ELSEVIER

Journal of Chromatography A, 829 (1998) 29–40

JOURNAL OF CHROMATOGRAPHY A

Comparison of solute descriptors for predicting retention of ionic compounds (phenols) in reversed-phase liquid chromatography using the solvation parameter model

Martí Rosés[a,*], David Bolliet[b], Colin F. Poole[b]

[a]*Departament de Química Analítica, Universitat de Barcelona, Diagonal 647, E-08028 Barcelona, Spain*
[b]*Department of Chemistry, Wayne State University, Detroit, MI 48202, USA*

Received 23 June 1998; received in revised form 7 September 1998; accepted 7 September 1998

Abstract

Two solute descriptors that account for the ionization of phenols under reversed-phase liquid chromatographic conditions are compared for the prediction of retention of phenols and neutral compounds at different mobile phase pH values using the solvation parameter model. One of the descriptors is the P descriptor (a scaled effective acid dissociation constant, $P=(14-pK^*)/10$), proposed in a previous work. The other descriptor [$\log (1-D(1-f)]$ is based on the degree of ionization (D) of the phenol and a retention derived parameter (f) with the value $f=10^{-1.80}$ for the chromatographic system studied. Calculation of the P descriptor is straightforward since its value is constant for all mobile phase pH*, but estimation of retention requires a different correlation equation for each mobile phase pH*. In contrast, the $\log [1-D(1-f)]$ descriptor is pH* dependent, but it allows the same correlation equation to be used for the estimation of retention at any mobile phase pH*. The D derived descriptor can be successfully applied to the estimation of retention of basic and amphiprotic compounds, for which the P descriptor has yet to be applied.

Keywords: Solvation parameter model; Retention; Phenols

1. Introduction

Linear free energy relationships (LFERs) are very useful for characterizing physicochemical processes, including solute retention in reversed-phase liquid chromatography. Amongst the host of LFERs established, the solvation parameter model, proposed by Abraham [1–3], has been demonstrated to provide an adequate description of the retention of neutral organic compounds under reversed-phase liquid chromatographic conditions on many different columns [4–21].

The solvation parameter model can be set out as

$$\log k = c + mV_X + rR_2 + s\pi_2^H + a\Sigma\alpha_2^H + b\Sigma\beta_2^0 \quad (1)$$

where k is the observed solute retention factor. The solute descriptors are McGowan's characteristic volume V_X (in $cm^3\ mol^{-1}/100$), an excess molar refraction R_2 (in $cm^3/10$), the solute dipolarity/polarizability π_2^H, and the solute's effective hydrogen-bond acidity and hydrogen-bond basicity $\Sigma\alpha_2^H$ and $\Sigma\beta_2^0$, respectively. The solute descriptors V_X and

*Corresponding author. Tel.: +34 93 4021796; fax: +34 93 4021233; e-mail: marti@zeus.qui.ub.es

0021-9673/98/$ – see front matter
PII: S0021-9673(98)00746-8

R_2 are easily calculated by addition of fragments while the other descriptors are obtained experimentally from liquid–liquid distribution and gas and liquid chromatographic systems [1–3]. Nowadays, the solute descriptors are available for more than 2000 compounds and can be estimated for many more. The solute descriptors used here are given in Table 1.

The coefficients in Eq. (1) are calculated by the method of multiple linear regression and are charac-

Table 1
Solute descriptors for ionic and non-ionic compounds

Compound	V_x	R_2	π_2^H	$\Sigma\alpha_2^H$	$\Sigma\beta_2^0$	P
4-Chloro-2-nitrophenol	1.072	1.125	1.24	0.10	0.30	0.719
2,4,6-Trichlorophenol	1.142	1.010	1.01	0.82	0.08	0.702
2-Nitrophenol	0.949	1.015	1.05	0.05	0.37	0.614
4-Nitrophenol	0.949	1.070	1.72	0.82	0.26	0.636
4-Hydroxybenzaldehyde	0.932	1.010	1.01	0.77	0.44	0.575
2,4-Dichlorophenol	1.020	0.960	0.84	0.53	0.19	0.547
2-Chlorophenol	0.898	0.853	0.88	0.32	0.31	0.469
Vanillin	1.131	1.040	1.04	0.32	0.67	0.598
3-Bromophenol	0.950	1.060	1.15	0.70	0.16	0.429
2-Naphthol	1.144	1.520	1.08	0.61	0.40	0.375
4-Chlorophenol	0.898	0.915	1.08	0.67	0.21	0.385
4-Chloro-3-methylphenol	1.034	0.920	1.02	0.65	0.23	0.370
1-Naphthol	1.144	1.520	1.05	0.60	0.37	0.367
4-Hydroxybenzyl alcohol	0.975	0.998	1.15	0.88	0.85	0.343
2-Aminophenol	0.875	1.110	1.10	0.60	0.66	0.354
3,5-Dimethylphenol	1.057	0.820	0.84	0.57	0.36	0.284
3,4-Dimethylphenol	1.057	0.830	0.86	0.56	0.39	0.285
m-Cresol	0.916	0.822	0.88	0.57	0.34	0.300
2,5-Dimethylphenol	1.057	0.840	0.79	0.54	0.37	0.280
o-Cresol	0.916	0.840	0.86	0.52	0.30	0.271
3-Aminophenol	0.875	1.130	1.15	0.65	0.79	0.343
Eugenol	1.354	0.946	0.99	0.22	0.51	0.323
4-Aminophenol	0.875	1.150	1.20	0.65	0.83	0.292
Phenol	0.775	0.805	0.89	0.60	0.31	0.320
p-Cresol	0.916	0.820	0.87	0.57	0.32	0.294
2,6-Dimethylphenol	1.057	0.860	0.79	0.39	0.39	0.260
Pyridine	0.675	0.631	0.84	0.00	0.52	0.000
Cyclohexanone	0.861	0.403	0.86	0.00	0.56	0.000
Benzyl alcohol	0.916	0.803	0.87	0.33	0.56	0.000
Acetanilide	1.114	0.870	1.40	0.50	0.67	0.000
Pentan-2-one	0.829	0.143	0.68	0.00	0.51	0.000
Hexan-2-one	0.968	0.136	0.68	0.00	0.51	0.000
2-Phenylethanol	1.057	0.811	0.91	0.30	0.65	0.000
3-Methylbutan-2-one	0.829	0.134	0.65	0.00	0.51	0.000
4-Methylpentan-2-one	0.968	0.111	0.65	0.00	0.51	0.000
Dibromomethane	0.600	0.714	0.67	0.10	0.10	0.000
Benzamide	0.973	0.990	1.50	0.49	0.67	0.000
4-Nitrobenzyl Alcohol	1.090	1.064	1.39	0.44	0.62	0.000
Benzene	0.716	0.610	0.52	0.00	0.14	0.000
Benzaldehyde	0.873	0.820	1.00	0.00	0.39	0.000
Benzonitrile	0.871	0.742	1.11	0.00	0.33	0.000
Octan-2-one	1.252	0.108	0.68	0.00	0.51	0.000
Acetophenone	1.014	0.818	1.01	0.00	0.49	0.000
Nitrobenzene	0.891	0.871	1.11	0.00	0.28	0.000
Chlorobenzene	0.839	0.718	0.65	0.00	0.07	0.000

