ELSEVIER

Prediction of internal standards in reversed-phase liquid chromatography
III. Evaluation of an alternative solvation parameter model to correlate and predict the retention of ionizable compounds

Jianwei Li*

Transdermal Drug Delivery, 3M Drug Delivery Systems, 3M Center, Building 235-BE-45, St. Paul, MN 55144, USA

Abstract

This paper describes the results of the evaluation of an alternative solvation parameter model for ionizable compounds. The new model is described as $\text{Log}(k) = Int + rR_2 + s\pi_2^{\text{H}} + a\Sigma\alpha_2^{\text{H}} + b\Sigma\beta_2^{\text{H}} + mV_x + \dfrac{U}{1 + V10^{\mp(\text{pH}-\text{p}K)}}$. The first six terms are the usual solvation parameter equation for neutral solutes, and the last term represents the contribution to retention from the ionization of solutes. Retention data obtained for 30 solutes in acetonitrile/aqueous buffer mobile phases are used to evaluate the capability of the function using different pH/pK scales. Because the function is not linear, nonlinear least-squares analysis is used to perform the data processing. It is concluded that the model function describes similarly the retention of ionizable compounds to the literature model without the need to accurately measure the mobile phase pH and solute's pK. Accordingly, the function simplifies the application of linear solvation energy relationships (LSERs) to ionizable compounds, and allows us to easily predict their retention for chromatographic optimization.
© 2002 Elsevier Science B.V. All rights reserved.

Keywords: Linear solvation energy relationships; Retention prediction; Internal standard; Solvation parameter model

1. Introduction

The optimization of selectivity is a major step in chromatographic method development. Depending on the complexity of a separation system, this step can be accomplished by either logical approaches (such as computer simulations [1–4], statistical designs [5–8], and quantitative structure–retention relationships [8]) or empirical trial-and-error. Obviously, the more we can understand the change in retention with the solute's properties and separation conditions, the more efficient is the selectivity optimization.

In fact, the change in retention and thus selectivity for neutral solutes can sometimes be modeled using the quantitative structure–retention relationships such as linear solvation energy relationships (LSERs) [9] and other theories [10–17]. The LSER model is probably the most successful one, and has been

*Tel.: +1-651-737-0468; fax: +1-651-737-7918.
E-mail address: jli7@mmm.com (J. Li).

applied to a number of diverse separation systems (for example, RPLC [18–24], NPLC [25–28], MLC/MEKC [29–32], chiral separation [33], and CE [34]). This model correlates chromatographic retention with the solute's properties (as descriptors) and separation conditions (as coefficients). We have shown previously that the change in selectivity with mobile phase composition can be computed for ODS columns when the molecular descriptors can be determined or estimated [9]. The computed results were very consistent with the experimental ones. Furthermore, the prediction of internal standards using neutral solutes in two earlier publications [9,35] was primarily based on the selectivity optimization by the LSER approach. Thus, if the ionizable compounds are used as internal standard candidates, their retention must be well correlated with their molecular properties. It is also noted that LSERs have been applied to many other systems [36].

However, there are rather limited studies on the application of LSERs to ionizable solutes [37,38], although most pharmaceutical compounds are ionic or ionizable in nature. The first successful LSER model for the ionizable compounds, in our opinion, is recommended by Martí Rosés et al. and represented as follows [38]:

$$\text{Log } (k) = Int + rR_2 + s\pi_2^{\text{H}} + a\sum\alpha_2^{\text{H}} + b\sum\beta_2^{\text{H}}$$
$$+ mV_X + d\log[1 - D(1 - f)] \tag{1}$$

where Int, r, s, a, b, and m are the intercept and LSER coefficients; R_2, π_2^{H}, $\Sigma\alpha_2^{\text{H}}$, $\Sigma\beta_2^{\text{H}}$ and V_X are the solutes' descriptors, representing their physicochemical properties; d is a coefficient; D is a parameter to describe the degree of the ionization of the solute; and f is the retention ratio of the ionized over neutral forms of a solute. The solutes' descriptors are the excess molar refraction (R_2); dipolarity/polarizability (π_2^{H}); "overall" or "effective" hydrogen bond acidity ($\Sigma\alpha_2^{\text{H}}$) and basicity ($\Sigma\beta_2^{\text{H}}$); and McGowan characteristic molar volume (V_X). Moreover, $\log[1 - D(1 - f)]$ is used as a single term (or parameter) to describe the contribution or correction to the overall retention. Eq. (1) contains seven coefficients, and has successfully correlated the retention of neutral, acidic and basic compounds.

However, the calculation of the D parameter requires the accurate determination of the mobile phase pH and solute's dissociation constant (pK) [39]:

$$D = \frac{10^{\text{pH} - pK}}{1 + 10^{\text{pH} - pK}} \tag{2}$$

Several pH/pK scales have been established, depending on how they are measured [40–44]. In $_{\text{w}}^{\text{w}}\text{pH}/_{\text{w}}^{\text{w}}pK$ scale, both the calibration and measurement are carried out in aqueous environment; however, the measurement in $_{\text{w}}^{\text{s}}\text{pH}/_{\text{w}}^{\text{s}}pK$ scale is carried out in the mobile phase (but still calibrated in water). It is reported that the D parameter should be determined using the rigorous $_{\text{w}}^{\text{s}}\text{pH}/_{\text{w}}^{\text{s}}pK$ scale [39], and it is thus necessary to determine the $_{\text{w}}^{\text{s}}pK$ values for solutes of interest under the mobile phase condition. Although the $_{\text{w}}^{\text{s}}pK$ values can be estimated [41,44], their availability still limits the application of the LSER model to ionizable compounds.

The goal of the study is to develop and evaluate an alternative LSER model that does not need solutes' $_{\text{w}}^{\text{s}}pK$ values. This model can use the $_{\text{w}}^{\text{w}}pK$ values taken from the literature, thereby significantly simplifying the application of LSERs to ionizable compounds. A relatively thorough development of the model is given in the next section.

2. Theoretical

The retention of ionizable compounds by RPLC is usually described as follows [39]:

$$k = \frac{k_{\text{HA}} + k_{\text{A}}10^{\pm(\text{pH} - pK)}}{1 + 10^{\pm(\text{pH} - pK)}} \tag{3}$$

where k_{HA} and k_{A} are the retention factors of the neutral and ionized forms, respectively. The positive sign refers to acidic compounds and the negative sign to basic compounds (however, k_{HA} refers to the neutral unprotonated form of bases and k_{A} refers to the acidic protonated form). From Eq. (3), we can easily derive the following:

$$k = \frac{k_{\text{HA}} + k_{\text{A}}10^{\pm(\text{pH} - pK)}}{1 + 10^{\pm(\text{pH} - pK)}}$$
$$= k_{\text{HA}}\left[\frac{1 + \left(\dfrac{k_{\text{A}}}{k_{\text{HA}}}\right)10^{\pm(\text{pH} - pK)}}{1 + 10^{\pm(\text{pH} - pK)}}\right] \tag{4}$$

Then the logarithm of the retention factor (Eq. (4)) is given as:

$$\text{Log}(k) = \text{Log}\left\{ k_{\text{HA}}\left[\frac{1 + \left(\dfrac{k_{\text{A}}}{k_{\text{HA}}}\right)10^{\pm(\text{pH}-\text{p}K)}}{1 + 10^{\pm(\text{pH}-\text{p}K)}} \right] \right\}$$

$$= \text{Log}(k_{\text{HA}}) + \log\left[\frac{1 + \left(\dfrac{k_{\text{A}}}{k_{\text{HA}}}\right)10^{\pm(\text{pH}-\text{p}K)}}{1 + 10^{\pm(\text{pH}-\text{p}K)}} \right]$$

$$= \text{Log}(k_{\text{HA}}) + \log\left(\frac{1 + f10^{\pm(\text{pH}-\text{p}K)}}{1 + 10^{\pm(\text{pH}-\text{p}K)}} \right) \qquad (5)$$

where f is ratio $(k_{\text{A}}/k_{\text{HA}})$ of retention factors. It is reported that f can be approximately regarded as a constant for different acids and bases [39]. $\text{Log}(k_{\text{HA}})$ is the retention of the neutral form of a solute, and can then be described by the usual LSER equation:

$$\text{Log}(k_{\text{HA}}) = Int + rR_2 + s\pi_2^{\text{H}} + a\sum\alpha_2^{\text{H}} + b\sum\beta_2^{\text{H}} + mV_X \qquad (6)$$

