Journal of Chromatography, 592 (1992) 183–195
Elsevier Science Publishers B.V., Amsterdam

CHROMSYMP. 2440

Computer simulation for the prediction of separation as a function of pH for reversed-phase high-performance liquid chromatography

I. Accuracy of a theory-based model

J. A. Lewis, D. C. Lommen, W. D. Raddatz, J. W. Dolan and L. R. Snyder*

LC Resources Inc., 2930 Camino Diablo, Walnut Creek, CA 94596 (USA)

I. Molnar

Institut für angewandte Chromatographie, Berlin (Germany)

ABSTRACT

Computer simulation software (DryLab I/mp) is described for predicting high-performance liquid chromatographic separation as a function of changes in mobile phase pH. Three experimental runs with pH (only) varied are used to derive values of pK_a plus capacity factors (k') for the ionized and non-ionized form of each ionizable solute. Various tests of the experimental data then allow classification of each solute as acidic, basic, neutral (including strong or weak acids or bases) and amphoteric. Experimental data are reported for the separation of several substituted anilines as a function of pH and solvent composition (%B). Experimental requirements for the accurate prediction of separation (*ca.* ±2–4% in α) as a function of pH are discussed. The reliability of the software is demonstrated for three different samples: mixtures of (a) substituted benzoic acids, (b) substituted anilines and (c) catecholamine-related compounds.

INTRODUCTION

A considerable literature exists for resolution mapping in high-performance liquid chromatography (HPLC) method development [1,2]. In most instances, empirical fitting functions are used with a small number of experimental measurements in order to predict separation as a function of some mobile phase variable or variables. A well known example of this approach is the work of Glajch *et al.* [3] for mapping resolution as a function of mixtures of methanol, acetonitrile and tetrahydrofuran, based on only seven experimental runs. Sachok *et al.* [4] first decribed a similar approach for predicting separation as a function of pH and ion-pair reagent concentration.

The usual goal of resolution-mapping procedures is to achieve accurate predictions of separation for a reasonably broad range of conditions, with a minimum number of initial experimental runs. When only one variable is to be mapped, the number of required experiments may be reasonably small. However, the simultaneous variation of two or more variables leads to a rapid increase in the number of experiments that are necessary. This problem is exacerbated by a need for peak tracking [1,2], *i.e.*, the matching of bands for a given sample component among the various experimental runs.

We have recently developed a software package (DryLab I/mp) which allows the chromatographer to use computer simulation for the mapping of any mobile phase parameter, including temperature. In

0021-9673/92/$05.00

this paper we examine the use of this software for predicting sample retention and separation as a function of mobile phase pH. In Part II [5] this discussion is extended to separation as a function of simultaneous change in pH and solvent strength (%B).

THEORY

General considerations

Resolution vs. retention: allowable errors in predicted values of α. Most comparisons of experimental data with predictions from resolution mapping emphasize the accuracy of predicted retention times for individual solutes. However, the chromatographer is usually more interested in accurate predictions of resolution, R_s, or *differences* in retention time. The effect on predicted values of resolution of a $\pm 3\%$ error in α is illustrated in Fig. 1 for two different (average) values of k': $k' = 1$ and 20. Generally, we aim for a resolution of $R_s = 1.5$ or greater, as in the example in Fig. 1. If the experimental separation is within ± 0.5 resolution units of the predicted value ($R_s = 1.5$), then computer simulation can be useful in method development. For $k' = 20$, a $\pm 3\%$ error in α results in an error in R_s of ± 0.4 units, *i.e.*, an acceptable prediction. For $k' = 1$, the error in R_s that results from a $\pm 3\%$ error in α is even smaller (only ± 0.2 units), and the allowable error in α for $k' = 1$ is about $\pm 7\%$. The fundamental resolution equation[2]

$$R_s = \tfrac{1}{4}(\alpha - 1)\, N^{1/2}\, [k'/(1+k')] \qquad (1)$$

allows us to generalize the example in Fig. 1 as follows. For an acceptable error in R_s of ± 0.5 units, the allowable fractional error in α ($\delta\alpha$) is

$$\delta\alpha \approx 2\, N^{-1/2}\, [(k'+1)/k'] \qquad (2)$$

that is, the accuracy in predicted values of α become more important as (a) k' increases and (b) N be-

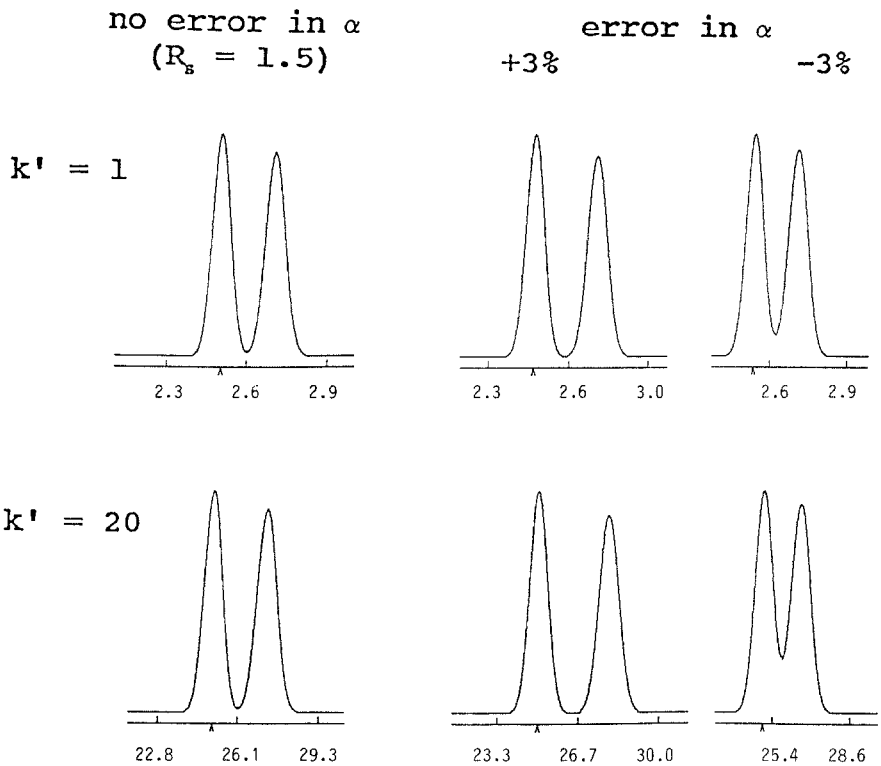

Fig. 1. Effect of error in the prediction of α on a resulting computer-simulated separation. A plate number of $N = 5000$ is assumed; the value of α ("no error") is 1.18 for $k' = 1$ and 1.09 for $k' = 20$.

comes larger (N = 5000 in the example in Fig. 1). Sample resolution is usually more critical for small values of k'; for representative values of k' = 2 and N = 10 000, eqn. 2 suggests that errors as large as ±4% in predicted values of α should be acceptable in the use of computer simulation for HPLC method development. However, when the column plate number is large, *e.g.*, N = 20 000, the required accuracy of predicted values of α may increase significantly (*e.g.*, allowable errors no larger than ±2%). The preceding analysis assumes that $\alpha \approx 1.0$, which will be the case for the critical (most overlapped) pair of bands in the chromatogram of a "difficult" separation.

