ELSEVIER

Journal of Chromatography A, 857 (1999) 41–68

JOURNAL OF CHROMATOGRAPHY A

www.elsevier.com/locate/chroma

Reversed-phase liquid chromatographic separation of complex samples by optimizing temperature and gradient time III. Improving the accuracy of computer simulation

J.W. Dolan[a], L.R. Snyder[a,*], R.G. Wolcott[b], P. Haber[c], T. Baczek[c], R. Kaliszan[c], L.C. Sander[d]

[a]*LC Resources Inc., 2930 Camino Diablo, Suite 110, Walnut Creek, CA 94596, USA*
[b]*Department of Chemistry, Linfield College, McMinnville, OR 97128, USA*
[c]*Department of Biopharmaceutics and Pharmacodynamics, Medical University of Gdansk, Gen. J. Hallera 107, 80-416 Gdansk, Poland*
[d]*National Institute of Standards and Technology, Gaithersburg, MD 20899, USA*

Received 27 November 1998; received in revised form 24 June 1999; accepted 25 June 1999

Abstract

Previous studies have shown that four experimental runs, where both temperature T and gradient time t_G are varied, can be used for the reliable prediction of separation as a function of these two variables (two-dimensional optimization). Computer simulation (e.g., DryLab) can then be used to predict ''optimized'' conditions for maximum sample resolution using either isocratic or gradient elution. Samples that contain a large number of components (e.g., $n>15–20$) present a greater challenge. Resolution for these more complex samples is often quite sensitive to small changes in T or t_G, in turn requiring greater accuracy in predictions that result from computer simulation. In the present study of several samples, we have examined computer simulation errors that can arise from inexact expressions for retention time as a function of T, t_G or isocratic %B. Resulting conclusions are applicable to both complex and simpler samples, in either one- or two-dimensional optimization. Means to anticipate and minimize the impact of these predictive errors are examined.

Keywords: Optimization; Temperature effects; Gradient elution; Computer simulation; Mobile phase composition; Retention times

1. Introduction

Previous work [1–3] has shown that reversed-phase liquid chromatographic separation can be predicted as a function of temperature T, gradient time t_G (and other gradient conditions), column size and flow-rate, if four experiments are first carried out for temperatures T_1 and T_2 and gradient times t_{G1} and t_{G2} (column size, flow-rate and other conditions fixed). Other studies [4,5] suggest that the simultaneous variation of T and t_G [two-dimensional (2D) optimization] generally results in significant changes in sample selectivity, as a result of which a majority of samples with $n<20$ can be adequately separated

*Corresponding author.

0021-9673/99/$ – see front matter
PII: S0021-9673(99)00766-9

[2,3,6–8]. The convenient selection of ''best'' values of T, t_G, and other conditions is made possible by the use of computer simulation [2,3,6–8], which can provide maps of resolution R_s as a function of T and t_G (Fig. 1).

Fig. 1a shows a resolution map for the separation of a mixture of eight substituted benzoic acids as a function of T and t_G, while Fig. 1b shows a similar map for the separation of a mixture of 40 toxicology standards. For the eight-component sample of Fig. 1a, maximum resolution (±5%) as indicated by the white region of the map extends over a considerable range in T (±4°C) and t_G (±5 min). Thus, if the prediction of optimum conditions of T and t_G for this sample were in error by a few degrees and/or a few minutes, the resulting separation would not be significantly compromised. In the case of the 40-component sample of Fig. 1b, the white region for maximum R_s (t_G=30 min, 40°C) is limited to a fraction of a °C and a fraction of a minute. In this case, errors of a few minutes or degrees in the predicted optimum values of T and/or t_G would result in much reduced sample resolution. Fig. 1 appears representative of other reported separations [2,3], and we conclude that ''complex'' samples containing, e.g., >15–20 components will generally require (a) more accurate predictions when computer simulation is used and (b) a more careful control of separation conditions.

Previous comparisons of computer simulation (varying T and/or t_G, as in Fig. 1) versus confirmatory experimental separations suggest that these predictions can be reasonably accurate [1–3,9,10]. However, no systematic study has been carried out on the accuracy of these gradient predictions as a function of conditions (T, t_G, column size, flow-rate) for the input runs versus conditions for predicted runs. The linear-solvent-strength (LSS) model [11] upon which DryLab computer simulation software is based also allows the prediction of isocratic separation from two gradient runs where only t_G is varied. Various errors can be introduced in these isocratic predictions [12,13], and such errors will be more important for complex samples. Finally, no systematic attention has been given to computer simulation errors that result when T is varied (e.g., molecules of varying shape, as well as occasional ''outliers'' noted in [1]).

A primary goal of the present study was the evaluation and possible control or correction of various errors that can arise during computer simulation, when the initial input runs involve change in T and/or t_G. Similar errors are found in other applications of computer simulation which assume some relationship for the dependence of chromatographic retention on experimental conditions (e.g. [14–16]), so that conclusions reached in the present study should be widely applicable. Preceding papers (Part I [17], Part II [18]) have addressed other issues related to the challenge of separating complex samples.

2. Background and theory

2.1. Errors in retention time and resolution

The basis of computer simulation for optimizing T and t_G has been discussed [1]. In the 2D optimization of T and t_G, experimental retention times t_R for each sample component of interest are first obtained from four linear-gradient chromatograms. Computer simulation can then be used to predict t_R, bandwidth W and resolution R_s for other experimental conditions, including changes in T, t_G, initial and final %B in the gradient, gradient shape, column dimensions, particle size and flow-rate [1,11,16], as well as corresponding isocratic separations where %B is varied instead of t_G. Errors in predicted values of t_R can arise from various causes, as summarized in Table 1 and discussed below. Previous discussions of such errors [10,12,13,19–21] have emphasized predicted values of t_R; however, errors in retention-time differences Δt_R for adjacent bands i and j are of greater concern, because Δt_R values directly affect predicted values of resolution R_s (as in Fig. 1).

The LSS model used for DryLab simulations assumes for isocratic reversed-phase retention that

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

where φ [equal to 0.01(%B)] is the volume fraction of the B solvent, k_w is the (extrapolated) value of k for water as mobile phase (φ=0), and S is a constant for the solute (conditions other than φ held constant). Similarly, for varying temperature,

$$\log k = A + B/T_K \quad (2)$$

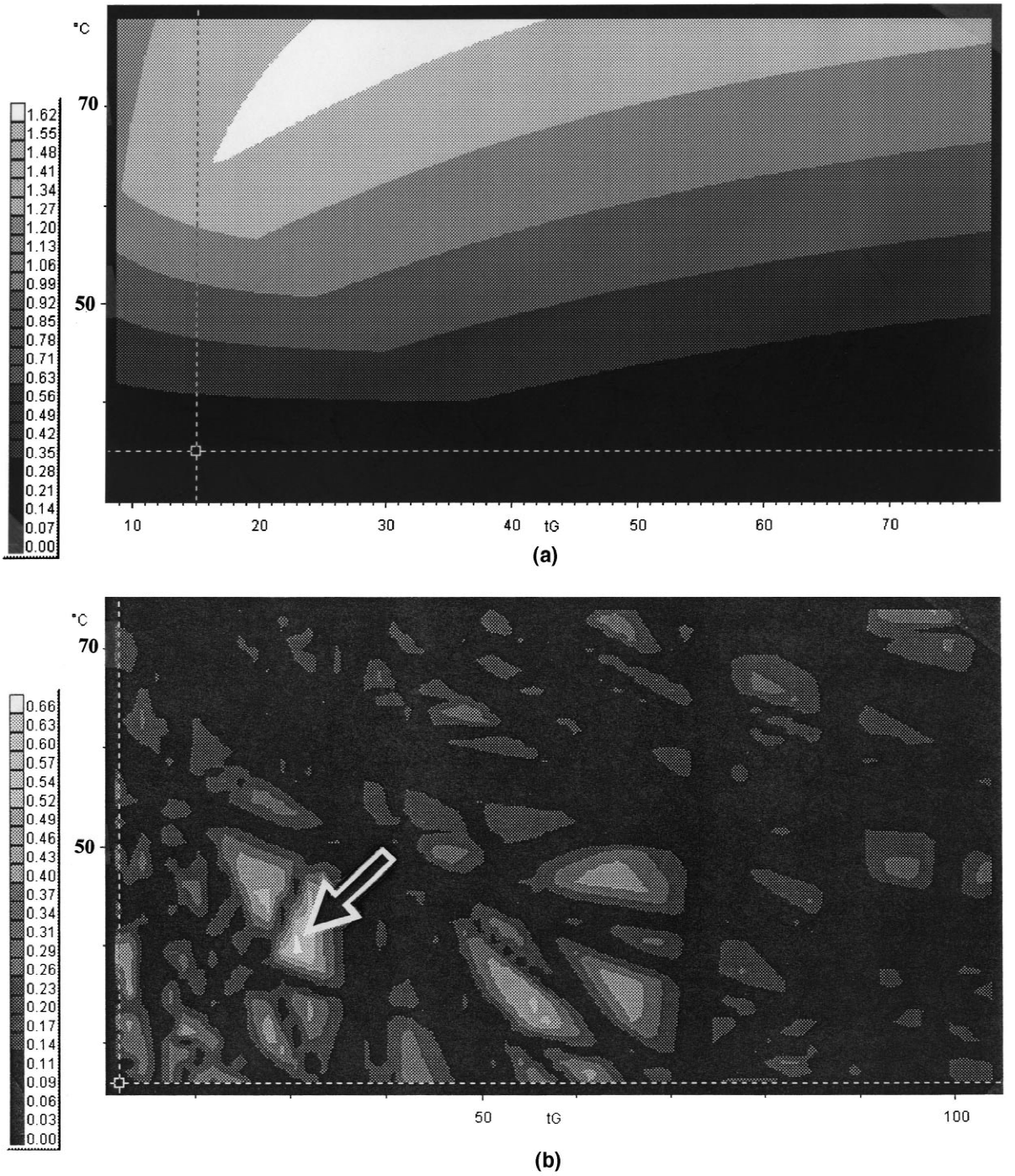

Fig. 1. Resolution maps: R_s versus temperature T and gradient time t_G. (a) Substituted benzoic acid sample; (b) toxicology-standards sample. Reprinted with permission from [2]; see [2] for experimental conditions.

Table 1
Possible contributions to error in predictions of retention time t_R by means of computer simulation

Error source	Ref.	Comment
HPLC equipment	[12]	Gradient dispersion, flow-rate inaccuracy, and errors in gradient formation
Random variations in separation	[10,13]	Changes in column retention, variability of conditions in separation
Solvent demixing	[13]	Mobile phase changes its composition due to preferential uptake by the column of the organic solvent
Failure of model equations	[1,13], [21]	Plots of log k versus φ or t_R versus T generally show slight curvature

where A and B are constants for a given solute when other conditions are constant; T_K is temperature in K. In the case of gradient elution, retention time t_R can be derived (with certain assumptions) as (1)

$$t_R = A' + B'/T_K \quad (3)$$

However, actual values of t_R appear to be described more accurately [1] by the empirical relationship

$$t_R = A'' + B'' T \quad (4)$$

which is the basis of computer simulations carried out with DryLab software (here and later, T is in °C). Quantities A', B', A'' and B'' are constants for a given solute (other conditions constant).

Errors in predicted values of t_R or Δt_R can be expressed in various ways; e.g., errors in k or α. We have found it advantageous to define these errors in terms of an equivalent change (or error) $\delta\varphi$ in the volume-fraction of the B solvent φ for the predicted separation. That is, a predicted value of t_R for a mobile phase composition $(\varphi - \delta\varphi)$ will equal the correct experimental value of t_R for a mobile phase composition φ.

Likewise, errors in Δt_R for adjacent peaks i and j, as a result of errors in individual values of t_R, can be expressed as $\delta\delta\varphi = (\delta\varphi)_j - (\delta\varphi)_i$. Appendix A discusses these errors $\delta\varphi$ and $\delta\delta\varphi$ in more detail and extends their application to both gradient and isocratic elution. It is shown in Appendix A that for both isocratic and gradient elution, an error $\delta\delta\varphi = 0.001$ typically corresponds to an average error in resolution of about 0.2 units, which for complex samples we regard as an acceptable *average* error. For less complex (more common) samples, an average error $\delta\delta\varphi = 0.002$ (with error in $R_s \leq 0.4$) may be tolerable.

2.2. Individual error sources of Table 1

Equipment-related errors can arise from inaccurate performance of the gradient equipment, dispersion (''rounding'') of the gradient as it passes through the HPLC system, and errors in flow-rate that result from inadequate compensation for mobile phase compressibility or other causes. The contribution of the equipment to errors in t_R can be systematically reduced in various ways [12,22].

Random variations in separation can arise from unanticipated fluctuations in separation conditions (temperature, flow-rate, mobile phase composition), along with changes over time in column retention characteristics (e.g., from a loss of bonded phase, column contamination by sample, etc.). The present study found a random error of $\delta\delta\varphi \approx 0.0003$–$0.0005$ for each of three different samples, each separated in different laboratories; i.e., well below the acceptable total error of $\delta\delta\varphi = 0.001$–$0.002$.

Solvent demixing refers to changes in the mobile phase during gradient elution, as a result of a preferential uptake by the column of the organic solvent (''B solvent''). The mobile phase composition φ during the separation will generally lag the value calculated from the gradient program [13], leading to positive errors $\delta\varphi$. Values of $\delta\varphi$ due to solvent demixing are generally small (<0.01) and can be reduced further by appropriate experimental conditions [13].

Preceding errors $\delta\varphi$ can be minimized to some extent by attention to experimental conditions and

procedures. These errors are in most cases also similar for adjacent peaks *i* and *j* and therefore largely cancel; consequently, these errors are expected to have only a minor effect on values of $\delta\delta\varphi$. The remaining discussion is aimed primarily at retention errors that arise from a failure of the model equations used for computer simulation (Eqs. (1), (2), (4)). Also discussed are errors introduced when gradient retention data are used to predict isocratic retention times.

