Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Chromatography A, 987 (2003) 29–37
JOURNAL OF CHROMATOGRAPHY A
www.elsevier.com/locate/chroma

Predictive approaches to gradient retention based on analyte structural descriptors from calculation chemistry

Tomasz Bączek, Roman Kaliszan*
Department of Biopharmaceutics and Pharmacodynamics, Medical University of Gdańsk, Gen. J. Hallera 107, 80-416 Gdańsk, Poland

Abstract

Quantitative structure retention relationships (QSRRs) were applied to predict reversed-phase HPLC gradient retention. The performance of the recently recommended QSRR models was compared. One tested model is based on structural descriptors from molecular modeling. To quantitatively characterize the structure of analytes the following three structural descriptors are employed: total dipole moment, electron excess charge of the most negatively charged atom and water-accessible molecular surface area. Reliability of the resulting gradient retention time predictions was compared to that provided by the models relating retention to the theoretically calculated logarithm of *n*-octanol–water partition coefficient, log *P*. The requested values of log *P* were obtained using three commercially available softwares. The predicted retention parameters were compared for a series of structurally diversified small molecular mass analytes. It has been demonstrated that the retention predictions from both the molecular modeling descriptors-based and the log *P*-based QSRR are characterized by similar errors. It has been hypothesized that the optimization of separation based on QSRRs and the linear solvent strength theory might be of practical analytical value.

Keywords: Retention prediction; Molecular descriptors; Gradient elution; Linear solvent strength theory; Molecular modeling; Structure–retention relationships

1. Introduction

An a priori prediction of properties, either biological or physicochemical, of chemical substances from their structural formulas is a fundamental, however still quite unrealistic (at least in quantitative terms), task of chemistry. Starting conditions for deriving quantitative structure–property relationships, allowing for reliable property predictions, are determined by the accurate, reproducible property measures on one hand, and the exact, unambiguously defined structural features of the chemical entities under consideration encoding specific information on their individual property aspects, on the other hand. Chromatography may obviously be an excellent source of quantitatively comparable property measures that can conveniently be collected for representative series of analyte structures. Therefore, quantitative structure–retention relationships (QSRRs) have, since their introduction in the late 1970s, been considered a model approach to establish strategy of property predictions, to test the performance of various chemometric data processing methods as well as property predictor potency of theoretically unlimited number of structural descriptors offered by computational chemistry [1–3].

Previous studies in this laboratory [4–7] demonstrated a good performance, especially in quantitative

*Corresponding author. Tel.: +48-58-349-3260; fax: +48-58-349-3262.
E-mail address: romankal@farmacja.amg.gda.pl (R. Kaliszan).

0021-9673/02/$ – see front matter
PII: S0021-9673(02)01701-6

comparison of retention properties of diverse HPLC columns, of the general QSRR model employing the following analyte descriptors: (i) total dipole moment, μ; (ii) electron excess charge of the most negatively charged atom, δ_{Min}; (iii) water-accessible molecular surface area, A_{WAS}. The following physical meaning of individual descriptors is assumed: μ accounts for the dipole–dipole and dipole-induced dipole attractive interactions of the analyte with the components of the competing mobile and stationary phase; δ_{min} reflects ability of analytes to participate in polar interactions with the phases of the charge-transfer and hydrogen-bonding type; A_{WAS} describes the strength of dispersive interactions (London–Hall interactions) of the analyte with the molecules forming the chromatographic phases. The general QSRR equation based on these molecular modeling-derived descriptors has the form:

$$\text{retention parameter} = k_1 + k_2\mu + k_3\delta_{\mathrm{Min}} + k_4 A_{\mathrm{WAS}} \quad (1)$$

where retention parameter may be either isocratic log k_{w} or gradient retention time, t_{R}, and k_1–k_4 are regression coefficients.

Several other QSRR models have been reported in the literature. Best known is the model based on solvatochromic or LSERs (linear solvation energy relationships) analyte structure descriptors assumed to account for differences among the analytes regarding their ability to take part in intermolecular interactions with the components of a chromatographic system of the following types: cavity formation, polarizability, hydrogen bond donation, hydrogen bond acceptance and dispersive attractions [8–14]. The model shows good retention prediction potency but requires empirically determined structural parameters that obviously are not available for all the possible analytes. As a matter of fact, Wilson et al. [15] have recently elaborated a LSER-based procedure of retention prediction based on limited amount of experimental measurements. The approach, however comprehensive, appears rather complex for routine retention prediction purposes.

An a priori prediction of retention from structure of analyte offers a model proposed by Galushko et al. [16,17]. The model employs a term accounting for dispersive interactions and a term supposed to differentiate the analytes with regards to their so-called polarity. Our comparative study [18] showed that the Galushko approach is quite approximate, however.