teristic of the difference in solvation properties of both phases forming the chromatographic system. The *r* constant determines the difference in capacity of the solvated stationary and mobile phases to interact with solute π- and n-electrons; the *s* constant is a measure of the difference in dipolarity/polarizability between the two phases; the *a* and *b* constants measure the difference in the phases hydrogen-bond basicity and acidity, respectively; and the *m* constant is a measure of the relative ease of forming a cavity for the solute in the solvated stationary phase and mobile phase.

The solvation parameter model was established for neutral solutes and without modification does not apply to ionic or partially ionized solutes. In reversed-phase chromatography, ionic solutes are retained to a lesser extent than neutral solutes and the descriptors set out for a neutral species (e.g. a phenol) cannot describe the solute–solvent interactions of the corresponding ionic species (a phenolate ion). Moreover, the retention of an ionizable compound to a great extent depends on those factors that affect the extent of ionization, such as pH and composition and ionic strength of the mobile phase.

However, in a previous study [21] we proposed a solute descriptor for ionizable compounds, *P*, which depends on the dissociation pK^* value of the solute, that in conjunction with the solute descriptors of the neutral form of the compound (V_X, R_2, π_2^H, $\Sigma\alpha_2^H$, and $\Sigma\beta_2^0$) was able to accurately predict the retention of neutral and ionizable (phenols) compounds over a wide mobile phase pH* range. Notice that we denote the pH of the mobile phase measured after mixing the aqueous buffer with the organic modifier as pH* to distinguish it from the pH of the aqueous buffer before mixing with the organic solvent (pH), which is commonly used by many chromatographers. In consonance, pK^* refers to the pK value of the phenol in the particular mobile phase used (50% methanol), not to the pK value of the phenol in water.

Some other descriptors based on the degree of ionization of the solute, *D*, at the pH of the mobile phase were considered in our previous study, but none of them gave better predictions than the *P* solute descriptor. Further work has led to the derivation of another descriptor based on the degree of ionization that gives predictions as good as those obtained from the *P* solute descriptor as well as having a more consistent theoretical basis. In this work, we shall compare both solute descriptors and show the advantages and handicaps of each descriptor.

2. Theory

For a weak acid (HX) with an ionization process such as

$$HX \Leftrightarrow H^+ + X^- \qquad K^* = [H^+][X^-]/[HX] \tag{2}$$

the *P* solute descriptor was defined from the effective acid dissociation constant of the acid at the mobile phase composition (pK^*) as [21]

$$P = (14 - pK^*)/10 \tag{3}$$

This definition applied to a neutral solute (which has $K^* = 0$), however, would lead to a *P* value tending to minus infinity. To avoid this problem an effective p$K^* = 14$ was assigned to neutral solutes to achieve a *P* value of zero for these compounds. The values of 14 and 10 in Eq. (3) are arbitrary, but they provide a reasonable zero point for the scale and a reasonable numerical range of *P* values similar to the range of values for the other solute descriptors used in the model [21].

Inclusion of the *P* solute descriptor in the solvation parameter model led to the main correlation equation

$$\log k = c + mV_X + rR_2 + s\pi_2^H + a\Sigma\alpha_2^H + b\Sigma\beta_2^0 + pP \tag{4}$$

Eq. (4) was successfully applied to the retention of the compounds identified in Table 1 on a column containing a polymeric sorbent with methanol–water (50:50, v/v) at pH* values of 4, 7, 9, 11 and 12 as mobile phase.

However, Eq. (4) was not the only equation tested to account for ionization of the phenols. In the first instance, the degree of ionization, *D*, of the phenol was considered as a solute descriptor. *D* is calculated from the pK^* value of the acid and the pH* value of the mobile phase from

$$D = [X^-]/([HX] + [X^-]) = K^*/([H^+] + K^*)$$
$$= 10^{(pH^* - pK^*)}/[1 + 10^{(pH^* - pK^*)}] \quad (5)$$

A general model for the influence of ionization on retention of weak acids has been proposed earlier [21–27]. The observed retention factor for the weak acid (k) is taken as an average of the retention factors of the neutral (k_{HX}) and ionic (k_{X^-}) species according to the fraction of each species present at the mobile phase pH [$(1-D)$ and D, respectively].

$$k = (1 - D)k_{HX} + Dk_{X^-} \quad (6)$$

Eq. (6) describes a sigmoidal relationship of k as a function of mobile phase pH commonly observed in reversed-phase liquid chromatography of weak acids. The ionic form of the compound exhibits a much weaker retention than the neutral form, $k_{X^-} \ll k_{HX}$, and if the retention of the ionized form is assumed to be insignificant compared to the neutral form, Eq. (6) can be simplified to

$$\log k = \log k_{HX} + \log (1 - D) \quad (7)$$

where k_{HX} is the retention factor of the neutral form of the acid and is linearly related to the solute descriptors through Eq. (1). In this instance, $\log (1-D)$ can be considered a solute descriptor that accounts for the ionization of the acid [21].

As a solute descriptor, however, $\log (1-D)$ has certain disadvantages. For a neutral compound, $D=0$, $\log (1-D)=0$ and $\log k$ follows Eq. (1), but for a completely ionized compound (that is to say when $pH^* \gg pK^*$), $D \rightarrow 1$ and $\log (1-D) \rightarrow -\infty$. At high pH^* values several phenols used in the study are almost completely ionized, and therefore, their $\log (1-D)$ descriptors have a high negative value that does not allow good correlations with the $\log k$ values.