2.3. Errors that result from non-linear plots of log k versus %B

Isocratic elution. Eq. (1) is commonly used as an approximation for isocratic retention *k* as a function of %*B* (or φ) in reversed-phase systems [23]. However, a number of studies have shown that plots of log *k* versus φ can be noticably concave, rather than linear, as illustrated in Fig. 2 for five dialkylphthalates separated on a C_8 column with acetonitrile/water mobile phases. Johnson et al. [24] analyzed plots of log *k* versus φ for several hundred studies from the literature (methanol and acetonitrile as *B* solvents) and concluded that the data were better fit (versus Eq. (1)) by

$$\log k = a' + b'E_T(30) \quad (5)$$

$E_T(30)$ is a solvatochromic parameter that yields concave curves when plotted versus φ for different *B* solvents; a' and b' are constants for a given solute and solvent. A re-examination [23] of results from [24] found that Eq. (5) gives an improved fit to experimental data for acetonitrile as *B* solvent (i.e., curved plots of log *k* versus φ), but Eq. (1) appears more reliable for methanol as solvent (i.e., more nearly linear plots of log *k* versus φ). Other workers have proposed a 3-parameter fitting equation [25]:

$$\log k = A + B\varphi + C\varphi^2 \quad (6)$$

Since $E_T(30)$ for acetonitrile (ACN) as *B* solvent can

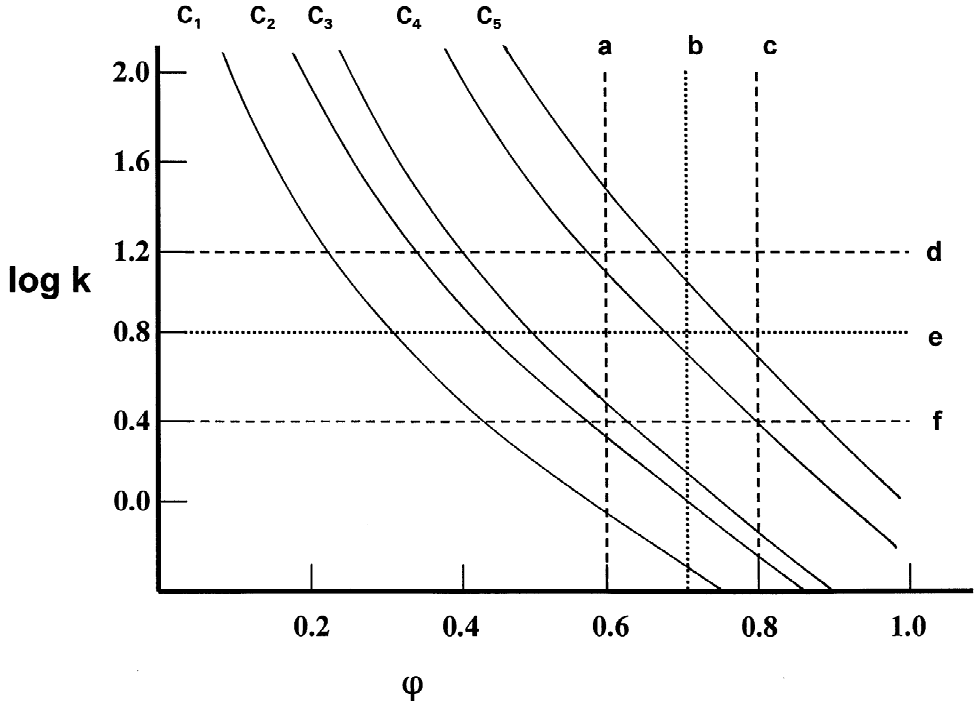

Fig. 2. Log *k* versus %*B* for alkylphthalate sample. C_1, dimethyl phthalate; C_2. diethyl; C_3, diallyl; C_4, dibutyl; C_5, dipentyl. Data from [12], recalculated for $k = (t_R - t_o)/t_o$. Conditions: water (A) and acetonitrile (B) mixtures as mobile phase; 35°C; 25×0.46 cm column; 2.0 ml/min; see Section 3 for other details.

be represented (±0.1 unit, 1 SD) over the range $0 < \varphi < 0.8$ by

$$E_T(30) = 63.1 - 18.1\varphi + 10.3\varphi^2 \quad (7)$$

systems (with ACN as *B* solvent) that are described by Eq. (5) will be described equally well by Eq. (6).

When only %*B* or t_G is varied, computer simulation based on Eq. (1) (two experimental input runs, DryLab software) has proven generally reliable as a basis for predictions of retention and resolution as a function of isocratic %*B* or gradient time; e.g., see 20 studies cited in pp. 157–160 of Ref. [11]. However, this is less likely to be true for the case of (a) complex samples, (b) simultaneous changes in *T* and t_G, or (c) $N \gg 10\,000$ [e.g., for capillary electrochromatography (CEC); see Appendix A]. Curvature of log *k* versus φ plots (and resulting errors in predicted t_R values) can also be more pronounced when φ is either small ($\varphi < 0.3$) or large ($\varphi > 0.7$), e.g., see Fig. 1 of [23].

Fig. 3 illustrates the origin of predictive errors that can arise when Eq. (1) is applied to samples which exhibit curved plots of log *k* versus φ (i.e., isocratic separation). It is assumed in this example that two experimental runs were carried out for $\varphi = 0.45$ and 0.55 (closed circles), and these data are then used to determine the coefficients of Eq. (1). Predictions based on Eq. (1) are illustrated in Fig. 3 by the straight (dashed) line through the experimental data points. For estimation of k at $\varphi = 0.5$, the dashed curve is seen to predict a value of *k* that is too large relative to the experimental value (open circle). This error in *k* can also be described in terms of an error $\delta\varphi = 0.015$, corresponding to the difference in φ values for the correct value of $k = 1.7$. Predictions of retention for conditions intermediate between the two experimental (input) runs ($\varphi = 0.45$ and 0.55 in Fig. 3) will be referred to as interpolations; for such predictions, values of $\delta\varphi$ are usually positive. For predictions outside the range covered by experimental runs (extrapolations), errors in $\delta\varphi$ are usually negative, as illustrated in Fig. 3 for mobile phase compositions of $\varphi = 0.43$ and 0.57 (open circles). Extrapolative errors are also potentially larger than are interpolative errors.

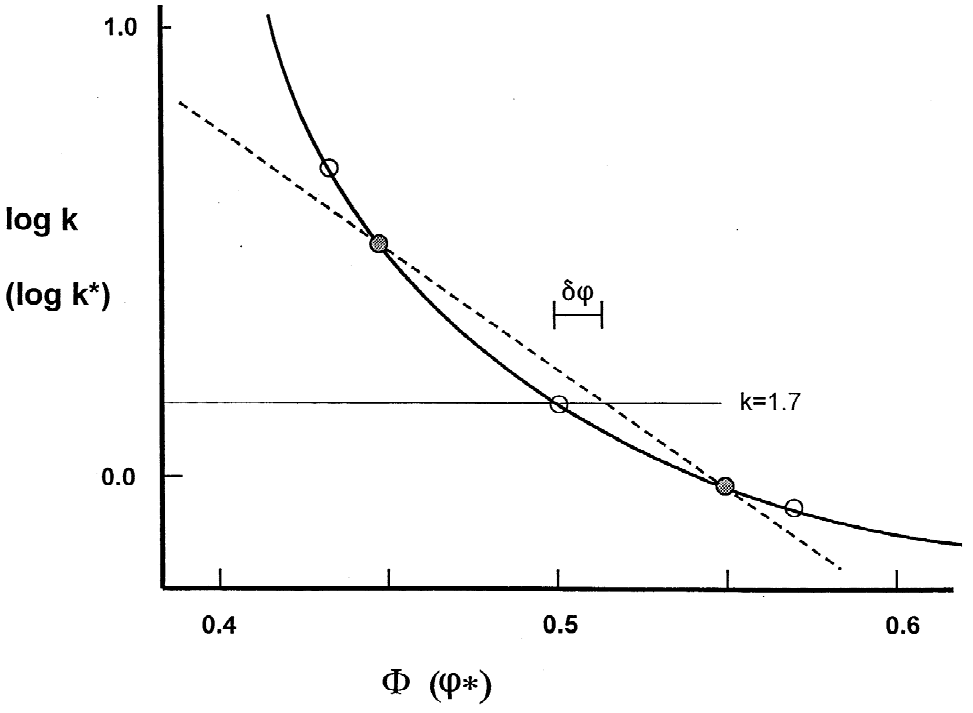

Fig. 3. Prediction of retention from Eq. (1) for hypothetical solute. See text for details.

Temperature. Plots of gradient retention t_R versus T are often slightly concave, representing a modest failure of Eq. (4) (see later discussion). This can lead to small errors $\delta\varphi$, comparable to those when φ is varied as in Fig. 3 (in a small number of cases summarized in [1], larger deviations from Eq. (4) were observed). Deviations from Eq. (2) for isocratic separations are usually smaller, at least for a restricted range in T (e.g., $<50°C$).

2.4. Equivalence of ''corresponding'' isocratic and gradient separations

''Corresponding'' isocratic and gradient separations are those that are carried out with the same sample and other conditions (e.g., acetonitrile-water), such that average values of the retention factors k (isocratic) and k^* (gradient) are equal for an adjacent pair of bands of interest [11]. For ''corresponding'' separations, similar ''random'' errors $\delta\varphi$ should result for either isocratic or gradient elution. In one study (pp. 168–170 of Ref. [11]), retention reproducibility was the same for a given sample in ''corresponding'' isocratic and gradient separations (acetonitrile/water mobile phases): $\delta\varphi = 0.002$ units (1 SD) over a 3-month period with the same column for both isocratic and gradient separations. If the column plate number N is also similar for the two separations (generally the case), the resolution of two adjacent bands in such corresponding separations will be approximately equal. Gradient retention k^* can be related to experimental conditions via the LSS model [11]

$$k^* = 1/(1.15\, b + [1/k_o]) \tag{8}$$

where k_o is the value of k at the start of the gradient, and b is a gradient steepness parameter given by

$$b = V_m\, \Delta\varphi\, S/(t_G\, F) \tag{8a}$$

Here, V_m is the column dead-volume and F is the flow-rate. The gradient retention parameter k^* is equal to the instantaneous value of k when the solute band in gradient elution has migrated halfway through the column. A quantity φ^* also can be defined, corresponding to the instantaneous value of φ for $k = k^*$ (i.e., solute at the column midpoint). Values of k versus φ and k^* versus φ^* (for ''corresponding'' conditions) plotted on the same graph should fall on a single curve that can be approximated by Eq. (1); see experimental examples in Fig. 3 of [11]). Thus, when using gradient input data to predict isocratic retention, values of k^* versus φ^* derived from the gradient runs (for varying t_G) can be equated (approximately) to k versus φ for the isocratic runs. When the initial solute retention in gradient elution is large (and therefore $1/k_o$ is small),

$$k^* = 1/1.15b \tag{9}$$

Returning to Fig. 2, the vertical dashed lines (a,c) correspond to isocratic separations carried out with $\varphi = 0.6$ and 0.8, for prediction of separation for $\varphi = 0.7$ (dotted line b). The horizontal dashed lines (d,f) of Fig. 2 correspond to gradient separations carried out with $\log k^* = 0.4$ and 1.2, where one of the variables on the right-hand-side of Eq. (8a) (usually gradient time t_G) is changed, so as to change k^*. The horizontal dotted line (e) then corresponds to a separation to be predicted on the basis of runs d and f. From Fig. 2 we see that sample retention in isocratic elution (k) is varied by changing φ, while sample retention in gradient elution (k^*) can be changed by varying column size, flow-rate, or (more often) t_G (Eqs. (8,9)).

3. Experimental

The present study, involving several samples, has made use of both new and previously reported experimental data.

Alkylphthalate sample. Ref. [13] reports isocratic and gradient retention data for five *o*-dialkyl phthalates: C_1, dimethyl; C_2, diethyl; C_3, diallyl; C_4, di-*n*-butyl; C_5, di-*n*-pentyl, separated with acetonitrile–water mobile phases for a wide range of conditions. For isocratic separations, conditions were: 25×0.46 cm C_{18} column, $10<\%B<90$, 2.0 ml/min and 35°C. For gradient separations, conditions were 25 and 50×0.46 cm C_{18} columns, 10–100% B gradients, $2.5<t_G<320$ min, $0.2<F<4$ ml/min, and 35°C. Fig. 2 shows plots of $\log k$ versus φ for this sample at 35°C. Fig. 4a shows a typical separation of this sample (computer simulation). Isocratic values of k reported in [13] are based on $k = (t_R - t_{sec})/t_{sec}$, rather than the usual $k = (t_R - t_o)/$

t_o. We have recalculated these data by means of the latter (more conventional) equation, using values of t_{sec} and t_o versus φ reported in [13]. The quantity t_{sec} refers to the ''equivalent'' dead-time of a retained solute, as opposed to the dead-time for an unretained, small-molecule solute. Values of t_{sec}/t_o decrease as solute molecular mass increases, and $t_{sec} = t_o$ for a very small molecule.

Shape-selective sample (LCS). From a nine-compound mixture used earlier [1] to evaluate molecular-

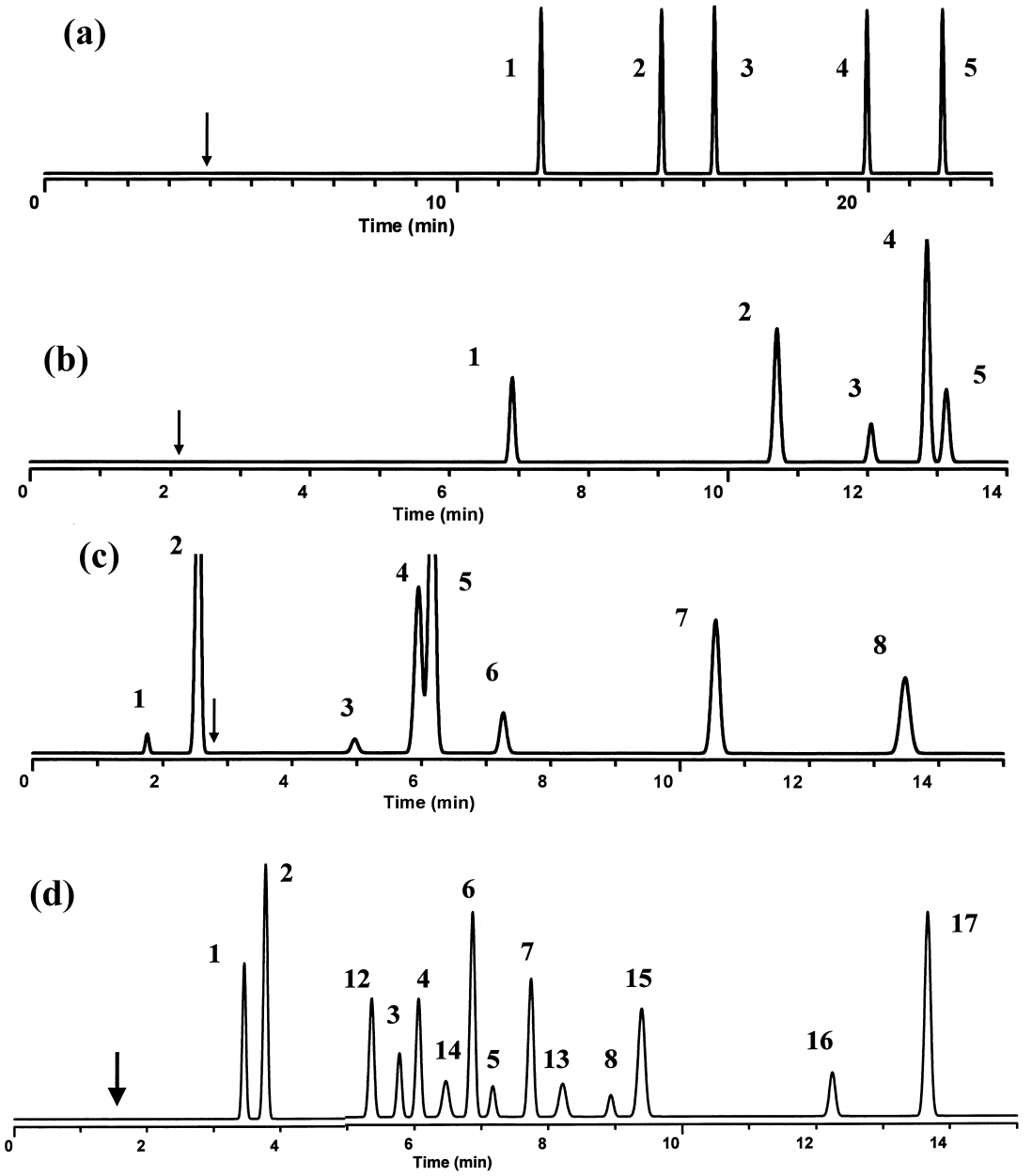

Fig. 4. Representative chromatograms for four samples from the present study. Conditions: 20-min gradients; other conditions as in Section 3 except where noted otherwise; all peaks numbered in order of their listing in the Section 3. (a) Alkyl phthalate sample; (b) shape-selective sample, 45°C; (c) hydrophilic drug sample; (d) mixture of substituted benzoic acids and anilines, 79°C. Computer simulations based on gradient input data. Arrows in each chromatogram indicate arrival of gradient at column outlet.

shape selectivity, a five-compound sub-set was selected that contained trypticene, triphenylene, tetraphenylmethane, 1,3,5-triphenylbenzene, and 1,6-diphenylhexatriene. Fig. 4b shows a typical separation of this sample (computer simulation). Equipment, materials, procedures and conditions are as described for Laboratory E of [1]; a 25×0.46 cm Hypersil Green PAH clumn (Shandon) was used, and 50–100% acetonitrile-gradients were carried out with a flow-rate of 2.0 ml/min. Temperature was varied in 5°C increments between 15 and 65°C, and gradient time was varied from 20 to 120 min. The dwell volume was 1.9 ml.