In reversed-phase liquid chromatography retention parameters are since the time of Martin and Synge [19] known to correlate with partition coefficients. The reference liquid–liquid partition parameter is the logarithm of *n*-octanol–water partition coefficient, log *P* [20]. Values of log *P* can be calculated from structural formulas employing commercially available softwares. Next, for retention prediction a simple regression equation can be applied:

$$\text{retention parameter} = k_1 + k_2 \log P \quad (2)$$

where k_1 and k_2 are regression coefficients.

Using Eqs. (1) and (2) retention parameters for a structurally representative and sufficiently large (for meaningful statistics) model series of analytes chromatographed in a given HPLC system can be described. A model series of 18 analytes was previously designed to compare retention properties of various stationary phase materials [5,6]. It was later found [7] that the model series could be shortened to 15 compounds without meaningful loss of statistical significance of the resulting QSRRs.

The pH of the buffers used in experiments was chosen to minimize the dissociation of the analytes. Otherwise pH would had to be considered as an additional variable affecting gradient retention time. Predictions of retention time would then require knowledge of pK_{a} of analytes [21,22].

The values of gradient retention times and of the structural descriptors of the analytes used to derive QSRR equations characterizing the HPLC system studied are presented in Table 1.

The aim of this work was to check and compare the goodness of predictions based on Eqs. (1) and (2). Structural descriptors used in Eq. (1) were derived by the standard molecular modeling. The log *P* values used in Eq. (2) were calculated with use of three softwares: ACD (Advanced Chemistry Development, Toronto, Canada), HYPERCHEM with the extension CHEMPLUS (Hypercube, Waterloo, Canada) and CLOGP (BioByte, Claremont, CA, USA). By means of the ACD software it was possible to predict directly the gradient retention times of analytes based on their log *P* values. On the other hand, the log *P* data

Table 1
Experimental gradient retention times, $t_{R\ exp}$, along with the total dipole moment, μ, electron excess charge of the most negatively charged atom, δ_{Min}, water-accessible molecular surface area, A_{WAS}, all from molecular modeling and the log P values calculated with the use of ACD, HYPERCHEM and CLOGP computer programs, for the set of analytes employed to derive QSRR models. Linear gradient of methanol 10–90% at $t_G = 10$ min

No.	Analyte	$t_{R\ exp}$ (min)	μ (D)	δ_{Min} (electron)	A_{WAS} (Å^2)	log P		
						ACD	HYPERCHEM	CLOGP
1	Benzamide	6.84	3.583	−0.4333	293.46	0.74	1.05	0.65
2	4-Cyanophenol	7.93	3.311	−0.2440	290.90	1.60	1.80	1.60
3	Indazole	9.05	1.547	−0.2034	284.44	1.82	1.14	1.63
4	Benzonitrile	9.01	3.336	−0.1349	279.14	1.65	2.08	1.57
5	Indole	9.77	1.883	−0.2194	292.38	2.14	1.82	2.13
6	2-Naphthol	10.35	1.460	−0.2518	323.16	2.71	2.76	2.65
7	Anisole	10.62	1.249	−0.2116	288.94	2.13	1.79	2.06
8	Benzene	10.69	0.000	−0.1301	245.21	2.22	2.05	2.14
9	1-Naphthylacetonitrile	10.60	3.031	−0.1381	364.26	2.68	3.34	2.74
10	Benzyl chloride	11.12	1.494	−0.1279	296.17	2.49	2.66	2.70
11	Naphthalene	12.14	0.000	−0.1277	311.58	3.45	3.05	3.32
12	Biphenyl	12.65	0.000	−0.1315	358.08	3.98	3.73	4.03
13	Phenanthrene	13.12	0.020	−0.1279	374.73	4.68	4.05	4.95
14	Pyrene	13.61	0.000	−0.1273	392.41	5.17	4.37	4.49
15	2,2′-Dinaphthyl ether	13.79	1.463	−0.1606	510.36	6.67	5.48	6.59

provided by HYPERCHEM and CLOGP were used to derive model QSRR equations which served next to calculate t_R based on the linear solvent strength (LSS) relationships. Retention predictions based on the structural descriptors from molecular modeling (total dipole moment, electron excess charge of the most negatively charged atom and water-accessible molecular surface area) and on the three kinds of log P parameters were discussed in terms of relative error in gradient retention coefficient, k^*, which was calculated after Snyder and Dolan [23].

2. Experimental

2.1. Equipment

Chromatographic measurements were made with an HPLC apparatus (Waters, Milford, MA, USA) equipped with a pump, variable-wavelength UV–Vis detector, autosampler and thermostat. Data were collected using the Waters MILLENNIUM 2.15 software. Supelcosil LC_{18} column, 15.0×0.46 cm I.D., particle size 5 μm (Supelco, Bellefonte, PA, USA), packed with octadecyl-bonded silica was employed.

