ELSEVIER

Journal of Chromatography A, 672 (1994) 67–85

JOURNAL OF CHROMATOGRAPHY A

Computational strategy for solvent strength optimization in reversed-phase liquid chromatography

P. Chaminade*, A. Baillet, D. Ferrier

Laboratoire de Chimie Analytique III, Faculté de Pharmacie, 1 Avenue Jean-Baptiste Clément, F-92296 Chatenay-Malabry Cedex, France

(First received November 23rd, 1993; revised manuscript received January 28th, 1994)

Abstract

Computer algorithms for iterative solvent strength optimization are presented together with a practical development strategy consistent with other widely accepted schemes. This optimization strategy appears to be flexible enough to make use of all the experiments realized during the mobile phase development. The computer algorithm accepts any case of solute elution, during or after the gradient rise for linear or multi-linear gradients, as input or output parameters. The computer routine is able to compute solute retention from two gradients differing in their slopes and also slight changes in solvent composition. The iterative computation of the linear model allows a further minimization of the error in prediction. If some additional experiments are provided, the retention model can be extended to a more accurate quadratic form. From its computation, which uses the Nedler and Mead simplex controlled by an error reduction criterion, this model allows a further improvement in retention time prediction. This permits low prediction errors in the case of difficult developments that need several chromatographic runs.

1. Introduction

In the last 10 years, particular attention has been paid to the use of computer-assisted method development in liquid chromatography. Numerous approaches of great help for chromatographers have been published, recently summarized by Tchapla [1]. However, some drawbacks have been pointed out, particularly the rigidity of the procedure and the possible "overkilling" of easy separations.

In reversed-phase high-performance liquid chromatography (RP-HPLC), the primary goal is to optimize the solvent strength in order to achieve a good elution within a reasonable analysis time. When the technique is to be used routinely and one or more compounds are to be determined, isocratic elution is the preferred method as column equilibration between each run is not needed. In this case, if the required analysis time and/or resolution are not met, the mobile phase development can be followed by a selectivity optimization with a change of the organic modifier and, further, the use of ternary and quaternay mobile phases.

The use of a particular organic modifier may be preferred to improve solute detection (hyperchromic effect in UV detection or organic modifier limit of oxidation in electrochemical detection) or to ensure solute stability (peptides

* Corresponding author.

0021-9673/94/$07.00
SSDI 0021-9673(94)00099-U

analysis). In those cases, the only method is to perform a gradient elution to achieve the elution of both polar and non-polar solutes within a reasonable analysis time. The development of binary isocratic mobile phases is advantageous as only one parameter (the organic content of the mobile phase) is to be set and the solute retention *vs.* mobile phase composition can be conveniently depicted by a graphical procedure. The advantages of computer-aided development become clear for gradient profile optimization as the chromatographer must simultaneously consider at least three parameters: the initial and final composition and the gradient slope.

However, the common drawback of the software involved is that it only solves the retention equations which depict a particular kind of initial experiment (*e.g.*, unique gradient, pair of linear gradients, pair of isocratic experiments). Further, the software is unable to take advantage of subsequent experiments if their corresponding experimental conditions do not match the required input.

Our purpose was to write an adaptable computer program usable as a specific tool designed for the chromatographer. This program is based on the two main strategies proceeding from linear gradients but may also use isocratic experiments and/or multilinear gradients.

2. Theoretical background

In RP-HPLC with binary mobile phases, the solute capacity factor (k') is accurately described by a quadratic relationship involving the volume fraction (ϕ) of organic modifier [2]:

$$\ln k' = A\phi^2 + B\phi + C \tag{1}$$

As three experiments are needed to compute the parameters A, B and C of the model, a simplified equation is widely preferred:

$$\ln k' = \ln k_w - S\phi \tag{2}$$

where k_w is the hypothetical value of the solute capacity factor for an aqueous mobile phase and S is the slope of the relationship. However, this model often fails to describe the solute retention for wide variations of the organic content of the mobile phase.

In the specific case of linear gradient elution, this last relationship allows the algebraic development of the fundamental equation of gradient elution [3]:

$$\int_0^{V_g} \frac{\delta V}{V_a} = 1 \tag{3}$$

where V_g is the retention volume under gradient conditions and V_a is the retention volume under isocratic elution. This equation can be expressed as a function of time and solute capacity factor:

$$\int_0^{t'_g} \frac{\delta t}{k'_{(\phi)}} = t_0 \tag{4}$$

where t_0 is the column dead time, t'_g is the net retention time under gradient conditions and $k'_{(\phi)}$ is the solute capacity factor expressed as a function of the volume fraction of organic modifier ϕ and therefore as a function of time since ϕ varies with time under gradient conditions. Consequently, by solving Eq. 4 for S and $\ln k_w$, it becomes possible to obtain the solute retention time for any composition of the mobile phase [4–8].