To avoid this problem, the degree of ionization (D) was directly used as a solute descriptor by using Eq. (8) [21]

$$\log k = c + mV_X + rR_2 + s\pi_2^H + a\Sigma\alpha_2^H + b\Sigma\beta_2^0 + dD \quad (8)$$

The correlations obtained, however, were worse than those obtained by using Eq. (4) and P was selected as the best solute descriptor to account for solute ionization [21].

A more realistic approach is to consider the retention of the ionized species to be smaller than the retention of the neutral species, but not insignificant. In this instance, k_{X^-} is a fraction of k_{HX} and both retention factors can be related through an f parameter defined in Eq. (9).

$$f = k_{X^-}/k_{HX} \quad (9)$$

Eq. (6) can then be rewritten as given below

$$\log k = \log k_{HX} + \log [1 - D(1 - f)] \quad (10)$$

and the $\log [1-D(1-f)]$ parameter taken as a solute descriptor that accounts for the ionization of the compound.

As a solute descriptor $\log [1-D(1-f)]$ has advantages over $\log (1- D)$. For a neutral compound, $D=0$ and both descriptors become zero, but for a completely ionized compound (for which $\log (1-D) \rightarrow -\infty$), $D=1$ and $\log [1-D(1-f)]=\log f$. That is to say, the value of the descriptor for a weak acid is limited between 0 and $\log f$ depending on the difference between the pK^* of the acid and the pH^* of the mobile phase (Eq. (5)). In addition, Eq. (10) is more rigorous than Eq. (7) because the retention of the ionized form of the solute has not been neglected.

Since the $\log k_{HX}$ value is linearly related to the solute descriptors for the neutral compound, the final correlation equation that will be tested is

$$\log k = c + mV_X + rR_2 + s\pi_2^H + a\Sigma\alpha_2^H + b\Sigma\beta_2^0 + d\log [1 - D(1 - f)] \quad (11)$$

In fact, Eq. (10) predicts the d coefficient to be 1.00, but we choose to calculate it in order to check the validity of the theoretical model.

We shall apply Eq. (11) to the $\log k$ data obtained for phenols and neutral compounds obtained in the previous study [21] and compare the results with those obtained in the correlation with the P solute descriptor, Eq. (4).

3. Results and discussion

*3.1. Variation of retention with mobile phase pH**

The application of the $\log [1-D(1-f)]$ descriptor

in the general Eq. (11) requires calculation of the f parameter. This parameter is obtained from the retention factor of the neutral and ionized forms of the acid (Eq. (9)), and in principle, it should be different for each compound. However, the calculation would be enormously simplified if the same f value could be applied to all compounds. To calculate and analyze the f values, we have fitted the log k values of the phenols at the different pH values to Eq. (6) by non-linear regression.

The pK^*, k_{HX}, and k_{X^-} values for the various phenols after analysis of the retention factor data at the pH* values of 2, 4, 7, 9, 11 and 12 are summarized in Table 2. For the aminophenols, the data at pH 2 and 4 have been removed since at these pH* value the amino group will be partially protonated. Nor was 4-phenylphenol at pH* 11 and 12 included because the large retention of the neutral form prevented determination of the retention factors at pH* values lower than 11, and therefore estimation of pK^* and k_{HX}.

Some graphical examples of the fits obtained are presented in Fig. 1. The most acidic phenols, such as 4-chloro-2-nitrophenol and 4-nitrophenol are almost completely ionized at pH* 11 and 12, whereas the less acidic (e.g. 4-chloro-3-methylphenol, 2,6-dimethylphenol, and 4-aminophenol) are only partially ionized. At acidic pH* (2 and 4) all phenols are completely in the neutral form except for those phenols containing a basic group (aminophenols). Compounds with only the acidic phenolic group show the typical sigmoidal k vs. pH* shape, but the shape for aminophenols (which have an acidic and a basic group) is a composite of two sigmoides with a plateau at the pH* values where the neutral form of the compounds predominates. We shall discuss

Table 2
Parameters for variation of the retention of ionizable solutes with mobile phase pH* (Eqs. (6) and (9)) for methanol–water (50:50, v/v) and a polymeric PLRP-S100 stationary phase [21]

Compound	p$K^*_{lit.}$	pK^*	S.D.	k_{HX}	S.D.	k_{X^-}	S.D.	log f
4-Chloro-2-nitrophenol	6.81	7.13	0.01	160.21	0.45	1.18	0.37	−2.13
2,4,6-Trichlorophenol	6.98	7.46	0.09	311.19	15.53	2.22	0.72	−2.15
2-Nitrophenol	7.86	7.77	0.20	102.68	3.59	0.84	3.18	−2.09
4-Nitrophenol	7.64	7.96	0.13	22.98	0.47	0.17	0.26	−2.13
4-Hydroxybenzaldehyde	8.25	8.21	0.10	5.92	0.09	0.22	0.10	−1.44
2,4-Dichlorophenol	8.53	8.55	0.01	166.78	0.58	1.03	0.59	−2.21
2-Chlorophenol	9.31	9.28	0.03	25.09	0.22	0.56	0.37	−1.65
Vanillin	8.02	9.34	0.11	10.17	0.29	0.05	0.50	−2.31
3-Bromophenol	9.71	10.11	0.11	54.64	0.64	0.28	1.20	−2.30
2-Naphthol	10.25	10.38	0.22	120.31	3.34	3.64	7.49	−1.52
4-Chlorophenol	10.15	10.45	0.06	33.47	0.32	0.55	0.65	−1.79
4-Chloro-3-methylphenol	10.30	10.50	0.24	74.74	2.60	2.62	6.05	−1.45
1-Naphthol	10.33	10.54	0.20	142.65	4.47	0.70	10.55	−2.31
4-Hydroxybenzyl alcohol	10.57	10.60	0.08	0.71	0.01	0.001	0.01	−2.85
2-Aminophenol	10.46	10.92	0.16	1.91	0.08	0.17	0.14	−1.06
3,5-Dimethylphenol	11.16	10.95	0.02	49.21	0.23	1.54	0.55	−1.51
3,4-Dimethylphenol	11.15	11.08	0.01	44.76	0.10	2.21	0.25	−1.31
m-Cresol	11.00	11.13	0.06	17.76	0.20	0.75	0.63	−1.37
2,5-Dimethylphenol	11.20	11.14	0.01	58.95	0.07	4.32	0.19	−1.13
o-Cresol	11.29	11.16	0.00	22.40	0.01	1.10	0.02	−1.31
3-Aminophenol	10.57	11.17	0.14	0.82	0.04	0.001	0.02	−2.91
Eugenol	10.77	11.18	0.06	142.80	1.77	0.22	5.76	−2.82
4-Aminophenol	11.08	11.19	0.13	0.48	0.01	0.08	0.03	−0.80
Phenol	10.80	11.22	0.05	7.06	0.06	0.07	0.20	−2.00
p-Cresol	11.06	11.39	0.04	16.08	0.10	1.01	0.45	−1.20
2,6-Dimethylphenol	11.40	11.85	0.10	54.52	0.29	4.01	4.99	−1.13
							Mean	−1.80
							S.D.	0.59