The mobile phase contained methanol and 100 m*M* Tris buffer of pH 2.5 and 7.2 necessary for suppression of dissociation of individual analytes. The buffer was prepared by dissolving tris(hydroxymethyl)aminomethane (P.C. Odczynniki, Gliwice, Poland) in water and adjusting the pH with 1 *M* HCl (Fluka, Buchs, Switzerland). The pH of the buffer was measured at 21 °C before mixing with the organic modifiers. The pH measurements were done with an HI 9017 pH meter (Hanna Instruments, Bedfordshire, UK).

All the chromatographic measurements were done at 35 °C with eluent flow-rate of 1 ml/min. The injected sample volume was 20 μl.

2.2. Chemicals

Methanol was from P.C. Odczynniki. Water was prepared with a Milli-Q Water Purification System (Millipore, Bedford, MA, USA).

The following test analytes (Table 1) were selected to derive model QSRR equations: benzamide, indazole, benzonitrile, 2-naphthol, anisole, 1-naphthylacetonitrile, benzyl chloride, naphthalene, biphenyl, pyrene, 2,2′-dinaphthyl ether, all from Lancaster (Newgate, UK); indole and benzene, both from P.C. Odczynniki; 4-cyanophenol from Aldrich (Gillingham, UK) and phenanthrene from Koch-Light Labs. (Koinbrook, UK).

Table 2
Experimental, $t_{R\ exp}$, and calculated gradient retention times, $t_{R\ calc}$, along with structural descriptors and the relative errors in gradient retention coefficient, k^*, for the set of analytes employed to test the retention prediction potency of QSRR models derived for test analytes from Table 1 and described in Table 3. Linear gradient of methanol 10–90% at $t_G = 10$ min

No.	Analyte	μ (D)	δ_{Min} (electron)	A_{WAS} (Å^2)	log P			$t_{R\ exp}$ (min)	Eq. (4) (Table 3)		Eq. (5) (Table 3)		Eq. (6) (Table 3)		Eq. (7) (Table 3)	
					ACD	HYPERCHEM	CLOGP		$t_{R\ calc}$	Relative error in k^*	$t_{R\ calc}$	Relative error in k^*	$t_{R\ calc}$	Relative error in k^*	$t_{R\ calc}$	Relative error in k^*
1	1-Bromonaphthalene	1.414	−0.1540	340.71	4.22	3.84	4.18	12.75	11.28	0.27	12.29	0.12	12.35	0.11	12.33	0.11
2	Cumene	0.247	−0.2057	322.15	3.56	3.24	3.57	12.37	11.49	0.18	11.49	0.18	11.47	0.18	11.59	0.17
3	*n*-Propylbenzene	0.336	−0.2118	329.97	3.74	3.31	3.70	12.53	11.50	0.21	11.71	0.18	11.58	0.20	11.75	0.17
4	Anthracene	0.000	−0.1267	379.15	4.68	4.05	4.49	13.01	13.21	0.08	12.84	0.06	12.66	0.11	12.71	0.10
5	*n*-Hexylbenzene	0.349	−0.2106	421.46	4.80	4.10	4.76	13.79	13.01	0.26	12.98	0.27	12.73	0.32	13.03	0.26
6	*n*-Butylbenzene	0.341	−0.2107	360.86	4.27	3.70	4.23	12.96	12.02	0.23	12.35	0.17	12.15	0.20	12.39	0.16
7	*n*-Amylbenzene	0.349	−0.2107	391.43	5.34	4.50	5.29	13.36	12.51	0.24	13.63	0.13	13.31	0.02	13.68	0.15
8	2-Ethyltoluene	0.468	−0.2106	323.11	3.67	3.38	3.62	12.48	11.30	0.22	11.63	0.18	11.68	0.17	11.65	0.18
9	1,3,5-Trimethylbenzene	0.000	−0.1786	332.12	3.60	3.45	3.64	12.72	12.05	0.15	11.54	0.23	11.78	0.19	11.67	0.21
10	1,2,3-Trimethylbenzene	0.487	−0.1807	319.85	3.60	3.45	3.54	12.56	11.45	0.21	11.54	0.20	11.78	0.17	11.55	0.20
11	1-Methylnaphthalene	0.274	−0.1811	335.17	3.91	3.52	3.81	12.51	11.87	0.15	11.91	0.14	11.88	0.15	11.88	0.15
12	*o*-Xylene	0.437	−0.1804	297.13	3.14	2.98	3.09	12.00	11.12	0.16	10.99	0.18	11.10	0.17	11.00	0.18
13	*m*-Xylene	0.258	−0.1790	302.97	3.14	2.98	3.14	12.11	11.36	0.15	10.99	0.20	11.10	0.19	11.06	0.19
14	*p*-Xylene	0.000	−0.1780	303.57	3.14	2.98	3.14	12.13	11.58	0.12	10.99	0.20	11.10	0.19	11.06	0.19
15	3-Cyanobenzoic acid	3.907	−0.3554	322.24	1.48	1.78	1.55	8.32	7.54	0.17	9.00	0.25	9.35	0.44	9.13	0.32
16	3-Fluorobenzoic acid	2.759	−0.3568	293.89	2.16	1.88	2.13	9.49	7.95	0.23	9.81	0.09	9.49	0.00	9.84	0.10
17	o-Toluic acid	2.077	−0.3695	308.71	2.35	2.21	2.38	9.89	8.62	0.23	10.04	0.05	9.97	0.02	10.14	0.08
18	*p*-Toluic acid	2.809	−0.3670	316.80	2.35	2.21	2.38	10.05	8.21	0.29	10.04	0.00	9.97	0.02	10.14	0.03
19	4-Ethylbenzoic acid	2.889	−0.3672	343.94	2.89	2.61	2.91	10.77	8.59	0.36	10.69	0.03	10.56	0.07	10.78	0.01
20	3-Hydroxybenzoic acid	3.496	−0.3584	299.84	1.50	1.46	1.56	7.44	7.46	0.01	9.02	0.68	8.88	0.58	9.14	0.77
21	4-Hydroxybenzoic acid	3.010	−0.3682	300.43	1.42	1.46	1.56	6.80	7.77	0.36	8.93	1.25	8.88	1.20	9.14	1.52
22	Benzoic acid	2.418	−0.3651	287.97	1.89	1.75	1.88	9.20	8.05	0.19	9.49	0.08	9.30	0.03	9.53	0.09
23	1-Naphthylacetic acid	2.028	−0.3742	376.83	3.13	2.68	2.59	10.45	9.75	0.22	10.98	0.29	10.66	0.10	10.40	0.02
24	Acetylsalicylic acid	5.816	−0.3321	353.76	1.19	1.24	1.02	8.53	6.76	0.31	8.65	0.04	8.56	0.01	8.49	0.01
25	Naproxen	2.346	−0.3584	446.68	3.00	2.99	2.82	11.01	10.77	0.14	10.82	0.12	11.11	0.07	10.68	0.19