Based on Eqs. (5) and (6), the retention of ionizable compounds is described by the LSER model as follows:

$$\text{Log}(k) = Int + rR_2 + s\pi_2^{\text{H}} + a\sum\alpha_2^{\text{H}} + b\sum\beta_2^{\text{H}}$$
$$+ mV_X + \log\left(\frac{1 + f10^{\pm(\text{pH}-\text{p}K)}}{1 + 10^{\pm(\text{pH}-\text{p}K)}} \right) \qquad (7)$$

The last term of Eq. (7) is the modification to the LSER equation due to the ionization of acids and bases, and it is represented by $d\log[1 - D(1-f)]$ in Eq. (1). The function in parentheses is a hyperbolic function. The logarithm of this function is a sigmoidal function. Thus, if the log() function is simulated

by a sigmoidal function, the following equation is obtained:

$$\text{Log}\left(\frac{1 + f10^{\pm(\text{pH}-\text{p}K)}}{1 + 10^{\pm(\text{pH}-\text{p}K)}} \right) = \frac{U}{1 + V10^{\mp(\text{pH}-\text{p}K)}} \qquad (8)$$

where U and V are two fitting coefficients. However, the negative refers to acids and the positive sign to bases. When Eqs. (7) and (8) are combined, the following equation is obtained:

$$\text{Log}(k) = Int + rR_2 + s\pi_2^{\text{H}} + a\sum\alpha_2^{\text{H}} + b\sum\beta_2^{\text{H}}$$
$$+ mV_X + \frac{U}{1 + V10^{\mp(\text{pH}-\text{p}K)}} \qquad (9)$$

This is the model to be evaluated in this study to correlate the retention of ionizable compounds. There are eight coefficients (or variables) in Eq. (9). It is noted that the model is not linear anymore, and the data processing will be different from the usual multivariate linear regression analysis, as will be explained later. Moreover, f value is assumed a constant in Eq. (9).

Eq. (8) indicates that the log () function can be described approximately by a sigmoidal function. To evaluate this assumption, we generated data by the left side of Eq. (8) at different f values. These data were then fitted by the right side of Eq. (8) (a sigmoidal function) via a nonlinear least-squares fitting. Table 1 summarizes the fitting results, and Fig. 1 visually illustrates the goodness-of-fits. It can be seen from both Table 1 and Fig. 1 that the fits are excellent, particularly at large f values. Although there is a slight discrepancy between the fitted and raw data at small f values (0.03 and 0.04), the sigmoidal function overall describes the log () function very well. It can then be concluded that the

Table 1
The fitting results of the right side of Eq. (8) to the data simulated with the left side

Coefficient	"f" value					
	0.03	0.04	0.05	0.1	0.2	0.3
U	−1.505	−1.385	−1.291	−0.995	−0.697	−0.522
	(0.004)	(0.003)	(0.003)	(0.001)	(0.000)	(0.001)
V	5.560	4.849	4.357	3.118	2.221	1.819
	(0.120)	(0.089)	(0.070)	(0.031)	(0.011)	(0.005)
χ^2	6.3E−4	3.9E−4	2.6E−4	6.0E−5	7.1E−6	1.3E−6

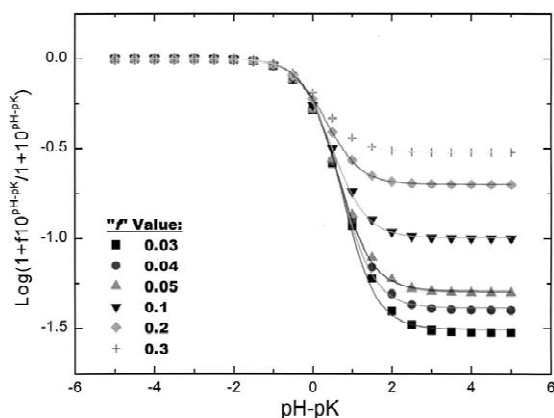The values in parentheses are the standard errors of the coefficients.

Fig. 1. Illustration of the fits with the sigmoidal equation (the right side of Eq. (8)) to the data generated by the log () function (the left side). The data (symbols) were generated by Excel in the $(pH - pK)$ range of -6 to 6 with an increment of 0.1. The fits are indicated by the solid lines. One out of every five symbols is plotted.

sigmoidal function can be used to model the retention of ionizable compounds.

It is emphasized at this point that the difference between Eqs. (1) and (9) is the mathematical form of the ionization term. It has been assumed that $d \log[1 - D(1 - f)]$ in Eq. (1) is equivalent to $[U/(1 + V10^{\mp(pH - pK)})]$ in Eq. (9). A comparison of the two terms indicates that $U = \log(f)$ and V is related to (or corrects) the change in $pH - pK$ when the mobile phase contains an organic solvent [39]. This change may be buffer-dependent.

As mentioned earlier, the mobile phase pH can be measured in any of the rigorous ${}_w^s pH$ and ${}_s^s pH$ (both calibration and measurement are carried out in the mobile phase condition) scales, and the pK values can be determined similarly. In Eq. (1), both ${}_w^s pH$ and ${}_w^s pK$ are needed to obtain the D parameter [39]. It is thus logical to assume that ${}_w^s pH$ and ${}_w^s pK$ values should be used in Eq. (9). However, the ${}_w^w pK$ values are readily available from the literature, and an ideal retention model should be able to use this parameter because the measurement of ${}_w^s pK$ is not needed. We expect that the V coefficient in the proposed model should be able to compensate for the shift between $({}_w^w pH - {}_w^w pK)$ and $({}_w^s pH - {}_w^s pK)$, and ${}_w^w pK$ values can be used in the model.

This study will evaluate the use of both ${}_w^s pH/{}_w^s pK$ and ${}_w^w pH/{}_w^w pK$ scales in Eq. (9) to correlate the retention of ionizable compounds with an emphasis on the latter.

3. Experimental

3.1. Chromatographic conditions

The retention data were obtained directly from Professor Martí Rosés at the University of Barcelona through private communication, and they were used to generate the results and conclusions presented in Ref. [39]. The chromatographic retention was measured on a polymer reversed-phase column (150 × 4.6 mm I.D.) using acetonitrile (ACN)/aqueous buffer as the mobile phase. The composition of ACN was 20, 40 and 60% (v/v), respectively. The mobile phase pH at each ACN composition was varied from 2 to 12. The solutes used and related properties are indicated in Table 2.

3.2. Linear regression and nonlinear least-squares analysis of retention data

Because Eq. (9) is not a linear equation anymore, the usual linear regression analysis cannot be applied. It is a nonlinear equation, and the nonlinear least-squares analysis should be used to determine the coefficients. Microsoft Solver, a very powerful routine based on the Marquardt–Levenberg algorithm, was used for the nonlinear least-squares analysis.

Three fitting schemes are used in this study, and they are shown in Table 3. In fitting scheme A, the linear regression analysis was carried out on the retention data of the neutral solutes (16 solutes listed in Table 2) at ${}_w^w pH$ of 2. The LSER coefficients obtained were then used as the initial estimates in the nonlinear fitting when all solutes (both neutral and ionizable) were included. Theoretically, the coefficients by the linear regression of 16 solutes should be consistent with those obtained for all solutes by the nonlinear analysis because the contribution to retention from the degree of ionization can be considered independent for ionizable compounds. However, this assumption was not used and the LSER coefficients were obtained by the best fits.