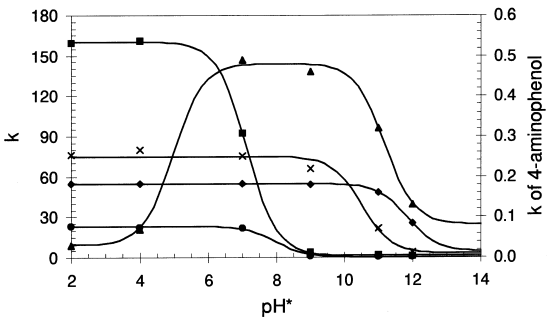

Fig. 1. Variation of retention factors for some phenols with mobile phase pH*. (■) 4-Chloro-2-nitrophenol, (×) 4-chloro-3-methylphenol, (♦) 2,6-dimethylphenol, (●) 4-nitrophenol, (▲) 4-aminophenol (scale on right-hand ordinate). The lines have been plotted from the parameters in Table 2 using Eq. (6) (Table 6 and Eq. (25)) for 4-aminophenol.

generalization of the equations and models to these amphiprotic and to basic compounds in the last part of this paper.

The results in Table 2 illustrate that the k_{HX} values can be accurately obtained with low standard deviations, except for 2,4,6-trichlorophenol. The retention of the neutral form of this compound is very large and k could not be measured at the acidic pH* values of 2 and 4. Since many phenols have high pK^* values, and are not completely in the basic X^- form at pH*=11 and even at pH*=12, the standard deviations of k_{X^-} values are, in general, larger than the standard deviations of k_{HX}. In two cases (4-hydroxybenzyl alcohol and 3-aminophenol), we have obtained k_{X^-} values almost equal to zero and to avoid infinite values in the log f calculation, they have been set to 0.001. Table 2 shows that the log f values obtained, although not identical, are reasonably constant, The average value (log $f=-1.80$) is used in all further correlations involving the log $[1-D(1-f)]$ solute descriptor.

Table 2 also presents the pK^* value (p$K^*_{lit.}$) used for the phenols in a previous work [21], which was estimated from literature pK values in pure water and in mixtures of methanol and water. There is only partial agreement between the two sets of pK^* values. The possible reasons for the discrepancy are several. On the one hand, the uncertainty associated with the estimation of p$K^*_{lit.}$ values [21] in 50% methanol from aqueous pK values or from pK^*/pH* measurements in water–methanol mixtures can be quite large. On the other hand, the uncertainty associated with pK^* determinations from retention data (partially reflected in the standard deviation value) also can be quite large because of the low number of $k-$pH* data analyzed and the errors associated with calibration and pH* measurement in mixed solvents. Since the log $[1-D(1-f)]$ descriptor is mobile phase pH* dependent, we have calculated this solute descriptor from the experimental pK^* values to average out errors associated with the pH* measurements. For the P solute descriptor, however, we have used the p$K^*_{lit.}$ values (already used in the previous study [21]) because the P solute descriptor in a particular mobile phase composition is independent of mobile phase pH*.

3.2. Comparison of retention models for neutral and ionic compounds

Table 3 summarizes the values of the log $[1-D(1-f)]$ solute descriptor for phenols at the different pH* values studied, taking log $f=-1.80$. Since the solute descriptor depends on the extent of ionization, and this changes with the mobile phase pH*, the descriptor is also pH dependent, and its value decreases with an increase in pH*. When the compound is almost completely in the neutral form (pH 2, 4 and even 7 for most phenols), the value of the descriptor is zero. When the phenol is highly ionized (pH 11 and 12 for the most acidic phenols), the descriptor tends to the log $f=-1.80$ value. For neutral compounds the solute descriptor is always zero. Therefore, the value of the descriptor is limited to values between 0 and -1.80.

The results obtained by applying Eq. (11) to the phenols and to the phenols combined with the neutral compounds for the different pH* values are presented in Table 4. At pH* 2 and 4, all the phenols are in the neutral form (except aminophenols) and only one correlation is presented. At these pH* values the log $[1-D(1-f)]$ solute descriptors are all zero (except for the aminophenols and pyridine, which are excluded from the correlations). This prevents calculation of the d coefficient. Nevertheless, good correlations are obtained using the model already established for neutral compounds [Eq. (1)]. At pH* values of 7, 9, 11, and 12, two correlations are presented. The first correlation includes only the

Table 3
Compilation of the log[1−D(1−f)] solute descriptor for phenols at different pH* values for methanol–water (50:50, v/v)