26	Ketoprofen	2.779	−0.3581	481.35	2.81	3.46	2.76	10.75	11.01	0.29	10.59	0.13	11.80	2.13	10.60	0.12
27	Fenbufen	4.028	−0.3584	490.29	2.93	2.83	3.14	11.17	10.19	0.57	10.74	0.35	10.88	0.26	11.06	0.11
28	Diclofenac	1.783	−0.3754	462.21	3.28	3.97	4.32	11.87	11.34	0.30	11.16	0.36	12.54	0.78	12.50	0.71
29	2-Chloroaniline	1.676	−0.4010	284.98	1.91	1.78	1.91	9.31	8.30	0.17	9.51	0.05	9.35	0.01	9.57	0.07
30	2-Methoxyaniline	0.802	−0.4035	306.41	1.09	1.01	1.18	9.12	9.31	0.06	8.53	0.14	8.22	0.19	8.68	0.11
31	3,4-Dichloroaniline	3.707	−0.4025	309.72	2.51	2.30	2.59	10.19	7.13	0.31	10.23	0.01	10.10	0.02	10.40	0.05
32	3,5-Dichloroaniline	2.989	−0.4026	312.41	2.70	2.30	2.71	10.72	7.73	0.34	10.46	0.07	10.10	0.14	10.54	0.05
33	3,5-Dimethylaniline	1.274	−0.4137	322.95	1.86	2.20	1.91	9.76	9.14	0.15	9.45	0.09	9.96	0.07	9.57	0.06
34	3-Chloroaniline	2.603	−0.4073	288.71	1.81	1.78	1.91	9.17	7.60	0.22	9.39	0.05	9.35	0.04	9.57	0.10
35	3-Methylaniline	1.469	−0.4131	293.50	1.40	1.73	1.41	8.67	8.51	0.04	8.90	0.06	9.27	0.19	8.96	0.08
36	4-Chloroaniline	3.086	−0.4066	289.22	1.76	1.78	1.91	9.09	7.24	0.23	9.33	0.06	9.35	0.06	9.57	0.12
37	*N*-Ethylaniline	1.867	−0.3605	327.54	2.13	1.96	2.17	10.16	9.16	0.22	9.78	0.11	9.61	0.14	9.89	0.08
38	4-Methoxyaniline	1.966	−0.4157	309.35	0.74	1.01	1.00	7.17	8.37	0.55	8.11	0.39	8.22	0.46	8.46	0.62
39	Coumarin	4.818	−0.28799	310.80	1.39	1.82	1.41	8.69	7.15	0.23	8.89	0.05	9.41	0.22	8.96	0.07
40	Phthalimide	3.348	−0.40254	306.23	1.15	1.22	1.15	7.73	7.35	0.08	8.60	0.30	8.53	0.27	8.65	0.32
41	Phthalonitrile	5.298	−0.1134	308.61	1.25	2.12	1.01	7.79	8.06	0.08	8.72	0.35	9.84	1.22	8.48	0.23
42	1,4-Naphthoquinone	1.332	−0.2698	324.50	1.79	1.04	1.93	9.49	10.21	0.29	9.37	0.03	8.27	0.24	9.59	0.03
43	Phenylacetylene	0.257	−0.1964	290.81	2.40	2.23	2.41	10.67	11.03	0.10	10.10	0.11	10.00	0.13	10.18	0.10
44	Carbazole	1.206	−0.2449	361.17	2.67	2.94	3.52	11.23	11.10	0.05	10.43	0.21	11.04	0.06	11.53	0.12
45	9,10-Anthraquinone	0.003	−0.2863	388.67	2.44	2.44	2.62	11.84	12.17	0.16	10.15	0.36	10.31	0.34	10.43	0.33
46	Xanthene	1.146	−0.1523	376.35	3.93	3.51	4.40	12.83	12.09	0.21	11.94	0.23	11.87	0.25	12.60	0.08
47	Hexachlorobutadiene	0.000	−0.0730	340.60	3.98	2.61	4.90	13.12	12.98	0.04	12.00	0.22	10.56	0.33	13.20	0.03
									Mean	0.21	Mean	0.19	Mean	0.27	Mean	0.19