The initial estimates for U and V were set -1 and

Table 2
The properties of solutes used in the study

Solute	Assigned #	R_2	π_2^H	$\Sigma\alpha_2^H$	$\Sigma\beta_2^H$	V_X	$(_w^w pK)^a$	$_w^s pK$ (40% ACN)
Naphthoic acid	1	1.46	1.3	0.6	0.45	1.301	4.16	5.10
2-Nitrobenzoic acid	2	0.99	1.1	0	0.7	1.172	2.18	3.60
3-Nitrobenzoic acid	3	0.99	1.41	0.7	0.56	1.172	3.46	4.40
4-Nitrobenzoic acid	4	0.99	1.43	0.68	0.51	1.172	3.441	4.31
Benzoic acid	5	0.73	0.9	0.59	0.4	0.932	4.204	5.30
Resorcinol	6	0.98	1	1.1	0.58	0.834	9.32	10.99
Phenol	7	0.805	0.89	0.6	0.3	0.775	9.99	11.55
2,4-Dichlorophenol	8	0.96	0.84	0.53	0.19	1.02	7.95	8.88
2,4-Dinitrophenol	9	1.2	1.5	0.1	0.55	1.124	4.08	4.37
β-Naphthol	10	1.52	1.08	0.61	0.4	1.144	9.57	11.18
2-Nitrophenol	11	1.015	1.05	0.05	0.37	0.949	7.222	7.92
3,5-Dichlorophenol	12	1.02	1	0.91	0	1.02	8.179	9.33
3-Bromophenol	13	1.06	1.15	0.7	0.16	0.95	9.031	10.32
4-Chlorophenol	14	0.915	1.08	0.67	0.2	0.898	9.43	10.76
m-Cresol	15	0.822	0.88	0.57	0.34	0.916	10	11.59
Benzene	16	0.61	0.52	0	0.14	0.716	20[b]	20[b]
Acetophenone	17	0.818	1.01	0	0.48	1.014	20[b]	20[b]
Benzaldehyde	18	0.82	1	0	0.39	0.873	20[b]	20[b]
Nitrobenzene	19	0.871	1.11	0	0.28	0.891	20[b]	20[b]
Methylphenyl ether	20	0.708	0.75	0	0.29	0.916	20[b]	20[b]
Benzonitrile	21	0.742	1.11	0	0.33	0.871	20[b]	20[b]
2,4,6-Trimethylpyridine	22	0.634	0.69	0	0.6	1.098	7.43	6.58
4-Chloroaniline	23	1.06	1.13	0.3	0.31	0.939	3.99	3.11
Aniline	24	0.955	0.96	0.26	0.41	0.816	4.6	3.96
N-Ethylaniline	25	0.945	0.85	0.17	0.51	1.098	5.11	4.57
N,N-Dimethylbenzylamine	26	0.668	0.8	0	0.69	1.239	9.02	8.15
p-Toluidine	27	0.923	0.95	0.23	0.45	0.957	5.08	4.58
Pyridine	28	0.631	0.84	0	0.52	0.675	5.17	4.61
2,6-Dimethylaniline	29	0.972	0.89	0.2	0.46	1.098	3.95	3.22
3-Aminophenol(Amino)	30	1.13	1.15	0.65	0.78	0.875	4.37	3.68
3-Aminophenol(Phenol)	30	1.13	1.15	0.65	0.78	0.875	9.83	11.43

The data were taken from Ref. [39]. Solutes 6–21 are considered neutral solutes at pH 2, and then used to determine the initial estimates of LSER coefficients.

[a] Taken from Ref. [45]. They were obtained at 25 °C in water (ionic strength, $\mu = 0$).

[b] Arbitrarily assigned to 20.

4, respectively. Based on the initial estimates, the predicted log (*k*) was computed with Eq. (9), and the residual for each solute was computed as the difference between the predicted and measured log (*k*). The residuals were then squared and summed. The best fits were obtained by minimizing the sum of squares of residuals (SSR).

Microsoft Solver was then started with the default options. Three steps were followed during the fitting scheme. First, the variables *U* and *V* were varied. Then the other six LSER coefficients were optimized. Finally, all eight variables were allowed to change simultaneously. This procedure was repeated

for the retention data at different mobile phase $_w^w pH$. Although this procedure usually provided the best fits, the coefficients obtained may not sometimes be meaningful or consistent (at different pH, for example), depending on the quality of the retention data. Moreover, the quality of the data in general affects the consistency of the coefficients by the linear regression.

In the second fitting scheme B (Table 3), the LSER coefficients obtained by scheme A at different $_w^w pH$ were averaged at each ACN composition, and the averages were used as the initial estimates. However, the ratio of *b/m* was fixed during the

Table 3
Fitting procedures

Fitting scheme	Initial estimates	Procedure	Remarks
A	*Int*, *r*, *s*, *a*, *b* and *m* by linear regression on 16 neutral solutes at lowest pH $U = -1$ $V = 4$	Vary *U* and *V* Vary *Int*, *r*, *s*, *a*, *b* and *m* Vary *Int*, *r*, *s*, *a*, *b* and *m*, and *U* and *V*	All coefficients were allowed to vary
B	*Int*, *r*, *s*, *a*, *b* and *m* by averages of coefficients over all pH from scheme A *U* and *V* by fitting results of scheme A	Vary *U* and *V* Vary *Int*, *r*, *s*, *a*, *b* and *m* Vary *Int*, *r*, *s*, *a*, *b* and *m* and *U* and *V*	All coefficients were allowed to vary, but *b*/*m* ratio was fixed in last two steps
C	*Int*, *r*, *s*, *a*, *b* and *m* by averages of coefficients over all pH from scheme B *U* and *V* by fitting results by scheme B	Vary *U* and *V*	Only *U* and *V* coefficients were allowed to vary

second and third steps. The fixed ratio of b/m was implemented using a Solver constraint. This constraint improved the consistency and physical significance of the LSER coefficients.

The third fitting scheme C is very similar to scheme B, but all LSER coefficients were taken as the averages from the fitting scheme B, and they were fixed. Only U and V coefficients were allowed to vary.

It should be emphasized at this point that, although schemes B and C are somewhat complicated, they are designed based on the available data. If the retention of a number of the neutral solutes (e.g. 20) is collected together with the ionizable solutes, the retention of the neutral solutes can be used to obtain the LSER coefficients. The coefficients are then used as the initial estimates when all solutes are fitted together. It is believed that the proper selection of solutes and use of scheme B or C should be a good approach to obtaining both meaningful coefficients and acceptable statistics.

After each fitting scheme was completed, the predicted log (k) was regressed with the measured value to obtain the statistics. The regression statistics should be comparable to the results obtained by the linear regression analysis based on Eq. (1), and can be used to compare the quality of the fit with

different models. If a standardized residual was beyond ± 3, the solute was considered an outlier. The data point was removed, and the fitting scheme was repeated. Usually no more than two outliers were observed for each data set.

To further evaluate the independence of the $[U/(1 + V10^{\mp(\text{pH}-\text{p}K)})]$ term in Eq. (9), it was computed for each solute based on the final variable (U and V) values. This term was then plotted against pH $-$ pK (acids) or p$K -$ pH (bases) to examine if a sigmoidal shape was obtained. Furthermore, the ionization term was subtracted from log (k) [log $(k) - (U/1 + V10^{\mp(\text{pH}-\text{p}K)})$]. The resultant values should be the contribution to retention from neutral forms of the compounds, and were correlated as follows:

$$\left[\text{Log}(k) - \frac{U}{1 + V10^{\mp(\text{pH}-\text{p}K)}} \right] = Int + rR_2 + s\pi_2^H$$
$$+ a\sum \alpha_2^H + b\sum \beta_2^H + mV_X \qquad (10)$$

The correlation should be acceptable, and the obtained LSER coefficients should be meaningful. If Eq. (9) describes the retention data well, the shape between $[U/(1 + V10^{\mp(\text{pH}-\text{p}K)})]$ and $\mp(\text{pH} - \text{p}K)$ should be sigmoidal, and the correlation by Eq. (10) should be acceptable.

The fitting results and associated statistics are

shown in Tables 4, 5 and 6 using mobile phase ${}_{w}^{w}pH$ and solute ${}_{w}^{w}pK$ scale for the three ACN mobile phase compositions. Table 7 shows the results using mobile phase ${}_{w}^{s}pH$ and solute ${}_{w}^{s}pK$ scale at 40% ACN composition. It is noted that ${}_{w}^{w}pK$ and ${}_{w}^{s}pK$ values for the neutral solutes are arbitrarily assigned to 20 (Table 2).

Finally, Origin software (Microcal Software, Northampton, MA, USA) was used to perform the nonlinear least-squares curve-fitting for the data in Fig. 1.

4. Results and discussion

4.1. The independence of model terms

One of the assumptions of Eq. (9) is the mathematical separation of retention of ionizable compounds into two individual components: one due to the neutral unionized molecules and the correction to it due to ionization. The first component is described by the usual linear LSER equation (Eq. (6)), while the second part is described by the right side of Eq. (8), which is a sigmoidal function. Therefore, the contribution to retention due to ionization is relatively independent because the last term may not be convoluted significantly with the other terms. Consequently, this may improve the significance and consistency of the LSER coefficients.