Predicting retention as a function of pH

A general theory for the reversed-phase retention of monoprotic acidic or basic solutes as a function of pH is available [6–10]. It can be assumed that a given solute exists in ionized (±) and non-ionized (^{0}) forms, in which case the solute capacity factor value k' is given by

$$k' = k^0 (1 - F^{\pm}) + k^+ F^{\pm} \quad (3)$$

where k^0 and k^+ refer to k' values for the non-ionized and ionic forms and $F^{\pm}$ is the fraction of solute molecules that are ionized. The fraction $F^{\pm}$ of ionized molecules is

$$F^+ = 1/\{1 + (K_a/[H^+])\} \quad (4)$$

for the case of a basic solute and

$$F^- = 1/\{1 + ([H^+]/K_a)\} \quad (5)$$

for an acidic compound.

Several studies [7,8,11] have compared experimental data for acidic solutes with eqns. 3 and 5 (corresponding comparisons for basic solutes with eqns. 3 and 4 are less common); data for a reasonably wide pH range were fitted to eqns. 3–5 with good agreement. However, significant deviations from theory sometimes occur, especially for pH values that are either $\gg pK_a$ or $\ll pK_a$. Resolution mapping has also been applied as a means of optimizing mobile phase pH in reversed-phase HPLC. Eqns. 3–5 [4,8,9,12], polynomial equations [13] or hybrid relationships [14] have been used as fitting functions. The use of eqns. 3–5 allows predictions of separation as a function of pH, based on three initial experiments [three unknowns exist for each solute: k^0, k^+ (or k^-) and K_a].

Potential errors in the use of eqns. 3–5. The use of eqns. 3–5 for mapping retention *vs.* pH is based on certain approximations. Error in predicted separations can be expected as a result of various effects:

(i) retention of solutes (especially protonated bases) by processes other than solvophobic interaction; *e.g.*, interactions with exposed silanols or metal contaminants [15], whose pK_a or complexing constants can also vary with pH;

(ii) change in K_a values as a function of ionic strength; *i.e.*, buffer concentration is usually maintained constant, but the fraction of the buffer that is ionized then varies with pH;

(iii) solvophobic effect of ionic strength on solute retention (hydrophobic interaction);

(iv) ion-pair interaction of sample ions with ionized buffer species;

(v) change in the microscopic nature (and sorption properties) of the stationary phase (C_8 or C_{18}) as a result of changing ionization of silanols;

(vi) a change in buffer type, when more than one buffer is needed to cover a given pH range.

Additional complications in pH mapping are presented by the presence of neutral, polybasic and/or amphoteric sample components [9,16]. A retention-mapping scheme must be able to recognize these various solute types and treat them differently. Finally, the retention of neutral species is in some instances dependent on sample pH, although this dependence is much less pronounced than for ionizable species. The questions then are the extent to which eqns. 3–5 are useful for mapping resolution *vs.* pH, assuming only three experimental runs as a starting point, and the means that can be used to avoid major errors as a result of this approach.

Software development

Identifying solute type. Prior to the use of eqns. 3–5 for mapping retention *vs.* pH, it is necessary to distinguish among four possible situations, as summarized in Fig. 2. From the experimental retention times of a solute at three different pH values, it is possible to recognize each solute type: acid, base, amphoteric or neutral. The use of retention data for this purpose is complicated, however, by various effects, including experimental error in these retention time measurements. As seen in the idealized plot of retention *vs.* pH for the basic solute in Fig. 3, the major change in solute retention (and resulting

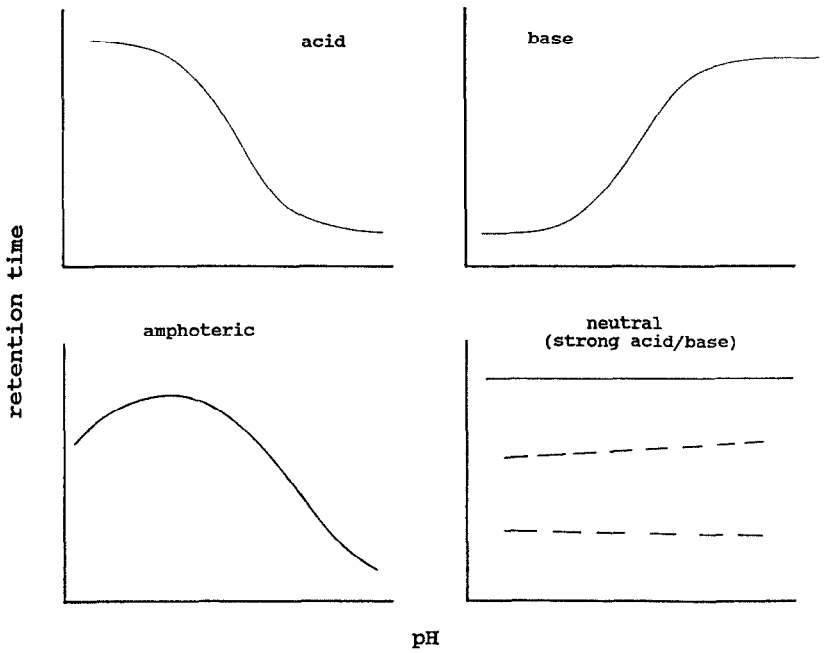

Fig. 2. Illustration of the dependence of retention time on pH for different solute types.

values of α) occurs for pH values that are close to the pK_a value of the solute. For this reason, pH values that bracket the pK_a range of the sample offer the greatest opportunity for varying selectivity and optimizing separation. The open circles in Fig. 3 represent possible experimental data to be used as input for computer simulation. Derived values of K_a, k^+ and k^0 will be most accurate when the experimental input data bracket the pK_a value of the solute, *e.g.*, pH range B in Fig. 3 or pH = 3, 4 and 5 in this example (pK_a = 4).

