ELSEVIER

Journal of Chromatography A, 913 (2001) 3–13

JOURNAL OF CHROMATOGRAPHY A

www.elsevier.com/locate/chroma

Comparison of selected retention models in reversed-phase liquid chromatography

Jaeju Ko, John C. Ford*

Department of Chemistry, Indiana University of Pennsylvania, 975 Oakland Avenue, Indiana, PA 15705, USA

Abstract

Retention models are usually compared by how well the model equation fits retention data for one solute taken over a range of mobile phase compositions. Even when retention data for multiple solutes are used, the quality of the fit is often judged by the statistical goodness-of-fit alone. This study compared four different RPLC retention models, encompassing three distinct mathematical forms. Each model was fit to the retention data of multiple solutes and the sets of best-fit parameters were examined in terms of the underlying physico-chemical assumptions of the models. Next, for the linear and quadratic models, some of the model parameters were calculated a priori and the rest of the model parameters were then obtained in subsequent fittings. The sets of best-fit parameters obtained in this manner were more consistent with the underlying assumptions of these models than were the sets of parameters obtained entirely through regressions to the experimental data. Thus, the extraction of parameters by fitting a model to the retention data of a single solute may result in unreliable values for those parameters, even in the case of a fit that would be considered good when judged by conventional statistical criteria. That is, although parameters extracted in such a fashion may be suitable for optimization or similar uses, they may not be suitable for determining the appropriateness of the underlying assumptions of retention models.

Keywords: Retention models; Reversed-phase liquid chromatography; Liquid chromatography; Mathematical modelling; Alkylbenzamides

1. Introduction

Any successful theory for reversed-phase liquid chromatography (RPLC) should predict the effect of the mobile phase composition on retention. Numerous chromatographic models have been proposed for this purpose and particularly for mobile phases consisting of binary mixtures of organic solvents and water. Excellent reviews of these models are available [1–4].

Retention models are often evaluated by the statistical goodness-of-fit, such as the value of the reduced χ^2, the correlation coefficient, or the residuals themselves, to retention data using the parameters from each model as adjustable fitting parameters (e.g., Refs. [5–7]). The objective of this study is to show that the statistical goodness-of-fit for the retention data of a single solute alone is not a reliable criterion in assessing the appropriateness of the underlying assumptions of retention models.

In this work, retention models involving three different functional forms were selected from among the many models reported in the literature. Each model was then fit to several sets of retention data and the resulting sets of best-fit parameters were examined in terms of the model's underlying assumptions. Next, some model parameters were calcu-

*Corresponding author. Fax: +1-724-357-5700.
E-mail address: jford@grove.iup.edu (J.C. Ford).

0021-9673/01/$ – see front matter
PII: S0021-9673(01)00605-7

lated a priori, thus reducing the number of adjustable parameters for each model. and the fittings then repeated. These new sets of best-fit parameters were also examined in terms of the model's underlying assumptions.

The most widely applied model in the optimization of separations, particularly for gradient elution, is the linear solvent strength (LSS) model developed by Snyder and co-workers [8–11] According to the LSS model, the retention factor, k, of a solute is related to the solvent composition, φ, by:

$$\ln k = \ln k_w - S\varphi \quad (1)$$

where S is the solvent strength parameter, φ is the volume fraction of the stronger solvent and k_w is the retention factor of that solute in pure water. (Note that Eq. (1) is usually presented as $\log k$ rather than $\ln k$.) Generally, S increases with increasing solute size for a given mobile phase and column [12–14]. Wang et al. have shown that S can be considered to be the standard free energy of solute transfer from pure water to pure organic eluent divided by $-RT$ [15]:

$$S = \frac{\Delta G_s^o}{-RT} \quad (2)$$

where ΔG_s^o is the standard free energy of transfer of solute from pure water to pure organic eluent, R is the gas constant and T is the absolute temperature. Carr notes that this requires the assumption that ''the stationary phase is not modified by sorption of mobile phase'' [15].

Several quadratic models have been proposed [16–24]:

$$\ln k = A\varphi^2 + B\varphi + C \quad (3)$$

where the model parameters A, B and C can be calculated from interaction indices [20], from solubility parameters [16–19], or from normalized contact free energies [21–24]. In order to determine interaction indices [20], it is necessary to define a scale based on standard compounds. Since our study used published retention data and since the retention data for these standard compounds was not available in that study, we did not consider the interaction indices model in our comparison.

In the quadratic solubility parameter model [17–19], retention depends on the solubility parameters of the solute, the stationary phase, and the mobile phase components. The coefficients of this model are given by the following expressions:

$$A = (V_i/RT)(\delta_o - \delta_w)^2 \quad (4)$$

$$B = -(2V_i/RT)(\delta_w - \delta_o)(\delta_w - \delta_i) \quad (5)$$

$$C = \ln k_w = (V_i/RT)[(\delta_w - \delta_i)^2 - (\delta_s - \delta_i)^2] + \ln(n_s/n_m) \quad (6)$$

where δ is the solubility parameter; i, s, o and w denote the solute, stationary phase, organic modifier, and water, respectively; V_i is the molar volume of the solute; R is the gas constant; T is the absolute temperature; and n_m and n_s are number of moles of solute in the mobile (m) and stationary (s) phases.