To confirm that Eq. (9) works in the way anticipated, Fig. 2 shows the fitting results for 40% ACN at ${}_{w}^{w}pH$ of 3 (fitting scheme A). ${}_{w}^{w}pH$ and ${}_{w}^{w}pK$ were used to carry out the data analysis. Fig. 2A is a plot of the predicted $\log(k)$ against measured $\log(k)$ to illustrate the overall fit quality by the model equation. It can be clearly seen from the figure that the predicted retention based on Eq. (9) is linearly correlated very well with the measured. The slope of the correlation is essentially unity (close to 45° line with a slope $+1$ through (0,0)). Thus, the estimated model function gives an accurate prediction of the values actually observed, and Fig. 2A provides a visual assessment of model effectiveness in making predictions. Table 5 shows the model coefficients and related statistics. It can be seen in Table 5 that

the correlation coefficient is 0.985 and the standard error (SE) is 0.095, thereby confirming the very good fit by Eq. (9).

Fig. 2B shows the linear fitting results when $\left[\log(k) - (U/1 + V10^{pH-pK})\right]$ is correlated with LSER parameters (Eq. (10)). All data points are again positioned along the diagonal line, indicating the effectiveness of the model. The correlation coefficient is 0.966, and the SE is 0.104 (Table 5). It is noted that the SE is about 9% larger than that in Fig. 2A, thereby indicating that about 91% of the model error is contributed by the LSER coefficients.

Fig. 2C demonstrates the dependence of $(U/1 + V10^{\mp(pH-pK)})$ versus ${}_{w}^{w}pH - {}_{w}^{w}pK$ (or ${}_{w}^{w}pK - {}_{w}^{w}pH$ for bases). It can be clearly seen that the shape is exactly sigmoidal. Overall, Fig. 2 clearly shows that the retention model of Eq. (9) works as expected.

4.2. Comparison of three fitting schemes

Three fitting schemes are evaluated in this study, and the purpose is to evaluate which one provides both the best fit and meaningful LSER coefficients. Fig. 3 shows the comparison of the three fitting methods (40% ACN). Fig. 3A is a plot of the SE from the neutral component (Eq. (10)) against the overall SE when all coefficients are allowed to vary (scheme A). It is seen in Fig. 3A that there is a perfect linear relationship, and the slope is 1.11 (about 11% larger). This indicates that the neutral component of the model contributes to about 89% of the error, the remaining 11% is from the ionization term.

Fig. 3B is a plot of SE by keeping b/m ratio constant (-1.119, Table 5, scheme B) against SE obtained by scheme A. The slope is about 1.05, indicating that the model errors are very similar by the two schemes. Fig. 3C is a plot of SE by scheme C, and the slope is 1.12, thereby indicating that the model error is about 12% more on average.

It can be concluded from Fig. 3 that the model errors with scheme A and B are similar, while scheme C offers the largest model error due obviously to the rigidity of the scheme. It is noted that SE is used to compare the quality of the fits because it is a more sensitive parameter than the correlation coefficient.

Table 4
Nonlinear least-squares fitting results at 20% ACN using ^w_wpH and ^w_wpK

^w_wpH	2	3	3.99	5	6.01	7.01	8.02	10.03	11.01	12.04	Average
^s_wpH	2.07	3.24	4.31	5.38	6.49	7.43	8.41	9.78	10.84	12.38	
^s_spH	2.1	3.27	4.34	5.41	6.52	7.46	8.44	9.81	10.87	12.41	
All parameters allowed to change											
Int	0.698	0.723	0.658	0.045	−0.235	−0.062	−0.175	0.095	−0.020	0.093	0.182
r	0.612	0.933	0.693	−0.114	−0.181	−0.009	−0.044	0.073	0.153	0.183	0.230
s	0.038	−0.203	−0.788	−0.287	0.066	0.160	0.215	0.130	0.152	0.233	−0.028
a	−0.954	−0.893	−0.923	−0.696	−0.705	−0.741	−0.746	−0.620	−0.613	−0.500	−0.739
b	−2.633	−2.582	−2.552	−2.389	−2.546	−2.815	−2.521	−2.693	−2.698	−2.398	−2.583
m	1.498	1.361	2.250	3.063	3.096	2.715	2.710	2.446	2.464	2.122	2.373
U	−1.705	−1.449	−1.810	−2.078	−1.976	−1.868	−1.870	−1.579	−1.896	−1.913	$b/m =$
V	1.709	14.083	18.430	6.929	14.308	46.481	13.822	29.600	67.669	41.731	−1.089
R	0.989	0.985	0.982	0.985	0.982	0.975	0.971	0.967	0.973	0.977	0.979
R^2	0.978	0.970	0.964	0.970	0.965	0.950	0.942	0.935	0.947	0.954	0.958
Adjusted R^2	0.977	0.969	0.963	0.969	0.964	0.948	0.940	0.933	0.945	0.953	0.956
Standard error	0.143	0.146	0.147	0.147	0.165	0.197	0.204	0.190	0.198	0.212	0.175
Observations	29	30	30	30	30	29	30	30	30	30	
$(R)^a$	0.949	0.957	0.963	0.961	0.954	0.915	0.924	0.931	0.928	0.900	0.938
$(R^2)^a$	0.900	0.916	0.928	0.924	0.910	0.837	0.855	0.867	0.860	0.810	0.881
Adjusted $(R^2)^a$	0.878	0.899	0.913	0.908	0.891	0.801	0.824	0.839	0.831	0.771	0.856
Standard error^a	0.156	0.160	0.162	0.161	0.181	0.219	0.227	0.212	0.220	0.234	0.193
All parameters allowed to change, fixed b/m ratio											
Int	0.353	0.478	0.634	0.373	0.224	0.005	0.071	0.078	−0.028	0.047	0.223
r	0.479	0.633	0.667	0.136	0.115	0.033	0.078	0.065	0.149	0.162	0.252
s	−0.173	−0.447	−0.812	−0.160	−0.043	0.161	0.189	0.130	0.151	0.226	−0.078
a	−0.913	−0.925	−0.921	−0.716	−0.768	−0.751	−0.778	−0.620	−0.612	−0.492	−0.750
b	−2.239	−2.237	−2.518	−2.617	−2.737	−2.849	−2.609	−2.684	−2.695	−2.382	−2.557
m	2.057	2.055	2.313	2.404	2.514	2.617	2.397	2.466	2.476	2.188	2.349
U	−1.810	−1.828	−1.847	−1.772	−1.754	−1.852	−1.792	−1.582	−1.897	−1.919	
V	1.044	32.748	18.671	4.980	12.921	51.981	15.657	30.099	67.977	42.811	
R	0.983	0.989	0.982	0.990	0.980	0.975	0.970	0.967	0.973	0.977	0.979
R^2	0.967	0.977	0.964	0.980	0.961	0.950	0.940	0.935	0.947	0.954	0.958
Adjusted R^2	0.966	0.977	0.963	0.980	0.960	0.948	0.938	0.933	0.945	0.953	0.956
Standard error	0.173	0.129	0.147	0.120	0.173	0.197	0.207	0.190	0.198	0.212	0.175
$(R)^a$	0.941	0.969	0.963	0.973	0.949	0.914	0.922	0.931	0.928	0.901	0.939
$(R^2)^a$	0.885	0.938	0.928	0.946	0.901	0.836	0.850	0.867	0.861	0.811	0.882
Adjusted $(R^2)^a$	0.860	0.925	0.913	0.934	0.880	0.800	0.818	0.839	0.832	0.772	0.857
Standard error^a	0.165	0.137	0.162	0.131	0.187	0.219	0.228	0.212	0.220	0.234	0.190
U and V parameters allowed to change, fixed Int, s, a, b and m											
U	−1.730	−1.747	−1.693	−1.767	−1.762	−1.823	−1.773	−1.550	−1.843	−1.806	
V	1.271	39.291	13.607	4.616	12.847	38.705	13.506	35.352	73.678	50.648	
R	0.981	0.984	0.967	0.990	0.980	0.973	0.970	0.965	0.972	0.974	0.976
R^2	0.962	0.968	0.935	0.980	0.961	0.946	0.940	0.932	0.944	0.949	0.952
Adjusted R^2	0.961	0.967	0.933	0.980	0.960	0.944	0.938	0.929	0.942	0.947	0.950
Standard error	0.185	0.154	0.194	0.120	0.173	0.204	0.207	0.198	0.209	0.231	0.188
$(R)^a$	0.947	0.969	0.960	0.972	0.949	0.915	0.922	0.929	0.926	0.900	0.939
$(R^2)^a$	0.896	0.940	0.921	0.945	0.901	0.836	0.850	0.863	0.858	0.810	0.882
Adjusted $(R^2)^a$	0.874	0.926	0.905	0.933	0.881	0.801	0.819	0.835	0.829	0.770	0.857
Standard error^a	0.158	0.139	0.166	0.132	0.187	0.219	0.228	0.212	0.221	0.238	0.190

^{a} Regression statistics from Eq. (10).