Retention data collected in pH ranges A or C in Fig. 3 may not be accurate enough to use for deriving values of K_a, k^+ and k^o, as small errors in retention can result in large errors in the derived solute parameters and in related predictions outside this pH range. Further, it may not be possible to distinguish acidic or basic solutes from neutral compounds when there is little change in retention as the pH is varied. For this reason, it is necessary to set up test conditions that arbitrarily assign different solutes to different groups (acid, base, neutral, etc.) on the basis of changes in retention with pH. This is discussed further in the Appendix.

EXPERIMENTAL

Equipment and software

An LC Analyst Expert Method Development System (Perkin-Elmer, Norwalk, CT, USA) was used to carry out the HPLC experiments. Computer simulations were performed with DryLab I/mp

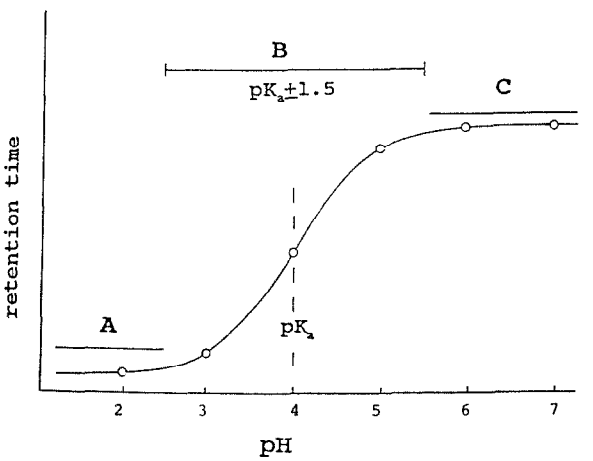

Fig. 3. Retention time *vs.* pH for a basic solute: effect of experimental pH values on accuracy of predicted separation.

TABLE I

RETENTION TIMES FOR SUBSTITUTED ANILINES AS A FUNCTION OF pH AND METHANOL CONCENTRATION

Conditions: 25 × 0.46 cm I.D. StableBond CN column; methanol–buffer mobile phases, 25 m*M* sodium citrate (pH ⩾ 4.0) or potassium phosphate buffer (pH < 4.0); flow-rate, 1 ml/min; temperature, 35°C.

Methanol concentration (%)	pH	Retention time (min)[a]									
		A	B	C	D	E	F	G	H	I	J
25	2.00	2.84	3.01	5.84	7.19	3.58	4.27	3.60	3.01	13.84	29.03
	2.25	2.93	3.13	6.80	8.54	3.94	5.09	3.92	3.15	18.09	33.8
	2.50	3.13	3.43	8.22	10.43	5.00	6.95	4.30	3.40	24.85	39.1
	2.75	3.13	3.52	8.71	11.09	5.80	8.21	4.46	3.51	28.19	40.7
	3.00	3.28	3.80	9.13	11.57	6.86	9.52	4.85	3.71	30.25	41.30
	3.25	3.31	4.04	9.35	11.88	8.11	10.68	5.35	3.90	32.68	42.94
	3.50	3.53	4.58	9.42	12.12	9.54	11.79	6.23	4.39	33.41	42.67
	4.00	3.85	5.53	9.75	12.39	11.54	12.87	8.26	5.56	34.63	43.32
	4.50	4.75	7.33	9.64	12.49	12.52	13.44	10.88	8.74	34.35	42.25
	5.00	5.67	8.36	9.80	12.59	13.11	13.77	13.10	11.77	35.15	43.13
	5.50	6.59	8.98	9.87	12.73	13.37	13.94	14.39	14.2	35.50	43.52
	6.00	7.07	9.31	10.05	12.97	13.65	14.19	15.15	15.5	36.38	44.57
	6.50	7.28	9.47	10.16	13.14	13.82	14.36	15.54	15.9	36.98	45.30
35	2.0	2.59	2.77	5.19	6.29	3.308	4.02	3.08	2.77	10.84	18.92
	2.5	2.90	3.16	6.41	8.02	4.753	6.33	3.65	3.09	16.41	22.56
	3.0	3.05	3.55	6.66	8.36	6.228	7.80	4.21	3.41	17.85	22.66
	3.5	3.19	4.38	6.78	8.58	7.903	8.86	5.63	4.24	19.12	23.58
	4.0	3.68	5.14	6.86	8.69	8.719	9.27	6.56	5.41	19.08	23.25
	4.5	4.20	5.91	6.67	8.50	8.665	9.07	7.90	7.27	18.25	22.14
	5.0	4.73	6.40	6.79	8.70	8.929	9.40	9.00	8.80	18.84	22.86
	5.5	5.11	6.70	6.83	8.73	9.081	9.40	9.40	9.83	19.03	23.09
	6.0	5.30	6.82	6.93	8.89	9.252	9.59	9.73	10.32	19.51	23.70
	6.5	5.36	6.92	6.97	8.96	9.321	9.71	9.88	10.51	19.72	23.95

[a] Solutes: A, 4-methoxyaniline; B, 3-methylaniline; C, 3-cyanoaniline; D, 2-chloroaniline; E, 4-chloroaniline; F, 3-chloroaniline; G, 3,5-dimethylaniline; H, N-ethylaniline; I, 3,4-dichloroaniline; J, 3,5-dichloroaniline.

software (LC Resources, Lafayette, CA, USA). This program allows the user to begin with a single isocratic experiment and then map separation as a function of different experimental conditions with one or two additional runs for each parameter (*e.g.*, percent organic component, ternary solvent mixtures, temperature, pH, ion-pair reagent or buffer concentrations).