The following analytes (Table 2) were used to test the retention prediction potency of the model QSRR equations: 1-bromonaphthalene, cumene, *n*-propylbenzene, anthracene, *n*-butylbenzene, *n*-amylbenzene, 2-ethyltoluene, 1,3,5-trimethylbenzene, 1,2,3-trimethylbenzene, 1-methylnaphthalene, *o*-xylene, *m*-xylene, *o*-toluic acid, *p*-toluic acid, 4-ethylbenzoic acid, 3-hydroxybenzoic acid, 4-hydroxybenzoic acid, 1-naphthylacetic acid, phenylacetylene, carbazole, xanthene, 9,10-anthraquinone, hexachlorobutadiene, 1,4-naphthoquinone, coumarin, phthalimide, phthalonitrile, all from Lancaster (Newgate, UK); 3-cyanobenzoic acid, 3-fluorobenzoic acid, 2-chloroaniline, 3,4-dichloroaniline, 3,5-dichloroaniline, 3,5-dimethylaniline, 3-chloroaniline, 3-methylaniline, 4-chloroaniline, *N*-ethylaniline, 4-methoxyaniline, all obtained from LC Resources (Walnut Creek, CA, USA); *n*-hexylbenzene from Aldrich; *p*-xylene from Romil (Shepshed, UK); benzoic acid from Merck (Darmstadt, Germany); 2-methoxyaniline from P.C. Odczynniki; acetylsalicylic acid, diclofenac, ketoprofen, fenbufen, naproxen, all from the drug and reagent collection of the Medical University of Gdańsk (Gdañsk, Poland).

2.3. Determination of retention parameters for QSRR studies

Gradient retention times, $t_{R\ exp}$, of the model series of analytes from Table 1 were measured on Supelcosil LC_{18} column washed with linear gradient of 10–90% of methanol at gradient time, t_G, of 10 min. The data from these gradient experiments were used to derive model QSRRs.

2.4. Structural descriptors of analytes

Molecular structure descriptors of the analytes which were employed in QSRR analysis, i.e. total dipole moment, μ, electron excess charge of the most negatively charged atom, δ_{Min}, and water-accessible molecular surface area, A_{WAS}, were calculated by standard molecular modeling. The HYPERCHEM program for personal computers with the extension CHEMPLUS was used for the calculations of these parameters. The software performed geometry optimization by the molecular mechanics MM+ force field method followed by quantum chemical calculations according to the semiempirical AM1 method [24,25]. The log *P* values were calculated with use of the following softwares: ACD, HYPERCHEM with the extension CHEMPLUS and CLOGP. The structural descriptors for the QSRR model analytes are collected in Table 1 and for the test analytes the respective data are in Table 2.

2.5. QSRR analysis

2.5.1. Multiple regression analysis

Multiple regression analysis equations were derived employing Microsoft EXCEL software (Microsoft, Redmond, WA, USA) run on a personal computer. Regression coefficients (±standard deviations), multiple correlation coefficients, *R*, standard errors of estimate, *s*, significance levels of each term and of the whole equations, *P*, and values of the *F*-test of significance, (*F*) were calculated and are reported in Table 3.

To derive model QSRR equations to be used for retention predictions, gradient retention times, $t_{R\ exp}$, for analytes from Table 1 were regressed against the three structural descriptors obtained from molecular modeling: total dipole moment, μ, electron excess charge of the most negatively charged atom, δ_{Min}, and water-accessible molecular surface area, A_{WAS}, and against individual log *P* values calculated by the three softwares studied. The resulting QSRR equations [Eqs. (4)–(7)], characterizing the HPLC system studied, are collected in Table 3.