Hydrophilic drug sample (PH, TB and RK). This is an eight-component mixture of compounds that are weakly retained in reversed-phase HPLC at low pH: nitromethane, sulfanilamide, p-phenetidine, acetaminophen, aminophenazone, codeine, caffeine, tripelennamine. Fig. 4c shows a typical separation of this sample (computer simulation). The HPLC system was an LC Model 1 (Waters Associates, Milford, MA, USA) with a dwell volume of 4.3 ml. Conditions were as follows: 15×0.46-cm Inertsil C_{18} column (GL Sciences); A-solvent, 25 m*M* KH_2PO_4 adjusted with H_3PO_4 (pH 2.75 at 21°C); *B* solvent, acetonitrile; 35°C, 2 ml/min; 10 μl injection containing 2 m*M* each solute. Gradients were 3–23% *B* in 20 and 80 min, and isocratic separations were carried out for 3, 6, 9 and 12%*B*.

Benzoic acid plus aniline derivatives (RGW). A previous study [1] examined the separation of mixtures of (a) eight benzoic acid and (b) nine aniline derivatives as a function of T and t_G. This study was repeated here for similar conditions and compared with these previous results. A 14-component sample composed of substituted benzoic acids 1–8 plus anilines 12–17 (see later Table 6 for solute numbering) was separated using the following conditions, which result in partial ionization of the various sample components: column, 15×0.46 cm Zorbax SB C_{18} (Hewlett-Packard); 5–65% *B* gradients; A solvent is phosphate buffer, pH 2.6; *B* solvent is acetonitrile; gradient times of 13, 26 and 39 min; temperatures of 35, 55 and 75°C; 1.5 ml/min. The dwell volume was 2.4 ml. Fig. 4d shows a typical separation of this sample (computer simulation).

Calculations. Computer simulations for prediction of retention were carried out with DryLab for Windows, Version 2.0 (LC Resources). Final values of $\delta\delta\varphi$ reported here for each of the above samples, e.g., as in Table 5, are often averages of several different conditions for input and predicted separations, as a means of reducing the scatter of final values of $\delta\delta\varphi$ and simplifying their interpretation.

Individual retention times are not reported here but can be obtained from one of the authors (L.R.S.).

4. Results and discussion

In this section, we will address the following questions:

(1) How does separation error differ for predictions of (a) isocratic retention from isocratic input data, (b) gradient retention from gradient data, and (c) isocratic retention from gradient data.

(2) How can conditions for the input runs be selected so as to yield acceptable predictive errors for complex ($\delta\delta\varphi \leq 0.001$) and other ($\delta\delta\varphi \leq 0.002$) samples? Similarly, what limits on T, t_G or %*B* should be placed on predicted separations?

(3) Are predictive errors a function of the sample? Can we anticipate samples for which predictive errors will be larger?

(4) If it is known or suspected that predictive errors will be unacceptable, how can we minimize or (better) correct such errors?

We first examine errors for predictions of retention as a function of gradient time t_G or isocratic mobile phase composition %*B* (%$B = 100\varphi$); i.e., errors that arise from a failure of Eq. (1). Next, we consider errors for predictions of retention as a function of temperature T (i.e., failures of Eq. (4)), followed by errors for predictions of retention as a function of simultaneous change in T and t_G. Finally, we look at means for minimizing or correcting these predictive errors.

4.1. Errors in predicted retention when varying isocratic %B (alkylphthalate sample)

Consider first the prediction of isocratic retention (e.g., for $\varphi = 0.7$), using two experimental isocratic runs (e.g., $\varphi = 0.6$, 0.8) as inputs for computer simulation; note the similar example of Fig. 3 for errors that result from a failure of Eq. (1).

Predictions from smoothed data. Values of k versus φ for the alkylphthalate sample can be described by Eq. (5) ($r^2 > 0.979$ for each solute), as summarized in Table 2. For solute C_2, which we will use as example, retention is given as $\log k = -24.959 + 0.4498\ ET(30)$, which with Eq. (7) becomes

$$\log k = 3.4234 - 8.1413\,\varphi + 4.6329\varphi^2 \qquad (10)$$

Values of k as a function of φ for C_2 from Eq. (10) will be used initially in place of raw (''unsmoothed'') data, for a preliminary examination of retention errors $\delta\varphi$ that result solely from the use of Eq. (1) in computer simulation. This approach has certain advantages. First, the smoothed data from Eq. (10) are not complicated by random experimental error, thus simplifying the interpretation of errors due to a failure of Eq. (1). Second, values of k can be predicted for any value of φ in the range $0.2 < \varphi < 0.8$. Finally, because Eq. (5) is a good approximation for a wide range of solutes with acetonitrile as B solvent, any conclusions resulting from the use of Eq. (10) for solute C_2 should be widely applicable (errors for methanol as B solvent are expected to be similar in nature, only smaller). When k versus φ for two solutes in the same sample is described *exactly* by Eq. (5), it can be shown (Appendix B) that $\delta\delta\varphi = 0$. That is, if the data exactly fit Eq. (5), and Eq. (1) is used to predict retention, errors $\delta\varphi$ will be the same for each solute and therefore cancel-because of similar curvature of log k versus φ plots for different solutes over a common range in φ. As a result, predictions of resolution based on Eq. (1) will not be in error for this case. However, because Eq. (5) is *not* an exact relationship for any experimental system, actual (unsmoothed) values of $\delta\delta\varphi$ are not expected to equal zero, as will be seen.

Fig. 5 (curve a) summarizes values of $\delta\varphi$ versus φ for the case of input φ values of 0.45 and 0.55. As expected (see Fig. 3), $\delta\varphi = 0$ for the latter (input) values of φ. For interpolated predictions ($0.45 < \varphi < 0.55$), the maximum error $\delta\varphi = 0.003$, and the average interpolated error will be approximately half this value (0.0015). Errors (absolute values of $\delta\varphi$) are larger for even modest extrapolation; e.g., $\delta\varphi = -0.010$ for $\varphi = 0.40$ or 0.50. Fig. 5 (curve b) shows a corresponding plot for (more widely spaced) input values of $\varphi = 0.40$ and 0.60. Compared with the example of curve a, errors are larger for both interpolation (maximum $\delta\varphi = 0.013$ for $\varphi = 0.50$) and extrapolation ($\delta\varphi = -0.017$ for $\varphi = 0.35$ and 0.55). A practical conclusion from Fig. 5 is that both interpolation and extrapolation errors $\delta\varphi$ increase, as the difference in φ for the two input runs increases. Conversely, an increase in the difference in input-φ values allows predictions of k for a wider range in φ. Thus, the choice of preferred values of φ for the input runs represents a necessary compromise between greater predictive accuracy versus a reduced range of φ values for which acceptable predictions of k (e.g., $\delta\varphi < 0.005$) are possible. The desired range in φ is related to the practical requirement for $0.5 < k < 20$ in the final separation (p. 34 of Ref. [9]). When values of input φ other than those of Fig. 5 were chosen, but difference in the two input values of φ was the same (e.g., equal 0.1 in Fig. 5, curve a), identical errors in k (δ log k) were obtained for

Table 2
Correlation of data for phthalate sample in terms of Eq. (5): $\log k = a + bE_T(30)$; conditions: 25×0.46 cm C_{18} column, acetonitrile–water mobile phases ($0.1 < \varphi, 0.9$), 2.0 ml/min, 35°C (data taken from [13]

Solute	Correlation results[a]					Range in φ[b]
	a	b	r^2	SD[c]	n[d]	
C_1	−20.473	0.3639	0.998	0.04	11	0.10–0.80
C_2	−24.959	0.4498	0.991	0.07	10	0.20–0.80
C_3	−29.825	0.5389	0.986	0.09	7	0.25–0.80
C_4	−36.892	0.6768	0.985	0.07	8	0.40–0.80
C_5	−46.328	0.8514	0.979	0.07	7	0.50–0.80

[a] Best fit of k versus φ data to Eq. (2) (least squares).
[b] Range of experimental φ values used in correlation versus Eq. (5).
[c] Standard deviation of fit of log k values.
[d] Number of data points.

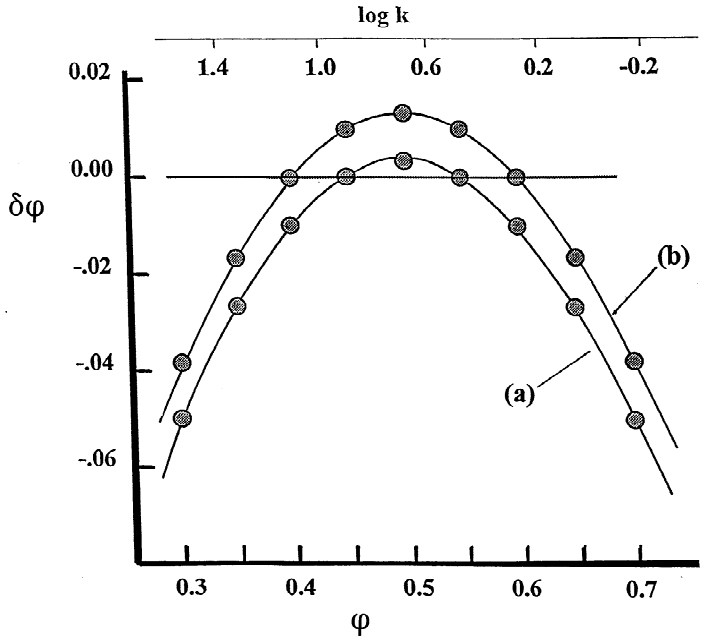

Fig. 5. Error $\delta\varphi$ in predicted retention as a function of φ for solute C_2; assumes true retention given by Eq. (5) with values of a and b from Table 2 (smoothed retention data). (a) input values of $\varphi = 0.45$ and 0.55; (b) input values of $\varphi = 0.4$ and 0.6.

similar interpolation or extrapolation; e.g., for input values of $\varphi = 0.2$ and 0.3, $\varphi = 0.25$ for interpolation, $\varphi = 0.15$ or 0.35 for extrapolation.

Values of log k corresponding to the φ values of Fig. 5 are indicated at the top of this figure, as a reminder that plots of $\delta\varphi$ versus φ can also be presented as plots of $\delta\varphi$ versus log k (cf. Eq. (1)). This is relevant to the discussion in a following section of errors $\delta\varphi$ for predictions of retention in gradient elution.

Predictions from unsmoothed data. Predictions of retention based on Eq. (1) were next carried out for the five solutes of the alkylphthalate sample, using unsmoothed (''raw'') data. Input conditions were initially selected for $\varphi = 0.6$ and 0.7, comparable to the conditions for Fig. 5 (curve a) in that the median value of k for the entire sample ≈ 5 for an intermediate $\varphi = 0.65$. Fig. 6a shows the resulting plot of average values of $\delta\varphi$ versus φ, with the corresponding plot of Fig. 5 (curve a, shifted by 0.15 φ-units) superimposed onto these data. The curves describing the data of Fig. 5 (curve a) and 6a should be roughly similar, which appears to be the case. Fig. 6b shows a corresponding plot of $|\delta\delta\varphi|$ versus φ; here, and later, we plot values of $|\delta\delta\varphi|$, rather than $\delta\delta\varphi$, because it is the absolute error in $\delta\delta\varphi$ or R_s which is of interest. As expected, the general shape of this plot is similar to that of Fig. 6a; i.e., errors $\delta\delta\varphi$ track errors $\delta\varphi$, but are smaller (because of partial cancellation in errors $\delta\varphi$ for adjacent bands). The greater scatter of data points for $\varphi > 0.7$ in Fig. 6 may be due to experimental uncertainty in the smaller values of k ($0.2 < k < 6$) for these data.