Materials and procedures

All solvents were of HPLC grade (Burdick & Jackson, Muskegon, MI, USA). The substituted-aniline sample (Table I) was formulated from compounds supplied by Aldrich (Milwaukee, WI, USA). StableBond CN and C_8 columns (25 × 0.46 cm I.D.) MacMod Analytical, Chadds Ford, PA, USA) were used. These new stationary phases are synthesized by the reaction with silica of a sterically protected, monofunctional silane: $ClSi(i\text{-}C_3H_7)_2(CH_2)_3CN$ or $ClSi(i\text{-}C_3H_7)_2(CH_2)_3C_8$. The silica contains undetectable amounts (<1 ppm) of heavy metals such as iron or aluminium and is processed so as to avoid the presence of "acidic" silanols [17,18].

All separations were carried out at 35°C with a flow-rate of 1.0 ml/min. UV detection at 255 nm was used. The mobile phase was formulated from methanol (B) plus aqueous buffer. The buffer was variously 25 m*M* sodium citrate (pH 4–6.5) or 25 m*M* sodium phosphate (pH 2–4). When a change in mobile phase pH was made, equilibration with the new mobile phase was carried out for at least 20 min

(eight column volumes) before injecting the next sample. An equilibration time of 10 min was found to achieve 95% equilibration (as measured by retention time values) for a change in pH from 3 to 6 for the substituted-aniline sample; smaller changes in pH from run to run were the rule.

The precision of reported values of retention time and α was determined from replicate separations during the time the data in Table I were collected. Retention times were repeatable within ± 0.8% (1 S.D.) and values of α had a precision of ± 1.5% (1 S.D.).

RESULTS AND DISCUSSION

Accuracy of computer simulation for a change in mobile phase pH

Assume that three mobile phases of different pH ($pH_1 < pH_2 < pH_3$) are selected for the initial experimental runs with a given sample; we need to know the likely error in predicted values of α for other pH values. In this way, computer simulations can be restricted to a range in pH that allows sufficiently accurate predictions. Presumably the error in the predicted values of α will depend on the values of pH_1, pH_2 and pH_3.

Returning to Fig. 3, assume that the pH values for the initial experimental runs are 3, 4 and 5. There will then be some error in the predicted values of α for other pH values. In general, predictions should be more reliable for the case of interpolation ($3 < pH < 5$) than for extrapolation ($pH < 3$ or > 5). Similarly, as the difference between pH_3 and pH_1 becomes smaller, interpolation should become more accurate and extrapolation less accurate. The most useful predictions will generally be possible when $pH_2 \approx pK_a$ for a given solute, as in this pH region α and resolution usually change the most as pH is varied.

Substituted aniline sample. A mixture of ten substituted anilines was used as sample for these studies. The mobile phase pH was varied from 2 to 6.5 in 0.25- or 0.50-unit increments for mobile phases that contained either 25 or 35% methanol. Table I summarizes the retention data for each of these sample components, using a StableBond CN column. A 25 m*M* phosphate buffer was used to control pH for pH < 4.0, and citrate buffer was used for higher pH values. Retention times (t_R) for pH = 4.00 (25% and 35% methanol) were similar for either citrate or phosphate buffer. Thus, we observed that at pH 4.0

$$t_R \text{ (phosphate)} = 1.02\ t_R \text{ (citrate)} \quad (6)$$

within ±2% (1 S.D.). At pH 3.0, there was a greater deviation between retention times for the two buffers:

$$t_R \text{ (phosphate)} = 1.11\ t_R \text{ (citrate)} \quad (7)$$

within ±6% (1 S.D.). The greater difference in retention at lower pH for citrate *vs.* phosphate buffers is expected, as the sample becomes significantly ionized at pH < 4.

Our present software was used with these data to predict separation (resolution, retention times, etc.) and values of α. Input data were selected for three pH values, and values of α for other pH values were subsequently obtained by computer simulation. Predicted and experimental results were then com-

TABLE II

PREDICTED *vs.* EXPERIMENTAL SEPARATION OF SUBSTITUTED ANILINE SAMPLES

Conditions as in Table I, except pH = 3.5 and 25% B. Computer simulation based on input data for pH of 3, 4 and 5.

Solute[a]	Retention time (min)		α		pK_a
	Expt.	Calc.	Expt.	Calc.	
4-Methoxy	3.54	3.47	1.88	1.95	4.54
N-Ethyl	4.40	4.33	1.10	1.07	4.15
3-Methyl	4.58	4.46	1.82	1.91	Neutral[b]
3,5-Dimethyl	6.24	6.19	1.86	1.91	Neutral[b]
3-Cyano	9.42	9.46	1.02	1.03	3.26
4-Chloro	9.54	9.65	1.32	1.31	2.96
3-Chloro	11.8	11.8	1.03	1.02	4.08
2-Chloro	12.1	12.0	3.21	3.25	4.55
3,4-Dichloro	33.2	33.3	1.31	1.30	Acid[c]
3,5-Dichloro	42.6	42.4			Neutral[b]
Av. error		±1%		±2%	

[a] Substituent on aniline listed, *e.g.*, 4-methoxyaniline.
[b] DryLab I/mp classifies solute as a neutral species as described in the Appendix.
[c] Data cannot be fitted to eqns. 3–5; linear-segment fit to data.

pared in order to assess the accuracy of computer simulation. This procedure is illustrated in Table II.

The data of Table II show good agreement between experimental and predicted retention times (± 1%) and values of α (± 2%). Fig. 4 compares the predicted and actual separations for pH 3.5 and 25% B. In this example, the accuracy of predicted values of α is excellent. [Note also that three solutes are classified as "neutral", because retention did not change much with pH. Also, one solute (3,4-dichloroaniline) required adjustment of the experimental retention times (by the computer) in order to obtain a fit of the data to eqns. 3–5; see the discussion in the Appendix]. The procedure in Table II was re-