Many workers have pointed out the usefulness of gradient-based development procedures [9–11]: first, as the sample is subjected to wide variations of mobile phase composition, the polar and less polar compounds of the sample can be measured on the same chromatogram; and moreover, linear gradients allow the solute band width to remain nearly constant for the overall chromatogram [9,12].

From a general point of view, the initial gradient experiment is designed to appreciate the sample complexity. Consequently, this first gradient will preferably cover a wide range of elution strength. However, gradients which meet these requirements are also subjected to "non-ideal" processes generated by the gradient equipment or the retention mechanism in gradient elution itself. An extensive review of these phenomena was published by Quarry *et al.* [13,14].

An important equipment-related cause of non-ideality that contributes greatly to the error in prediction is the gradient delay time, t_D. The delay time is generated by the time needed by the mobile phase to flow through the gradient mixer and tubing. However, t_D can be easily accounted for by considering an isocratic elution step before the actual start of the gradient (see Section 2.1). The gradient mixer also generates gradient profile distortions due to dispersion of the mobile phase in the mixer itself [13].

Chromatographic-related causes of error are quantitatively more important: the flow-rate during the gradient run can vary by as much as 5% [15,16], mainly because of volume contraction generated by the mixing of the aqueous and organic fractions of the mobile phase [17]; the column dead volume also varies during the gradient, owing to changes in spatial conformation of the alkyl chains [18] and uptake of the organic component of the mobile phase by the stationary phase [19]; this last phenomenon (solvent uptake) was also implicated in solute retention by Quarry *et al.* [14]; and non-equilibrium between the stationary and mobile phases during the gradient run is encountered with short gradients, which limits the calculation of isocratic retention times from gradient elution as this equilibrium is the basis of isocratic elution.

Two main strategies have been suggested for gradient-based mobile phase development: (1) the calculation of parameters S and $\ln k_w$ from two gradient runs with different slopes proposed by Quarry *et al.* [5] and Heinish and co-workers [6–8] permits one to solve the two equation–two unknowns system only in the specific case where the solute elution occurs before the end of the gradient rise; and (2) the strategy put forward by Schoenmakers *et al.* [4] based on a further correlation which links together the parameters S and $\ln k_w$ [2]:

$$S = p + q \ln k_w \quad (5)$$

where p (intercept) and q (slope) were obtained by regression analysis based on isocratic measurements. As Eq. 2 can be expressed as a function of only one parameter, both S and $\ln k_w$ can be computed from a single gradient run.

The application fields and the accuracy of these two methods are different:

(i) The simultaneous estimate of S and $\ln k_w$ requires a prior knowledge of the p and q constants from Eq. 5. These two constants depend on the organic modifier, the stationary phase [2] and also the chemical structure of the solute [20]. As these constants are the result of a linear regression, these reflect the average behaviour of a set of solutes. Consequently, even in the most favourable case where all the solutes to be separated are structurally related, it is possible that at least one of them will exhibit a particular retention behaviour. This limits the present strategy to be a rapid procedure designed to provide a rough estimate of the first- and last-eluting peak retention times.

(ii) In the case of the dual gradients-based strategy, the necessity for solute elution before the end of the gradient rise may require several experiments to be performed before this condition is met. Nevertheless, this calculation method is independent of the organic modifier in use, the stationary phase and the solute structure and, therefore, is suitable for a simultaneous estimate of analysis time and resolution.

The common drawback of the preceding methods is that they are non-iterative. After the initial gradient or pair of gradients has been applied, the proposed mobile phase or gradient profile must be adjusted without any assistance from the computer program. The computer program we propose was designed to perform an iterative solvent strength optimization. The basic requirement was to accept data from any kind of experiment, *i.e.*, including both isocratic and linear and multi-linear gradients, although a linear gradient is the preferred initial experiment. Consequently, a numerical integration of Eq. 4 is used to facilitate the calculation of retention times both in isocratic and gradient conditions. The algorithm used to compute the values of S and $\ln k_w$ from one experiment consists in a classical bisection method [21]. The bisection method is a simple and robust algorithm for root finding. When solving a function

such as $y = f(x)$ the principle of this algorithm is to bracket the zero value of $f(x)$ using lower and upper values of x, x_a and x_b, that lead to $f\langle x_a\rangle < 0$ and $f(x_b) > 0$. The next step, that will be repeated until completion, will be to test the value of $x_c = (x_a + x_b)/2$. If, for example, $f(x_c) < 0$, the value of x_a will be replaced by x_c until the zero value of $f(x)$ is approximated with sufficient precision The linear model (2) and Eq. (5) allow one to express the capacity factor as a function of only one parameter:

$$\ln k' = \frac{S(1 - q\phi) - p}{q} \tag{6}$$

As soon as two experiments are available, the values of parameters S and $\ln k_w$ are computed using a numerical method. A Monte Carlo-based algorithm permits one to take into account gradients with differences in initial and/or final compositions and also in gradient slope. When more than two experiments are performed, the Nedler and Mead simplex [22] is used to compute the linear model by χ^2 fitting. Further, when a sufficient number of experiments are performed, the linear model is extended to a quadratic relationship by adding the term $A\phi^2$ to Eq. 2. The parameters A, S and $\ln k_w$ are fitted to data using the same Nedler and Mead simplex algorithm [22].

2.1. Calculation of retention times

Retention times under isocratic or gradient conditions are calculated by numerical integration of a rearranged form of Eq. 4:

$$\frac{1}{t_0}\int_0^{t_g} \frac{\delta t}{k'_{(\phi_{i,i+1})}} = 1 \tag{7}$$

This calculation method assumes that a gradient step is similar to a sequence of short isocratic steps of increasing solvent strength. This calculation is similar to the algorithm proposed by Tomellini *et al.* [23] but uses a dynamic calculation of the step duration.

Typically three cases must be taken into account in a linear gradient: solute elution can occur during (1) the gradient delay time, this case being identical with isocratic elution arising in the initial mobile phase of the gradient; (2) the gradient rise, where the contribution of the previous isocratic step must also be accounted for, especially for polar solutes; and (3) after the gradient end, corresponding to a late isocratic elution in the final mobile phase after the two preceding steps.

Isocratic elution

An isocratic step between times t_i and t_{i+1} contributes to the integration of the fundamental equation for a quantity x given by

$$\frac{1}{t_0}\int_{t_i}^{t_{i+1}} \frac{\delta t}{k'_{(\phi_{i,i+1})}} = x$$

$$\frac{1}{t_0}\cdot\frac{t_{i+1} - t_i}{k'_{(\phi_{i,i+1})}} = x \tag{8}$$

Gradient elution

A gradient step is considered as a sequence of short isocratic steps. The number of isocratic steps n is calculated as a function of the gradient slope b to be $n = 100b$. Consequently, n is greater for a short gradient than for a gradient with a low slope. The contribution of each short isocratic step is calculated using Eq. 8. The instantaneous capacity factor is averaged between its value in the current and the next step. The whole equation for a complete gradient step is

$$\frac{1}{t_0}\sum_{j=1}^{j=n-1} \frac{t_{j+1} - t_j}{\dfrac{k'_{(\phi_j)} + k'_{(\phi_{j+1})}}{2}} = x \tag{9}$$

The contributions of each isocratic (and therefore gradient) step are summed in addition to the time intervals of each step. The integration is complete when the sum of contributions (denoted Σx) exceeds unity. The excess time Δt can then be calculated according to the following equation derived from Eq. (8):

$$\Delta t = \left[\left(\sum x\right) - 1\right] k'_{(\phi_{i,i+1})} t_0 \tag{10}$$

This excess time must then be subtracted from

the sum of the time intervals to obtain the net retention time.

In all the preceding equations, the capacity factor can be expressed as a log–linear (Eq. 2) or log–quadratic (Eq. 1) function of the volume fraction of organic modifier ϕ.

2.2. Numerical solution for a pair of gradients

The algorithm used to compute the two equations describing the two experiments is based on a Monte Carlo simulation [21]. The basis of this technique is to submit an equation to a set of random numbers until the solution is reached with reasonable precision. In our case, the two experimental runs are described by the numerical integration of Eq. 4 that uses Eq. 2 to express the solute capacity factor for a given composition of the mobile phase. Solving this system of two equations that include two parameters (S and $\ln k_w$) by searching for which pair of random numbers satisfies both equations would be easy to program, but would also lead to excessive calculation times.

Our algorithm consists in a few steps designed to minimize the calculation time:

(i) Rather than solving the two equations simultaneously, each of them is solved separately and the common solutions of S and $\ln k_w$ are then computed. For each equation, the result of the Monte Carlo simulation is a set of S and $\ln k_w$ values. By plotting S as a function of $\ln k_w$, the two sets of solutions lead to two straight lines. The common solution, *i.e.*, the values of S and $\ln k_w$ which satisfy both equations, are at the intersection point of these two lines.