In the quadratic interphase model [21,22], the parameters A, B and C depend on the normalized contact free energies (χ) as follows:

$$A = \chi_{ab} \quad (7)$$

$$B = \chi_{sb} - \chi_{sa} - \chi_{ab} \quad (8)$$

$$C = \ln k_0 \quad (9)$$

where a and b denote the two mobile phase solvents, s denotes the solute, and k_0 is the k when the volume fraction of solvent b is zero. The normalized contact free energy can be estimated from the free energy of transfer, i.e., χ_{xy} is the free energy of transfer of solute x from pure x to infinitely dilute y divided by the product of the Boltzmann's constant and the absolute temperature [22]. A subsequent modification to this theory [23,24] includes a size correction factor in A and B: $A = n\chi_{ab}$ and $B = n(\chi_{sb} - \chi_{sa} - \chi_{ab})$ where n is the ratio of the size of a solute molecule to the size of a molecule of either solvent a or b.

A logarithmic dependence of retention to modifier concentration has been also proposed [25]. This stoichiometric displacement (SD) model assumes the existence of five equilibria involving the solute, the alkyl bonded phase on the support surface, and the free organic solvent. The retention factor expression derived from these five equilibria can be written as:

$$\ln k = \ln I + Z \ln(1/[D_o]) \quad (10)$$

where Z is the number of solvent molecules required to displace the solute from the surface, $[D_o]$ is the molarity of the organic modifier and $\ln I$ contains the equilibrium constant for solvation of bound solute, the concentration of ligand–solvent complex, and the phase ratio. The authors state, ''The physical meaning of Z is the total moles of the displacing agent [organic modifier] released ... as one mole of solute is adsorbed ...'' and ''$\log I$ contains a number of constants which relate to the affinity of one mole of the solute to the stationary phase'' [26]. The SD model has been applied to proteins [25,27] and small solutes [26] in RPLC and to nucleic acids in anion-exchange chromatography [28].

Eq. (10) can be recast as:

$$\ln k = I - Z \ln \varphi \quad (11)$$

since $\varphi = [D_o]V$ when the volume change upon mixing is negligible and where V is the molar volume of the modifier. In Eq. (11), the parameter I contains the $\ln I$ term of Eq. (10) and the molar volume term above. The functional form of Eq. (11) (rather than Eq. (10)) is used in a recent comparison study [5].

In the current study, the models were compared by examining the results of two methods of fitting model parameters as well as two methods of direct estimation of selected model parameters. These methods were

1. *unconstrained fittings*, where all model parameters were used as fitting parameters. In unconstrained fittings, the linear regressions of Eq. (1) (the LSS model), Eq. (3) (the solubility parameter model and the interphase model), and Eq. (11) (the SD model) onto a set of retention data were performed;
2. *constrained fittings*, where the parameter A of the quadratic models was estimated using Eq. (4) or Eq. (7) for each solute prior to the fitting to the same set of retention data by the equation below:

$$Y = \ln k - A\varphi^2 = B\varphi + C \quad (12)$$

3. estimating the LSS model parameter S using Eq. (2) and
4. estimating the interphase model parameter B (in addition to A) using Eq. (8).

Note that the SD model (Eqs. (10) or (11)) does not contain parameters that can be estimated a priori.

Wells and Clark's published isocratic retention data for 18 homologous *N*-alkylbenzamides [29] were used in this study. These data were used because the numerical values of the retention factors are reported (rather than an equation describing a fit to the retention factors or as a figure), the retention factors span the range of mobile phase compositions from 0 to 100% acetonitrile in water, and the retention factors are all greater than 1.5 (minimizing the error in k associated with the choice of void marker). Also, more importantly for the purposes of the current study, this set of data, being the retention data for a homologous series taken over a wide range of mobile phase compositions, allowed a closer examination of the extracted model parameters. That is, the obtained best-fit parameters must show systematic relationships and Wells and Clark's data are particularly suitable for examining underlying assumptions of these models. However, there is no estimate of the uncertainty in the retention factors.

In the unconstrained fittings, all three sets of regressions gave small values of reduced χ^2 (χ_r^2). The quadratic equation showed the best overall performance as judged from χ_r^2. The resulting sets of best-fit parameters, S, A and B values for the homologous series did *not* show the trends predicted. The $\ln k_w$ parameters and the Z parameters did show the predicted trends.

In the constrained fittings, where the model parameter A for each of the two quadratic models was fixed prior to fitting, the interphase model gave smaller residuals than the solubility parameter model. The trend in the set of best-fit C values of the interphase model was consistent with the underlying assumptions of that model. The trend in the set of best-fit B values of the solubility parameter model was found to be consistent with the underlying assumptions of that model.

When the estimated S parameters from Eq. (2) were used for the LSS model, the resulting average $\ln k_w$ values increased regularly with the solute size as expected. Similarly, when the estimated A and B parameters (from Eqs. (7) and (8), respectively) were used for the interphase model, the resulting average $\ln k_w$ values increased regularly with the size of the solutes as expected.

2. Calculations and data analysis

Both the unconstrained and constrained fittings were performed using the published retention data of Wells and Clark [29]. Their measurements were performed on an Altex Ultrasphere ODS column using acetonitrile–water mixtures as the mobile phase.

The regressions were performed using Axum (version 5.0, MathSoft, Cambridge, MA, USA, 1996). The goodness-of-fit is discussed in terms of the χ_r^2, which is the sum of the squares of the residuals divided by the degrees of freedom.