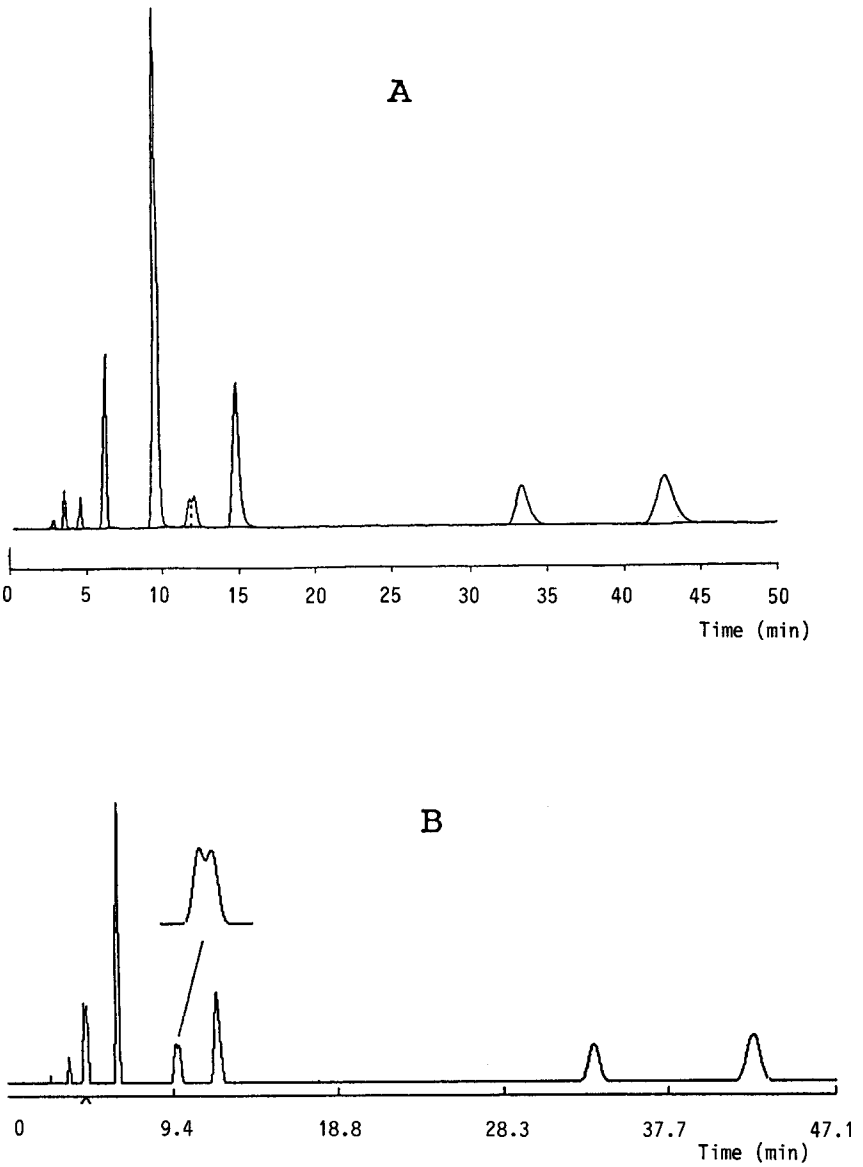

Fig. 4. Experimental and predicted chromatograms for separation of a substituted aniline sample. Conditions as in Table II (pH 3.5, 25% B).

TABLE III

SUMMARY OF EXPERIMENTAL *vs.* SIMULATED RETENTION DATA FOR SUBSTITUTED ANILINES (AS IN TABLE II)

Based on experimental data in Table I (25% and 35% B data averaged together).

ΔpH	Error in retention time (%)					Error in α (%)				
	Int.	Extrapolated[a]				Int.	Extrapolated[a]			
		−1.0	−0.5	0.5	1.0		−1.0	−0.5	0.5	1.0
0.5	–		25	7				13	10	
1.0	1	21	4	3	8	2	15	8	8	21
2.0	2	23	8	2	4	2	15	8	3	6
3.0	2		23	2	4	3		15	2	2
4.0	7					5				

[a] Numbers (−1.0, −0.5, etc.) refer to extrapolation range; *e.g.*, if pH_1 is 3.0, −0.5 refers to a pH of 2.5.

peated for various combinations of input runs and for different pH values for the predicted separation, *e.g.*, pH 2, 3 and 4 for input runs with prediction of separation for pH = 2.25, 2.5, 2.75, 3.25, etc. Table III summarizes the results of these computer simulations as a function of (a) the range of pH values used as input ($\Delta pH = pH_3 - pH_1$) and (b) whether interpolation or extrapolation was used to obtain the predicted separation.

The data in Table III for the substituted anilines show that interpolated values of α are accurate to better than ±3% (values of α were generally less reliable when $k' < 1$ for one or both bands; data for $k' < 1$ are not included in the summaries in Tables III and IV, because $k' > 1$ for all solutes is a general goal of method development [19]), when ΔpH is no greater than 3.0. For ΔpH = 4 the accuracy of interpolated predictions is ±5%, which is marginal for the use of computer simulation as an aid in HPLC method development. This is intuitively reasonable. For extrapolation to higher values of pH, small values of ΔpH give unacceptable accuracy in α (±8–10%) for an extrapolation of 0.5 units, while an extrapolation of 1 unit gives even larger errors (6–21%). Extrapolation to lower pH values is in every instance unreliable, yielding errors of 8% or larger in α for extrapolation by −0.5 units or more. The reason for the poorer predictions when extrapolating to lower pH values (and lower retention times) is the result of several factors: generally smaller values of k', larger changes in k' and α for this sample and a failure of eqns. 3–5 to apply exactly. The accuracy of simulated separations was generally better at pH values which correspond to decreased ionization of the sample.

To conclude, the data in Table III for this aniline sample suggest that ΔpH should be no larger than 3 pH units for this sample, interpolated predictions of

TABLE IV

SUMMARY OF EXPERIMENTAL *vs.* SIMULATED RETENTION DATA FOR SUBSTITUTED BENZOIC ACIDS (AS IN TABLE III)

Based on experimental data in ref. 6.

ΔpH	Error in retention time (%)		Error in α (%)	
	Int.	Extrapolated[a]	Int.	Extrapolated[a]
Without phthalic acid				
1.6	–	7	–	3
1.2	1	4	0	2
1.8	1	–	1	–
With phthalic acid[b]				
0.6	–	11	–	6
1.2	2	7	1	7
1.8	2	–	3	–

[a] Extrapolated 0.3–0.6 pH units.

[b] Retention time data for all solutes averaged together under "with phthalic acid" category.

α will then be reliable to ±2% and extrapolation is only recommended for pH values that are higher (by no more than 1.0 pH unit) than the pH values used as input for computer simulation.

Substituted benzoic acid sample. We have previously reported retention *vs.* pH data [6] for several substituted benzoic acids, similar to the data of Table I for the substituted anilines. The pH range covered in that study [6] was 2.6–4.4, in increments of 0.3 pH units. Table IV summarizes the application of computer simulation to these data, as in Table III for the anilines. Errors in predicted retention times range from 1 to 7% for all monoprotic solutes, including both interpolated retention times and values extrapolated by as much as 0.6 pH units. Predicted values of α are of greater interest: if phthalic acid is excluded, interpolated values of α are predicted with an average accuracy of ±1%, *i.e.*, quite good. Values of α obtained by extrapolation by as much as ±0.6 pH units are also reliable (±2–3%), in contrast to the case with the aniline sample.