As seen in Fig. 5 (curve b versus curve a), the maximum interpolative error $\delta\varphi$ increases as the input-values of φ become more different, corresponding to increasingly different values of k for the two input runs (k_1 and k_2). This should also be apparent from the example of Fig. 3. It is useful to plot maximum interpolated values of $|\delta\delta\varphi|$; (i.e., for φ halfway between input φ-values φ_1 and φ_2) versus the ratio k_1/k_2 for the input runs (see Fig. 7, circles; alkylphthalate sample). In the following discussion, we will emphasize interpolated values of $|\delta\delta\varphi|$; values of this quantity reported here are in most cases maximum values (corresponding to the halfway point between input conditions) for the case of gradient predictions from gradient input runs, or isocratic predictions from isocratic input runs. For predictions of isocratic retention from gradient input

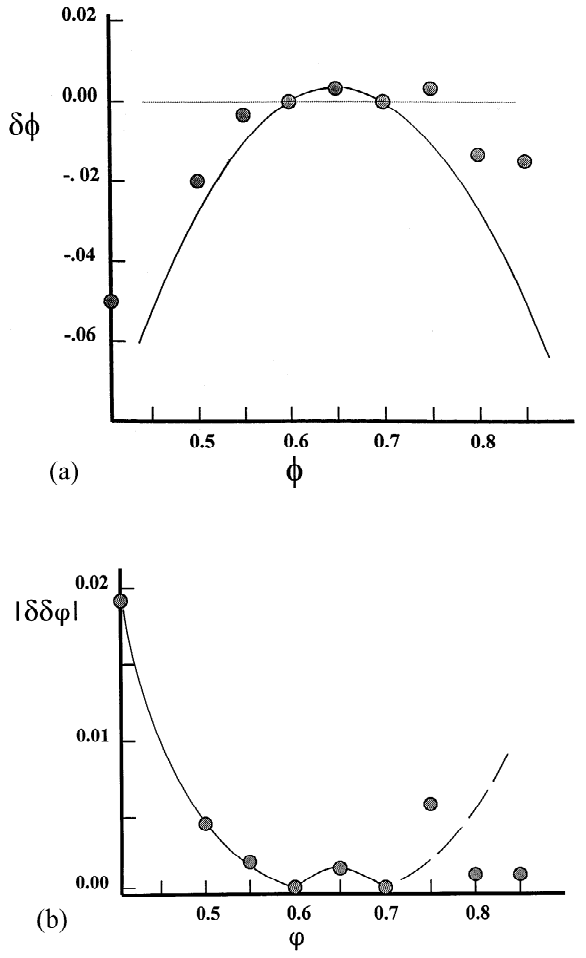

Fig. 6. Average errors in predicted retention (*a*, $\delta\varphi$) or resolution (*b*, $|\delta\delta\varphi|$) as a function of φ for all five components of alkylphthalate sample (actual experimental data). Assumes input values of $\varphi = 0.6$ and 0.7; other conditions as in Fig. 2. See text for further discussion.

runs, interpolated values of $|\delta\delta\varphi|$ are usually *average* values (equal to about one-half of the maximum value; see Fig. 6b). In the following discussion, we define maximum interpolated errors as $\delta\delta\varphi(m)$ and average interpolated errors as $\delta\delta\varphi(a)$, and assume $\delta\delta\varphi(a) \approx 0.5\delta\delta\varphi(m)$.

Fig. 7 (circles) suggests that for a *maximum* error $\delta\delta\varphi(m) \leq 0.002$ (or $\delta\delta\varphi[a] \leq 0.001$), the ratio k_1/k_2 should be ≤ 4, corresponding to a maximum recommended difference in input-φ values of about 10% *B* for this sample (note that $S \approx 6$ for this sample). Since $S \approx 4$ for most small ($M_r < 500$ Da) analytes [30], this suggests input conditions differing by 15% *B* (rather than 10% in the present case). Fig. 6b suggests that extrapolation outside this interpolative range by ±5% *B* will not greatly reduce predictive accuracy; i.e., acceptable predictions over a range of 25% *B* for most small-molecule samples.

4.2. Errors in predicted retention when using gradient runs as input

Errors in predicting gradient elution retention as t_G is varied. Because of the interrelationship of gradient and isocratic separation [11], a similar behavior as in Fig. 5 (curve a) can be expected for predictions of gradient retention by computer simulation. For gradient elution, log k^* can replace φ in plots as in Fig. 6b ($|\delta\delta\varphi|$ versus φ or log k). Also, if only t_G is varied during computer simulation, a change in t_G is equivalent to a change in k^* (Eq. (5)). Thus plots of $\delta\delta\varphi$ versus t_G or log k^* can be considered similar to plots of these errors versus φ as in Fig. 6b (note equivalence of φ and log k in Fig. 5 or Eq. (1) for a given solute). Similarly, the ratio of input values of t_G (t_{G1}/t_{G2}) is equal to the ratio $k_1^*/k_2^* = k_1/k_2$ for isocratic separation (as in Fig. 7), as long as k_o in Eq. (8) is large, and therefore Eq. (9) applies. (A later section examines the case of early-eluting compounds where $k_o < 50$, and Eq. (9) is no longer reliable.)

Errors $|\delta\delta\varphi|$ for the prediction of gradient retention from two gradient input runs are summarized in Fig. 8 (alkylphthalate sample), for input conditions of $t_G = 20$ and 80 min, $F = 2$ ml/min, and column length $L = 25$ cm ($k^* = 2.9$ and 11.5, respectively). Predictions were carried out for several conditions other than those of the input runs: (a) only gradient time varying, $5 \leq t_G \leq 320$ min (circles); (b) only flow-rate varying, $0.2 \leq F \leq 4$ ml/min (squares); (c) flow-rate varying and column length different, $L = 50$ cm (triangles). As expected, it is the value of k^* (or φ^*) that determines predictive error $\delta\delta\varphi$ as in the case of Fig. 6b for isocratic separation, regardless of whether t_G, F or L is varied. That is, all of the data of Fig. 8 fall on a common curve whose shape resembles that of Fig. 6b. This again reflects the similarity of isocratic and gradient elution with respect to the origin of these predictive errors $\delta\delta\varphi$.

As in the case of isocratic predictions from

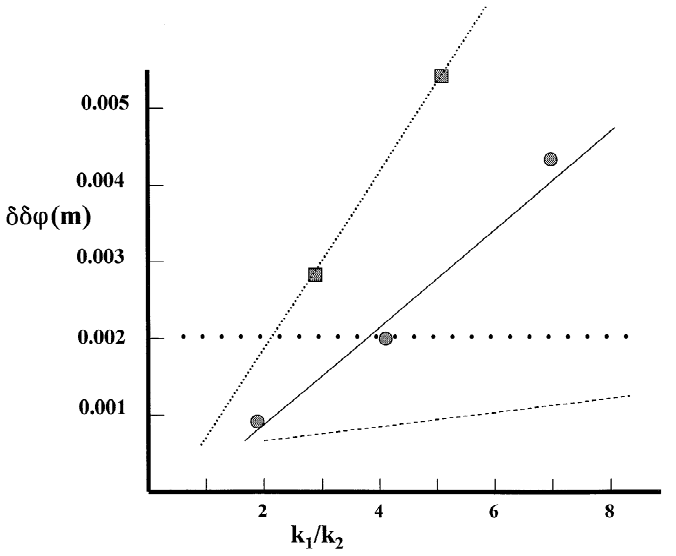

Fig. 7. Maximum interpolative error $\delta\delta\varphi(m)$ for the prediction of isocratic retention from isocratic input runs, as a function of the ratio of k values for the two input runs (k_1 and k_2). (squares, ...), hydrophilic drug sample; (circles, —), alkylphthalate sample; (----), curve from Fig. 9 for predictions of gradient retention from gradient input data. See text for further discussion.

isocratic data, the maximum interpolative error $\delta\delta\varphi$ for gradient predictions from gradient data increases, as the ratio k_1^*/k_2^* for the two input runs increases; this is shown in Fig. 9. A study of predictive errors was carried out for the shape-selective sample (squares), and these results are compared in Fig. 9 with corresponding errors for the alkylphthalate sample (circles). The two studies of Fig. 9, which involve two very different samples and conditions, provide values of $\delta\delta\varphi(m)$ as a function of t_{G2}/t_{G1} which fall close to a common curve. Data are reported in later sections for two other samples (benzoic acids + anilines, nitroaromatics), where for $k_2^*/k_1^* = 4$, $\delta\delta\varphi(m) = 0.001$ (i.e., close to the curve of Fig. 7) for each sample. Data for these four different samples therefore suggest that the relationship of Fig. 9 is fairly general when acetonitrile is used as B solvent, a conclusion which is supported by the similar curvature of log k versus φ plots when using this solvent (i.e., approximate applicability of Eq. (5)). Smaller values of $\delta\delta\varphi$ are expected for methanol as B solvent, because of the closer fit of retention data to Eq. (1).

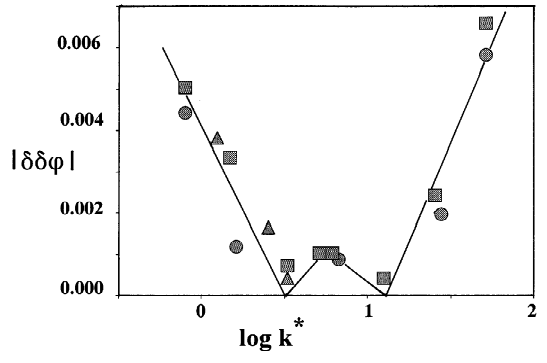

Fig. 8. Plot of error $|\delta\delta\varphi|$ versus log k^* for alkylphthalate sample. Input t_G values are 20 and 80 min. ($k^* = 3.2$ and 12.8); (circles) only t_G varies; (squares) only flow-rate varies (triangles) only column length varies; other conditions as in Fig. 2. See text for details. Note that extrapolated values are included, for which the terms $\delta\delta\varphi(a)$ or $\delta\delta\varphi(m)$ do not apply.

The maximum allowable value of $\delta\delta\varphi(m) = 0.002$ (corresponding to $\delta\delta\varphi[a] = 0.001$) is indicated by the dotted line in Fig. 9, suggesting that $t_{G2}/t_{G1} = k_1^*/k_2^* \leq 15$ should provide generally acceptable errors $\delta\delta\varphi$ in computer predictions of resolution for complex samples. This maximum value of k_1^*/k_2^* is much greater than the value $k_1/k_2 \leq 4$ found in Fig. 7 for corrresponding predictions of isocratic retention from isocratic input runs and reflects an apparent greater reliability of gradient versus isocratic predictions of retention. The curve of Fig. 9 is replotted in

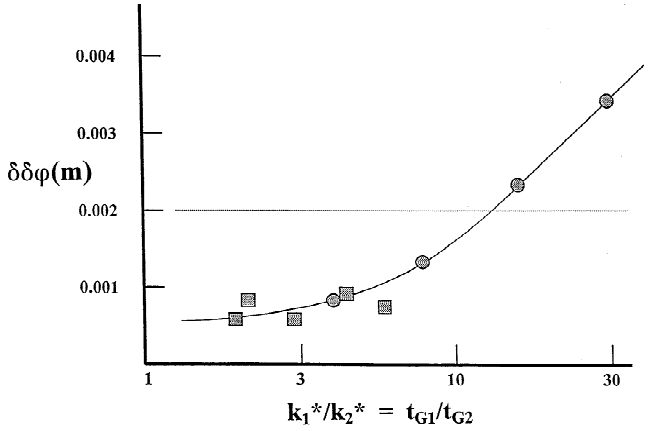

Fig. 9. Plot of maximum interpolative error $\delta\delta\varphi(m)$ as a function of the ratio of k^* or t_G values for the two input runs used for computer simulation (values of t_G for each prediction are the geometric mean of the input values t_{G1} and t_{G2}). (○) Data for phthalate sample, $T=35°C$, other conditions as in Fig. 2; (□) data for shape-selective sample, $T=25°C$, other conditions as in Section 3.

Fig. 7 as the lower curve (----), showing that values of $\delta\delta\varphi$ for gradient predictions are smaller versus isocratic predictions (o, __), especially for larger values of k_1/k_2. A similar result will be shown (below) for other samples, suggesting that gradient predictions (from gradient data) are generally more accurate than are isocratic predictions from isocratic data. Returning to Fig. 2, the latter observation suggests that the curvature of these plots for different compounds in the samples of this study is more similar over a common range in k^* (or k) values than over a common range in φ.

It might be expected that the error $\delta\delta\varphi$ for interpolated predictions of gradient retention time would approach zero as k_1^*/k_2^* approaches one. As seen in Fig. 9, however, this is not the case. Rather, the limiting value of $\delta\delta\varphi(m)$ is about 0.0005, which probably corresponds to contributions from random experimental error (''random variations in separation'', Table 1). This value can be compared with the long-term repeatability of both isocratic and gradient separations in the study of [13]: $\delta\varphi = 0.002$. As expected, $\delta\delta\varphi < \delta\varphi$ (cf. results of Fig. 6a versus b).

Errors in predicting isocratic retention when using gradient runs as input (alkyl-phthalate sample). Values of k^* versus φ^* can be predicted from the two gradient input runs. The present LSS model for computer simulation assumes that isocratic values k and φ can be substituted for values of k^* and φ^*, thereby allowing the prediction of isocratic retention from two gradient input runs. Data for the alkylphthalate sample were used to evaluate errors $\delta\delta\varphi$ that result when gradient data are used to predict isocratic retention. Fig. 10 shows results for two band-pairs from the alkylphthalate sample: C_1/C_2 (closed circles) and C_4/C_5 (closed squares). Data for gradient predictions (only t_G varying, from Fig. 8) are also superimposed on these plots (o, lower curve). The input runs had $t_G=20$ and 80 min for all predictions, as in Fig. 8, and these input conditions

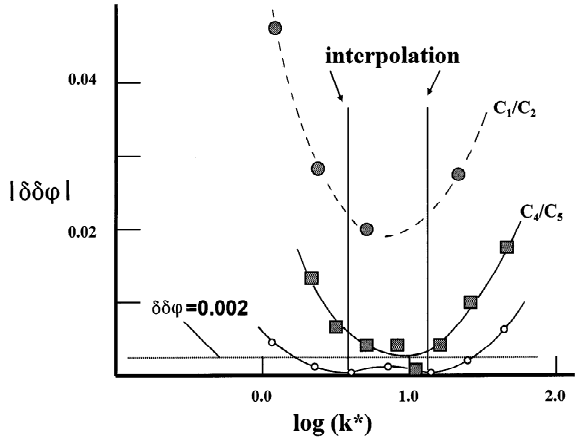

Fig. 10. Errors in predicted isocratic retention $|\delta\delta\varphi|$ for alkylphthalate sample based on gradient input runs ($t_G=20$ and 80 min, $k^*=4.2$ and 13.6). Closed circles and squares, isocratic data; open circles, gradient data from Fig. 9a. For conditions, see Section 3. See also text for details. Other conditions as in Table 4.

correspond to $k^*=3.2$ and 12.8 (vertical lines in Fig. 10 marked ''interpolation'').

It is seen in Fig. 10 that predictive errors $\delta\delta\varphi$ are considerably larger for isocratic versus gradient predictions from gradient input runs. In the case of the C_1/C_2 data, values of $\delta\delta\varphi$ are greater by an order of magnitude (versus predictions of gradient retention in Fig. 8), whereas for the C_4/C_5 data average interpolative errors $\delta\delta\varphi(a)$ are closer to the allowed limit of $\delta\delta\varphi(a)\leq 0.001$ (but still too large by a factor of about two).

Table 3 summarizes a number of examples for the alkylphthalate sample of average interpolative errors $\delta\delta\varphi(a)$ that result when gradient input data are used for isocratic predictions. These values are themselves averages from predictions for different pairs of input runs (varying in t_G), and thus summarize a large number of experiments. While Table 3 confirms the greater error of isocratic predictions for the C_1/C_2 band-pair, it is seen that the average errors for the three remaining band-pairs (C_2/C_3, C_3/C_4, C_4/C_5) are always much in excess of $\delta\delta\varphi(a)=0.001$, regardless of the ratio of t_G (or k^*) for the two input runs. These observations imply some other contribution to $\delta\delta\varphi$ than a failure of Eq. (1); i.e., for a failure of Eq. (1), errors $\delta\delta\varphi$ should increase as k_1^*/k_2^* increases. Possible contributions include errors in measured values of t_o and V_D, equipment imperfections, etc., although such errors are believed to be small in the study of [12,13].