	pK*	pH*=2	pH*=4	pH*=7	pH*=9	pH*=11	pH*=12
4-Chloro-2-nitrophenol	7.13	0.000	0.000	−0.236	−1.536	−1.792	−1.796
2,4,6-Trichlorophenol	7.46	0.000	0.000	−0.127	−1.361	−1.788	−1.795
2-Nitrophenol	7.77	0.000	0.000	−0.067	−1.150	−1.780	−1.794
4-Nitrophenol	7.96	0.000	0.000	−0.044	−1.008	−1.772	−1.793
4-Hydroxybenzaldehyde	8.21	0.000	0.000	−0.026	−0.814	−1.755	−1.792
2,4-Dichlorophenol	8.55	0.000	0.000	−0.012	−0.563	−1.710	−1.787
2-Chlorophenol	9.28	0.000	0.000	−0.002	−0.180	−1.463	−1.748
Vanillin	9.34	0.000	0.000	−0.002	−0.160	−1.431	−1.741
3-Bromophenol	10.11	0.000	0.000	0.000	−0.032	−0.892	−1.545
2-Naphthol	10.38	0.000	0.000	0.000	−0.017	−0.685	−1.408
4-Chlorophenol	10.45	0.000	0.000	0.000	−0.015	−0.634	−1.367
4-Chloro-3-methylphenol	10.50	0.000	0.000	0.000	−0.013	−0.598	−1.336
1-Naphthol	10.54	0.000	0.000	0.000	−0.012	−0.570	−1.310
4-Hydroxybenzyl alcohol	10.60	0.000	0.000	0.000	−0.011	−0.528	−1.270
2-Aminophenol	10.92	–	–	0.000	−0.005	−0.335	−1.038
3,5-Dimethylphenol	10.95	0.000	0.000	0.000	−0.005	−0.319	−1.015
3,4-Dimethylphenol	11.08	0.000	0.000	0.000	−0.004	−0.257	−0.915
m-Cresol	11.13	0.000	0.000	0.000	−0.003	−0.236	−0.876
2,5-Dimethylphenol	11.14	0.000	0.000	0.000	−0.003	−0.232	−0.869
o-Cresol	11.16	0.000	0.000	0.000	−0.003	−0.224	−0.853
3-Aminophenol	11.17	–	–	0.000	−0.003	−0.220	−0.845
Eugenol	11.18	0.000	0.000	0.000	−0.003	−0.216	−0.838
4-Aminophenol	11.19	–	–	0.000	−0.003	−0.212	−0.830
Phenol	11.22	0.000	0.000	0.000	−0.003	−0.201	−0.807
p-Cresol	11.39	0.000	0.000	0.000	−0.002	−0.146	−0.678
2,6-Dimethylphenol	11.85	0.000	0.000	0.000	−0.001	−0.056	−0.373

Table 4
Fit of the of the solvation parameter model with the log[1−D(1−f)] solute descriptor (Eq. (11)) under different pH* conditions for phenols and phenols combined with neutral (marked as 'all') compounds[a]

System constants								Statistics[b]				
pH*	c	r	s	a	b	m	d	ρ	S.E.	F	n	Solutes
2	0.07	0.46	−0.28	−0.97	−2.87	2.85	–	0.980	0.120	167	40	All
4	0.06	0.44	−0.29	−0.96	−2.88	2.88	–	0.981	0.112	174	40	All
7	−0.11	0.61	−0.26	−0.79	−2.90	2.75	0.98	0.990	0.123	151	26	Phenols
7	0.02	0.44	−0.26	−0.96	−2.89	2.89	1.22	0.987	0.117	240	45	All
9	0.23	0.77	−0.63	−0.82	−2.79	2.54	0.95	0.992	0.109	195	26	Phenols
9	0.13	0.50	−0.40	−1.03	−2.76	2.81	1.01	0.989	0.111	270	45	All
11	0.22	0.53	−0.89	−0.60	−2.21	2.69	0.88	0.963	0.248	41	26	Phenols
11	−0.21	0.68	−0.31	−0.63	−2.83	2.79	1.19	0.992	0.126	156	23	Phenols
11	0.09	0.51	−0.37	−0.85	−2.82	2.84	1.18	0.989	0.124	263	42	All
12	0.56	0.77	−0.45	−0.58	−3.25	2.56	1.40	0.981	0.193	80	26	Phenols
12	0.24	0.51	−0.51	−0.57	−3.06	2.91	1.27	0.987	0.169	237	45	All
Mean	0.11	0.57	−0.38	−0.82	−2.90	2.78	1.15					
S.D.	0.21	0.13	0.12	0.17	0.15	0.13	0.16					

[a] Correlations for pH*=11 have been repeated after elimination of three outliers.
[b] ρ=overall correlation coefficient; S.E.=standard error in the estimate; F=F-statistic; n=number of solutes.

phenols, which are partially ionized at these pH* values, and the second correlation includes the phenols and the neutral compounds, that is to say all compounds of Table 1. The coefficients and statistics obtained with both sets of compounds are quite good and very similar. Only three outliers (vanillin, 4-hydroxybenzaldehyde and 2-chlorophenol at pH* 11) were detected, and the correlation for this pH* was repeated after elimination of the three outliers.

For comparison, we present the results obtained with Eq. (4) and the *P* descriptor in Table 5 [21]. There are a few small differences in the data set analyzed in the previous work in comparison with the data set analyzed here. Correlations at pH* 11 and 12 included 4-phenylphenol which is not included in Table 4 because the sparse data available prevented calculation of k_{X^-}, k_{HX}, and pK^* from Eq. (6). Correlations at pH* 2, 4, 7 and 9 included nitromethane, 1-nitropropane, and 1-nitrobutane which decompose at higher pH values. In order to maintain a similar set of compounds for all mobile phase pH* values, we have eliminated these compounds in the correlations presented in Table 4. In general, three correlations are presented in Table 5. The first correlation is for all phenols; the second for phenols after elimination of some outliers; and the third correlation, is for phenols and neutral compounds after elimination of the outliers. The statistics of the correlations are similar to those of Table 4, but the number of outliers is larger. If we consider the correlations with phenols: 2-nitrophenol at pH* = 7 (aminophenols excluded); 4-chloro-2-nitrophenol and vanillin at pH* = 9 and 11; and 4-chloro-2-nitrophenol, 2,4,6-trichlorophenol and 4-hydroxybenzyl alcohol at pH* = 12 were identified as outliers [21]. In the correlations with phenols and neutral compounds, the outliers were: 2-nitrophenol at pH* = 7; 4-chloro-2-nitrophenol; 2,4,6-trichlorophenol, and 4-hydroxybenzaldhyde at pH* = 9; 4-chloro-2-nitrophenol, 2,4-dichlorophenol, 2,4,6-trichlorophenol and 2-nitrophenol at pH* = 11; and 2,6-dimethylphenol and 4-hydroxybenzyl alcohol at pH* = 12.