The *A* parameter for each solute given by Eq. (4) was calculated using the molar volume of the solute and $\delta = 24.55$ $(\text{cal}/\text{cm}^3)^{1/2}$ for water and $\delta = 12.50$ $(\text{cal}/\text{cm}^3)^{1/2}$ for acetonitrile [16] (1 cal = 4.18 J). The molecular volume of each solute was obtained by the AM1 [30] geometry optimization routine in Spartan PC Plus (Wavefunction, Irvine, CA, USA). The molecular volume was assumed to be the volume contained in the isodensity surface of 0.002 electrons/bohr^3, which is equivalent to the volume defined by the van der Waals radii [31]. Molar volumes obtained by this method were lower than the experimental molar volumes estimated from the densities and the molecular masses of the molecules tested. For example, the molar volume of naphthalene is 125 cm^3/mol at 20°C, calculated from the density and the molecular mass. The molar volume calculated by AM1 geometry optimization was 86.2 cm^3/mol. All other small molecules we tested showed similar trend and thus, our calculated molar volumes are likely to be about 70% of the actual molar volumes.

A detailed description of estimating the normalized contact free energy (χ) from the free energy of transfer is given by Dill [22]. Following the description given in Ref. [22], *A* (Eq. (7)) and *B* (Eq. (8)) were estimated from the standard free energies of solvation.

The standard free energy of solvation was calculated by the SM5.0R model [32–34] using the AM1 optimized gas phase geometries. The SM5.0R model achieves a mean unsigned error in the calculated standard free energy of solvation of about 0.57 kcal/mol using the gas-phase AM1 optimized geometries when water is the solvent [32] and about 0.4 kcal/mol from either Hartree–Fock level optimized geometries or from AM1 optimized geometries [33]. The SM5.0R model is discussed elsewhere [32,33].

The χ_{ab} (Eq. (7)) for acetonitrile (ACN)–water mixtures was calculated to be 2.37. This value was obtained using the AM1 optimized geometry of acetonitrile and calculating the free energy of transfer of acetonitrile from pure acetonitrile to water, divided by *RT*. This value is very close to the experimental value of 2.21 given in Dill's work [22]. Similarly, the normalized free energy of transfer $(\chi_{sb} - \chi_{sa})$ was taken to be the difference in the calculated free energies of solvation of each solute in two solvents (ACN and water) divided by *RT*.

In a similar fashion, the *S* value for each solute was estimated using Eq. (2), where ΔG_s^o was taken to be the standard free energy of solvation of the solute in acetonitrile minus the standard free energy of solvation of the solute in water.

3. Results and discussion

3.1. Unconstrained fitting

The results of the unconstrained fittings of Eqs.(1), (3), and (11) to each solute's retention data are presented in Table 1. The fitting results given in Table 1 indicate that all three equations are excellent in terms of fitting the data points with relatively small residuals. Plots of representative regressions and the corresponding residuals are shown in Figs. 1–3. All three equations render noticeable trends in their residuals in the low φ region. Trends are also apparent in the residuals of the linear and the quadratic equations in the intermediate region, although these residuals are smaller in magnitude than in the low φ region. Trends in the residuals indicate a failure of the model to account completely for the observed behavior.

The linear and quadratic models allow several useful indices for direct comparison of the models to experimental data. The α values (in water) calculated from the best-fit k_w values serve as a convenient index for comparison of the models. The experimentally determined $\ln k_w$ values for *N*-methylbenzamide and *N*-ethylbenzamide are 4.506 and 5.439, respectively [29]. The $\ln k_w$ values for solutes larger than

Table 1
Unconstrained fitting results for linear (Eq. (1)), quadratic (Eq. (3)), and logarithmic (Eq. (11)) equations[a]

Number of alkyl carbons	1	2	3	4	5	6	7	8	9	10	11	12	13	14	16	17	18	19
Number of data points	10	9	8	8	7	6	6	6	7	7	7	6	10	10	9	8	7	7
φ range (%)	0–30	0–30	3–35	10–45	15–45	25–60	30–60	35–60	35–70	40–70	45–80	50–80	55–100	60–100	65–100	70–100	75–100	75–100
Best-fit S for Eq. (1)	13.7	14.6	12.8	10.7	11.6	9.92	10.5	11.0	10.5	10.7	9.65	9.46	7.68	7.97	8.90	9.29	9.79	10.5
Best-fit $\ln k_w$ for Eq. (1)	4.00	4.78	5.27	5.65	6.68	6.76	7.57	8.32	8.75	9.28	9.15	9.44	8.69	9.24	10.7	11.3	12.1	13.0
χ_r^2 for fit to Eq. (1)	0.177	0.226	0.130	0.040	0.025	0.041	0.037	0.013	0.037	0.023	0.031	0.013	0.042	0.027	0.017	0.015	0.015	0.014
Best-fit A for Eq. (3)	41.3	44.8	27.8	14.8	15.5	13.5	16.7	14.2	14.4	14.6	11.6	9.36	8.97	8.24	7.12	7.73	9.81	8.84
Best-fit $-B$ for Eq. (3)	25.1	27.3	23.1	18.9	20.9	21.4	25.6	24.5	25.7	26.7	24.1	21.6	21.7	21.3	20.8	22.5	27.1	26.0
Best-fit C for Eq. (3)	4.30	5.17	5.87	6.57	7.92	9.01	10.8	11.4	12.5	13.5	13.5	13.3	14.0	14.5	15.5	16.9	19.6	19.8
χ_r^2 for fit to Eq. (3)	0.024	0.032	0.009	0.002	0.000	0.001	0.000	0.001	0.001	0.001	0.003	0.002	0.005	0.007	0.008	0.010	0.012	0.012
Best-fit Z for Eq. (11)	0.971	1.03	1.65	2.56	3.22	4.03	4.60	5.15	5.40	5.80	5.90	6.06	5.91	6.36	7.34	7.90	8.58	9.18
Best-fit I for Eq. (11)	−0.410	−0.247	−0.341	−0.882	−0.890	−1.08	−0.947	−0.834	−0.414	−0.179	0.248	0.613	1.11	1.35	1.85	2.10	2.33	2.58
χ_r^2 for fit to Eq. (11)	0.156	0.117	0.069	0.011	0.006	0.000	0.003	0.002	0.003	0.003	0.004	0.001	0.012	0.010	0.008	0.009	0.010	0.010