4.3. Predictive errors for early-eluting bands (hydrophilic drug sample)

Bands that elute early in the gradient (small values of k_o) can result in either larger or smaller predictive errors $\delta\delta\varphi$. Fig. 4 shows representative chromatograms for the four samples used in the present study. The chromatograms for the alkylphthalate (Fig. 4a), shape-selective (Fig. 4b), and aniline/benzoic acid (Fig. 4d) samples do not involve early-eluting bands, whereas this is the case for the hydrophilic drug sample of Fig. 4c. The arrows in Fig. 4 indicate the arrival of the gradient at the column outlet at some time after sample injection (time zero). This gradient delay is the result of the hold-up volume of the equipment plus column (V_D+V_m).

Gradient pre-elution. During the time before the arrival of the gradient at the column inlet, the sample is eluted isocratically by the starting mobile phase (''pre-elution''). For the alkyl phthalate (Fig. 4a) and shape-selective (Fig. 4b) samples, the first peaks in the chromatogram appear at a time considerably later than the arrows, implying large values of k_o for all peaks. As a result, all compounds in the sample are held at the column inlet during pre-elution, with no resulting effect on their separation (other than an approximately equal increase in retention time for each band), and no special contribution to predictive errors $\delta\delta\varphi$. For the hydrophilic drug sample (Fig. 4c), peak 1 elutes before the gradient arrives at the column inlet, and peak 2 elutes before the gradient reaches the end of the column. As a result, both peaks elute under isocratic conditions. Computer simulation (based on two gradient runs where t_G is varied) normally allows a determination of values of k_w and S (Eq. (1)) for each sample band [11], thereby permitting the prediction of sample retention for both isocratic and gradient conditions. However, this is not possible when early peaks elute isocrati-

Table 3
Computer simulation for the phthalate sample where isocratic retention is predicted from gradient input runs: %B and t_G varied[a]

t_{G2}/t_{G1}	$\delta\delta\varphi$ (interpolation)				Average $\delta\delta\varphi$[b]
	C_1/C_2	C_2/C_3	C_3/C_4	C_4/C_5	
2	0.009±0.003	0.005±0.003	0.002±0.000	0.005±0.003	0.004
4	0.015±0.011	0.006±0.003	0.009±0.006	0.006±0.003	0.007
8	0.009±0.001	0.008±0.006	0.001±0.000	0.006±0.003	0.008
Average	0.011	0.006	0.004	0.006	0.006

[a] Conditions as in Experimental section. For input gradient runs, values of t_G varied; other conditions as in Fig. 2. Errors $\delta\delta\varphi$ are averages for several interpolated predictions that involve different input runs. Data are grouped by band-pair (e.g. C_1/C_2) and the ratio of input gradient times (t_{G2}/t_{G1}).

[b] C_1/C_2 data excluded.

cally, as in Fig. 4c. As long as the initial mobile phase composition φ_o is unchanged, the retention times of bands such as No.'s 1 and 2 in Fig. 4c will remain constant as t_G is varied. Consequently, accurate isocratic predictions for bands 1 and 2 at compositions other than φ_o are not possible.

Early-eluting bands and predictions of retention from gradient input runs. Bands 3–8 of Fig. 4c elute from the column under gradient conditions, but their values of k_o are still small ($6<k_o<37$) compared to the examples of Figs. 4a,b ($k_o>100$). As a result, Eq. (9) is a poor approximation for estimating the average retention k^* of these bands during gradient elution, and Eq. (8) must be used instead. We have seen that maximum predictive accuracy occurs for interpolative predictions, which are in turn defined by values of k^* for the two gradient runs used as input. So far as predictions for gradient retention are concerned, there is no practical effect of sample pre-elution (as long as φ_o remains constant). Thus, if the input runs are based on gradient times of 10 and 40 min, any gradient time within this range corresponds to an interpolation. In the case of isocratic predictions, however, the situation is quite different. This is illustrated in Table 4 for the hydrophilic drug sample, where calculated values of k^* (gradient) are shown for bands 3–8 and gradient times of 10–80 min (left side) and compared with experimental isocratic values of k for 3, 6, 9 and 12% B (right side).

If input runs of 10 and 40 min are assumed in Table 4, values of k^* for band 3 equal 3.4 (10 min) and 4.9 (40 min), respectively, for a ratio of 4.9/3.4=1.4. That is, interpolative predictions of isocratic retention for band 3 are only possible within a narrow range of k values (corresponding to a narrow range of %B values). Comparing this range in k^* ($3.4\leq k^*\leq 4.9$) with isocratic values of k for 3–12% B (band 3, Table 4), it is seen (values marked by ''a'') that only the value of $k=3.4$ for 6% B represents an interpolation of these k^* values. This situation contrasts with the situation of samples (or bands) where k_o is large. In that case, varying t_G by 4-fold would result in k^* values that also vary by four-fold (Eq. (9)); i.e., a much wider range of values of k (isocratic) can be predicted accurately (interpolation) from gradient input runs when k_o is large. Returning to the data of Table 4, the range in k^* is seen to increase for later peaks (with larger values of k_o); e.g., for input runs of 10 and 40 min, peak 8 has $k^*=5.9$–16, or a ratio of 16/5.9=2.7. As the value of k_o continues to increase (later eluting bands), this ratio will eventually approach the value of 4 predicted by Eq. (9), for input t_G values equal to 10 and 40 min.

Isocratic predictions from isocratic input data. The isocratic data of Table 4 for early-eluting bands were used to estimate the error $\delta\delta\varphi$ in isocratic predictions from isocratic data as follows. For input runs of 3 and 9% B, predictions were made for 6% B; similarly, for input runs of 6 and 12% B, predictions were made for 9% B (errors equal to $\delta\delta\varphi[m]$). Additionally, for input runs of 3 and 12% B, predictions were made for 6 and 9% B. Maximum interpolative errors $\delta\delta\varphi(m)$ were then obtained for a change in input %B of 6% (avg. $k_1/k_2=2.9$) or 9% (avg. $k_1/k_2=5.1$). The observed errors $\delta\delta\varphi$ for $k_1/k_2=5.1$ (6 and 9% B) are not *maximum* interpolated

Table 4
Retention for hydrophilic drug sample as a function of t_G (gradient) and %B (isocratic); other conditions as in Section 3

Band	Calculated k^* (gradient)				Experimental k (isocratic)			
	$t_G=10$ min	20 min	40 min	80 min	3% B	6% B	9% B	12% B
3	3.4	4.3	4.9	5.3	5.9	3.4[a]	2.2	1.9
4	4.0	5.3	6.3	7.0	7.9	4.4[a]	2.8	1.9
5	3.3	4.8	6.2	7.3	9.2	4.0[a]	2.2	1.3
6	3.5	5.6	7.8	9.8	13.4	5.5[a]	2.8	1.3
7	4.4	7.7	12.4	17.8	36.4	13.5[b]	6.4[a]	3.6
8	5.9	10.3	16.5	23.5		24.1	13.3[c]	8.1[a]

[a] Value of k corresponds to interpolated prediction for input runs of $t_G=10$ and 40 min.
[b] Value of k corresponds to interpolated prediction for input runs of $t_G=20$ and 80 min.
[c] Value of k corresponds to interpolated prediction for both 10/40 and 20/80 min input runs.

errors, which would only be the case for a predicted run with (intermediate) 7.5% *B*. The actual values of $\delta\delta\varphi$ were accordingly adjusted by an estimated factor of 1.5. Resulting values of $\delta\delta\varphi(m)$ are plotted versus k_1/k_2 as squares in Fig. 7, and are seen to be larger than for the alkyl phthalate sample (circles in Fig. 7). Two effects could explain the greater error of retention predictions for the hydrophilic drug sample. First, experimental values of *k* for the latter sample are on average smaller than for the alkyl phthalate sample, and therefore may be less accurate. Second, the hydrophilic drug sample exhibits reasonable retention ($k>1$) only for relatively low values of %*B*, and the curvature of plots of log *k* versus %*B* is usually greater at low %*B*. Whereas a value of $k_1/k_2<4$ provides acceptable accuracy ($\delta\delta\varphi[m]\leq 0.002$) for well-retained samples such as the alkylphthalates, the corresponding requirement for the hydrophilic drug sample is $k_1/k_2\leq 2$. For well-retained samples, Fig. 7 (circles) suggests a maximum difference in %*B* for the two input runs of 15%; for less-retained samples (Fig. 7, squares), the maximum difference is only about 5%. Allowing for modest extrapolation of predictions, the recommended predictive ranges are (assumes two input runs with %*B* varying): ≤25% *B* for well-retained samples; ≤10% *B* for samples that elute with $\varphi<$ 20% *B*. The lesser accuracy of predictions for $\varphi<$ 20% *B* may reflect steeper plots of log *k* versus φ for small φ, as predicted by Eqs. (5) and (7).

Gradient predictions from gradient input data. Computer simulations of gradient retention were carried out as follows. For input runs of $t_G=10$ and 40 min, separation was predicted for $t_G=20$ and 80 min; for input runs of $t_G=20$ and 80 min, separation was predicted for $t_G=10$ and 40 min. In each case, the ratio of input t_G values is 4. The resulting average value of $\delta\delta\varphi(m)$ was 0.0003 for peaks 1 and 2, while the average value for remaining peaks 3–8 was $\delta\delta\varphi(m)=0.0009$. As expected, the error for the two pre-eluting peaks (0.0003) is less and probably corresponds to random experimental error (since t_R should not vary with t_G for these peaks). The value of $\delta\delta\varphi(m)=0.0009$ for the remaining peaks can be compared to the average value found for the preceding two samples for $t_{G1}/t_{G2}=4$: $\delta\delta\varphi(m)=0.0008$ (Fig. 9). Thus, all three samples exhibit a similar error $\delta\delta\varphi(m)$ for gradient predictions from gradient input data, suggesting that the results of Fig. 9 are generally applicable. Extrapolative error for a two fold change in t_G outside the range of t_G for the input runs was $\delta\delta\varphi(m)=0.0017$, which is somewhat larger than observed for the alkylphthalate sample, but still acceptable (since $\delta\delta\varphi[a]\approx 0.008$).

Isocratic predictions from gradient input data. Because very few band-pairs fall within the interpolation limits of Table 4 for the hydrophilic drug sample (see entries marked by one or more *), the criterion for inclusion was broadened as follows. The *average* value of $k=k(\text{avg})$ was determined for each band-pair, and interpolation was then defined as the value of $k(\text{avg})$ falling within the limits defined in Table 4. For example, for 10- and 40-min input runs with prediction of retention for bands 6 and 7 and 6% *B*, the value of *k* for band 7 (13.5) falls outside the interpolation range of $4.4\leq k^*<12.4$. Therefore, the value of $\delta\delta\varphi$ for this case would be considered as an extrapolation (original definition). However, by accepting an average value of *k* for bands 6 and 7 [0.5(5.5+13.5)=9.5], and the average range in k^* for these two bands: [0.5(3.4+4.4)$<k^*<$0.5(7.8+12.4), or (4.0$<k^*<$10.1)], the value of $\delta\delta\varphi$ is now defined as an ''interpolation''. With the latter convention, 11 out of 46 predicted values of $\delta\delta\varphi(a)$ (based on both 10/40 and 20/80 input runs) became ''interpolations''. The resulting average value for interpolation was $\delta\delta\varphi(a)=0.002$ (i.e., marginal), while for extrapolation the average value of $|\delta\delta\varphi|=$ 0.006 (unacceptable).

4.4. Summary of predictive error when varying gradient time or isocratic %B

Values of maximum interpolated error $\delta\delta\varphi(m)$ are summarized in Table 5 for the above three samples plus two additional samples: ten nitro aromatics and 14 benzoic acid plus aniline derivatives (described in a following section). In each case, the two input runs have k_1/k_2 or $k_1^*/k_2^*=4$. The errors for gradient predictions from gradient input runs (middle column) are seen to be consistent and small: $\delta\delta\varphi(m)=0.001$. Average errors for such predictions are estimated as $\delta\delta\varphi(a)=0.0005$. Errors for the prediction of isocratic retention from isocratic input data show an average value of $\delta\delta\varphi(a)=0.0015$; i.e., three times

Table 5
Summary of maximum interpolative errors $\delta\delta\varphi(m)$ for predicted separations of different samples (change in gradient time or isocratic %B); assumes that input values of k_1/k_2 or $k_1^*/k_2^* = 4$

Sample	$\lvert\delta\delta\varphi\rvert(m)$		
	Isocratic predictions from isocratic data	Gradient predictions from gradient data	Isocratic predictions from gradient data[a]
Alkylphthalates	0.002	0.001	0.012[b]
Shape selective	–	0.001	–
Hydrophilic drugs	0.004	0.001	0.004
Benzoic acid plus aniline derivatives	–	0.001	–
Nitroaromatics (Table 7 of [21])	–	–	0.004
Average $\lvert\delta\delta\varphi\rvert(m)$	0.003	0.001	0.007
Average $\lvert\delta\delta\varphi\rvert(a)$	0.0015	0.0005	0.0035

[a] Assumes $\delta\delta\varphi(m) = 2\delta\delta\varphi(a)$.
[b] Excludes outlier values for C_1/C_2 band-pair (Table 3).

less accurate than for predictions of gradient retention. However, only two samples were studied, one of which (the hydrophilic sample) may be atypical because of elution by low-φ mobile phase. Errors for isocratic retention from gradient data vary markedly between the alkylphthalate and other two samples, but give an average value of $\delta\delta\varphi(a) = 0.003$ (excluding C_1/C_2 alkylphthalate data).

Values of $|\delta\delta\varphi|$ were also obtained for the samples of Table 5 with extrapolated predictions, where the extrapolation was extended by a factor of two in k or k^* at both ends of the interpolative range $k_1 \le k \le k_2$; i.e., a predictive range of $0.5k_1 \le k \le 2k_2$. Fig. 6b suggests that the average error $\delta\delta\varphi(a)$ should not increase for this modest extrapolation, compared to errors from interpolation only. This conclusion (modest extrapolation OK) was confirmed for the other samples studied, for predictions of isocratic retention from isocratic input data, and for gradient retention from gradient input data. Larger extrapolative errors were sometimes found, however, for isocratic predictions from gradient input runs.