In Table 2 the relative errors in gradient retention coefficients, *k**, are given to quantify the prediction potency of the QSRR models here derived. The calculations of the errors were done according to the following equation [23]:

$$\text{relative error in } k^* = (t_0/2.3\,b)(\delta k/k) \quad (3)$$

where t_0 is dead time, *b* is gradient steepness parameter, δk is the absolute difference between the experimental and the calculated gradient retention coefficient, *k* is the experimental gradient retention coefficient.

To illustrate the gradient retention prediction capabilities of the QSRR models specified in Table 3 the respective experimental t_R data are plotted against the calculated ones in Fig. 1.

Table 3
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 forms: $t_R = k_1 + k_2\mu + k_3\delta_{Min} + k_4 A_{WAS}$, and $t_R = k_1 + k_2$ log *P*, respectively, for the series of analytes from Table 1 employed to derive the QSRR models studied

k_1	k_2	k_3	k_4	*R*	*s*	*F*	*P*	Eq. no.
QSRR based on analyte descriptors from molecular modeling								
7.9076 (±0.6208)	−0.7723 (±0.0918) (P = 4E−06)	7.5117 (±1.4875) (P = 0.0004)	0.0165 (±0.0016) (P = 5E−07)	0.9870	0.3727	138	5E−09	(4)
QSRR based on log *P* from ACD								
7.2209 (±0.4580)	1.2005 (±0.1382) (P = 9E−07)	–	–	0.9236	0.8170	75	9E−07	(5)
QSRR based on log *P* from HYPERCHEM								
6.7529 (±0.6039)	1.4574 (±0.2011) (P = 6E−06)	–	–	0.8953	0.9493	53	6E−06	(6)
QSRR based on log *P* from CLOGP								
7.2476 (±0.4517)	1.2156 (±0.1420) (P = 1E−06)	–	–	0.9216	0.8271	73	1E−06	(7)

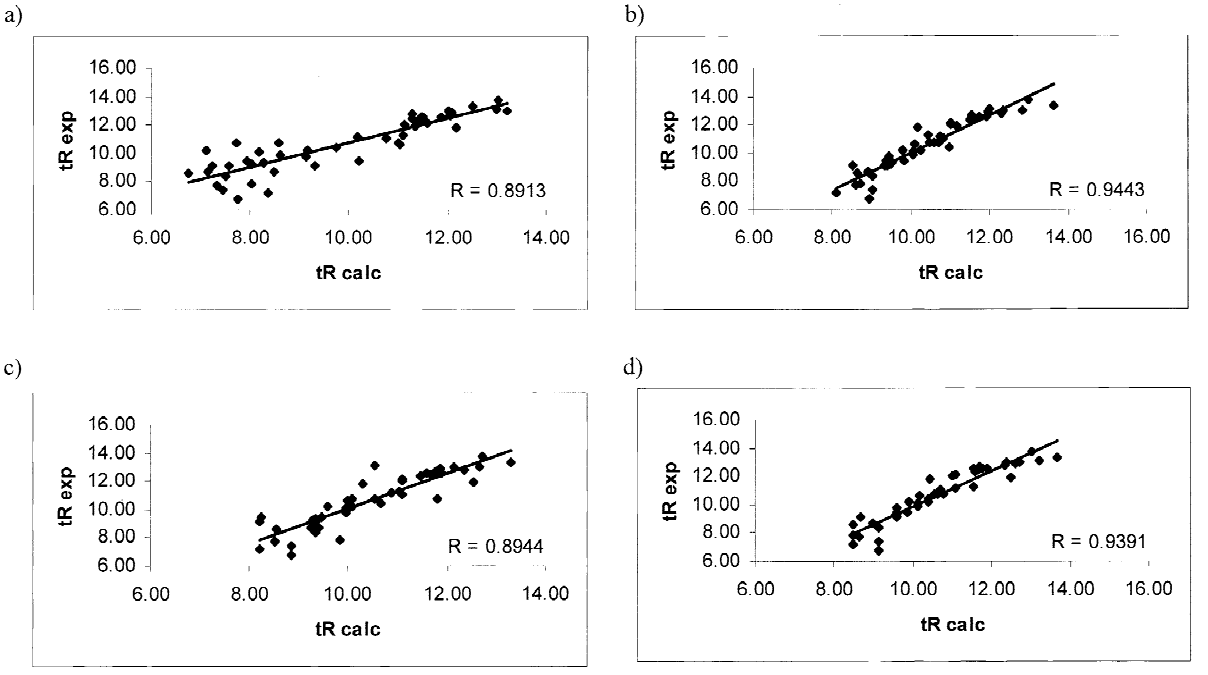

Fig. 1. Correlations between the calculated from individual QSRR models and the experimental gradient retention times for a set of test analytes from Table 2: (a) model based on total dipole moment, electron excess charge of the most negatively charged atom and water-accessible molecular surface area as the structural descriptors, (b) model based on log *P* values calculated with use of ACD software, (c) model based on log *P* values calculated with use of HYPERCHEM software, (d) model based on log *P* values calculated with use of CLOGP software.