(ii) The random values of S and $\ln k_w$ are enclosed within boundaries to avoid testing the lowest or highest values of S and $\ln k_w$. The algorithm assumes that the "true" values of S and $\ln k_w$ are near by the values of the estimate S_e and $\ln k_{w_e}$ computed for each run using the correlation $S = p + q \ln k_w$ (Eq. 5). Consequently, for each equation, the variation interval is:

(1) run 1:

$S : [0.50S_{e1} \ldots 1.50S_{e1}]$

$\ln k_w : [0.50 \ln k_{w_{e1}} \ldots 1.50 \ln kw_{e1}]$

(2) run 2:

$S : [0.50S_{e2} \ldots 1.50S_{e2}]$

$\ln k_w : [0.50 \ln k_{w_{e2}} \ldots 1.50 \ln k_{w_{e2}}]$

Once the values of S and $\ln k_w$ satisfying the two equations have been calculated, the validity of the solution is checked as follows:

(i) As considered previously, the values of the estimate S_e and $\ln k_{w_e}$ are assumed to be close to the true solution. For example, if the value of S lies outside the interval $[0.50S_{e1} \ldots 1.50S_{e1}]$ and $[0.50S_{e2} \ldots 1.50S_{e2}]$, this solution is rejected. In this case, it is considered that the two experiments are incompatible and do not admit a common solution. This last point will be discussed later.

(ii) A second Monte Carlo simulation computes which variation of the parameters S and $\ln k_w$ still lets them verify the two equations with a 1% precision. This last step allows one to appreciate the precision of the calculation of S and $\ln k_w$ for each solute.

2.3. Iterative calculation and linear model extension

The simplex algorithm [21,22] is used to perform the iterative calculation of S and $\ln k_w$ and to extend the linear model to a quadratic form. The simplex rules will not be discussed as this algorithm, first designed for non-linear regression, is a well known and widely used optimization procedure [24–27].

Although its convergence is slow, this algorithm is a powerful means of iterative calculation as the criterion used to fit the model (noted χ^2) is based on the reduction of the squared difference between experimental and calculated values. Starting from the values of S and $\ln k_w$ obtained by Monte Carlo simulation, this algorithm is used to compute the linear model from more than two experiments.

When three or more experiments are available, the linear model can be extended to Eq. 2. The quadratic extension is given by

$$\ln k' = A\phi^2 - S\phi + \ln k_w \qquad (11)$$

Considering Eq. (11), we prefer to call this model "extended linear" to avoid confusion with the true quadratic model. In the case of Eq. 1, the three parameters A, B and C are computed simultaneously from three or more isocratic measurements. In the case of the extended linear model, the starting equation is Eq. 2: S and $\ln k_w$ are equal to the values calculated in the preceding step and the parameter A is initially set to zero. The use of the simplex algorithm allows one to perform smooth variations of A, S and $\ln k_w$ enclosed within boundaries. These boundaries are of particular importance as they ensure that the simplex will smoothly extend the linear model to approach the flexibility of the quadratic model and to improve the calculation accuracy by the mean of the χ^2 reduction. It should be noted that the values of the linear model are kept by the program if the calculation accuracy cannot be improved by adding this $A\phi^2$ term.

3. Experimental

The programming and calculations were carried out with a Chronosoft 486-33 IBM-AT compatible microcomputer with built-in math coprocessor. The computer program is written in Pascal using Borland's Turbo Pascal. The cluster analysis was performed using the MVSP statistical package from W.L. Kovach (Kovach Computing Services, Pentraeth, UK).

Measurements were carried out with a Spectra-Physics XR 8700 ternary gradient liquid chromatograph equipped with a Rheodyne injection valve with a 10-μl sample loop (Spectra-Physics, Les Ulis, France). Detection was performed with a Shimadzu SPD-2A UV detector (Touzat et Matignon, Vitry sur Seine, France). The chromatograms were recorded with a Spectra-Physics ChromJet integrator or a Kontron PC-Integration pack Rev. 3.90 (Kontron Instruments, St. Quentin Yvelines, France). The flow-rate was set at 1 ml min^{-1}. The column dead time was measured by injection of a 50 mg l^{-1} solution of sodium nitrate (Merck, Nogent sur Marne, France) diluted in the mobile phase. The gradient delay time (t_D) was measured by the observation of the rise in the baseline while running a gradient from a 100% methanol to a methanol–0.2% dimethyl ketone mobile phase. The experimental value of t_D was 2.90 min for a flow-rate set at 1 ml min^{-1} for the part of the study concerning phenolic antioxidants and 3.40 min (flow-rate 1 ml min^{-1}) for benzodiazepines as an on-line filter was added to the pumping device.