The results of Table 5 suggest that computer simulation accuracy is acceptable for both complex and simpler samples, when gradient retention is predicted from gradient input runs ($\delta\delta\varphi[a] = 0.0005$ for $k_1^*/k_2^* \le 4$). Fig. 7 further suggests that input values of k_1^*/k_2^* as large as 15 will still meet the requirement of $\delta\delta\varphi(a) \le 0.001$; e.g., input values of t_G can be varied by 15-fold, allowing prediction of an even wider (but impractical) range in k^* by modest extrapolation. For predictions of isocratic retention from isocratic input data, the average error $\delta\delta\varphi(a) \approx 0.0015$, which is marginal for complex samples. However, complex samples generally require gradient elution for their separation. Predictions of isocratic retention from gradient data have an average error $\delta\delta\varphi(a) = 0.0035$, corresponding to an error in resolution of 0.7 units. Errors in predicted resolution of this size are unacceptable for complex samples, and are marginal to unacceptable for simpler samples. However, a later section will show that initial predictions that are less reliable can be corrected by carrying out one additional experimental run.

4.5. Errors in predicted retention when varying temperature (gradient elution only)

Isocratic retention varies with absolute temperature T_K as

$$\log = A + B/T_K \quad (2)$$

and gradient retention can be approximated (1) by

$$t_R = A' - B'T_K \quad (4)$$

Modest curvature of plots of t_R versus T_K in gradient elution (Eq. (4)) was noted in [1], as illustrated in Fig. 11 for three representative examples. For each compound in Fig. 11, a straight line joins the two end points. In most cases, the curvature of data as in Fig. 11 has an acceptably small effect on predictive accuracy (average errors in predicted values of α

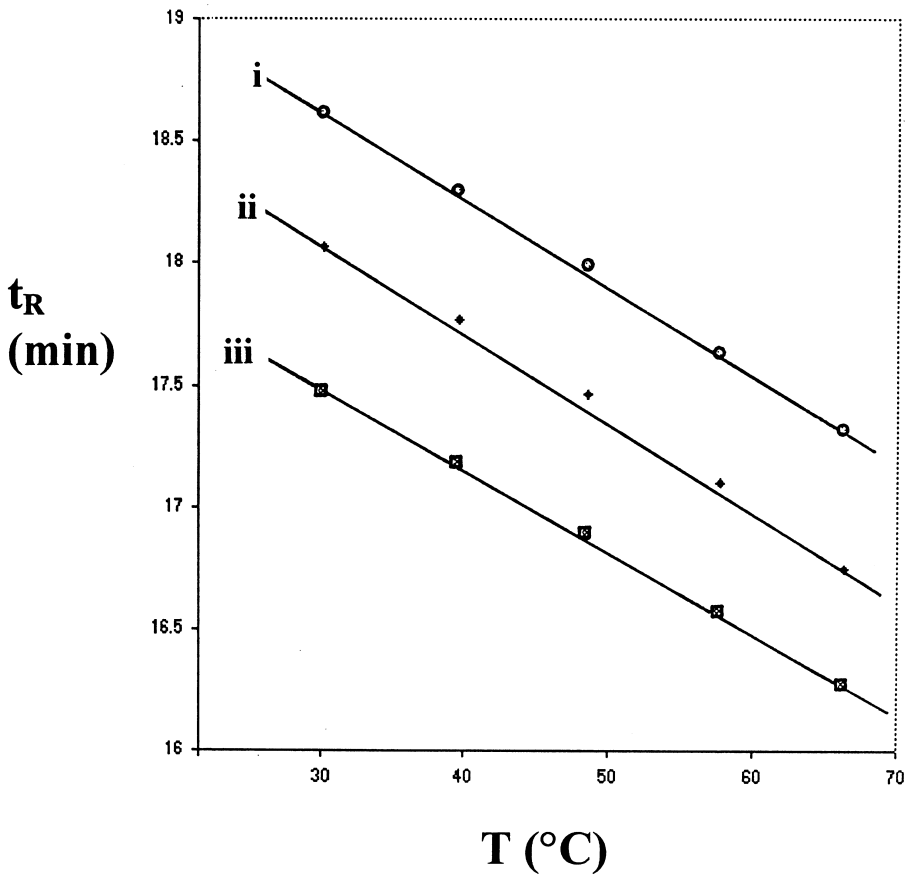

Fig. 11. Gradient retention time versus temperature for ''typical'' compounds. Data from laboratory B of (1); 25×0.46 C_{18} column, $t_G = 20$ min, 5–100% *B* gradients, 2.0 ml/min; sample compounds are danthron (i), 1-nitrooctane (ii) and diflunisal (iii).

equal 0.4–1.1% for six samples of Table 8 of [1]). A corresponding average value of $\delta\delta\varphi(a) = 0.0008$ can be calculated (Eq. (A.3) of Appendix A) for the latter separations, where the difference in T for the input runs (ΔT) varied from 17 to 40°C.

Shape-selective sample. Larger errors in predicted retention (as T was varied) were observed in (1) for the case of certain ''shape-selective'' compounds separated with a polymeric C_{18} column. Shape-selective molecules vary greatly in length-to-width ratio or in three dimensional ''bulkiness'' (e.g. nonplanar *o*-terphenyl versus planar triphenylene). We have re-examined the separation of some of these compounds (comprising the shape-selective sample) as a function of temperature. Fig. 12 shows representative plots (gradient t_R versus T) obtained in the present study for three of these compounds. Not only are the deviations from linearity (Eq. (4)) larger than in Fig. 11, but both concave and convex plots were observed, which can result in much larger errors $\delta\delta\varphi$. These more serious deviations from Eq. (4) can serve as a model for evaluating the potential seriousness of errors $\delta\delta\varphi$ that can arise from a failure of Eq. (4), as well as for exploring means to minimize or correct such errors.

Maximum interpolative errors $\delta\delta\varphi(m)$ for the shape selective sample were determined as a function of ΔT for different pairs of input runs: $\Delta T = 10$, 20, 30, 40, 50°C; $\delta\delta\varphi(m) = 0.0007$, 0.0017, 0.0035, 0.0080 and 0.0138, respectively. As expected from the examples of Fig. 12, values of $\delta\delta\varphi$ increase rapidly with ΔT. For $\Delta T = 20$°C, the average interpolative value of $\delta\delta\varphi(a) \approx (0.00017/2) = 0.0008$. Values of $\Delta T > 20$°C for this sample lead to unacceptable errors $\delta\delta\varphi(a)$, but this suggests that ΔT can be at least as large as 20°C for this sample.

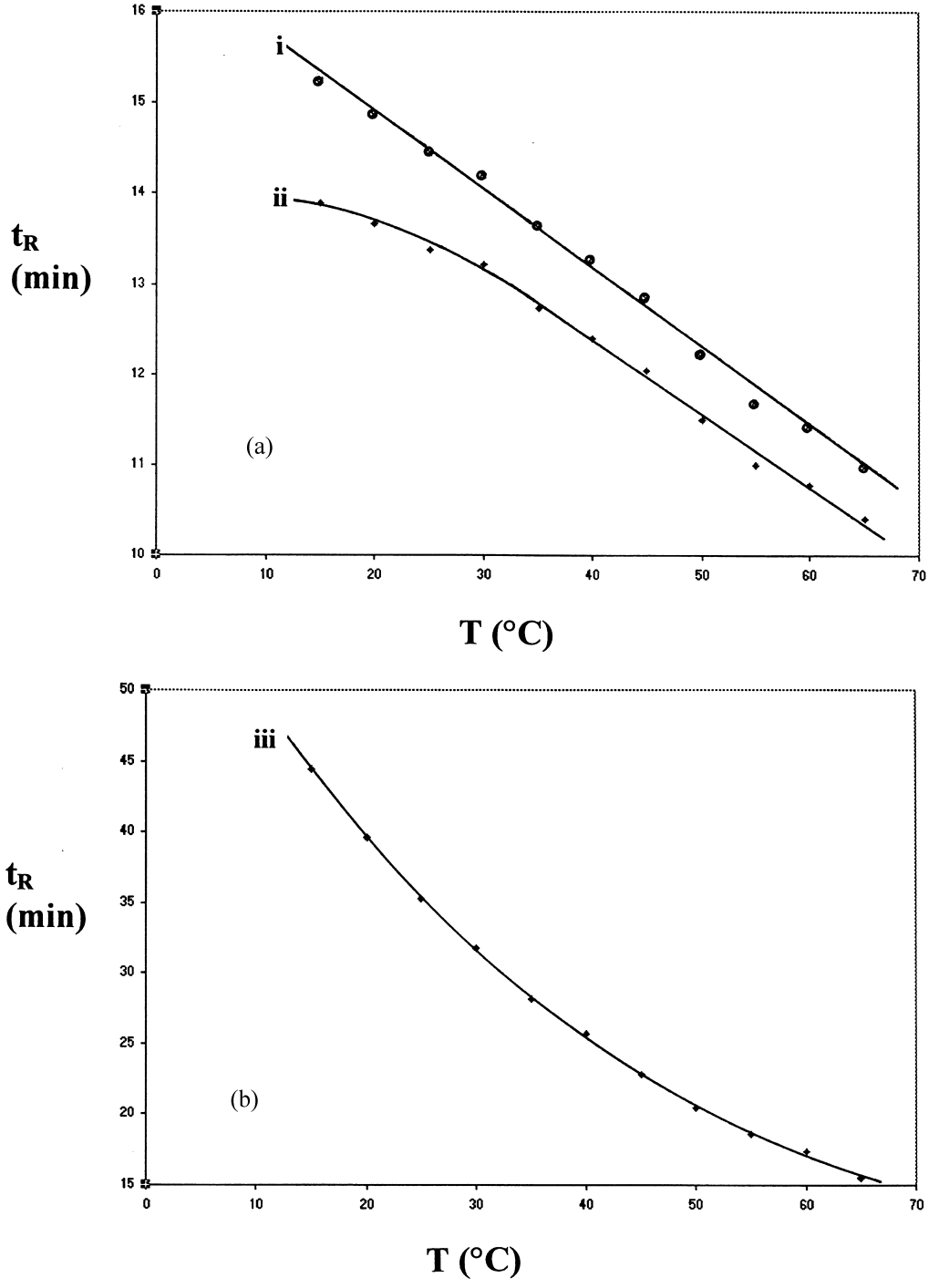

Fig. 12. Gradient retention time versus temperature for ''shape selective'' sample from present study. Sample: 1,3,5-triphenylbenzene (i), tetraphenylmethane (ii) and 1,6-diphenylhexatriene (iii); gradient time of 20 min, other conditions as in Section 3.

Because the shape-selective sample is intentionally atypical, $\Delta T > 20$°C should be acceptable for most samples and conditions, as confirmed by prior examples cited above from [1].

Interconvertable species. A solute X may undergo an equilibrium change to species Y in the mobile phase: X⇔Y. In general, the temperature dependence of retention $d(\log k)/d(1/T_K)$ will differ for species X and Y, and the equilibrium constant K for this reaction will also be temperature dependent. As a result, when both X and Y are present in the mobile phase in comparable concentrations (e.g., 10–90% X or Y), a failure of Eqs. (2) and (4) becomes possible. Some examples of this kind have been reported for the use of cyclodextrin (CD) as a solute-complexing agent in reversed-phase HPLC [26–28], where X corresponds to the free solute and Y to its complex with CD.

A more common example of solute interconversion, with the possibility of failure of Eqs. (2) or (4), arises for acid or base solutes that undergo reversible protonation [29]. For mobile phase pH values that result in >90% conversion of the solute to either its acid or base conjugate, retention will be determined by the dominant species, and any failure of Eqs. (2) or (4) should be minor. The situation is different, however, when the mobile phase pH is close to the pK_a value of the solute. That is, failure of Eqs. (2) or (4) is likely in the case of partly-ionized solute molecules where the relative amounts of X and Y ≫ 10%. An example of this is provided by the separation of the benzoic acids/anilines sample of Section 3. Using input runs with temperatures of 35 and 75°C, predictions of retention were made for 55°C. The resulting resolution error (average of results for three different gradient times) was $\delta\delta\varphi(m) = 0.0026$, which is larger than the value of $\delta\delta\varphi(m) \approx 0.001$ suggested by other samples reported in [1]. However, if data for two of the 14 solutes in the benzoic acids/anilines sample (3- and 4-chloroaniline) are excluded, $\delta\delta\varphi\phi(m) = 0.0013$ (a reduction in average error by half). The pK_a values of the latter two solutes (3.0 and 3.3, respectively [29]) are reasonably close to the mobile phase pH = 2.6, which suggests partial ionization of the solute as a possible contribution to these larger errors $\delta\delta\varphi$ (see further discussion below).

A preceding study [1] reported retention as a function of temperature, gradient time and pH for the 14 substituted benzoic acids and anilines in the present sample, plus three additional anilines. Table 6 summarizes retention errors $\delta\varphi$ for predictions at a temperature (51°C) which is intermediate between the two temperatures (32 and 70°C) used for the two input runs. The resulting errors of Table 6, for mobile phases of varying pH, are averages for

Table 6
Failure of Eq. (4) for prediction of retention t_R versus temperature T as a function of the relative ionization of the solute; substituted benzoic acids and anilines; errors in retention $\delta\varphi$ calculated from data of [1]; bolded values are for partially ionized solutes (see text)

Sample	Mobile phase pH	Average error $\delta\varphi(m)$ in predicted retention for 50.9°C[a]		
		$2.0 \le pK_a \le 2.5$[b]	$3.0 \le pK_a \le 3.5$[c]	$3.5 \le pK_a \le 4.5$[d]
Substituted anilines	2.6	−0.0018	**−0.0038**	−0.0005
	3.6	−0.0012	**−0.0023**	**−0.0016**
	4.6	−0.0015	−0.0013	**−0.0045**
	5.6	−0.0011	−0.0014	−0.0010
Substituted benzoic acids	2.6	**−0.0017**	**−0.0016**	−0.0021
	3.2	−0.0012	**−0.0016**	−0.0024
	3.7	−0.0004	−0.0009	**−0.0024**
	4.3	−0.0003	−0.0007	**−0.0020**

[a] Input run temperatures of 32.1 and 69.7°C; average results for two different t_G values; pK_a values in this mobile phase estimated from data of [29].

[b] 2-chloro (No. 15), 3,4-dichloro (No. 16) and 3,5-dichloro (No. 17) anilines; 2-nitrobenzoic acid (No. 2).

[c] 4-chloro (No. 14) and 3-chloro (No. 13) anilines; phthalic acid (No. 1), 3-cyano (No. 3), 2-fluoro (No. 4), 2-chloro (No. 5) and 3-nitro (No. 6) benzoic acids.

[d] 4-methoxy (No. 9), 3-methyl (No. 10), *N*-ethyl (No. 11) and 3,5-dimethyl (No. 12) anilines; 3-fluoro (No. 7) and 2,6-dimethyl (No. 8) benzoic acids.

groups of solutes having similar pK_a values (and therefore a similar extent of ionization for a mobile phase of given pH). The bolded values in Table 6 correspond to errors $\delta\varphi > 0.0015$. From the preceding discussion, we would expect larger errors when values of the solute pK_a and mobile phase pH are similar, and this is observed in almost every case. The one exception is for the less acidic benzoic acids (p$K_a > 3.5$), where $\delta\varphi > 0.0015$ for all mobile phase pH values. Errors are also seen to be generally larger for the anilines than for the benzoic acids, which may reflect differences in the dependence of pK_a on temperature for basic versus acidic solutes.