[a] The retention data are from Wells and Clark [29]. The original retention data are reported as log k values and to the third decimal place. The χ_r^2 is the reduced chi-squared, which is the sum of the squares of the residuals divided by the degrees of freedom. The values are reported here to the third decimal place due to limited space. Estimates of the standard deviations of the fitted parameters are not reported, again due to limited space. These can be provided to interested readers.

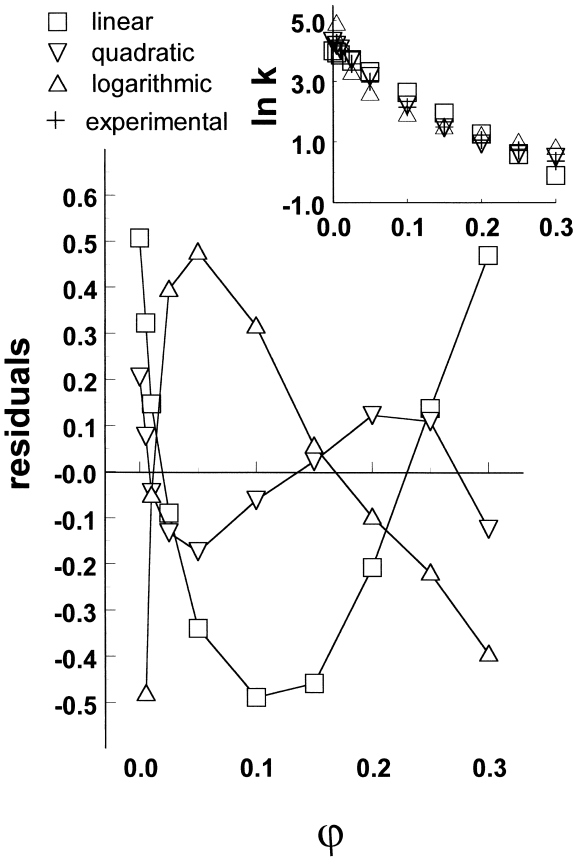

Fig. 1. Representative unconstrained regressions and corresponding residuals in the low φ region, from the fitting of Eqs. (1) (linear), (3) (quadratic), and (11) (logarithmic) to the measured ln k of *N*-methylbenzamide.

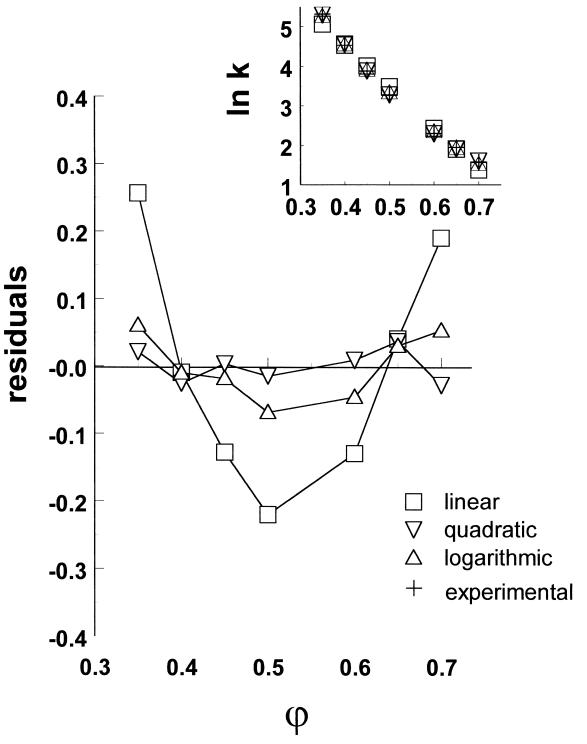

Fig. 2. Representative unconstrained regressions and corresponding residuals in the intermediate φ region, from the fitting of Eqs. (1), (3), and (11) to the measured ln k of *N*-nonylbenzamide.

N-ethylbenzamide were not reported [29]. The best-fit $\ln k_w$ values for the linear and the quadratic equations were 4.00 and 4.78 (linear) and 4.30 and 5.17 (quadratic), respectively. The experimental α value from the original retention data is 2.54. The corresponding calculated α values for the linear and quadratic models are 2.18 and 2.39, respectively. The SD model does not contain $\ln k_w$ as a model parameter.

It might be argued that this comparison gives no insight beyond that afforded by simple comparison of the χ_r^2 for the two models; clearly, the quadratic model has smaller residuals than does the linear model. However, testability is a fundamental tenet of the scientific method; whenever possible, testing a model by direct comparison to experiment is desirable. The limited testing thus far described suggests that both the linear and the quadratic models fit the experimental data fairly well, although the quadratic model offers slightly better performance than the linear model.

The advantage of evaluating the fits to the retention data of a relatively significant number of homologs is that the consistency of the sets of fitted parameters to the underlying physicochemical assumptions of each model can provide additional insights into the appropriateness of the models.