Benzodiazepine standards were obtained by courtesy of Professor Farinotti from the toxicological department of X. Bichat C. Bernard Hospital (Paris, France). The hydrolysis procedure was adapted from the method of Maurer and Pfleger [28] used to identify 1,4- and 1,5-benzodiazepines in urine by gas chromatography–mass spectrometry: 1.0 ml of a stock solution of each compound at a concentration of 1 mg ml^{-1} in methanol was evaporated under a stream of nitrogen at ambient temperature. The residue was then refluxed with 10 ml of 37% hydrochloric acid (Merck) for 30 min and then neutralized with 90 ml of 1 M Na_2CO_3 (Merck) solution in acetonitrile–water (50:50). The volume was adjusted at 100 ml with acetonitrile–water (50:50) after the excess of CO_2 had been removed. The resulting solution containing 10 mg l^{-1} of each standard was injected directly into the chromatograph. Peak identification was done using solutions of each hydrolysed benzodiazepine prepared at the same level of 10 mg l^{-1}. UV detection was performed at 230 nm. The column was Sup.Rs (250 × 4.6 mm I.D.) packed with LiChrospher C_{18} with a particle size of 5 μm (Prolabo, Paris, France). The experimental valuc of the column dead time was 2.10 min for a flow-rate set at 1 ml min^{-1}. Acetonitrile (Merck) and water (Baker) were of HPLC quality.

Phenolic antioxidants were obtained as pure compounds from Merck. Standard solutions of approximately 0.1 mg ml^{-1} of each compound were prepared in the mobile phase (or initial mobile phase in the case of gradient elution). Methanol of HPLC gradient quality was purchased from Prolabo. Water was glass-distilled. Acetic acid of analytical-reagent grade was obtained from Prolabo. UV detection was per-

formed at 230 nm. The column was SFCC Spherisorb ODS-2 (150 × 4.6 mm I.D.) with a particle size of 3 μm (SFCC–Shandon, Eragny, France). The experimental value of the column dead time was 1.12 min for a flow-rate set at 1 ml min^{-1}.

4. Results and discussion

4.1. Iterative solvent strength optimization

The two main strategies for solvent strength optimization were described in section 2. They are summarized and compared with the methodology we propose in Fig. 1. Our strategy is consistent with both of the two preceding approaches but uses an iterative calculation algorithm. This feature is illustrated hereafter by the separation of a complex mixture of a hydrolysis product of benzodiazepines presented in Table 1. The aim of this separation was to investigate how HPLC could be used as a screening procedure before a gas chromatographic–mass spectrometric analysis.

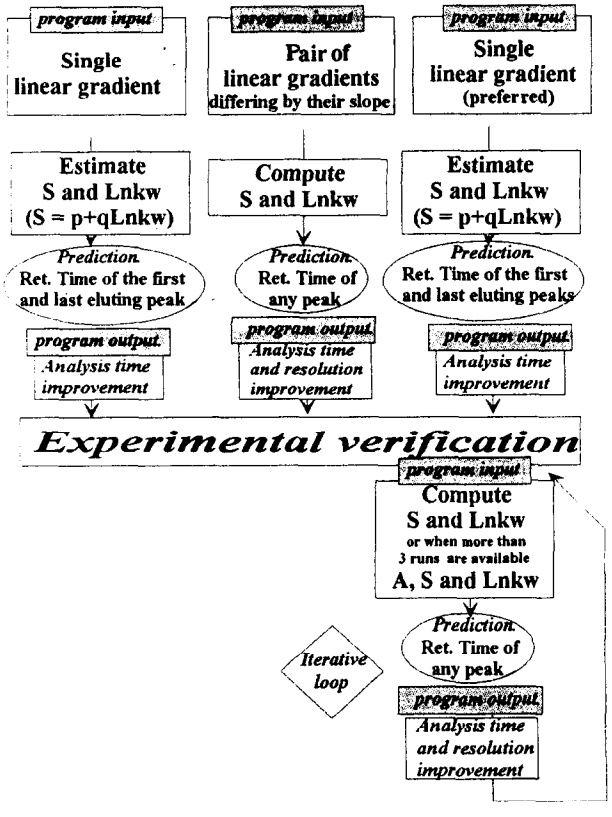

Fig. 1. Schematic representation of the optimization procedure.

Iterative solvent strength optimization using the linear model

The methodology used by Schoenmakers *et al.* [4] uses only one gradient to compute an estimate of the S and $\ln k_w$. The two gradients-based method uses a pair of gradient runs with the same initial and final compositions but a difference in slope to calculate S and $\ln k_w$. In our approach, we use the results from the first gradient to estimate S and $\ln k_w$ and then to compute the conditions of the second gradient. The aim of this step is to try to adjust the gradient profile in order to maximize the peak separation. To illustrate this step, two gradient pairs were designed in order to compare the mean (accuracy) and the standard deviation (precision), σ, of the error in prediction obtained with these two data sets: two gradients A and B, ranging from 10 to 100% of acetonitrile with a gradient duration of 30 and 90 min, respectively, that are consistent with the experiments needed with the two gradient based method; the gradient A and a gradient B′ computed from gradient A, using the one gradient-based estimation of S and $\ln k_w$ (gradient B′ = 30–70% acetonitrile in 40 min).