3. Results and discussion

3.1. QSRR equations characterizing the HPLC column studied

The model QSRR equation relating t_R of the training set of analytes to their total dipole moment, electron excess charge of the most negatively charged atom and water-accessible molecular surface area is of excellent statistical quality [Eq. (4) in Table 3]. All the coefficients at the three parameters are statistically significant ($P \leq 0.0004$) as is the whole equation ($P = 5E-09$). Multiple correlation coefficient ($R = 0.9870$), standard error of estimate ($s = 0.3727$), and the value of the F-test of significance ($F = 138$), are all very good.

Statistically significant QSRR equations were also found to describe gradient retention in terms of the log P values calculated with the use of the ACD, HYPERCHEM and CLOGP computer programs (Eqs. (5)–(7) in Table 3). The coefficients at the log P parameters are at the level of $P \leq 1E-06$ in the case of all the three equations derived. Correlation coefficients, R, are for all the three equations lower than for Eq. (4) as are the values of the F-test of significance, F; higher are the standard errors of estimate, s. The best QSRR equation obtained with use of log P values was that calculated with the parameters provided by the ACD software ($R = 0.9236$). Nearly the same quality possesses the QSRR equation relating t_R to log P derived by CLOGP software ($R = 0.9216$). Of evidently lower quality was the equation employing log P values obtained by means of the HYPERCHEM software ($R = 0.8953$).

3.2. Testing of retention prediction potency of the derived QSRR models

In Table 2 the retention parameters calculated by means of Eqs. (4)–(7) and the experimental gradient retention times are collected for a large test series of structurally diverse analytes, which had not been used to derive the QSRR models. Goodness of predictions is illustrated in Fig. 1.

Standard deviations in the predicted gradient retention times were converted into relative errors in the gradient retention coefficient k^*. These errors are listed in Table 2 for a series of 47 test analytes chromatographed on the Supelcosil LC_{18} column with linear gradient 10–90% of methanol at gradient time $t_G = 10$ min. As seen from Table 2 the present QSRR molecular modeling approach gives mean relative error in prediction of k^* of 19–27%. It is assumed that to be practically useful for optimizing resolution, the errors in k^* should be no greater than about 5% [23]. Therefore, the retention predictions obtained by the approaches here discussed may only be treated as a first approximation, which is still better than just a guess.

The relative errors calculated according to Snyder and Dolan [23] are generally larger than conventional errors defined as $(\Delta k/k) \cdot 100$.

Gradient retention data predicted by log P values derived by means of three computer programs (ACD, HYPERCHEM and CLOGP) differ. Fig. 1b demonstrates the predictions based on log P values obtained with the use of ACD software. These predictions appear better ($R = 0.9443$) than those resulting from the molecular modeling-based QSRR (Fig. 1a, $R = 0.8913$). On the other hand, the error of prediction expressed as the absolute relative error in k^* (Table 2) is very similar in case of Eq. (4) and both the Eqs. (5) and (7), its mean being 21, 19 and 19%, respectively.

The predictions based on log P values obtained with HYPERCHEM [Eq. (6)] are less accurate than those provided by Eqs. (4), (5) and (7), however. Also, in the case of Eq. (6) the correlation ($R = 0.8944$) between the calculated and the experimental gradient retention times is lower and comparable to the correlation ($R = 0.8913$) observed for the data observed and predicted by Eq. (4). The mean relative error in k^* produced by Eq. (6) is larger (27%) than that resulting from application of Eqs. (4), (5) and (7).

Predictions based on log P values calculated with the use of CLOGP software are comparable to those achieved with the use of ACD software. In the former case, the correlation between the calculated and the experimental gradient retention times is described by $R = 0.9391$. The mean relative error in k^* (19%) is like in the case of the ACD-derived log P.

4. Conclusions

Results of present study provide additional evidence to our hypothesis [7] that QSRRs combined

with the LSS model allow for approximate prediction of gradient reversed-phase HPLC retention time of any analyte on a once characterized column. Information that can guide further optimization of the analytical procedure can thus be obtained. Gradient experiments carried out for a relatively short series of 15 model analytes serve to derive model QSRR equations. These equations, once established for a given column/eluent system, are next used to evaluate retention parameters for any analyte of a known molecular structure to be chromatographed in the given HPLC system. Consequently, the starting chromatographic conditions providing requested separation could be set a priori and then adjusted experimentally. The approach should be of help to finally optimize the separation of various analytes of actual analytical interest.