Table 5
Nonlinear least-squares fitting results at 40% ACN using ^w_wpH and ^w_wpK

^w_wpH	1.99	2.99	3.99	5.03	6.04	7.01	7.99	10.01	11.03	12	Average
^s_wpH	2.2	3.53	4.7	5.99	6.89	7.8	8.62	9.52	10.73	12.7	
^s_spH	2.34	3.67	4.84	6.13	7.03	7.94	8.76	9.66	10.87	12.84	
All parameters allowed to change											
Int	0.526	0.365	0.791	0.021	−0.299	−0.089	−0.158	0.029	−0.061	0.074	0.120
r	0.393	0.444	0.623	−0.161	−0.342	−0.120	−0.146	0.141	0.107	0.101	0.104
s	−0.129	−0.420	−0.652	−0.493	−0.155	−0.008	−0.008	−0.010	0.032	0.075	−0.177
a	−0.755	−0.790	−0.768	−0.618	−0.578	−0.599	−0.578	−0.600	−0.437	−0.277	−0.600
b	−1.904	−1.609	−1.653	−1.309	−1.441	−1.755	−1.543	−1.855	−1.761	−1.147	−1.598
m	0.796	1.148	0.757	2.063	2.228	1.723	1.732	1.408	1.442	0.981	1.428
U	−1.162	−1.299	−1.246	−1.549	−1.476	−1.385	−1.350	−1.271	−1.520	−1.451	$b/m =$ −1.119
V	31.856	81.930	17.968	5.216	13.014	85.270	14.640	149.674	147.481	33.100	
R	0.988	0.985	0.981	0.976	0.979	0.966	0.961	0.970	0.972	0.969	0.975
R^2	0.976	0.970	0.963	0.953	0.958	0.933	0.923	0.941	0.944	0.938	0.950
Adjusted R^2	0.975	0.969	0.962	0.951	0.956	0.931	0.920	0.939	0.942	0.936	0.948
Standard error	0.097	0.095	0.107	0.136	0.130	0.164	0.172	0.150	0.161	0.182	0.139
Observations	29	29	29	30	30	29	30	30	30	30	
$(R)^a$	0.955	0.966	0.959	0.934	0.940	0.866	0.879	0.913	0.889	0.750	0.905
$(R^2)^a$	0.912	0.934	0.921	0.872	0.884	0.749	0.773	0.833	0.790	0.563	0.823
Adjusted $(R^2)^a$	0.893	0.920	0.903	0.846	0.860	0.695	0.726	0.798	0.747	0.472	0.786
Standard error^a	0.107	0.104	0.118	0.151	0.143	0.184	0.193	0.168	0.179	0.203	0.155
All parameters allowed to change, fixed b/m ratio											
Int	0.358	0.287	0.629	0.467	0.275	0.005	0.060	−0.130	−0.137	0.049	0.186
r	0.406	0.354	0.522	0.200	0.025	−0.069	−0.035	0.069	0.069	0.090	0.163
s	−0.413	−0.494	−1.009	−0.554	−0.288	−0.020	−0.029	0.007	0.028	0.071	−0.270
a	−0.724	−0.781	−0.769	−0.679	−0.657	−0.614	−0.606	−0.581	−0.434	−0.272	−0.612
b	−1.086	−1.520	−1.511	−1.561	−1.675	−1.798	−1.619	−1.775	−1.729	−1.139	−1.541
m	0.970	1.358	1.351	1.395	1.497	1.606	1.447	1.586	1.545	1.018	1.377
U	−1.327	−1.358	−1.393	−1.271	−1.233	−1.327	−1.266	−1.252	−1.426	−1.325	
V	77.340	93.245	14.806	4.397	11.872	78.752	14.242	178.072	224.352	53.498	
R	0.983	0.984	0.962	0.972	0.973	0.966	0.959	0.969	0.972	0.969	0.971
R^2	0.966	0.969	0.926	0.945	0.946	0.933	0.920	0.939	0.944	0.938	0.943
Adjusted R^2	0.965	0.967	0.924	0.943	0.944	0.931	0.918	0.937	0.942	0.936	0.941
Standard error	0.121	0.097	0.148	0.147	0.146	0.165	0.174	0.153	0.161	0.182	0.149
$(R)^a$	0.938	0.966	0.933	0.923	0.926	0.865	0.875	0.911	0.890	0.752	0.898
$(R^2)^a$	0.881	0.933	0.871	0.852	0.858	0.749	0.765	0.830	0.792	0.565	0.810
Adjusted $(R^2)^a$	0.855	0.919	0.845	0.821	0.828	0.694	0.716	0.795	0.748	0.474	0.770
Standard error^a	0.120	0.106	0.165	0.160	0.155	0.184	0.194	0.168	0.179	0.203	0.164
U and V parameters allowed to change, fixed Int, s, a, b and m											
U	−1.327	−1.358	−1.393	−1.271	−1.233	−1.327	−1.266	−1.252	−1.426	−1.325	
V	77.340	93.245	14.806	4.397	11.872	78.752	14.242	178.072	224.352	53.498	
R	0.977	0.978	0.955	0.982	0.971	0.966	0.958	0.965	0.967	0.959	0.968
R^2	0.955	0.956	0.913	0.964	0.942	0.933	0.917	0.930	0.935	0.921	0.937
Adjusted R^2	0.953	0.954	0.910	0.963	0.940	0.931	0.914	0.928	0.932	0.918	0.934
Standard error	0.145	0.112	0.161	0.115	0.148	0.165	0.181	0.163	0.177	0.219	0.159
$(R)^a$	0.943	0.967	0.952	0.953	0.928	0.866	0.877	0.910	0.881	0.771	0.905
$(R^2)^a$	0.889	0.934	0.907	0.908	0.861	0.751	0.769	0.828	0.776	0.594	0.822
Adjusted $(R^2)^a$	0.865	0.920	0.887	0.888	0.832	0.697	0.721	0.792	0.730	0.510	0.784
Standard error^a	0.117	0.105	0.125	0.119	0.153	0.184	0.194	0.168	0.183	0.211	0.156

[a] Regression statistics by Eq. (10).

Table 6
Nonlinear least-squares fitting results at 60% ACN using $_w^w$pH and $_w^w$pK