The retention times of benzodiazepines are summarized for the different gradient profiles in Table 1. The accuracy and precision of the calculated retention times for a further gradient denoted C are presented in Table 2. The elution parameters S and $\ln k_w$ were computed by the Monte Carlo simulation algorithm using either the AB or AB′ pair of gradients. Both the gradient pairs AB and AB′ lead to an overestimate of the solute retention times. The average error ($\bar{m}$) induced by the computation based on gradients A and B is 9.16% with a standard deviation of 8.48%. This average error was found to be significantly higher than that observed for gradients A and B′ ($\bar{m}$ = 3.36%, σ =

Table 1
Retention times and gradient profiles for the hydrolysis products of benzodiazepines

No.	Parent compound	t_R (min)					
		A	B	B′	C	D	E
1	Tetrazepam	19.33	36.24	17.16	11.66	17.32	14.86
2	Triazolam	19.93	38.10	18.39	12.10	19.70	16.26
3	Tofisopam (1)	20.20	38.56	18.93	13.01	20.36	16.98
4	Medazepam	20.85	39.97	20.27	14.91	22.13	18.93
5	Tofisopam (2)	21.03	40.77	20.97	15.87	23.15	20.20
6	Bromazepam	22.31	42.70	23.16	20.08	25.25	23.07
7	Tofisopam (3)	22.95	45.84	25.89	26.49	28.72	26.13
8	Nitrazepam	23.32	46.65	26.70	28.55	29.42	26.65
9	Clonazepam (2)	24.04	49.04	28.97	33.23	31.82	28.59
10	Midazolam	24.75	48.23	28.23	30.59	30.76	27.73
11	Flunitrazepam	25.38	52.32	32.22	37.20	34.97	31.14
12	Flurazepam	25.93	53.57	33.49	38.42	36.11	32.04
13	Nordazepam	26.54	54.71	34.60	39.44	37.18	32.86
14	Tetrazepam	27.80	62.03	41.87	45.69	43.83	37.99
15	Medazepam	29.58	63.15	43.12	46.35	44.73	38.79
16	Diazepam	30.13	64.22	44.04	47.61	45.80	39.66
Initial % and duration		10% to 100% at 30 min	10% to 100% at 90 min	30% to 70% at 40 min	40% for 22 min to 70% at 40 min	35% for 10 min to 70% at 40 min	30% to 40% at 10 min to 40% at 15 min to 70% at 30 min

Table 2
Percentage error in retention time prediction for gradients C, D and E using the linear (L) or extended-linear (EL) model

No.[a]	Data set and target									
	AB C	ABC D	ABCD E	ABC D	ABCD E	AB′ C	AB′C D	AB′CD E	AB′C D	AB′CD E
1	10.98	−0.12	0.94	−0.12	2.56	7.29	−0.12	0.87	−0.29	0.94
2	22.81	−8.38	4.12	−8.63	4.12	15.87	−8.38	1.48	−8.38	1.48
3	19.83	−5.84	2.00	−16.80	3.83	13.37	−5.84	1.47	−7.17	−0.47
4	20.59	−4.52	−2.43	−5.11	−2.43	11.40	−4.52	1.80	−6.37	1.43
5	20.04	−5.36	−3.81	−5.36	−3.86	10.14	−6.00	1.24	−10.45	0.69
6	16.88	−3.41	−0.69	−4.87	−1.30	5.38	−7.68	−0.26	−4.75	−0.74
7	11.10	−3.24	−0.80	−3.17	−0.80	0.60	−3.24	−0.80	−3.24	−0.80
8	8.06	−2.21	0.41	−1.90	0.26	−0.56	−2.21	0.41	−2.21	0.38
9	3.04	−1.10	1.61	−1.16	1.47	−1.81	−1.10	0.80	−1.10	0.77
10	8.47	−1.72	2.88	−1.66	2.13	−1.67	−1.72	1.55	−1.72	1.55
11	1.45	−0.63	0.71	−0.66	0.67	−1.56	−0.63	0.29	−1.06	0.29
12	1.28	−0.47	0.47	−0.55	0.44	−1.33	−0.47	0.19	−0.50	0.16
13	1.24	−0.40	0.43	−0.75	0.33	−1.39	−0.40	0.18	−1.40	0.18
14	−0.33	−0.55	−3.13	−0.52	−3.13	−0.72	−1.05	−3.13	−1.05	−3.13
15	0.82	0.18	−0.46	0.40	−0.46	−0.09	−0.40	−0.72	−0.40	−0.77
16	0.25	0.20	0.38	0.33	0.38	−1.09	−0.90	0.20	−0.87	−1.94
$\bar{m}$	9.16	−2.35	0.16	−3.16	0.26	3.37	−2.79	0.35	−3.19	0.00
σ	8.48	2.55	2.08	4.44	2.28	6.21	2.80	1.22	3.23	1.28
Model	L	L	L	EL	EL	L	L	L	EL	EL

[a] See Table 1 for compound names.