Comparing the *d* and *p* system constants of Tables 4 and 5, two facts stand out. The value of the *d* coefficient is reasonably constant for all pH* values and compounds. In fact, it is close to the theoretically expected value of 1.00 (Eq. (10)). But the *p* system constant changes with the pH* value, and the compounds analyzed. Since the *P* solute descriptor is not pH dependent, it is expected to vary with changes in the mobile phase pH*, as observed. However, the robustness of the LFER model requires that it should not change significantly with the

Table 5
Fit of the solvation parameter model with the *P* solute descriptor (Eq. (4)) under different pH* conditions for phenols (with and without outliers) and phenols combined with neutral (after elimination of outliers, marked as 'all') compounds

System constants								Statistics				
pH*	*c*	*r*	*s*	*a*	*b*	*m*	*p*	*ρ*	S.E.	*F*	*n*	Solutes
2	−0.20	0.53	−0.29	−1.09	−2.86	3.14	−	0.989	0.093	302	40	All
4	−0.16	0.50	−0.34	−1.02	−2.83	3.12	−	0.989	0.096	300	41	All
7	0.00	0.66	−0.19	−0.77	−2.84	2.67	−0.44	0.982	0.127	71	23	Phenols
7	0.34	0.55	−0.22	0.48	−2.90	3.00	−0.49	0.991	0.091	139	22	Phenols
7	−0.06	0.47	−0.29	−0.70	−2.91	2.99	−0.45	0.991	0.098	335	46	All
9	0.86	0.98	−0.77	−0.69	−2.61	2.59	−2.77	0.964	0.229	42	26	Phenols
9	1.08	1.09	−0.53	−0.90	−2.93	2.36	−3.01	0.981	0.176	71	24	Phenols
9	0.24	0.72	−0.92	−0.52	−2.63	2.98	−1.21	0.980	0.144	154	45	All
11	0.46	0.00	0.00	−0.76	−2.16	3.34	−4.53	0.968	0.227	82	27	Phenols
11	0.96	0.65	−0.42	−0.66	−2.61	2.80	−4.72	0.990	0.137	155	25	Phenols
11	0.25	0.88	−1.18	−0.68	−2.34	3.01	−2.03	0.985	0.140	184	41	All
12	0.32	0.00	0.00	−0.85	−2.96	3.08	−4.11	0.963	0.225	70	27	Phenols
12	0.43	0.00	0.00	−0.64	−2.44	0.43	−4.80	0.985	0.154	158	24	Phenols
12	0.12	0.38	−0.32	−0.67	−2.82	2.85	−3.57	0.981	0.188	157	44	All
Mean	0.30	0.58	−0.45	−0.64	−2.73	2.67	−					
S.D.	0.43	0.29	0.35	0.43	0.21	0.82	−					

number of ionized solutes analyzed, but significant changes in the coefficient for the correlations with phenols or with phenols and neutral compounds are observed at pH* values of 9, 11 and 12, where the phenols are highly ionized. In this sense, the log [1-$D(1-f)$] solute descriptor performs much better than the P solute descriptor.

The constancy of the other coefficients of the correlations at the different pH* values is also better for the correlations with the log [1− $D(1-f)$] solute descriptor (Table 4) than for the correlations with the P solute descriptor (Table 5), as demonstrated by the average (mean) and standard deviations (S.D.) of the coefficients at the different pH* values, also presented in Tables 4 and 5.

3.3. Generalization of the models

The constancy of the coefficients in Table 4, allows accurate correlations to be obtained, including the log k data for the compounds at different pH* values. If the log k data for phenols at the different pH* values from 2 to 12 is correlated altogether using the model of Eq. (11), the following correlation is obtained.

$$\log k = 0.02 + 0.64R_2 - 0.39\pi_2^H - 0.79\Sigma\alpha_2^H - 2.79\Sigma\beta_2^0 + 2.69V_X + 1.00\log [1 - D(1-f)]$$
$$\rho = 0.982 \text{ S.E.} = 0.147 \; F = 654 \; n = 148 \qquad (12)$$

And if the neutral compounds are included in the same correlation

$$\log k = 0.09 + 0.48R_2 - 0.35\pi_2^H - 0.96\Sigma\alpha_2^H - 2.78\Sigma\beta_2^0 + 2.83V_X + 1.01\log [1 - D(1-f)]$$
$$\rho = 0.982 \text{ S.E.} = 0.130 \; F = 1153 \; n = 260. \qquad (13)$$

In these correlations only the log k data corresponding to partially protonated compounds (pyridine, 2-aminophenol, 3-aminophenol and 4-aminophenol at pH*=2 and pH*=4) have been excluded. Residual analysis of the data shows that there are five outliers in the correlations with standard deviations greater than three times the average standard error (S.E.). These ouliers are vanillin, 4-hydroxybenzaldehyde and 2-chlorophenol at pH*=11 and vanillin and 2-nitrophenol at pH*=12. If these outliers are eliminated, the correlation obtained for phenols is

$$\log k = 0.07 + 0.65R_2 - 0.40\pi_2^H - 0.78\Sigma\alpha_2^H - 2.86\Sigma\beta_2^0 + 2.67V_X + 1.09\log [1 - D(1-f)]$$
$$\rho = 0.989 \text{ S.E.} = 0.134 \; F = 999 \; n = 143 \qquad (14)$$

And the correlation for phenols and neutral compounds is

$$\log k = 0.11 + 0.48R_2 - 0.35\pi_2^H - 0.95\Sigma\alpha_2^H - 2.82\Sigma\beta_2^0 + 2.84V_X + 1.10\log [1 - D(1-f)]$$
$$\rho = 0.987 \text{ S.E.} = 0.122 \; F = 1622 \; n = 255 \qquad (15)$$

The plot of the log k predicted from this last correlation against the log k observed is presented in Fig. 2.

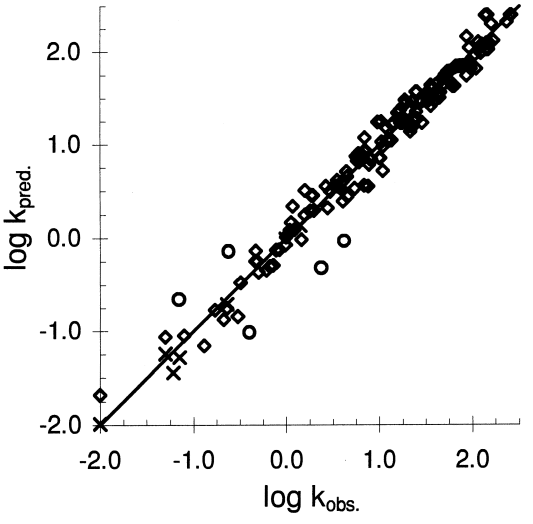

Fig. 2. Plot of log k predicted from Eq. (15) against experimental log k values for neutral, acidic, basic and amphiprotic compounds at several mobile phase pH* values. (◇) Neutral and acidic compounds, (×) basic and amphiprotic compounds (pyridine, 2-aminophenol, 3-aminophenol and 4-aminophenol at pH*=2 and pH*=4), (○) outliers (vanillin, 4-hydroxybenzaldehyde, and 2-chlorophenol at pH*=11 and vanillin and 2-nitrophenol at pH*=12).