Several QSRR approaches compared in this work are of similar gradient retention prediction potency. Either the approach based on the structural descriptors from molecular modeling (total dipole moment, electron excess charge of the most negatively charged atom and water-accessible molecular surface area), or the approach based on the log P values derived with the use of three computer programs (ACD, HYPERCHEM and CLOGP) give reasonable, however approximate, retention predictions. For the series of well known simple compounds used as the test analytes, the best predictive results provided the QSRRs based on log P values calculated by means of the ACD computer program. A good performance of ACD in the case of simple reagents of well characterized properties (including n-octanol–water partition accounted for by log P) must be confirmed for more complex compounds. For such compounds, which had not been included in the partition data bank employed for deriving the log P data by the calculating programs, the universal QSRR based on the molecular modeling-derived parameters (μ, δ_{Min}, A_{WAS}) might be more reliable.

The accuracy of a priori predictions of gradient retention from the computer-generated properties of analytes is limited. Selfevidently, the main limitation of such a prediction of retention of chemical compounds from their structure is inadequacy of the translation of structural formulas into sets of numerical descriptors. No doubt that such a translation that would reveal the properties encoded into the structure in a reliable manner is still lacking. Further efforts should hence be encouraged in the area of theoretical chemistry and molecular modeling that would result in better means of characterization of chemical entities and consequently, in more reliable predictions of their properties. QSRRs offer a unique tool to test the performance of new theoretical concepts and calculation procedures.

References

[1] R. Kaliszan, Quantitative Structure–Chromatographic Retention Relationships, Wiley, New York, 1987.
[2] R. Kaliszan, Structure and Retention in Chromatography—A Chemometric Approach, Harwood, Amsterdam, 1997.
[3] E. Forgacs, T. Cserhati, Molecular Bases of Chromatographic Separations, CRC Press, Boca Raton, FL, 1997.
[4] R. Kaliszan, M.A. van Straten, M. Markuszewski, C.A. Cramers, H.A. Claessens, J. Chromatogr. A 855 (1999) 455.
[5] M.A. Al-Haj, R. Kaliszan, A. Nasal, Anal. Chem. 71 (1999) 2976.
[6] M.A. Al-Haj, R. Kaliszan, B. Buszewski, J. Chromatogr. Sci. 39 (2001) 29.
[7] T. Bączek, R. Kaliszan, J. Chromatogr. A 962 (2002) 41.
[8] P.T. Jackson, M.R. Schure, T.P. Weber, P.W. Carr, Anal. Chem. 69 (1997) 416.
[9] J. Li, P.W. Carr, Anal. Chim. Acta 334 (1996) 239.
[10] J.H. Park, M.D. Jang, D.S. Kim, P.W. Carr, J. Chromatogr. 513 (1990) 107.
[11] M.J. Kamlet, R.W. Taft, J. Am. Chem. Soc. 98 (1976) 377.
[12] R.W. Taft, M.J. Kamlet, J. Am. Chem. Soc. 98 (1976) 2886.
[13] P.C. Sadek, P.W. Carr, R.M. Doherty, M.J. Kamlet, R.W. Taft, M.H. Abraham, Anal. Chem. 57 (1985) 2971.
[14] M.H. Abraham, Chem. Soc. Rev. 23 (1993) 660.
[15] N.S. Wilson, M.D. Nelson, J.W. Dolan, L.R. Snyder, R.G. Wolcott, P.W. Carr, J. Chromatogr. A 961 (2002) 171.
[16] S.V. Galushko, A.A. Kamenchuk, G.L. Pit, J. Chromatogr. A 660 (1994) 47.
[17] S.V. Galushko, J. Chromatogr. 552 (1991) 91.
[18] T. Bączek, R. Kaliszan, H.A. Claessens, M.A. van Straten, LC·GC Europe 14 (2001) 304.
[19] A.J.P. Martin, R.L.M. Synge, Biochem. J. 35 (1941) 1358.
[20] C. Hansch, A. Leo, Substituent Constants For Correlation Analysis in Chemistry and Biology, Wiley, New York, 1979.
[21] R. Kaliszan, P. Haber, T. Bączek, D. Siluk, Pure Appl. Chem. 73 (2001) 1465.
[22] R. Kaliszan, P. Haber, T. Bączek, D. Siluk, K. Valko, J. Chromatogr. A 965 (2002) 117.
[23] L.R. Snyder, J.W. Dolan, Adv. Chromatogr. 38 (1998) 115.
[24] N.L. Allinger, J. Am. Chem. Soc. 99 (1977) 8127.
[25] M.J.S. Dewar, E.G. Zoebisch, E.F. Healy, J.J.P. Stewart, J. Am. Chem. Soc. 107 (1985) 3902.