4.6. Errors in predicted retention when simultaneously varying gradient time and temperature

Several reported studies have used four runs with temperature T and gradient time t_G varying in order to optimize separation as a function of both t_G and T. For typical samples and representative changes in T (e.g. $\Delta T = 40°C$) and t_G (e.g. four-fold change) for the input runs, contributions $\delta\delta\varphi(a)$ are estimated above as ≈ 0.001 for a change in either T or t_G. If these contributions contribute independently, then we can estimate $\delta\delta\varphi(a) \approx 2^{1/2} \cdot 0.0001 \approx 0.0014$ for a simultaneous change in each variable – as in the simultaneous optimization of T and t_G (cf. Fig. 1). However, other evidence points to these separate values of $\delta\delta\varphi(a) \approx 0.001$ as arising partly from random errors (≈ 0.0003–0.0005) which would not add. This suggests that values of $0.001 < \delta\delta\varphi(a) < 0.0014$ are probably typical for the above range in T and t_G.

For eight different samples, Table 7 summarizes average errors $|\delta\delta\varphi|$ for predictions where both t_G and T differ from the input values (based on four input runs where T and t_G are varied, as in [2,3]). With the exception of separation No. 4, the average value of $|\delta\delta\varphi|$ for all remaining predictions was 0.0006; i.e., better than the expected value of $|\delta\delta\varphi(a) \approx 0.0010$–$0014$. Separation No. 4 features both a very short gradient time (5.5 min), as well as extrapolation outside the input run conditions ($t_G = 7$, 21 min), both of which conditions favor increased error $\delta\delta\varphi$. These results suggest that computer simulation involving interpolation or modest extrapolation for the simultaneous variation of T and t_G and the prediction of gradient retention should give generally acceptable results.

4.7. Improving the accuracy of retention predictions

The preceding analysis of computer simulation errors suggests that these errors may in some cases lead to unacceptable predictions of retention and resolution, for the case of both complex and other samples. Excessive predictive errors can be anticipated on the basis of guidelines developed so far and summarized in Tables 5 and 8. The question then is:

Table 7
Summary of average error $\delta\delta\varphi(a)$ for interpolated predictions of retention where temperature and/or gradient time differ from the input-run conditions; summary of data from indicated references (''Ref.'')

Sample	n	Ref.	Input conditions		Predicted conditions		$\delta\delta\varphi(a)$
			t_G (min)	T (°C)	t_G (min)	T (°C)	
(1) Herbicide impurities	9	[1]	10, 30	39.9, 57.3	20	48.4	0.0004
(2) Pharmaceuticals	9	[1]	30, 90	35, 75	60	55	0.0020
(3) Corticosteroids	9	[2]	20, 60	30, 60	52.5	30	0.0006
(4) Algal pigments	11	[3]	7, 21	40, 50	5.5	39	0.0100
(5) Synthetic organics	11	[2]	30, 60	30, 70	45	45	0.0008
(6) Herbicides	13	[2]	40, 120	30, 40	90	34	0.0001
(7) Herbicides, pH 2.7	19	[17]	40, 120	20, 35	80	25	0.0005
(8) Herbicides, pH 3.5	19	[17]	40, 120	20, 40	80	30	0.0002
(9) Algal pigments	29	[2]	17, 51	50, 60	54	55	0.0003

Table 8
Summary of rules to ensure reliable computer simulation

Type of prediction	Comment
Gradient retention from gradient input data (T constant)	Errors generally small ($\delta\delta\varphi[a] \leq 0.001$) for interpolation, when t_G or k^* values for input runs differ by no more than 15-fold (Fig. 9); modest extrapolation of t_G (by no more than a factor of 2) is also allowed (Fig. 8); predictions for changing flow-rate or column length must take into account the resulting change in k^* (Eqs. (8)–(10)), so that predicted values of k^* fall within the range of interpolation or modest extrapolation (Fig. 8)
Isocratic retention from isocratic input data (T constant)	Errors generally larger than for gradient predictions from gradient data; for interpolation and $\delta\delta\varphi$[a] <0.001, k for input runs should vary by no more than four-fold (Fig. 7); extrapolation can be carried out for an additional factor of 2 in k (Fig. 6b); for typical samples, this corresponds to input runs differing by 15 %B (e.g., 40 and 55 %B), with extrapolative predictions possible for 5% B lower and higher (e.g., 35 to 60% B)
Isocratic retention from gradient input data (T constant)	Errors generally larger than for isocratic predictions from isocratic data; average errors $\delta\delta\varphi(a) < 0.003$ may be difficult to achieve; extrapolation from input data may be risky
Gradient retention as a function of temperature (t_G constant)	Errors generally small ($\delta\delta\varphi[a] \leq 0.001$) for interpolation, and modest extrapolation is usually reliable; temperature of input runs can vary by 40°C and possibly more (Table 7)
Isocratic retention as a function of temperature (%B constant)	Not investigated, but assume similar guidelines as for prediction of gradient retention as a function of temperature
Gradient retention as a function of both T and t_G	For most samples, combined errors appear acceptable for a reasonable range in T (e.g., 40°C) and t_G (4-fold) (Table 7)

Exceptions to above rules; effect of sample type

Sample type	Comment
Weakly retained (hydrophilic) sample (mobile phases of 0–30% B)	Similar interpolative errors for gradient predictions from gradient data; somewhat larger (but still acceptable) extrapolative errors Isocratic predictions less reliable, and range of %B for accurate predictions is reduced; if predictions for a range in %B >10% is desired, three input runs varying in %B should be used (e.g. 5, 15 and 25% B)
Samples whose molecules have quite different shapes	See text for examples; these compounds show an unusual dependence of retention versus temperature T; if two input runs differing in T are used, predictions of retention versus T should be limited to a range in T of no more than 20°C; the use of three or more input runs varying in T allows reliable predictions for a wider range in T (e.g. 40°C or more)
Samples whose molecules can interconvert to two or more distinct species	The major example is partly ionized solutes (acids or bases, where mobile phase pH is within ±0.5 units of the pK_a value); predictions of retention versus T are less reliable, especially for basic samples; if two runs differing in T are used for input, the range in T should be no greater than 20°C; the use of three or more input runs varying in T allows reliable predictions for a wider range in T (see Table 6)
Complex samples	Samples with $n>20$ usually require errors $\delta\delta\varphi(a) \leq 0.001$; accurate predictions of gradient retention from gradient input runs are usually possible for a range in $t_G \leq 4$ and of $T \leq 40$°C (see Table 7)

can these errors be reduced to acceptable levels by appropriate steps? Two possible responses to this question are discussed next.

Use of additional input runs. Four input runs are required for the prediction of retention as a function of T and t_G. The preceding discussion suggests that additional runs will not usually be required to meet the accuracy requirement of $\delta\delta\varphi(a) \leq 0.001$ for complex samples and predictions of gradient retention. The range of values of T ($\leq$40°C) and t_G (at least ten-fold) that can be explored with only four runs is adequate for the purposes of method development. However, samples composed of molecules that are (a) of quite different shape or (b) partially ionized under the conditions of separation can exhibit an anomalous dependence of retention on temperature. For these situations, the use of three different temperatures for the input runs (six input runs total) may be required for acceptable predictive accuracy when the range in $T > 20$°C. We have found that a cubic spline fit to values of t_R versus T provides acceptable predictive accuracy, when data for three or more temperatures are available.

For isocratic predictions from isocratic input data,

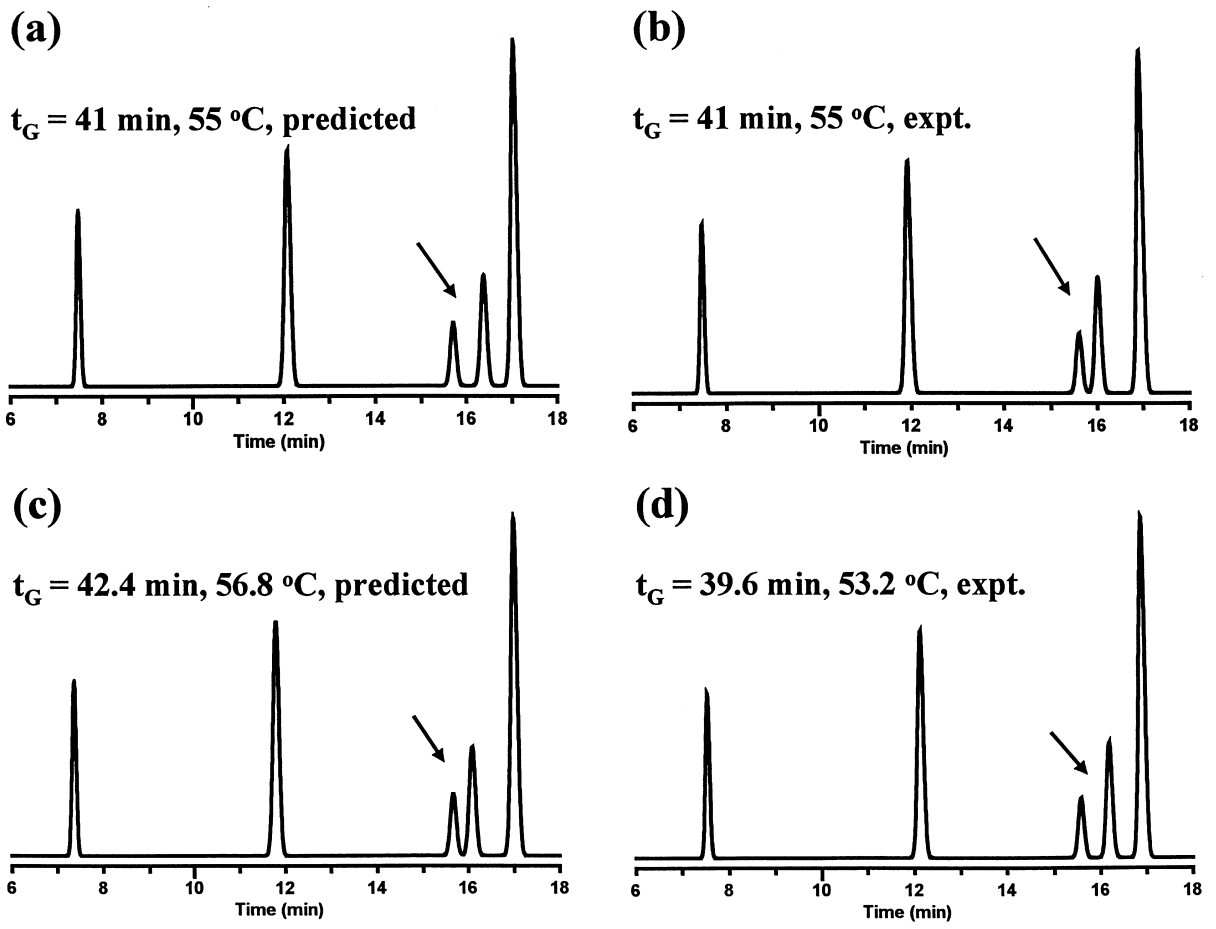

Fig. 13. Illustration of the correction of erroneous computer predictions by comparison of experimental and predicted chromatograms (''reflection procedure''). Conditions: shape-selective sample, 25×0.46 cm C_{18} column, 2 ml/min; 50–100% acetonitrile in water gradients; other conditions noted in figure. Input conditions: t_G=20 and 60 min, T=40 and 65°C. (a) Predicted (favored) separation for indicted t_G and T; R_s=2.6; (b) ''experimental'' separation for conditions of (a); (c) predicted separation (t_G=42.4 min, T=56.8°C) that matches that of (b); (d) adjustment of experimental conditions (''reflection'', t_G=39.6 min, T=53.2°C) to obtain desired separation of (a). Arrows indicate critical band-pair 3/4. See text for details.

the use of an additional run with %*B* varying should prove useful for predictions over a wide range in %*B* (>25% *B*), or for >10% *B* with less retained samples that elute with <30% *B*.

Correction of inaccurate predictions after the fact. Given that any failure of Eqs. (1) and (4) can limit the accuracy of computer simulation where *T* and/or t_G are varied, is it possible to correct for these errors $\delta\delta\varphi$ by the use of an additional run? For example, if separation is optimized in terms of *T* and t_G (based on four input runs), and it is found that the resulting experimental separation deviates significantly from the predicted run, can this fifth run be used for an improved prediction of optimized conditions (and a separation that matches the originally predicted optimum)? An answer to this question is provided by reference to Fig. 3, considering retention errors $\delta\varphi$ for a single compound in the sample. The predicted retention for $\varphi = 0.5$ will be in error by the quantity $\delta\varphi$. However, if the actual mobile phase were changed to $\varphi = 0.5 - \delta\varphi$, the resulting separation should be equivalent to that predicted (incorrectly) for $\varphi = 0.5$.

The foregoing observation can be translated into a strategy for correcting predictive errors, as illustrated by the following example for the shape-selective sample. This sample and input conditions ($t_G = 20$ and 60 min, $T = 40$ and 65°C) were selected in order to create significant error in predictions of retention. In order to obtain accurate (''experimental'') simulations for comparison with predictions based on the latter input runs (20, 60 min; 40, 65°C), input runs with a narrower range in temperature were used (50 and 60°C, for which errors in predicted resolution were expected to be <0.2 R_s units).

An initial simulation (based on 40 and 65°C inputs) resulted in a desired separation for a gradient time of 41 min and a temperature of 55°C; $R_s = 2.6$ for this predicted separation, shown in Fig. 13a. The ''experimental'' separation for these conditions (based on 50 and 60°C inputs) is shown in Fig. 13b, and it is significantly poorer; $R_s = 1.6$ for critical band-pair 4/5 (arrow). Trial-and-error computer simulation was next carried out in order to find conditions that would best match the ''experimental'' separation of Fig. 13b for all peaks; Fig. 13c for $t_G = 42.4$ min and 56.8°C gives a close fit to the ''experimental'' separation of Fig. 13b. From this, it appears that the initial prediction of optimized conditions (Fig. 13a) was ''in error'' by $\delta t_G = 1.4$ min and $\delta T = 1.8$°C. An experimental separation equivalent to the desired separation of Fig. 13a should therefore result by subtracting the latter values of δt_G and δT from the originally predicted optimum conditions of $t_G = 41$ min and $T = 55$°C (Fig. 13a). The ''experimental'' separation for these adjusted conditions ($t_G = 39.6$ min and $T = 53.2$°C) is shown in Fig. 13d. This separation agrees closely ($R_s = 2.5$) with that initially predicted, and the retention of all peaks in Figs. 13a and d is similar. Thus, when an experimental separation does not agree with one predicted by computer simulation, it should be possible to use the experimental run as a basis for correcting the prediction and arriving at the desired result. This effectively adds an additional experiment to method development (5 runs versus 4) as proposed here, but only for the case where the initial prediction (as in Fig. 13a) is sufficiently in error. The procedure of Fig. 13 compensates for predictive errors by a ''reflection'' of the changed conditions in Fig. 13c so as to estimate the corrected conditions of Fig. 13d. The automatic correction of predictive errors in this way (by computer) and an analysis of possible limitations of this procedure are currently under investigation.