6.21%) while the difference between the two standard deviations was not significant.

From these two pairs of gradients, it appears that by reducing the polarity range of the second gradient, the accuracy of retention time calculation increases while the precision remains about the same.

When computing a new gradient (D) on the basis of gradients ABC and AB'C, by Monte Carlo simulation followed by a simplex refinement of S and $\ln k_w$, the average error in prediction and the standard deviation are lowered to: $\bar{m} = -2.35\%$ and $\sigma = 2.55\%$ for ABC and $\bar{m} = 2.79\%$ and $\sigma = 2.80\%$ for AB'C. Considering the error in prediction obtained from AB and ABC, both the mean and the standard deviation were found to be significantly different. The same result is observed using AB' and AB'C. This shows the increase in accuracy and precision obtained with the iterative procedure when the results from a third solvent programme are entered as data. It is also noticeable that the two data sets ABC and AB'C now induce the same accuracy and precision.

Further, when computing the last gradient E (Fig. 2), on the basis of the previous four, the mean calculation error is lowered to 0.16% with a standard deviation of 2.1% for ABCD and 0.35% with a standard deviation of 1.20% for the AB'CD set.

This illustrates the need for an iterative strategy in solvent strength optimization: the use of two gradient programmes that differ in the slope and the initial and final compositions leads to a more efficient calculation of S and $\ln k_w$; and the iterative procedure allows the accuracy and precision of the calculated retention times to be increased.

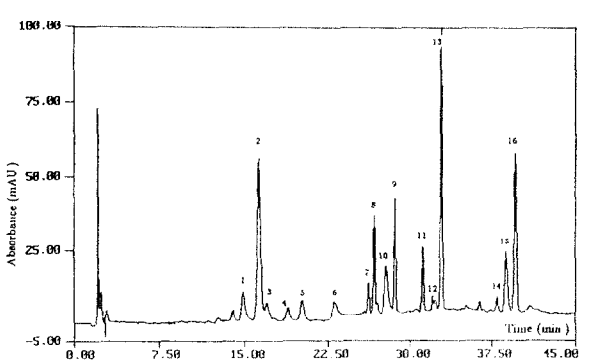

Fig. 2. Chromatogram of the hydrolysis products of benzodiazepines. See Table 1 for peak identification and gradient (E) conditions.

Linear model extension

All the preceding calculations were made using the linear model (Eq. 2). When at least three experiments are available, it appears possible to extended this linear model by adding an $A\phi^2$ term that allows a more adequate description of the solute retention behaviour. The results obtained with the model we call "extended linear" are presented in Table 2.

Calculation from four chromatographic runs. When performing the calculations using the four preceding experiments (data sets ABCD and AB'CD), the accuracy and precision of the extended linear model are now comparable to those obtained from the linear model.

Using the ABCD data set, the mean error and standard deviation of the predicted retention times obtained with the extended linear model are not significantly different from those computed with the linear model. The results obtained with the AB'CD data set show that the error in prediction tend to zero using the extended linear model, with a precision that is comparable to that for the linear model (linear model, $\bar{m} = 0.35$, $\sigma = 1.22$; extended linear model, $\bar{m} = 0.01$, $\sigma = 1.28$).

Two main conclusions can be drawn on comparing the results obtained from the two models: (i) Four experiments are needed for an effective computation of this model. Consequently, one must be conscious that we do not recommend performing four initial experiments to allow the computation of the extended linear model, but rather we recommend this computational method in iterative solvent strength optimization when four experiments are available. (ii) The differences in accuracy and precision observed between the two original data sets AB and AB' tend to vanish on adding some additional experiments. This shows that, even if one experiment

does not provide some adequate data, and leads to the adoption of an erroneous mobile phase composition and/or gradient profile for the following experiment, this inaccuracy may be corrected by the results of this last experiment itself. Considering this point, this method appears more rugged than non-iterative procedures.