Phthalic acid is a diprotic solute (pK_a values of 2.9 and 5.4) and as such its retention should not be described exactly by eqns. 3–5 (for pH values outside the range 2.9–5.4, the fit of retention data for phthalic acid to eqns. 3–5 should be better). The errors in predicted values of α (based on eqns. 3–5) for band pairs that include phthalic acid are generally about three times greater than for the remaining substituted benzoic acids in Table IV. This means that interpolated α-values are predicted with an accuracy of ±1–3%, depending on the value of ΔpH, which is acceptable for purposes of method development. Computer simulation based on extrapolation appears to be unreliable in the case of phthalic acid. By analogy we assume that this will also be true for other polyprotic acidic solutes.

Catecholamine metabolites. Data for the retention of seven catecholamine solutes as a function of pH are reported in ref. 20. Over the pH range 2.5–5.5, three of these compounds behave as acids (retention decreasing with pH) and the remaining four compounds are neutral. These data could be used to assess the accuracy of computer simulation, as in the prior examples summarized in Tables III and IV. An example is provided in Table V, and a summary as in Tables III and IV is given for the catecholamines in Table VI.

The nature of the solute (acidic or neutral) was correctly identified in every case (see Table V). The data in Table VI show that acceptable accuracy is obtained for interpolated values of α, when ΔpH < 3, but extrapolated values of α are at best marginally reliable. These results are probably representative of what can be expected by the average chromatographer.

Other errors. The components of the substituted aniline and benzoic acid samples are known to consist of either acids or bases. When the mobile phase

TABLE V

PREDICTED *vs.* EXPERIMENTAL SEPARATION OF CATECHOLAMINE METABOLITES

Conditions as in ref. 20, except pH = 3.0. Simulation based on input data for pH of 2.5, 3.5 and 4.5.

Solute[a]	Retention time (min)		α		pK_a
	Expt.	Calc.	Expt.	Calc.	
NA	1.56	1.55	3.20	3.21	Neutral
DA	2.94	2.91	2.23	2.29	Neutral
α-MDA	5.42	5.48	1.20	1.18	Neutral
DOPAC	6.32	6.31	1.24	1.25	Acid
5-HT	7.59	7.68	1.76	1.71	Neutral
5-HIAA	12.66	12.45	1.42	1.45	Acid
HVA	17.62	17.59			Acid

[a] NA = Noradrenaline; DA = Dopamine; α-MDA = α-methyldopamine; DOPAC = 3,4-dihydroxyphenylacetic acid; 5-HT = 5-hydroxytryptamine; 5-HIAA = 5-hydroxyindoleacetic acid; HVA = homovanillic acid.

TABLE VI

SUMMARY OF EXPERIMENTAL *vs.* SIMULATED RETENTION DATA FOR CATECHOLAMINES (AS IN TABLE III)

Based on experimental data in ref. 20.

ΔpH	Error in α (%)	
	Int.	Extrapolated[a]
1.0	–	6
2.0	1	6
3.0	3	–

[a] Extrapolation by ±0.5 pH units.

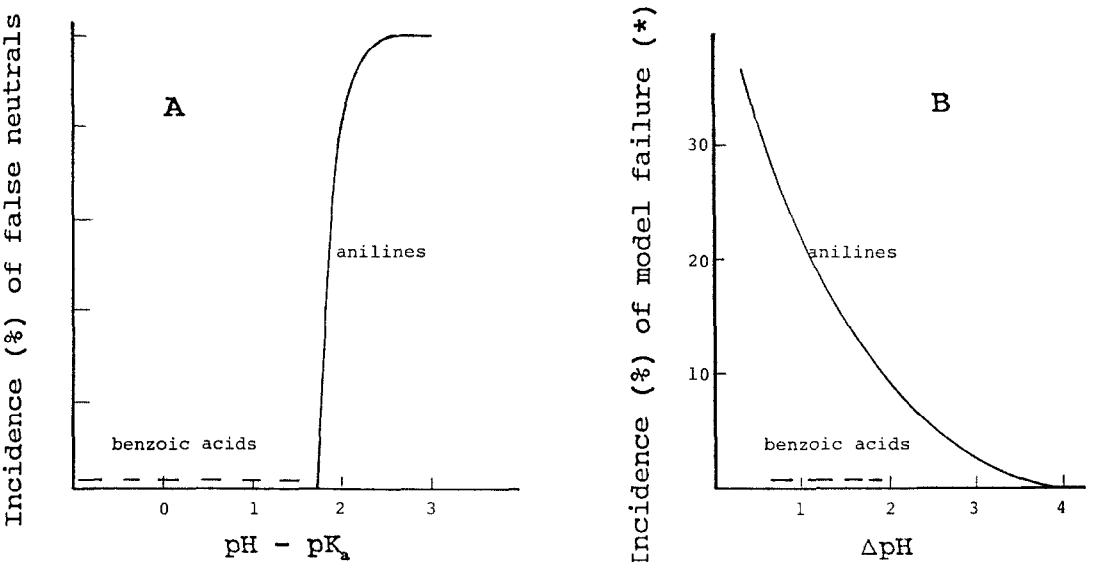

Fig. 5. Error rate in the application of computer simulation to the substituted aniline and benzoic acid samples. (A) Incidence of "false neutrals" as a function of pH; (B) incidence of error flags (*) as a function of ΔpH.

pH is far removed from the pk_a of a solute, the present software will classify that compound as neutral. We determined the incidence of "false neutrals" in these two samples as a function of pH $-$ pK_a, as summarized in Fig. 5A. The pH range over which the benzoic acids were studied was not sufficiently different from the pK_a values of these compounds to result in any "false neutrals". With the aniline sample, where pH was varied over wider limits, a sharp rise in the incidence of "false neutrals" is seen when pH $-$ pK_a exceeds 1.8. This is expected from the discussion of Fig. 3.