$_w^w$pH	2.01	3.00	4.00	5.02	6.02	7.02	8.01	10.01	11.02	12.00	Average
$_w^s$pH	2.24	3.77	5.13	6.35	7.11	8.02	8.99	9.36	10.42	13.19	
$_s^s$pH	2.70	4.23	5.59	6.81	7.57	8.48	9.45	9.82	10.88	13.65	
All parameters allowed to change											
Int	0.354	0.084	0.664	0.437	−0.624	−1.028	−0.377	−0.235	−0.090	−0.552	−0.137
r	0.302	0.327	0.533	0.547	−0.597	−0.035	0.251	0.576	0.501	−0.084	0.232
s	0.055	−0.107	−0.846	−1.450	−0.165	0.349	−0.090	−0.138	−0.228	0.385	−0.224
a	−0.746	−0.645	−0.708	−0.866	−0.565	−0.738	−0.750	−0.834	−0.814	−0.616	−0.728
b	−1.739	−1.678	−1.211	−0.693	−0.903	−1.114	−1.137	−1.347	−1.187	−1.797	−1.281
m	0.302	0.640	0.500	1.280	2.135	1.610	1.081	0.794	0.716	1.379	1.044
U	−0.992	−0.875	−1.093	−1.365	−1.727	−2.292	−1.669	−1.684	−2.371	−1.818	b/m = −1.227
V	99.603	4917.267	38.371	3.667	22.859	208.704	47.488	827.099	1675.971	41.270	
R	0.982	0.944	0.926	0.951	0.969	0.968	0.967	0.973	0.979	0.988	0.965
R^2	0.965	0.891	0.857	0.905	0.939	0.938	0.936	0.946	0.958	0.977	0.931
Adjusted R^2	0.964	0.887	0.852	0.901	0.937	0.936	0.933	0.944	0.956	0.976	0.929
Standard error	0.108	0.131	0.166	0.209	0.159	0.193	0.169	0.162	0.131	0.136	0.156
Observations	30	30	30	29	30	28	28	28	25	28	
$(R)^a$	0.950	0.916	0.882	0.891	0.904	0.812	0.847	0.878	0.915	0.837	0.883
$(R^2)^a$	0.902	0.839	0.778	0.795	0.817	0.660	0.717	0.771	0.838	0.700	0.782
Adjusted $(R^2)^a$	0.882	0.805	0.731	0.750	0.779	0.582	0.653	0.719	0.795	0.632	0.733
Standard error^a	0.119	0.149	0.194	0.238	0.178	0.219	0.196	0.183	0.147	0.150	0.177
All parameters allowed to change, fixed b/m ratio											
Int	0.007	−0.152	0.531	0.809	0.252	−0.535	−0.264	−0.464	−0.261	−0.607	−0.068
r	−0.060	0.157	0.389	0.858	−0.026	0.236	0.300	0.494	0.406	−0.111	0.264
s	−0.228	−0.270	−0.969	−1.512	−0.378	0.214	−0.116	−0.100	−0.199	0.381	−0.318
a	−0.662	−0.675	−0.694	−0.923	−0.683	−0.822	−0.758	−0.811	−0.778	−0.612	−0.742
b	−1.409	−1.404	−1.037	−0.882	−1.239	−1.295	−1.168	−1.283	−1.136	−1.793	−1.264
m	1.148	1.144	0.845	0.719	1.009	1.055	0.952	1.046	0.926	1.461	1.031
U	−1.261	−1.203	−1.290	−1.131	−1.298	−2.078	−1.628	−1.753	−2.380	−1.824	
V	226.170	4917.267	38.906	3.116	20.192	208.707	55.649	827.099	1675.971	42.651	
R	0.961	0.958	0.922	0.949	0.956	0.964	0.967	0.972	0.978	0.988	0.962
R^2	0.924	0.918	0.850	0.900	0.914	0.929	0.936	0.944	0.956	0.977	0.925
Adjusted R^2	0.921	0.915	0.845	0.897	0.911	0.927	0.933	0.942	0.954	0.976	0.922
Standard error	0.156	0.115	0.169	0.214	0.186	0.205	0.169	0.165	0.134	0.137	0.165
$(R)^a$	0.950	0.916	0.883	0.891	0.904	0.792	0.847	0.878	0.915	0.837	0.881
$(R^2)^a$	0.902	0.839	0.779	0.795	0.817	0.627	0.717	0.771	0.838	0.700	0.778
Adjusted $(R^2)^a$	0.882	0.805	0.733	0.750	0.779	0.542	0.653	0.719	0.795	0.632	0.729
Standard error^a	0.119	0.149	0.197	0.238	0.178	0.231	0.196	0.183	0.147	0.150	0.179
U and V parameters allowed to change, fixed Int, s, a, b and m											
U	−1.274	−1.182	−1.178	−2.604	−1.566	−1.857	−1.538	−1.623	−2.268	−1.612	
V	152.820	392.707	33.358	13.651	34.022	187.059	43.642	906.608	1371.507	30.444	
R	0.958	0.970	0.883	0.981	0.974	0.959	0.963	0.971	0.976	0.979	0.961
R^2	0.917	0.941	0.780	0.962	0.949	0.920	0.928	0.942	0.952	0.959	0.925
Adjusted R^2	0.914	0.938	0.772	0.960	0.947	0.917	0.925	0.940	0.950	0.958	0.922
Standard error	0.168	0.101	0.200	0.141	0.152	0.217	0.181	0.168	0.141	0.183	0.165
$(R)^a$	0.925	0.953	0.881	0.914	0.899	0.778	0.841	0.874	0.914	0.801	0.878
$(R^2)^a$	0.856	0.907	0.776	0.836	0.808	0.606	0.707	0.765	0.835	0.642	0.774
Adjusted $(R^2)^a$	0.827	0.887	0.729	0.797	0.767	0.516	0.640	0.711	0.791	0.561	0.723
Standard error^a	0.141	0.107	0.195	0.148	0.152	0.248	0.207	0.188	0.148	0.165	0.170

^{a} Regression statistics by Eq. (10).

Table 7
Nonlinear least-squares fitting results at 40% ACN using ^s_wpH and ^s_wpK

| ^w_wpH	1.99	2.99	3.99	5.03	6.04	7.01	7.99	10.01	11.03	12	Average
^s_wpH	2.2	3.53	4.7	5.99	6.89	7.8	8.62	9.52	10.73	12.7	
^s_spH	2.34	3.67	4.84	6.13	7.03	7.94	8.76	9.66	10.87	12.84	
All parameters allowed to change											
Int	0.529	0.487	−0.184	0.004	−0.020	0.179	0.135	0.120	−0.024	0.413	0.164
r	0.400	0.542	0.114	0.163	−0.014	0.155	0.090	0.115	0.047	0.124	0.174
s	−0.132	−0.220	0.006	−0.177	−0.060	0.105	0.093	0.075	0.132	0.070	−0.011
a	−0.755	−0.743	−0.690	−0.625	−0.623	−0.614	−0.627	−0.637	−0.520	−0.333	−0.617
b	−1.894	−1.564	−1.653	−1.787	−1.704	−1.860	−1.762	−1.820	−1.994	−1.016	−1.705
m	0.787	0.689	1.601	1.554	1.599	1.083	1.167	1.209	1.441	0.600	1.173
U	−1.158	−1.145	−1.393	−1.194	−1.230	−1.232	−1.241	−1.217	−1.454	−1.422	b/m =
V	7.731	6.257	5.369	4.350	2.177	18.897	17.027	31.751	34.117	3.160	−1.454
R	0.988	0.984	0.986	0.984	0.981	0.979	0.977	0.977	0.980	0.979	0.981
R^2	0.977	0.969	0.973	0.967	0.963	0.958	0.954	0.954	0.960	0.958	0.963
Adjusted R^2	0.976	0.968	0.972	0.966	0.962	0.956	0.952	0.952	0.958	0.957	0.962
Standard error	0.096	0.098	0.092	0.113	0.120	0.131	0.134	0.132	0.136	0.154	0.121
Observations	29	29	29	29	29	29	29	29	29	29	
$(R)^a$	0.956	0.957	0.967	0.952	0.943	0.931	0.925	0.931	0.933	0.756	0.925
$(R^2)^a$	0.914	0.915	0.935	0.907	0.890	0.866	0.856	0.866	0.871	0.571	0.859
Adjusted $(R^2)^a$	0.895	0.897	0.920	0.887	0.866	0.837	0.824	0.837	0.843	0.478	0.828
Standard error^a	0.105	0.108	0.101	0.124	0.133	0.145	0.148	0.146	0.150	0.170	0.133
All parameters allowed to change, fixed b/m ratio											
Int	0.327	0.376	0.020	0.222	0.228	0.065	0.104	0.076	0.018	0.274	0.171
r	0.246	0.416	0.278	0.285	0.134	0.094	0.077	0.102	0.067	0.037	0.174
s	−0.261	−0.341	0.039	−0.224	−0.074	0.104	0.096	0.094	0.136	0.124	−0.031
a	−0.716	−0.718	−0.703	−0.646	−0.649	−0.607	−0.623	−0.633	−0.525	−0.540	−0.636
b	−1.770	−1.463	−1.797	−1.859	−1.796	−1.825	−1.754	−1.824	−2.004	−1.230	−1.732
m	1.218	1.006	1.236	1.279	1.235	1.255	1.206	1.254	1.378	0.846	1.191
U	−1.231	−1.270	−1.215	−1.092	−1.131	−1.269	−1.249	−1.242	−1.446	−1.342	
V	10.298	7.793	5.030	3.829	2.984	13.017	17.184	30.681	33.353	8.306	
R	0.983	0.982	0.985	0.983	0.980	0.978	0.977	0.977	0.979	0.975	0.980
R^2	0.967	0.964	0.970	0.966	0.960	0.957	0.954	0.954	0.959	0.950	0.960
Adjusted R^2	0.965	0.963	0.969	0.965	0.959	0.956	0.952	0.952	0.958	0.948	0.959
Standard error	0.113	0.105	0.097	0.115	0.124	0.131	0.134	0.133	0.136	0.165	0.125
$(R)^a$	0.952	0.954	0.964	0.952	0.941	0.931	0.925	0.931	0.933	0.816	0.930
$(R^2)^a$	0.907	0.910	0.930	0.906	0.885	0.867	0.855	0.866	0.871	0.665	0.866
Adjusted $(R^2)^a$	0.887	0.890	0.914	0.885	0.860	0.838	0.824	0.837	0.842	0.593	0.837
Standard error^a	0.108	0.111	0.105	0.126	0.136	0.145	0.148	0.147	0.150	0.187	0.136
U and V parameters allowed to change, fixed Int, s, a, b and m											
U	−1.217	−1.178	−1.227	−1.173	−1.177	−1.270	−1.265	−1.242	−1.399	−1.342	
V	8.583	7.990	5.325	4.985	3.417	11.167	14.782	30.681	46.601	8.306	
R	0.978	0.975	0.983	0.981	0.979	0.978	0.976	0.976	0.977	0.968	0.977
R^2	0.957	0.951	0.967	0.962	0.959	0.956	0.952	0.952	0.954	0.938	0.955
Adjusted R^2	0.956	0.949	0.965	0.960	0.957	0.954	0.951	0.950	0.953	0.935	0.953
Standard error	0.129	0.124	0.100	0.120	0.127	0.134	0.137	0.135	0.144	0.191	0.134
$(R)^a$	0.954	0.957	0.964	0.952	0.941	0.931	0.925	0.931	0.929	0.816	0.930
$(R^2)^a$	0.910	0.916	0.930	0.907	0.885	0.866	0.856	0.866	0.862	0.665	0.866
Adjusted $(R^2)^a$	0.890	0.897	0.915	0.886	0.860	0.837	0.825	0.837	0.832	0.593	0.837
Standard error^a	0.107	0.110	0.105	0.125	0.135	0.146	0.149	0.147	0.152	0.187	0.136