For example, the *S* parameter in the linear model is expected to roughly increase with increasing solute size [12–14]. The best-fit *S* values in the linear equation do not exhibit an observable trend but instead fluctuate between 7.5 to 14.5 (see Table 1). This behavior is inconsistent with the idea that *S* increases with increasing solute size [12–14]. The earliest applications of the linear model indicated

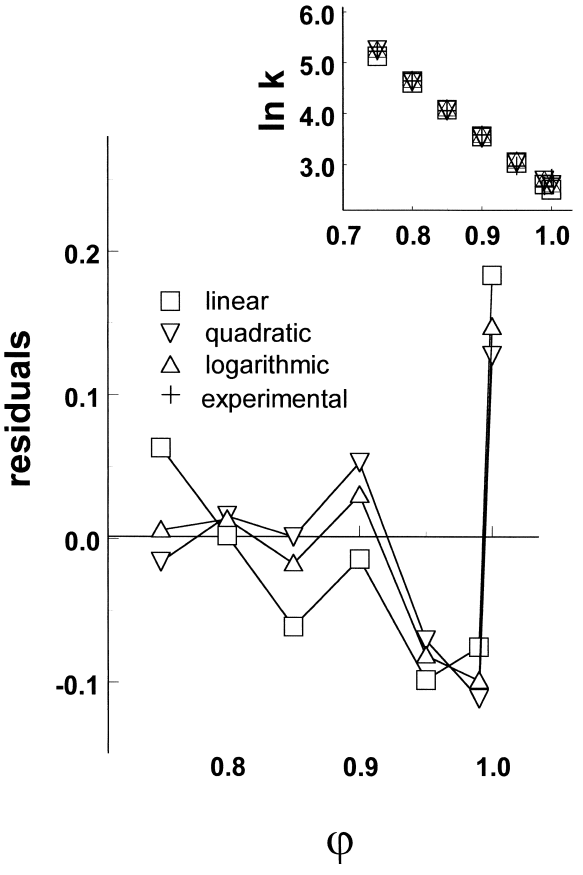

Fig. 3. Representative unconstrained regressions and corresponding residuals in the high φ region, from the fitting of Eqs. (1), (3), and (11) to the measured ln k of *N*-octadecylbenzamide.

that S was approximately independent of the solute, more consistent with the results seen here [8,9].

Similarly, according to the solubility parameter model, A is a function of solute size and should increase with increasing solute size while, according to the interphase model, A should be independent of solute in a given mobile phase mixture, i.e., it should be constant. In the modified version of the interphase model where the size correction factor is included [23,24], A is predicted to increase with increasing solute size also. However, as reported in Table 1, the best-fit A values decrease with increasing solute size, albeit with considerable fluctuations, in contradiction to the predictions of either model.

According to Eq. (5), B should become more negative as the size of the solute increases. Although δ_i for solutes are unknown, they are expected to decrease slowly as solute size increases within a homologous series as, for example, benzene ($\delta_i = 9.19$), toluene ($\delta_i = 9.09$), and ethylbenzene ($\delta_i = 8.96$) [16]. Thus, B is expected to become more negative as the size grows. This trend was not observed in our best-fit B values which, as shown in Table 1, were essentially constant.

A short communication by Kowalska and Prus points out the excellent applicability of the solubility parameter model, even when adsorption is the major retention mechanism [35]. As the solubility parameter model is based on a partitioning mechanism, this excellent fit must necessarily be considered fortuitous. Again, these workers do not attempt to relate the extracted parameters to the underlying assumptions of the model.

According to Eq. (11) and the assumptions made in the SD model [25], Z should increase with the surface area of the solute. This is observed in our best-fit Z values with only one exception, where *N*-dodecylbenzamide has a higher best-fit Z value than does *N*-tridecylbenzamide.

A study by Sadlej-Sosnowska and Sledzinska found that among the models they compared (which includes the linear and the quadratic equation but not the logarithmic equation), only the quadratic model fits the data satisfactorily for all five steroid hormones [6]. All other models in their study are unsatisfactory for one or more solutes. The criterion they employed is to examine the statistical improvement by adding an higher order term to a model equation, e.g., testing if the fit would statistically improve by adding a third-order term in the quadratic model. It should be noted that their retention data is not taken below 30% methanol or below 18% acetonitrile.

A recent comparison study where r^2 is used as the criterion includes a logarithmic model [5]. The retention data of small solutes including deoxyribonucleosides are fitted by the model equations compared. This logarithmic model performs most poorly in the low φ region (0.05–0.3 for methanol; 0.05–0.15 for acetonitrile), which is in agreement with what was found in the current study. In that same study [5], the quadratic model performs better than all other evaluated models compared over a wide range of φ, $\varphi = 0.05$ to $\varphi = 0.7$ for methanol–

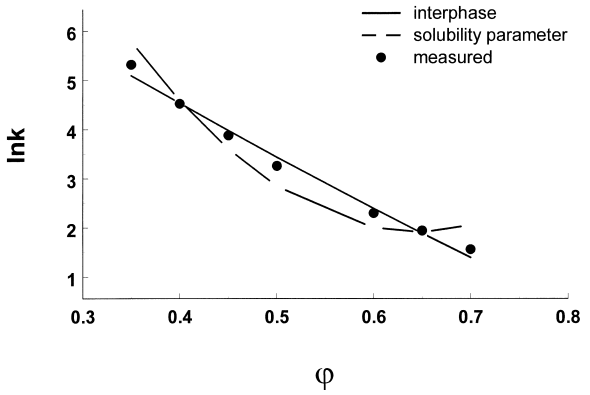

Fig. 4. Representative constrained regression of the interphase and solubility parameter models to the measured $\ln k$ (solid circles) of *N*-nonylbenzamide.

water mobile phase and $\varphi = 0.05$ to $\varphi = 0.5$ for acetonitrile–water mobile phase.