When a compound has been classified as either an acid or a base, but the experimental retention data cannot be fitted to eqns. 3–5 without adjustment (Appendix), DryLab I/mp flags these data with an asterisk. The incidence of error flags (asterisk) as a function of ΔpH is indicated in Fig. 5B for both the benzoic acid and aniline samples. It is seen that these errors are much more frequent for the aniline sample than for the benzoic acids. This observation together with the lower accuracy of predicted separations for the anilines suggests that computer simulation may be generally less reliable with basic samples (if the model of eqns. 3–5 is used). We also see in Fig. 5B that error flags become less frequent as ΔpH increases. The data in Tables III–V in conjunction with Fig. 5B suggest that ΔpH values between 1 and 3 are preferred for the most reliable predictions based on computer simulation (and eqns. 3–5). Smaller values of ΔpH make peak tracking easier, whereas larger ΔpH values allow computer simulation to be used over a wider range of pH values.

DryLab I/mp informs the user when a predicted separation for a requested pH value is likely to be unreliable, because of excessive extrapolation or the use of initial data with too large a value of ΔpH. Without these or similar computer advisories, the use of computer simulation (*i.e.*, based on eqns. 3–5 with three input runs) for the optimization of mobile phase pH is likely to be a frustrating experience for the user.

Derived values of pK_a: selection of the optimum pH range

A knowledge of the pK_a values of the components of a sample can be useful for a number of reasons, especially as an aid in selecting experimental conditions during method development [19]. If the initial three runs required for predictions of separation as a function of pH are found not to bracket the pK_a values of the sample, the predicted values of pK_a (from the first simulation) can be used to guide further experimentation: one or more additional runs with pH varied can be used to expand the pH

range accessible to accurate computer simulation. The question then arises of how accurate these predicted values of pK_a are. We can examine this question in two different ways: (a) by testing the consistency of pK_a values predicted from different sets of input data (different values of pH_1, pH_2 and pH_3) and (b) by comparison of derived values of pK_a with literature values.

Consistency of predicted values of pK_a. Table VII summarizes values of pK_a derived via computer simulation for the various benzoic acid samples. The pH values selected for the input data vary as shown. It is seen that the pK_a values predicted on the basis of runs with different pH (values of pK_1, pK_2, pK_3) are in reasonable agreement (±0.1–0.2 units), except for 2-nitrobenzoic and phthalic acids. The larger variation in the pK_a values for phthalic acid is expected, as the model of eqns. 3–5 is not applicable to this diprotic acid. The greater imprecision of predicted pK_a values for the case of 2-nitrobenzoic acid is also expected; the pH range used for the initial experimental runs does not overlap the (low) pK_a value of this compound, and k' values for this compound are generally small ($0.5 < k' < 2$).

Values of pK_a for the anilines in Table VI show greater variability than for the benzoic acids. There are also more cases where the model of eqns. 3–5 does not fit the input data (shown as footnote *c* in Table VII). This reflects the lower accuracy of predicted retention data for the substituted anilines *vs.* benzoic acids, as seen earlier in Tables III and IV.

Accuracy of predicted values of pK_a

Table VII lists approximate values of pK_a for the various substituted anilines ("Lit."), as estimated from data in refs. 21 and 22. The overal agreement between pK_a values predicted by computer simulation (eqns. 3–5) and these literature estimates is fairly good (±0.4 units). This is probably no worse than the uncertainty of these literature estimates, which are based on (1) the Hammett σ–ρ relationship and (2) approximations of the effect or %B on pK_a.

TABLE VII

VALUES OF pK_a FOR SUBSTITUTED ANILINES AND BENZOIC ACIDS AS DETERMINED FROM EQNS. 3–5

Compound[a]	pK_a[b] for indicated value of pH_2				
	3.2	3.5	3.8	S.D.	Lit.[e]
2-Nitrobenzoic acid	2.3	2.8	3.1	±0.4	2.2
Phthalic acid	2.9	3.2	3.6	±0.4	2.9
Impurity	3.3	3.5	3.5	±0.1	
2-Fluorobenzoic acid	3.5	3.6	3.7	±0.1	
3-Cyanobenzoic acid	3.4	3.5	3.6	±0.1	
2-Chlorobenzoic acid	3.0	3.2	3.3	±0.2	2.9
3-Nitrobenzoic acid	3.3	3.4	3.5	±0.1	3.5
3-Fluorobenzoic acid	3.8	–[c]	3.9	±0.1	
2,6-Dimethylbenzoic acid	3.4	3.6	3.6	±0.1	
	pK_a[b] for indicated value of pH_2				
	3.0	4.0	5.0	Lit.[d]	
4-Methoxyaniline	3.3	4.5	4.7	3.9	
3-Methylaniline	3.6	4.1	4.4	4.3	
3-Cyanoaniline	2.1	–[c]	–[c]	2.3	
2-Chloroaniline	2.1	–[c]	–[c]	–	
4-Chloroaniline	3.3	3.3	–[c]	3.6	
3-Chloroaniline	2.9	3.0	–[c]	3.1	
3,5-Dimethylaniline	3.7	4.1	4.3	3.8	
N-Ethylaniline	3.7	4.5	4.6	–	
3,4-Dichloroaniline	2.3	–[c]	–[c]	2.5	
3,5-Dichloroaniline	2.0	–[c]	–[c]	2.1	
Average difference (Lit. – calc.)	±0.4				

[a] Conditions: benzoic acids, see Table IV; anilines, cyano column, 25% B.

[b] Benzoic acids: $\Delta pH = 1.2$; *e.g.*, for $pH_2 = 3.2$, $pH_1 = 2.6$, $pH_3 = 3.8$. Anilines, $\Delta pH = 2.0$.

[c] DryLab I/mp did not accept a fit to eqns. 1–3.

[d] Literature values; data from ref. 21 (water as solvent) corrected for 25% methanol–water (ref. 22).

[e] Uncorrected literature values (in water, 25°C) [23].

CONCLUSIONS

Computer simulation based on a theoretical model (eqns. 3–5) is able to predict accurately retention and resolution for acidic and basic solutes as a function of pH. Three experimental runs with only pH varying are required as input for computer simulation. The most reliable predictions of retention are obtained for interpolations of the initial three experimental runs, and it is recommended that these runs span a range of no more than 2–3 pH units. Predicted retention time and α values were significantly more accurate for the case of acidic solutes (benzoic acids) than basic solutes (anilines). This

may be due to silanol effects, which play a more important role with basic solutes. The effective use of computer simulation for the prediction of separation as a function of pH requires that the computer flags predictions of marginal accuracy and recommends additional experimental data for more reliable predictions.