^{a} Regression statistics by Eq. (10).
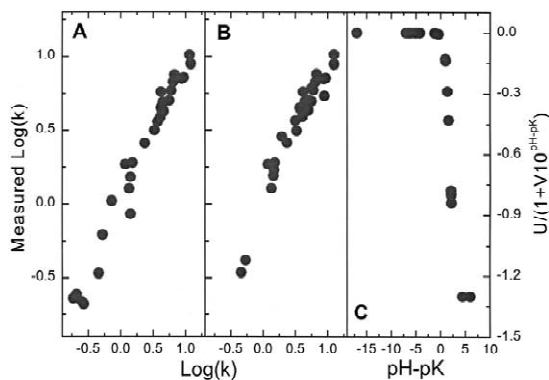
Fig. 2. Plots of the predicted versus measured log (k) at 40% ACN and $_w^w$pH of 3 to illustrate the model effectiveness. Plot A, overall predicted versus measured log (k); plot B, predicted log (k) (by Eq. (10)) versus $\left[\log(k) - (U/1 + V10^{\mp(\text{pH}-\text{p}K)})\right]$; plot C, $(U/1 + V10^{\mp(\text{pH}-\text{p}K)})$ versus $\mp(\text{pH} - \text{p}K)$.

4.3. Comparison of fitting quality with literature results

To compare the relative effectiveness of the current model (Eq. (9)) with the literature one (Eq. (1)), Fig. 4 shows the comparison of the fitting quality using SE. Fig. 4A plots the SE from literature (Table 5 of Ref. [39]) versus SE by scheme B at different pH (20% ACN). The slope of the plot is an
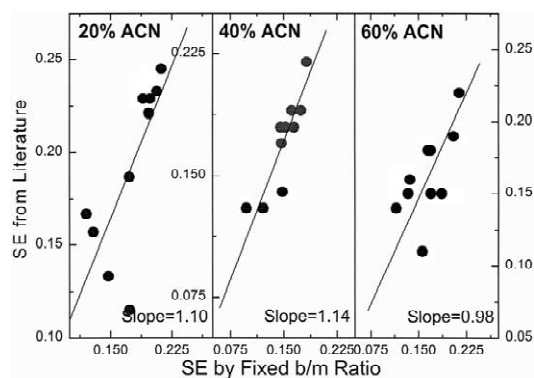

Fig. 4. Comparison of the quality of the fits with literature results. The abscissa is SE obtained by scheme B. The ordinate is SE obtained from Ref. [37] (Table 5). $_w^w$pH varies from 2 to 12.

indication of the relative fitting quality. The slope in Fig. 4A is 1.10, thereby indicating that the model error by Eq. (1) is on average about 10% larger than Eq. (9) at 20% ACN. Fig. 4B is a similar plot for 40% ACN. The slope is 1.14, and the model error by Eq. (1) is on average 14% larger. Fig. 4C indicates the model errors of Eqs. (1) and (9) are similar at 60% ACN. Overall, the performance of the two models is very similar. It is noted that Eq. (1) has seven coefficients, while Eq. (9) has eight.

4.4. Comparison of $_w^w$pH/$_w^w$pK and $_w^s$pH/$_w^s$pK scales in Eq. (9)

Eq. (1) needs to use $_w^s$pH/$_w^s$pK values; however, Eq. (9) can use both $_w^w$pH/$_w^w$pK and $_w^s$pH/$_w^s$pK values. To show the difference in the fitting quality due to the different pH/pK scales, Fig. 5 shows the comparison of SE at different pH values. Fig. 5A was generated by the fitting scheme A. It is seen that the SE values with $_w^s$pH/$_w^s$pK scale are usually smaller than those with $_w^w$pH/$_w^w$pK scale. This indicates that the model performance using $_w^s$pH/$_w^s$pK values is somewhat better. Fig. 5B shows the similar results with scheme B. Fig. 5C shows a similar plot with scheme C, and the difference in SE is even more obvious.

Fig. 5 clearly demonstrates that better fit is clearly obtained when the $_w^s$pH/$_w^s$pK scale is used in Eq. (9). This indicates that the V parameter in combination with $_w^w$pH/$_w^w$pK scale is unable to completely correct

Fig. 3. Comparison of the quality of the different fit schemes. The abscissa is the SE by the simultaneous change of all coefficients (scheme A). The ordinate in plot A is the SE from neutral component (Eq. (10)); that in plot B is the SE by fixed ratio of b/m (scheme B); and that in plot C is the SE by fixed LSER coefficients (scheme C). The ACN composition is 40%, and pH varies from 2 to 12.
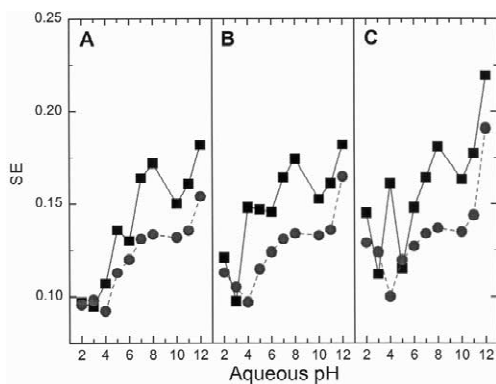
Fig. 5. Comparison of the quality of the fits using $_w^w\text{pH}/_w^w\text{p}K$ and $_w^s\text{pH}/_w^s\text{p}K$ scales. The ordinate is SE, while the abscissa is $_w^w\text{pH}$ value. Solid square: SE obtained by $_w^w\text{pH}/_w^w\text{p}K$ values; and solid circle, SE obtained by $_w^s\text{pH}/_w^s\text{p}K$ values. $_w^w\text{pH}$ varies from 2 to 12.

the shift in pH − pK. The V parameter in combination with $_w^s\text{pH}/_w^s\text{p}K$ scale better compensates the shift in pH − pK, thereby providing even better fits. However, the measurement of the $_w^s\text{p}K$ value for each solute is not easy, and the use of $_w^w\text{p}K$ is preferable. It is noted that the use of $_w^w\text{pH}/_w^w\text{p}K$ scale by Eq. (9) performs similarly to the use of $_w^s\text{pH}/_w^s\text{p}K$ by Eq. (1).

4.5. Comparison of linear and nonlinear analysis

Linear regression analysis is usually used to process the retention data based on Eq. (1). There are seven coefficients in Eq. (1), and they are linearly related. When multivariate linear regression analysis is carried out, the best fit is obtained without regarding the physical significance of each coefficient. In other words, the coefficients may not be consistent and meaningful. This is shown by the results in this study and Ref. [39] (Tables 4 and 5). The LSER coefficients are not consistent at different pH values.

However, the current model along with the nonlinear least-squares analysis allows us not only to improve the consistency of LSER coefficients, but also to keep the fit quality. The nonlinear least-squares curve-fitting offers more flexibility in controlling the data analysis. When a proper selection of solutes (at least 20 neutral solutes) and fitting scheme

B (or C) is made, Eq. (9) can be an excellent model to correlate the retention of ionizable compounds.