Neither of the above-mentioned studies [5,6], however, discusses any relationships between the best-fit parameters and the underlying assumptions of the models.

3.2. Constrained fitting

Fig. 4 compares typical constrained fittings. In these cases, the parameter *A* was first calculated according to the appropriate definition (Eqs. (4) or (7)). Then the *B* and *C* model parameters were determined by fitting to the data. The results of these constrained fittings are shown in Table 2. The χ^2_r values for the constrained fits to the interphase model were excellent for the solutes with alkyl moieties > propyl, and acceptable for the *N*-methyl, ethyl, and propyl homologs. The χ^2_r values for the constrained fits to the solubility parameter model were smaller for the *N*-methyl, ethyl, and propyl homologs than for the interphase model, but the fits of all other homologs gave larger χ^2_r values. Overall, the average χ^2_r values for the constrained fits to the solubility

Table 2
Constrained fitting results for two quadratic models (Eq. (12))[a]

CN[b]	Solubility parameter model					Interphase model, calc. $A = 2.37$[c]		
	Calc. A[d]	Best fit B	Best fit C	δ_i[e]	χ^2_r	Best fit B	Best fit C	χ^2_r
1	21.1	−19.5	4.15	19.2	0.058	−14.3	4.02	0.223
2	24.1	−21.4	4.91	19.4	0.069	−15.2	4.80	0.206
3	27.1	−22.8	5.85	19.7	0.008	−13.7	5.32	0.110
4	30.1	−27.3	7.52	19.3	0.051	−12.0	5.79	0.035
5	33.0	−31.7	9.44	19.0	0.049	−13.4	6.99	0.031
6	36.0	−40.6	12.8	18.0	0.148	−11.9	7.16	0.037
7	38.8	−45.4	15.0	17.7	0.086	−12.6	8.03	0.036
8	41.9	−50.8	17.5	17.5	0.061	−13.3	8.84	0.012
9	44.8	−57.7	20.5	17.1	0.201	−13.0	9.37	0.033
10	47.8	−63.2	23.2	16.9	0.146	−13.3	9.97	0.021
11	50.7	−72.9	28.2	16.2	0.401	−12.6	10.0	0.025
12	53.6	−79.2	31.6	16.0	0.345	−12.5	10.4	0.010
13	56.6	−96.3	42.0	14.7	1.070	−11.4	10.1	0.025
14	59.7	−105	47.2	14.4	0.840	−11.8	10.7	0.017
16	65.5	−118	55.2	14.1	0.685	−12.8	12.3	0.012
17	68.5	−127	60.9	13.8	0.444	−13.3	13.0	0.011
18	71.4	−136	66.9	13.5	0.253	−14.0	13.9	0.012
19	74.3	−141	70.0	13.5	0.285	−14.6	14.8	0.012

[a] The retention data are from Wells and Clark [29] and are for the retention of *N*-alkylbenzamides on an ODS column with water–acetonitrile mobile phases.

[b] Carbon number of the alkyl group on the *N*-alkylbenzamide.

[c] Calculated from Eq. (7), as explained in the text.

[d] Calculated from Eq. (4), as explained in the text.

[e] δ_i (in $(cal/cm^3)^{1/2}$) is extracted from the best-fit *B* value according to Eq. (5).

parameter and the interphase models were 0.289 and 0.048, respectively.

As shown in Fig. 4, the solubility parameter model predicted more curvature in $\ln k$ than experimentally observed, whereas the interphase model predicted more linear behavior than observed. From this observation, it appears that A given by Eq. (4) overestimates and A given by Eq. (7) underestimates the best value.

The overestimation of the solubility parameter model A parameter has been noted and various corrections to these values, including empirical corrections, have been suggested [17]. No such corrections were attempted in this study. However, the calculated molar volumes in Table 3 are expected to be about 70% of the actual volumes, as discussed in Section 2. Thus, the calculated A values should be lower than implied by Eq. (4). This underestimation of the A parameter should improve the fitting. We note that the intrinsic molar volume of Leahy et al. [36], which is utilized by Kamlet et al. [37] and by Hsieh and Dorsey [38], also tends to be lower than the actual molar volume. The intrinsic molar volumes of several small molecules in the data tables of Refs. [37,38] and the molar volumes calculated by AM1 geometry optimization were compared (data not shown). These were found to be similar, and thus other methods to calculate the molar volumes were not pursued.

Again, the experimentally determined $\ln k_w$ values for *N*-methylbenzamide and *N*-ethylbenzamide are 4.506 and 5.439, respectively [29], and the experimental α value is 2.54. The corresponding best-fit values for the solubility parameter model are 4.15,

Table 3
Calculated model parameters and extracted $\ln k_w$ for the LSS and the interphase models[a]