SYMBOLS

Symbols used in this paper and Part II [5] are summarized here.

A, B, C	constants in eqn. 1 in Part II
B	organic solvent used in the mobile phase (%B refers to solvent strength)
DMA	3,5-dimethylaniline
F^+, F^-	fraction of solute molecules that carry a positive or negative charge
$[H^+]$	hydrogen ion concentration
I/mp	isocratic, multi-parameter program for computer simulation
k'	solute capacity factor
k°, k^+, k^-	values of k' for non-ionized, positively charged, and negatively charged solute molecules, respectively
k_w	value of k' for water as mobile phase; see eqn. 2 in Part II
K_a	ionization constant for an acidic or basic solute
N	column plate number
pH_1, pH_2, pH_3	values of pH for mobile phases in initial experimental runs for pH mapping; $pH_1 < pH_2 < pH_3$
R_s	resolution of two adjacent bands
S	parameter that measures the change in solute retention as a result of change in %B (eqn. 2 in Part II)
S°	value of S for the non-ionized solute
t_R	solute retention time (min)
X, Y	hypothetical solutes in Fig. 8 in Part II
α	separation factor for two adjacent bands
$\delta\alpha$	error in a predicted value of α from computer simulation (eqn. 2)
ΔpH	pH range covered by initial experimental data used for pH mapping; equal to $pH_3 - pH_1$
φ	volume fraction of organic solvent in the mobile phase; equal to 0.01 · %B

ACKNOWLEDGEMENT

We are grateful for a grant from the Small Business Innovative Research Program of the National Institutes of Health in support of the work reported in this paper and Part II [5].

APPENDIX

DryLab I/mp Software

The DryLab I/mp software assigns different solutes to the various categories (acid, base, neutral, amphoteric) on the basis of retention times for a given solute at each of three pH values. Strong acids or bases, which exhibit little change in retention over the pH range 2–7, are classified as neutrals for the purpose of computer simulation. For solutes that are not classified as acidic or basic, DryLab I/mp uses a parabolic fit to the data for retention *vs.* pH. Solutes that are classified as acids or bases are fitted by eqns. 3–5.

The use of eqns. 3–5 for computer simulation is sensitive to errors in retention time or pH[a] for the input experimental data, resulting in some instances in negative values of k^+ or imaginary values of K_a. The present software assumes that small errors in retention time or mobile phase pH are likely. When unacceptable values of K_a, k^+ or k° are encountered, the pH value of the intermediate mobile phase is adjusted by ± 0.01 increments, until acceptable values of the solute parameters are obtained for all components of the sample. If no pH

[a] It should be noted that the measurement of mobile phase pH values with an accuracy of better than ± 0.05–0.1 unit is not always achieved in routine practice. This problem is considerably more serious if the pH of the final mobile phase (including organic solvent) is measured, compared with the recommended practice of measuring the pH of the aqueous buffer followed by addition of organic solvent. By taking special precautions, an accuracy of ± 0.02 pH units was achieved in this study. See also the discussion in ref. 14.

gives acceptable solute parameter values for all components, or if the required adjustment in pH is > -0.1 unit, problem solutes are flagged and a parabolic fit is used as with neutral of amphoteric solutes.

REFERENCES

1 P. J. Schoenmakers, *Optimization of Chromatographic Selectivity*, Elsevier, Amsterdam, 1986.
2 J. L. Glajch and L. R. Snyder (Editors), *Computer-Assisted Method Development for High-Performance Liquid Chromatography*, Elsevier, Amsterdam, 1990.
3 J. L. Glajch, J. J. Kirkland, K. M. Squire and J. M. Minor, *J. Chromatogr.*, 199 (1980) 57.
4 B. Sachok, R. C. Kong and S. N. Deming, *J. Chromatogr.*, 199 (1980) 317.
5 J. A. Lewis, L. R. Snyder and J. W. Dolan, *J. Chromatogr.*, 592 (1992) 197.
6 J. W. Dolan, D. C. Lommen and L. R. Snyder, *J. Chromatogr.*, 535 (1990) 55.
7 Cs. Horváth, W. Melander and I. Molnar, *Anal. Chem.*, 49 (1977) 142.
8 S. N. Deming and M. L. H. Turoff, *Anal. Chem.*, 50 (1978) 546.
9 F. Szokoli, Zs. Nemeth and J. Inczedy, *Chromatographia*, 29 (1990) 265.
10 B. Rittich and M. Pirochtova, *J. Chromatogr.*, 523 (1990) 227.
11 L. R. Snyder, J. W. Dolan and D. C. Lommen, *J. Chromatogr.*, 535 (1990) 75.
12 C. Herrenknecht, D. Ivanovic, E. G.-Nivaud and M. Guernet, *J. Pharm. Biomed. Anal.*, 8 (1990) 1071.
13 W. Lindberg, E. Johansson and K. Johansson, *J. Chromatogr.*, 211 (1981) 201.
14 R. M. L. Marques and P. J. Schoenmakers, *J. Chromatogr.*, 592 (1992) 157.
15 M. A. Stadalius, J. S. Berus and L. R. Snyder, *LC · GC*, 6 (1988) 494.
16 M. Patthy, T. Balla and P. Aranyi, *J. Chromatogr.*, 523 (1990) 201.
17 J. Kohler and J. J. Kirkland, *J. Chromatogr.*, 385 (1986) 125.
18 J. Kohler, D. B. Chase, R. D. Farlee, A. J. Vega and J. J. Kirkland, *J. Chromatogr.*, 352 (1986) 275.
19 L. R. Snyder, J. L. Glajch and J. J. Kirkland, *Practical HPLC Method Development*, Wiley-Interscience, New York, 1988.
20 M. Patthy and R. Gyenge, *J. Chromatogr.*, 449 (1988) 191.
21 L. P. Hammett, *Physical Organic Chemistry*, McGraw-Hill, New York, 1940, pp. 188–189.
22 B. L. Karger, J. N. LePage and N. Tanaka, in Cs. Horváth (Editor), *High-Performance Liquid Chromatography —Advances and Perspectives*, Vol. 1, 1980, p. 113.
23 *Handbook of Chemistry and Physics*, Chemical Rubber Co., Cleveland, OH, 52nd ed., 1972, D-117, D-120.