5. Conclusions

The present study has examined errors that can arise during computer simulation (e.g., current DryLab software, version 2.0), when temperature T, gradient time t_G, and/or isocratic %*B* is varied. The main source of error is usually a failure of the relationships that are assumed to describe retention as a function of %*B*, t_G or *T*; these or similar equations are assumed in other computer simulation software [14–16]. Errors in predicting resolution have received primary emphasis and can be expressed for all three variables (%*B*, t_G or *T*) in terms of equivalent errors in mobile phase composition $\delta\delta\varphi$. For complex samples containing 15–20 or more components, the average error in R_s should not exceed 0.2 units (or $\delta\delta\varphi \leq 0.001$). For simpler samples, with a smaller number of components, an

average error in R_s of as much as 0.4 units ($\delta\delta\varphi$ = 0.002) may be acceptable.

A number of conclusions can be drawn concerning both the prediction and correction of these errors $\delta\delta\varphi$ (see also the summary of Table 8):

(1) For the case where gradient separations are predicted on the basis of initial (input) gradient runs (and T is constant), the average error in resolution R_s is in most cases <0.2 units for all samples (i.e., acceptable). For gradient predictions where T also varies, the average error in resolution is also acceptable for most samples.

(2) Exceptions to No. 1 have been noted for the case of samples (a) composed of molecules of quite different shape (varying markedly in length-to-width ratio and/or planar versus three-dimensional molecules, and (b) containing partially-ionized acids or bases. For such samples, retention as a function of temperature is less predictable, and predictive errors can be unacceptably large when a wide range in T ($\Delta T > 20$°C) is explored. In this case, the use of input runs for three or more temperatures can increase predictive accuracy and/or the range in T for which predictions are possible.

(3) For predictions of isocratic retention from isocratic input runs, acceptable accuracy is possible if the range in predicted values of %B is held within limits; e.g., input runs of 40 and 55% B allow predictions for 35–60% B. The range in %B for reliable predictability can be extended by the use of a third input run; e.g., runs with 35, 50 and 65% B, for prediction of 30–70% B. For less-retained samples that require mobile phases of $<30\%$ B for acceptable retention, the predictive range must be reduced; e.g., 5–15% B for two input runs, or 5–25% B for three input runs. As a result, it is advisable to use three input runs for predictions of separation where $\%B < 30\%$. Complex samples with $n > 20$ will be predicted less reliably, but such samples normally require gradient elution.

(4) Isocratic predictions from gradient input runs (with T constant) are less reliable, and this situation may not improve much when a greater number (>2) of input runs used. Acceptable accuracy may be found for some simple samples, but this cannot be guaranteed.

(5) Rules have been deduced that will assure acceptable predictive accuracy for those cases where accurate prediction is possible (Table 8). Alternatively, it is possible to improve predictive accuracy by carrying out an additional separation which is intended to correct for initial errors in computer simulation (as in Fig. 13).

The present study was carried out with DryLab for Windows version 2.0. Conclusions drawn from this investigation are being incorporated into future versions of this software.

6. Symbols

See Glossary of terms section in Part I [17]. Initials J.W.D., L.R.S., R.G.W., P.F., T.B., etc. refer to the various authors of the present paper.

Acknowledgements

The present study, including work described in the previous two papers [17,18], was supported in part by a Small Business Innovation Research (SBIR) grant to LC Resources from the National Institutes of Health (US Department of Health and Human Services). Dr. John Dorsey of Florida State University provided helpful assistance. Certain commercial equipment, instruments or materials are identified in this report. Such identification does not imply recommendation or endorsement by any of the authors or their affiliates, nor is it implied that such equipment or materials are necessarily the best available for the purpose.

Appendix A. Acceptable predictive errors $\delta\delta\varphi$

From Eq. (1),

$$\delta\varphi = -[\log(k''/k)]/S \quad \text{(A.1)}$$

Here, k'' is the predicted (i.e., to some extent in error) value of k, and k refers to the true value. Thus, if the original value of φ were changed to $\varphi + \delta\varphi$, the predicted value of k would then be obtained ex-

perimentally. Similarly, for the value of φ when the band leaves the column in gradient retention, the error in the predicted value of φ is

$$\delta\varphi = \delta t_R(\Delta\varphi/t_G) \quad (A.2)$$

where δt_R is the error in predicted retention time t_R, and $\Delta\varphi$ is the change in φ during the gradient. It can also be shown that these errors $\delta\varphi$ in either isocratic or gradient elution refer to the same errors δk; i.e., $\delta k(isocratic) = \delta k^*(gradient)$, if δk is the same in corresponding isocratic and gradient separations [11].

Errors in Δt_R for peaks i and j, as a result of errors in individual values of t_R, can be expressed as $\delta\delta\varphi = (\delta\varphi)_j - (\delta\varphi)_i$. Errors in predicted values of resolution R_s as a result of errors in Δt_R (values of $\delta\delta\varphi$) can be related to errors in the separation factor α via Eq. (1):

$$\delta \log \alpha = S\delta\delta\varphi \quad (A.3)$$

The relationship between errors $\delta\delta\varphi$ and corresponding errors in resolution δR_s can be derived as follows. Resolution is given (p. 27 of Ref. [9]) as

$$R_s = (1/4)(\alpha - 1)\, N^{1/2}(k/[1+k]) \quad (A.4)$$

and an error in α ($\delta\alpha$) can be related to an error $\delta\delta\varphi$ as follows. For small values of α,

$$\delta(\log \alpha) = (1/2.3)\,\delta(\alpha - 1) = (1/2.3)\,\delta\alpha \quad (A.5)$$

Combining Eqs. (A.3) and (A.5) then gives

$$\delta\delta\varphi = (1/2.3\,S)\delta\alpha \quad (A.6)$$

The error in resolution δR_s is (Eqs. (1)–(4))

$$\delta R_s = (1/4)[\delta(\alpha-1)]N^{1/2}(k/[1+k]) = (1/4)\delta\alpha N^{1/2}(k/[1+k]) \quad (A.7)$$

so that Eqs. (A.6) and (A.7) give

$$\delta R_s = (1/4)N^{1/2}(k/[1+k])(2.3S\delta\delta\varphi) = 0.58N^{1/2}(k/[1+k])S\delta\delta\varphi \quad (A.8)$$

For an average value of S (small molecules, [30]) equal to 4.2, Eq. (A.8) becomes

$$\delta R_s = 2.4N^{1/2}(k/[1+k])\delta\delta\varphi \quad (A.9)$$

If we assume an average value of $N = 10\,000$ and $k = 3$, then

$$\delta R_s \approx 180\delta\delta\varphi \quad (A.10)$$

If an average error of 0.2 in R_s is acceptable for complex samples, then the maximum average value of $\delta\delta\varphi$ is 0.001, corresponding to an average value of $\delta\alpha = 0.01$. These estimates of maximum average error ($\delta\delta\varphi[a] \leq 0.001$) include the *combined* variation of T and t_G from the change of *either* of these two variables. The allowable error $\delta\delta\varphi$ in computer simulation varies inversely with $N^{1/2}$. Therefore, the extension of computer simulation to higher-efficiency separations by CEC will require still smaller errors $\delta\delta\varphi$ in predictions of sample retention.

Values of S generally increase with sample molecular mass according to the approximate relationship [31]

$$S = 0.48M_r^{0.44} \quad (A.11)$$

which suggests $S \cong 10$ for a sample molecular mass of 1000, with a corresponding maximum error $\delta\delta\varphi = \pm 0.0005$ units. Eq. (A.11) has been obtained from data for peptides, proteins and polystyrene fractions, eluted by water–acetonitrile gradients.

Appendix B. Errors $\delta\delta\varphi$ for samples whose retention can be described by Eq. (2)

It is assumed that k is given as a function of ET(30) by

$$\log k = a + b\,ET(30) \quad (2)$$

and $ET(30)$ can be expressed as (cf. Eq. (7))

$$ET(30) = a' + b'\varphi + c'\varphi^2 \quad (B.1)$$

Therefore,

$$\log k = a + ba' + bb'\varphi + bc'\varphi^2 \quad (B.2)$$

To simplify the following analysis, assume that input runs are carried out for φ-values of 0 and 1, giving experimental values k_1, $\varphi_1 = 0$ (run 1) and k_2, $\varphi_2 = 1$

(run 2). From Eq. (B.2), $\log k_1 = a + ba'$ and $\log k_2 = a + ba' + bb' + bc'$. A predicted value of $k(k_p)$ for a mobile phase composition φ is given by Eq. (1) as

$$\log k_p = \log k_w - S\varphi \quad \text{(B.3)}$$

From the values of $\log k$ for for values of $\varphi = 0$ and 1, we then calculate $\log k_w = a + ba'$ and $S = -bb' - bc'$. Let the true value of k for mobile phase φ be ''k''. The error in $\log k$ is then

$$\delta \log k = \log k_p - \log k \quad \text{(B.4)}$$

The value of $\log k$ is given by Eq. (B.2), and $\log k_p$ can be calculated from Eq. (1), where $\log k_w = a + ba''$ and $S = -bb' - bc'$, or

$$\delta \log k = bc'(\varphi^2 - \varphi) \quad \text{(B.5)}$$

and (Eq. (1))

$$\delta\varphi = (\delta \log k)/S = bc'(\varphi^2 - \varphi)/(-bb' - bc')$$
$$= (\varphi - \varphi^2)/[(b'/c') + 1] \quad \text{(B.6)}$$

The quantities b' and c' are constants (equal to -18.1 and 10.3, Eq. (7)), independent of the solute. For a given value of φ the retention error $\delta\varphi$ will therefore be the same for all solutes in the sample, so that $\delta\delta\varphi = 0$. This conclusion can be seen to be general for all input values of φ, since the choice of specific values $\varphi_1 = 0$ and $\varphi_2 = 1$ in the present example is arbitrary. Numerical simulations for several cases also confirmed this that this conclusion is general. That is, for solutes whose retention is described by Eq. (2), retention errors $\delta\varphi$ for a given value of φ will be the same for all solutes, and therefore resolution errors $\delta\delta\varphi$ (equal to the difference in $\delta\varphi$ values for two adjacent solutes) will equal zero.

References

[1] P.L. Zhu, L.R. Snyder, J.W. Dolan, N.M. Djordjevic, D.W. Hill, L.C. Sander, T.J. Waeghe, J. Chromatogr. A 756 (1996) 21.

[2] J.W. Dolan, L.R. Snyder, N.M. Djordjevic, D.W. Hill, D.C. Locke, D.L. Saunders, L. Van Heukelem, T.J. Waeghe, J. Chromatogr. A 803 (1998) 1.

[3] J.W. Dolan, L.R. Snyder, D.L. Saunders, L. Van Heukelem, J. Chromatogr. A 803 (1998) 33.

[4] P.L. Zhu, J.W. Dolan, L.R. Snyder, D.W. Hill, L. Van Heukelem, T.J. Waeghe, J. Chromatogr. A. 756 (1996) 51.

[5] P.L. Zhu, J.W. Dolan, L.R. Snyder, N.M. Djordjevic, D.W. Hill, J.-T. Lin, L.C. Sander, L. Van Heukelem, J. Chromatogr. A 756 (1996) 63.

[6] L.R. Snyder, J.W. Dolan, I. Molnar, N.M. Djordjevic, LC·GC 15 (1997) 136.

[7] I. Molnar, L.R. Snyder, J.W. Dolan, LC·GC Int. 11 (1998) 374.

[8] L.R. Snyder, J.W. Dolan, Chem. Anal. (Warsaw) 43 (1998) 495.

[9] L.R. Snyder, J.L. Glajch, J.J. Kirkland, Practical Hplc Method Development, 2nd ed, Wiley-Interscience, New York, 1997, 470–475.

[10] B.F.D. Ghrist, B.S. Cooperman, L.R. Snyder, J. Chromatogr. 459 (1989) 1.

[11] L.R. Snyder, J.W. Dolan, Adv. Chromatogr. 38 (1998) 115.

[12] M.A. Quarry, R.L. Grob, L.R. Snyder, J. Chromatogr. 285 (1984) 1.

[13] M.A. Quarry, R.L. Grob, L.R. Snyder, J. Chromatogr. 285 (1984) 19.

[14] D.E. Bautz, J.W. Dolan, W.D. Raddatz, L.R. Snyder, Anal. Chem. 62 (1990) 1561.

[15] N. Lundell, K. Markides, J. Chromatogr. 639 (1993) 117.

[16] L.R. Snyder, J.L. Glajch, J.J. Kirkland, Practical Hplc Method Development, 2nd ed, Wiley-Interscience, New York, 1997, Ch. 10.

[17] J.W. Dolan, L.R. Snyder, N.M. Djordjevic, D.W. Hill, T.J. Waeghe, J. Chromatogr. A 857 (1999) 1.

[18] J.W. Dolan, L.R. Snyder, N.M. Djordjevic, D.W. Hill, T.J. Waeghe, J. Chromatogr. A 857 (1999) 21.

[19] M.A. Quarry, R.L. Grob, L.R. Snyder, Anal. Chem. 58 (1986) 907.

[20] M.A. Quarry, R.L. Grob, L.R. Snyder, J.W. Dolan, M.P. Rigney, J. Chromatogr. 384 (1987) 163.

[21] L.R. Snyder, M.A. Quarry, J. Liq. Chromatogr. 10 (1987) 1789.

[22] T. Culley, J.W. Dolan, LC·GC 13 (1995) 456.

[23] K. Valko, L.R. Snyder, J.L. Glajch, J. Chromatogr. 656 (1993) 501.

[24] B.P. Johnson, M.G. Khaledi, J.G. Dorsey, Anal. Chem. 58 (1986) 2354.

[25] P.J. Schoenmakers, H.A.H. Billiet, L. de Galan, J. Chromatogr. 185 (1979) 179.

[26] H. Lamparczyk, P.K. Zarzycki, J. Nowakowska, J. Chromatogr. A 668 (1994) 413.

[27] P.K. Zarzycki, M. Wierzbowska, H. Lamparczyk, J. Pharm. Biomed. Anal. 14 (1996) 1305.

[28] J. Li, Anal. Chim. Acta 369 (1998) 21.

[29] J.A. Lewis, D.C. Lommen, W.D. Radatz, J.W. Dolan, L.R. Snyder, I. Molnar, J. Chromatogr. 592 (1992) 183.

[30] L.R. Snyder, J.W. Dolan, J. Chromatogr. A 721 (1996) 1.

[31] L.R. Snyder, M.A. Stadalius, Cs. Horváth (Ed.), High-Performance Liquid Chromatography – Advances and Perspectives, Vol. 4, Academic Press, New York, 1986, 242.