4.6. A practical comparison of the current (Eq. (9)) and previous (Eq. (1)) models

As indicated earlier, the major advantage of the current model is the use of $_w^w\text{pH}$ value of the mobile phase and $_w^w\text{p}K$ values of compounds. The V parameter is related to the difference between pH and pK. Thus, it may be buffer-dependent because both pH and pK are affected by the type of buffer used. The dependency of the model parameters on the type of buffer may limit the prediction of retention among different buffers. In that situation, the reference solutes should be rerun in a new buffer system to establish the model equation for the purpose of the prediction. Table 8 further compares the difference in the two models.

5. Conclusions

A solvation parameter model is derived and its application to retention data of ionizable compounds is evaluated. This model requires the use of nonlinear least-squares analysis, which can be easily implemented by Microsoft Excel Solver. Although $_w^s\text{pH}/_w^s\text{p}K$ scale provides smaller model error, both $_w^w\text{pH}/_w^w\text{p}K$ and $_w^s\text{pH}/_w^s\text{p}K$ scales can be used. The use of $_w^w\text{pH}/_w^w\text{p}K$ scale is a major advantage of the current model because the data can be obtained directly from the references or computed by software. When $_w^w\text{pH}/_w^w\text{p}K$ scale is used, the model function performs similarly to the literature model. Finally, the function simplifies the application of linear solvation energy relationships (LSERs) to ionizable compounds, and allows us to easily predict their retention and internal standards for chromatographic method development. The limitation of the model is the dependency of the coefficients on the type of buffer.

Table 8
Comparison of the current and previous models

Item	Current (Eq. (9))	Previous (Eq. (1))
Data processing	Nonlinear, some control on fitting procedure and coefficients. Effect of data quality on coefficients can be minimized	Linear, no control on fitting procedure and coefficients. Data quality affects fitting results significantly
pH	Both $_w^w\text{pH}$ and $_w^s\text{pH}$ can be used	$_w^s\text{pH}$ is needed
pK	Both $_w^w\text{p}K$ and $_w^s\text{p}K$ can be used	$_w^s\text{p}K$ is needed
Limitation	Dependency of coefficients on the type of buffer	Time needed to measure $_w^s\text{pH}$ and $_w^s\text{p}K$
Quality of fit	Eq. (9) using $_w^w\text{pH}/_w^w\text{p}K$ scale is similar to Eq. (1) using $_w^s\text{pH}/_w^s\text{p}K$ scale in terms of quality-of-fit. Better fits are obtained with Eq. (9) when $_w^s\text{pH}$ and $_w^s\text{p}K$ values are used	

References

[1] R.G. Wolcott, J.W. Dolan, L.R. Snyder, J. Chromatogr. A 869 (2000) 3.
[2] L.V. Heukelem, C.S. Thomas, J. Chromatogr. A 910 (2001) 31.
[3] L.R. Snyder, J.W. Dolan, J. Chromatogr. A 892 (2000) 107.
[4] J.R. Torres-Lapasio, M. Rosés, E. Bosch, M.C. Garcia-Alvarez-Coque, J. Chromatogr. A 886 (2000) 31.
[5] G. Vivo-Trayols, J.R. Torres-Lapasio, M.C. Garcia-Alvarez-Coque, J. Chromatogr. A 876 (2000) 17.
[6] S. Carda-Broch, J.R. Torres-Lapasio, J.S. Esteve-Romero, M.C. Garcia-Alvarez-Coque, J. Chromatogr. A 893 (2000) 321.
[7] V. Harang, A. Karlsson, M. Josefson, Chromatographia 54 (2001) 703.
[8] A.M. Siouffi, R. Phan-Tan-Luu, J. Chromatogr. A 892 (2000) 75.
[9] J. Li, D.S. Shah, J. Chromatogr. A 954 (2002) 159.
[10] S. Espinosa, E. Bosch, M. Rosés, J. Chromatogr. A 947 (2002) 47.
[11] J.R. Torres-Lapasio, M.C. Garcia-Alvarez-Coque, M. Rosés, E. Bosch, J. Chromatogr. A 955 (2002) 19.
[12] E. Estrada, Y. Gutierrez, J. Chromatogr. A 858 (1999) 187.
[13] T. Suzuki, S. Timotei, B.E. Iuoras, G. Uray, P. Verdino, W.M.F. Fabian, J. Chromatogr. A 922 (2001) 13.
[14] J.P. Wolbach, D.K. Lloyd, I.W. Wainer, J. Chromatogr. A 914 (2000) 299.
[15] E.B. Ledesma, M.J. Wornett, Anal. Chem. 72 (2000) 5437.
[16] P. Nikitas, A. Pappa-Louisi, P. Agfafiotou, J. Chromatogr. A 946 (2002) 9.
[17] P. Nikitas, A. Pappa-Louisi, P. Agfafiotou, J. Chromatogr. A 946 (2002) 33.
[18] F. Gritti, G. Felix, W.F. Achard, F. Hardouin, J. Chromatogr. A 922 (2001) 51.
[19] A. Wang, L.C. Tan, P.W. Carr, J. Chromatogr. A 848 (1999) 21.
[20] J. Zhao, P.W. Carr, Anal. Chem. 70 (1998) 3619.
[21] L.C. Tan, P.W. Carr, M.H. Abraham, J. Chromatogr. 752 (1996) 1.
[22] L.C. Tan, P.W. Carr, J. Chromatogr. 799 (1998) 1.
[23] J.H. Park, J.T. Chae, T.H. Nah, M.D. Jang, J. Chromatogr. A 664 (1994) 149.
[24] D.S. Seibert, C.F. Poole, Chromatographia 41 (1995) 51.
[25] J. Li, D.A. Whitman, Anal. Chim. Acta 368 (1998) 141.
[26] J. Li, T. Robison, Anal. Chim. Acta 395 (1999) 85.
[27] J.H. Park, M.H. Youn, Y.K. Ryu, B.E. Kim, J.W. Ryu, M.D. Jang, J. Chromatogr. A 796 (1998) 249.
[28] F.Z. Oumada, M. Rosés, E. Bosch, M.H. Abraham, Anal. Chim. Acta 382 (1999) 301.
[29] P.G. Muijselaar, H.A. Claessens, C.A. Cramers, Anal. Chem. 69 (1997) 1184.
[30] M.D. Trone, M.G. Khaledi, J. Chromatogr. A 886 (2000) 245.
[31] A. Detroyer, Y.V. Heyden, S. Carda-Broch, M.C. Garcia-Alvarez-Coque, D.L. Massart, J. Chromatogr. A 912 (2001) 211.
[32] M.A. Garcia, M.F. Vitha, J. Sandquist, K. Mulville, M.L. Marina, J. Chromatogr. A 918 (2001) 1.
[33] M.H. Abraham, Anal. Chem. 69 (1997) 613.
[34] M.H. Abraham, C. Treiner, M. Rosës, C. Rafols, Y. Ishihama, J. Chromatogr. A 752 (1996) 243.
[35] J. Li, J. Chromatogr. A 927 (2001) 19.
[36] M.H. Abraham, H.S. Chadha, in: V. Pliska, B. Testar, H. Vanda Waterbeemed (Eds.), Lipophilicity in Drug Action and Toxicology, VCH, Weinheim, 1996, p. 311.
[37] D. Bolliet, C.F. Poole, M. Rosés, Anal. Chim. Acta 368 (1998) 129.
[38] M. Rosés, D. Bolliet, C.F. Poole, J. Chromatogr. A 829 (1998) 29.

[39] S. Espinosa, E. Bosch, M. Rosés, J. Chromatogr. A 945 (2002) 83.

[40] E. Bosch, P. Bou, H. Allemann, M. Rosés, Anal. Chem. 68 (1996) 3651.

[41] E. Bosch, S. Espinosa, M. Rosés, J. Chromatogr. A 824 (1998) 137.

[42] I. Canals, J.A. Portal, E. Bosch, M. Rosés, Anal. Chem. 72 (2000) 1802.

[43] S. Espinosa, E. Bosch, M. Rosés, Anal. Chem. 72 (2000) 5193.

[44] S. Espinosa, E. Bosch, M. Rosés, Anal. Chim. Acta 454 (2002) 157.

[45] J.A. Dean, Handbook of Chemistry, 5th ed., McGraw-Hill, 1999, p. 8.24.