The above correlations are of great value because the same model can be used for any compound at any mobile phase pH* value. This is not possible with the *P* solute descriptor because it is not pH dependent and the *p* system constant changes with the pH* value of the mobile phase.

In addition, the *P* solute descriptor requires multiple definitions for different types of solutes. For a weak acid it is defined from the dissociation constant of the solute in the particular solvent used as the mobile phase (K^*) as $P=(14-pK^*)/10$. However, this definition cannot be applied to neutral solutes. A neutral solute cannot dissociate and therefore K^* is equal to zero, pK^* tends to infinity and *P* would become minus infinite. For these solutes, we assigned an effective p$K^*=14$. This provides a zero value for the *P* solute descriptor. Nor does the main definition hold for fully ionized solutes, such as strong acids, bases, or salts of strong acids and bases. For these solutes, K^* tends to infinity, pK^* tends to minus infinity and *P* would become infinite. Although we have not studied these kinds of solutes, in consonance with the definition for neutral solutes, probably an effective p$K^*=0$ could be assigned to ionized solutes, which would lead to a *P* value of 1.4 for them. The above triple definition for *P* is only theoretical, since from a practical point of view the working pH range for HPLC is restricted to between 2 and 12, approximately. This means that any solute with a pK^* larger than 14 would be less than 1% ionized at the highest pH possible, and may be considered a neutral solute with $P=0$. Also, any solute with a pK^* lower than 0 would be more than 99% ionized and *P* may be taken as equal to 1.4. Any weak acid with a pK^* between 0 and 14 would give a *P* value between 0 and 1.4. It is not clear what would happen with basic compounds and how *P* should be defined for them.

The log $[1-D(1-f)]$ solute descriptor has a unique definition for all compounds. For a neutral compound (which in practice can be considered an acid with p$K^* \gg$ pH*), $D=0$ and the descriptor becomes zero too (and log $k=\log k_{HX}$ according to Eq. (10)). For a fully ionized compound (that is to say an acid with p$K^* \ll$ pH*), $D=1$, and the descriptor becomes log *f* (and log $k=\log k_{X^-}$). For partially ionized compounds with pK^* values close to the working pH* range (±2 pH units approximately), $0<D<1$, the solute descriptor has a value between 0 and log *f*.

In contrast with the *P* descriptor, in principle the same descriptor definition can be applied to bases provided that *f* is taken as the ratio between the retention factor of the ionized and neutral species. That is to say, for a base X with the acid–base equilibria

$$HX^+ \Leftrightarrow H^+ + X \quad K^* = [H^+][X]/[HX^+] \tag{16}$$

f must be defined as

$$f = k_{HX^+}/k_X \tag{17}$$

and *D* as the degree of ionization

$$D = [HX^+]/([X]+[HX^+]) = [H^+]/([H^+]+K^*)$$
$$= 10^{(pK^*-pH^*)}/[1+10^{(pK^*-pH^*)}] \tag{18}$$

Therefore,

$$k = (1-D)k_X + Dk_{HX^+} \tag{19}$$

And these definitions lead to the same kind of equation as for the acids (Eq. (10)),

$$\log k = \log k_X + \log[1-D(1-f)] \tag{20}$$

where k_X is the retention factor of the neutral species X, which should be linearly related with the solute descriptors V_X, R_2, π_2^H, $\Sigma\alpha_2^H$ and $\Sigma\beta_2^0$.

For an amphiprotic solute, such as an aminophenol (HX), the model should be modified to include two protonation equilibria.

$$H_2X^+ \Leftrightarrow H^+ + HX \quad K^*1 = [H^+][HX]/[H_2X^+] \tag{21}$$

$$HX \Leftrightarrow H^+ + X^- \quad K^*2 = [H^+][X^-]/[HX] \tag{22}$$

which leads to a definition of two degrees of ionization (D_+ and D_-), which correspond to the two ionic forms of the solute (H_2X^+ and X^-, respectively).

$$D_+ = [H_2X^+]/([H_2X^+]+[HX]+[X^-])$$
$$= 10^{(pK^*1-pH^*)}/[10^{(pK^*1-pH^*)}+1$$
$$+10^{(pH^*-pK^*2)}] \tag{23}$$

Table 6
Parameters for variation of the retention of ionizable basic (pyridine) and amphiprotic (aminophenols) solutes with mobile phase pH* (Eq. (25))

Compound	pK^*1	S.D.	pK^*2	S.D.	$k_{H2X}+$	S.D	k_{HX}	S.D	k_{X-}	S.D.	log $k_{H_2X}+/k_{HX}$
Pyridine	4.18	0.10	–	–	0.03	0.16	3.50	0.09	–	–	−2.11
2-Aminophenol	4.00	0.12	10.92	0.16	0.04	0.11	1.91	0.08	0.17	0.15	−1.66
3-Aminophenol	4.42	0.14	11.24	0.13	0.01	0.04	0.81	0.03	0.001	0.07	−2.08
4-Aminophenol	4.97	0.34	11.18	0.15	0.03	0.02	0.48	0.02	0.08	0.04	−1.23

$$D_- = [X^-]/([H_2X^+] + [HX] + [X^-]) = 10^{(pH^* - pK^*2)}/[10^{(pK^*1 - pH^*)} + 1 + 10^{(pH^* - pK^*2)}] \quad (24)$$

Eq. (6) must be written as

$$k = D_+ k_{H_2}X^+ + (1 - D_+ - D_-)k_{HX} + D_- k_{X^-} \quad (25)$$

and if we assume the constancy of the f parameter

$$f = k_{H_2}X^+/k_{HX} = k_{X^-}/k_{HX} \quad (26)$$

The final equation obtained for correlation is

$$\log k = \log k_{HX} + \log [1 - D_+(1-f) - D_-(1-f)] \quad (27)$$

where the solute descriptor that accounts for ionization is log $[1-D_+(1-f)-D_-(1-f)]$. Eq. (27) becomes

$$\log k = \log k_{HX} + \log [1 - D_+(1-f)] \quad (28)$$

when $pH^* \ll pK^*2$ ($D_- \approx 0$), and

$$\log k = \log k_{HX} + \log [1 - D_-(1-f)] \quad (29)$$

when $pH^* \gg pK^*1$ ($D_+ \approx 0$).