CN[b]	AM1 geometry optimization		Calculated free energy of solvation		LSS model		Interphase model	
	Molecular volume (Å^3)	Molar volume (cm^3)	In ACN[c]	In water[c]	S[d]	$\ln k_w$[e] (S.D.)	B[f]	C[g] (S.D.)
1	143	86.2	−8.3	−9.1	−1.1	2.4 (1.67)	−1.3	2.6 (1.48)
2	163	98.4	−8.9	−9.0	−0.2	3.0 (1.68)	−2.2	3.2 (1.49)
3	184	111	−9.5	−9.0	0.9	3.1 (1.44)	−3.3	3.5 (1.26)
4	204	123	−10.0	−8.8	2.1	3.3 (1.07)	−4.4	3.7 (0.94)
5	224	135	−10.6	−8.7	3.3	4.2 (0.95)	−5.6	4.7 (0.85)
6	244	147	−11.1	−8.5	4.4	4.6 (0.74)	−6.8	5.1 (0.69)
7	263	158	−11.7	−8.3	5.6	5.4 (0.56)	−7.9	6.0 (0.53)
8	284	171	−12.2	−8.2	6.8	6.3 (0.41)	−9.2	6.9 (0.40)
9	303	183	−12.7	−8.3	7.4	7.1 (0.45)	−9.8	7.7 (0.45)
10	323	195	−13.3	−8.0	9.0	8.3 (0.23)	−11.3	8.9 (0.24)
11	344	207	−13.8	−7.6	10.6	9.7 (0.20)	−12.9	10.3 (0.14)
12	363	219	−14.3	−7.4	11.7	10.9 (0.26)	−14.1	11.4 (0.18)
13	383	231	−14.8	−7.2	12.8	12.8 (0.87)	−15.2	13.1 (0.65)
14	404	243	−15.4	−7.1	14.1	14.3 (0.89)	−16.4	14.6 (0.67)
16	443	267	−16.4	−6.9	16.0	16.7 (0.91)	−18.4	17.0 (0.71)
17	464	279	−17.0	−6.8	17.2	18.2 (0.89)	−19.6	18.5 (0.70)
18	484	291	−17.5	−6.6	18.4	19.8 (0.83)	−20.8	20.0 (0.66)
19	503	303	−18.1	−6.9	18.8	20.4 (0.80)	−21.2	20.7 (0.63)

[a] The retention data are from Wells and Clark [29] and are for the retention of *N*-alkylbenzamides on an ODS column with water–acetonitrile mobile phases.

[b] Carbon number of the alkyl group on the *N*-alkylbenzamide.

[c] Calculated standard free energies of solvation in corresponding solvent, as explained in the text.

[d] Estimated S from Eq. (2), as explained in the text.

[e] The average extracted $\ln k_w$ using the estimated S and the measured $\ln k$.

[f] Estimated B from Eq. (8), as explained in the text.

[g] The average extracted C value using the estimated A (=2.37), estimated B, and the measured $\ln k$.

4.91, and 2.13 (α) and for the interphase model, 4.02, 4.08, and 2.18. Thus, the C parameters extracted from the solubility parameter model were in somewhat better agreement with the experimental values.

With constrained fitting, the best-fit B of the solubility parameter model (see Table 2) behaved as expected from the assumptions of the model, i.e., B grew more negative as the size of the solute increases. From this best-fit B value, each solute's solubility parameter δ_i was extracted and is also shown in Table 2. The extracted δ_i values showed the decreasing trend expected. While the solubility parameters of the *N*-alkylbenzamides are not available in the literature and those shown in Table 2 appear large, the δ values of formamide and dimethylformamide are 17.9 and 11.5, respectively [39].

Again, with constrained fitting, the best-fit B values of the interphase model (Table 2) were nearly constant whereas the estimated B values (Table 3) from the calculated free energy of transfer grew more negative as the size of the solute increased, in agreement with the trend expected from the interphase model's definition of the parameter B.

Thus, although for constrained fitting, the χ^2_r generally favors the interphase model, examination of the self-consistency of the extracted B parameters and direct comparison to the available experimental $\ln k_w$ values and α values favor the solubility parameter model.

3.3. Estimating S of the LSS model

If the S value in the LSS model can be estimated for a solute using Eq. (2), then $\ln k_w$ can be obtained from a single retention measurement at any mobile phase composition and the retention factor at any mobile phase composition can then be predicted. As shown in Table 3, the S values estimated from the free energy of transfer calculation increased with increasing solute size, as expected. The $\ln k_w$ value of each solute was extracted using the estimated S and the measured retention data. This estimated S, the average $\ln k_w$ (because more than one $\ln k$ value per solute was available) and the standard deviation of $\ln k_w$ are given in Table 3. The LSS model predicts $\ln k_w$ of 2.4 and 3.0 for *N*-methylbenzamide and *N*-ethylbenzamide, respectively, and an α of 1.8. Again, given that the corresponding experimental data are 4.506, 5.439, and 2.54, respectively, this approach was less successful than the constrained fits to either of the quadratic models.

3.4. Estimating A and B of the interphase model

The C value ($=\ln k_w$) of the interphase model can be predicted from a single retention factor measurement at any mobile phase composition, if A and B are known. Thus, the C value of each solute can be extracted from retention data using the A and B values estimated as described above, and the measured retention data. Again, since there are multiple retention data for each solute, there are multiple extracted C values for each solute and Table 3 shows the average extracted C and its standard deviation. The $\ln k_w$ values that can be directly compared to experimental data are 2.6 and 3.2, and the predicted α was 1.8. These values are in only slightly better agreement with the experimental values than are the LSS predicted values.

4. Concluding remarks

In this work, four popular retention models were examined by fitting to the retention data of a set of homologous compounds. When the fittings were not constrained, all four models fit the experimental retention data of single solutes well. The extracted parameters, however, were not consistent with the underlying assumptions of the LSS, solubility parameter, and interphase models. Thus, the goodness-of-fit criteria were not sufficient for comparing these models. Direct comparison of predicted values to experimental values, evaluation of sets of extracted parameters for predicted trends, and, in particular, evaluation of constrained fits in which one or more parameters are obtained prior to the fitting, are all useful approaches when comparing retention models.

