.; Buciński, A.; Buszewski, B.; Sztupecka, M. Prediction of gradient retention from the linear solvent strength (LSS) model, quantitative structure-retention relationships (QSRR) and artificial neural networks (ANN). J. Sep. Sci. **2003**, *26*, 271–282.
5. Todeschini, R., Consonni, V. in. *Handbook of Molecular Descriptors, Methods and Principles in Medicinal Chemistry*; Mannhold, R., Kubinyi, H., Timmerman, H., Eds.; Wiley-VCH: Weinheim, 2000.
6. Randic, M. On characterization of molecular branching. J. Am. Chem. Soc. **1975**, *97*, 6609–6615.
7. Kier, L.B.; Hall, L.H. Derivation and significance of valence molecular connectivity. J. Pharm. Sci. **1981**, *70*, 583–589.
8. Kaliszan, R. *Structure and Retention in Chromatography. A Chemometric Approach*; Harwood Academic Publishers: Amsterdam, 1997.
9. Kaliszan, R.; Kaliszan, A.; Noctor, T.A.G.; Purcell, W.P.; Wainer, I.W. Mechanism of retention of benzodiazepines in affinity, reversed-phase and adsorption high-performance liquid chromatography in view of quantitative structure-retention relationships. J. Chromatogr. **1992**, *609*, 69–81.
10. Snyder, L.R.; Dolan, J.W. The linear-solvent-strength model of gradient elution. Adv. Chromatogr. **1998**, *38*, 115–187.

Received September 27, 2008
Accepted October 20, 2008
Manuscript 6408