We have checked the validity of these equations for pyridine and aminophenols. The results obtained from Eq. (25) are presented in Table 6. One graphical example for 4-aminophenol is given in Fig. 1. Table 6 shows that the results obtained for the neutral and anionic forms (pK^*2, k_{HX}, and k_{X^-}) for the analysis of the whole $k-pH^*$ data are comparable to those presented in Table 2 for the analysis of the data for pH* values of 7 or higher. In addition, Eq. (25) provides reasonably precise pK^* and k constants for the protonated forms of the compounds (pK^*1 and $k_{H_2X^+}$). It also shows that the ratio of the retention factors for the protonated and neutral forms of the compounds are similar to the ratios presented in Table 2 for the anionic and neutral forms. Therefore, in principle, the same average log $f=-1.80$ value is applicable to protonation of bases and amphiprotic compounds.

The applicability of the general correlation Eq. (15) developed for acids and applied to bases has been tested for pyridine, 2-aminophenol, 3-aminophenol and 4-aminophenol. By using the pK^*1 values of Table 6 and the log $f=-1.80$ value, we have calculated D_+ and log $[1-D_+(1-f)]$ for the basic compounds at $pH^*=2$ and $pH^*=4$. From these results and those for the neutral solutes (Table 1), we have estimated the log k values of the compounds for mobile phase pH* values of 2 and 4 by Eq. (15) and compared the calculated log k values with the experimental values. Table 7 shows that the calculated values are very good, except for 4-aminophenol

Table 7
Prediction of the retention of ionizable basic solutes with acidic mobile phases ($pH^*=2$ and $pH^*=4$) from Eq. (15)

Compound	$pH^*=2$				$pH^*=4$			
	D_+	log $[1-D_+(1-f)]$	log k pred.	log k exp.	log D_+	log $1-D_+(1-f)]$	log k pred.	log k exp.
Pyridine	0.993	−1.65	−1.24	−1.30	0.602	−0.39	0.14	0.15
2-Aminophenol	0.990	−1.59	−1.44	−1.22	0.500	−0.29	−0.01	−0.01
3-Aminophenol	0.996	−1.70	−1.99	−2.00	0.724	−0.54	−0.71	−0.64
4-Aminophenol	0.999	−1.77	−2.18	−1.52	0.903	−0.95	−1.28	−1.15

at $pH^* = 2$. These estimations for basic compounds are also presented in Fig. 2.

4. Conclusions

The retention of neutral and ionizable solutes at different mobile phase pH^* values can be accurately predicted by the solvation parameter model, provided that an appropriate solute descriptor for the ionization of the phenol is included in the correlation equation. At least two different solute descriptors, P and $\log [1-D(1-f)]$, are available to account for this ionization.

The P solute descriptor has the advantage of simplicity, since it is easily calculated from the pK^* value of the phenol at the mobile phase composition. However, generalization of the correlation equation to different mobile phase pH^* values and to basic compounds is not possible. Prediction of retention from the $\log [1-D(1-f)]$ solute descriptor requires accurate mobile phase pH^* measurements and solute pK^* estimation, but the same correlation equation can be used to estimate retention of acids or bases at any mobile phase pH^*.

References

[1] M.H. Abraham, Chem. Soc. Rev. 22 (1993) 73.
[2] M.H. Abraham, in: P. Politzer, J.S. Murray (Eds.). Quantitative Treatments of Solute/Solvent Interactions, Elsevier, Amsterdam, 1994, pp. 83–134.
[3] M.H. Abraham, H.S. Chadha, in: V.Pliska, B. Testa, H. van de Waterbeemed (Eds.), Lipophilicity in Drug Action and Toxicology, VCH, Weinheim, 1996, pp. 311–337.
[4] K.G. Miller, C.F. Poole, J. High Resolut. Chromatogr. 17 (1994) 125.
[5] C.F. Poole, S.K. Poole, D.S. Seibert, C.M. Chapman, J. Chromatogr. B 689 (1997) 245.
[6] M.H. Abraham, M. Rosés, C.F. Poole, S.K. Poole, J. Phys. Org. Chem. 10 (1997) 358.
[7] M.H. Abraham, M. Rosés, J. Phys. Org. Chem. 7 (1994) 672.
[8] M.H. Abraham, H.S. Chadha, A.J. Leo, J. Chromatogr. A 685 (1994) 203.
[9] L.C. Tan, P.W. Carr, M.H. Abraham, J. Chromatogr. A 752 (1996) 1.
[10] P.W. Carr, L.C. Tan, J.H. Park, J. Chromatogr. A 724 (1996) 1.
[11] D.S. Seibert, C.F. Poole, J. High Resolut. Chromatogr. 18 (1995) 226.
[12] D.S. Seibert, C.F. Poole, Chromatographia 41 (1995) 51.
[13] D.S. Seibert, C.F. Poole, M.H. Abraham, Analyst 121 (1996) 511.
[14] A. Nasal, P. Haber, R. Kaliszan, E. Forgacs, T. Cserhati, M.H. Abraham, Chromatographia 43 (1996) 484.
[15] J. Li, P.W. Carr, Anal. Chim. Acta 334 (1996) 239.
[16] M.H. Abraham, H.S. Chadha, R.A.E. Leitao, R.C. Mitchell, W.J. Lambert, R. Kaliszan, A. Nasal, P. Haber, J. Chromatogr. A 766 (1997) 35.
[17] D. Bolliet, C.F. Poole, Chromatographia 46 (1997) 381.
[18] D. Bolliet, C.F. Poole, Analyst 123 (1998) 295.
[19] P.T. Jackson, M.R. Schure, T.P. Weber, P.W. Carr, Anal. Chem. 69 (1997) 416.
[20] S.K. Poole, C.F. Poole, Anal. Commun. 34 (1997) 247.
[21] D. Bolliet, C.F. Poole, M. Rosés, Anal. Chim. Acta 368 (1998) 129.
[22] P.J. Schoenmakers, R. Tijssen, J. Chromatogr. A 656 (1993) 577.
[23] R.M. Lopes Marques, P.J. Schoenmakers, J. Chromatogr. 592 (1992) 157.
[24] J.A. Lewis, D.C. Lommen, W.D. Raddatz, J.W. Dolan, L.R. Snyder, I. Molnár, J. Chromatogr. 592 (1992) 183.
[25] C. Horváth, W. Melander, I. Molnár, Anal. Chem. 49 (1977) 142.
[26] S.N. Deming, M.L.H. Turoff, Anal. Chem. 50 (1978) 546.
[27] M. Rosés, I. Canals, H. Allemann, K. Siigur, E. Bosch, Anal. Chem. 68 (1996) 4094.