Acknowledgements

This work was supported by a Faculty Development Grant from the State System of Higher Educa-

tion of Pennsylvania and a Senate Fellowship and a Senate Research Award from Indiana University of Pennsylvania.

References

[1] C.H. Lochmuller, C. Reese, A.J. Aschman, S.J. Breiner, J. Chromatogr. A 656 (1993) 3.
[2] M. Jaroniec, J. Chromatogr. A 656 (1993) 37.
[3] R. Tijssen, P.J. Schoenmakers, M.R. Bohmer, L.K. Koopal, H.A.H. Billiet, J. Chromatogr. A 656 (1993) 135.
[4] A. Vailaya, Cs. Horváth, J. Chromatogr. A 829 (1998) 1.
[5] K.H. Row, J. Chromatogr. A 797 (1998) 23.
[6] N. Sadlej-Sosnowska, I. Sledzinska, J. Chromatogr. 595 (1992) 53.
[7] W. Prus, T. Kowalska, J. Planar Chromatogr. 9 (1996) 164.
[8] L.R. Snyder, J.W. Dolan, J.R. Gant, J. Chromatogr. 165 (1979) 3.
[9] L.R. Snyder, J.W. Dolan, J.R. Gant, J. Chromatogr. 165 (1979) 31.
[10] L.R. Snyder, in: Cs. Horváth (Ed.), High Performance Liquid Chromatography — Advances and Perspectives, Vol. 1, Academic Press, New York, 1980, pp. 207–320.
[11] L.R. Snyder, M.A. Stadalius, in: Cs. Horváth (Ed.), High Performance Liquid Chromatography — Advances and Perspectives, Vol. 4, Academic Press, New York, 1986, pp. 193–318.
[12] J.P. Larmann, J.J. DeStefano, A.P. Goldberg, R.W. Stout, L.R. Snyder, M.A. Stadalius, J. Chromatogr. 255 (1983) 163.
[13] M.A. Quarry, R.L. Grob, R.L. Snyder, Anal. Chem. 58 (1986) 907.
[14] M.A. Quarry, M.A. Stadalius, T.H. Mourey, R.L. Snyder, J. Chromatogr. 358 (1986) 1.
[15] A. Wang, L.C. Tan, P.W. Carr, J. Chromatogr. A 848 (1999) 21.
[16] R. Tijssen, H.A.H. Billiet, P.J. Schoenmakers, J. Chromatogr. 122 (1976) 185.
[17] P.J. Schoenmakers, H.A.H. Billiet, R. Tijssen, L. de Galan, J. Chromatogr. 149 (1978) 519.
[18] P.J. Schoenmakers, H.A.H. Billiet, L. de Galan, J. Chromatogr. 218 (1981) 261.
[19] P.J. Schoenmakers, H.A.H. Billiet, L. de Galan, J. Chromatogr. 282 (1983) 107.
[20] P. Jandeara, H. Colin, G. Guiochon, Anal. Chem. 54 (1982) 435.
[21] J.A. Marqusee, K.A. Dill, J. Chem. Phys. 85 (1986) 434.
[22] K.A. Dill, J. Phys. Chem. 91 (1987) 1980.
[23] P.T. Ying, J.G. Dorsey, K.A. Dill, Anal. Chem. 61 (1989) 2540.
[24] J.G. Dorsey, K.A. Dill, Chem. Rev. 89 (1989) 331.
[25] X. Geng, F.E. Regnier, J. Chromatogr. 296 (1984) 15.
[26] X.D. Geng, F.E. Regnier, Chromatographia 38 (1994) 158.
[27] M. Kunitani, D. Johnson, L.R. Snyder, J. Chromatgr. 371 (1986) 313.
[28] R.R. Drager, F.E. Regnier, J. Chromatogr. 359 (1986) 147.
[29] M.J.M. Wells, C.R. Clark, J. Chromatogr. 243 (1982) 263.
[30] M.J.S. Dewar, E.G. Zoebisch, E.F. Healy, J.J.P. Stewart, J. Am. Chem. Soc. 107 (1985) 3902.
[31] PC Spartan Plus Tutorial and User's Guide, version 1.0, Wavefuction, Irvine, CA.
[32] G.D. Hawkins, C.J. Cramer, D.G. Truhlar, J. Phys. Chem. B 101 (1997) 7147.
[33] G.D. Hawkins, D.A. Liotard, C.J. Cramer, D.G. Truhlar, J. Org. Chem. 63 (1998) 4305.
[34] G. D. Hawkins, D.A. Liotard, C. J. Cramer, D. G. Truhlar, OMNISOL-Version 1.0.1, University of Minnesota, Minneapolis, MN, 1997.
[35] T. Kowalska, W. Prus, Acta Chromatogr. 7 (1997) 210.
[36] D.E. Leahy, P.W. Carr, R.S. Perlman, R.W. Taft, M.J. Kamlet, Chromatographia 21 (1986) 473.
[37] M.J. Kamlet, R.M. Doherty, P.W. Carr, D. Mackay, M.H. Abraham, R.W. Taft, Environ. Sci. Technol. 22 (1988) 593.
[38] M.-M. Hsieh, J.G. Dorsey, J. Chromatogr. 631 (1993) 63.
[39] B.L. Karger, L.R. Snyder, Cs. Horváth, in: An Introduction to Separation Science, Wiley, New York, 1973.