Factors influencing retention time prediction

In our example, the error in prediction, initially high, decreases with increasing number of experiments involved in the data set. Two points must be considered to try to explain this initial error and to verify the influence of the input gradient profile on accuracy improvement: the experimentally determined gradient delay may not agree with its actual value; and the prediction accuracy may proceed from the similarity between experiments rather than from the iterative calculation procedure.

Prediction accuracy and experiments. Marengo *et al.* [29] reported that, when calculating retention times under gradient conditions from isocratic data, the gradient steepness directly influences the error in prediction. For a series of simulations recorded with the same column and organic modifier, the rate of variation of the mobile phase composition is mainly responsible for the error in prediction. In our example, gradients C, D and E look similar as the optimization process tends to lead to a multilinear gradient. Considering this point, one may believe that this similarity between the input and calculated conditions is mainly responsible for the improvement in accuracy and precision when supplementary runs are added. To elucidate this point, we decided to compare results obtained with the data sets ABC and AB'C with those obtained with ABB', ABD and AB'D for the calculation of gradient E using the linear model (Eq. 2). Among these gradients, B' and D are the more closely related to gradient E. Consequently, if this hypothesis of similarity of gradients is true, a set of experiments containing those two gradients (B' and D) would be assumed to lead to superior calculation accuracy and precision. Despite this, the AB'D data set did not provide better calculation results than the other four. The mean errors in prediction ranged from −1.28 and 2.16 for all these sets and the standard deviation ranged from 1.69 to 2.91. Significant differences in accuracy were found for data sets ABD and AB'D with respect to ABC and AB'C. The most accurate and precise results were obtained with the set ABB' ($\bar{m} = 0.37$, $\sigma = 1.69$), but significant differences in standard deviation were observed only for ABD ($\bar{m} = 2.16$, $\sigma = 2.67$) and AB'D ($\bar{m} = 1.94$, $\sigma = 2.91$). The mean errors in prediction for these two sets were also found to be significantly different from those for sets ABC ($\bar{m} = -0.99$, $\sigma = 1.98$) and AB'C ($\bar{m} = -1.28$, $\sigma = 2.13$), as the retention times are underestimated using ABC and AB'C and overestimated with ABD or AB'D. As stated previously, the best accuracy and precision observed with data set ABB' tend to confirm that the prediction accuracy and precision of this computational method are more closely related to differences in gradient profiles than to similarity between the input and predicted values.

Gradient delay time. The technique used to determine the gradient delay influences the calculation accuracy. In this study, the system dwell time is determined experimentally by adding a UV tracer (dimethyl ketone) to the mobile phase. However, according to some workers [8,30] and despite of the reliability of this method, a calculated value of t_D leads to superior accuracy in the prediction of retention times. This calculation [8] uses the results from three gradients of decreasing slope. The system dwell time is calculated by comparing the solute retention time recorded with the gradient of intermediate slope with the values predicted from the calculations based on the two remaining gradients. From the data provided [8], values of t_D calculated in this way are about 17–25% higher than those determined by the use of a UV tracer. In our study, the gradient delay was estimated to be 3.40 min. If we consider Fig. 3, which depicts the average error in prediction as a function of the gradient delay, the most appropriate value for t_D is 4.0 min. The use of this value, which is about 20% higher than the experimental t_D,

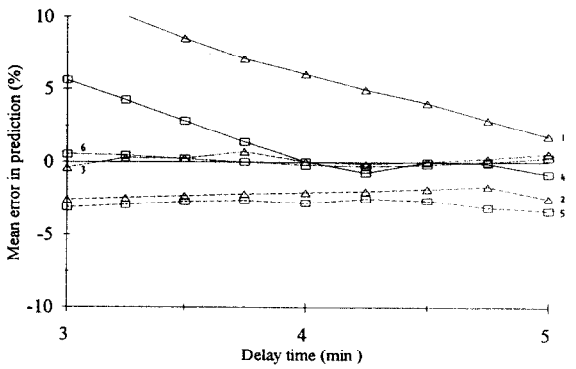

Fig. 3. Percentage error in retention time prediction as a function of the gradient delay time. 1 = AB; 2 = ABC; 3 = ABCD; 4 = AB′; 5 = AB′C; 6 = AB′CD.

lowers the error in prediction from 3.37% to nearly 0% for the AB′ pair and 9.16 to 6% for AB. The influence of the t_D value is less important when three or more experiments are included in the calculations of the linear model. Concerning the sets ABC, AB′C, ABCD and AB′CD, the prediction accuracy remains nearly constant for a large range of t_D values.

This illustrates the importance of using an iterative procedure in solvent strength optimization. The error-minimization algorithm, used for the calculation of elution parameters, is necessary as it allows one to compensate for the effect of partially accounted for parameters such as gradient delay.

4.2. Choice of initial gradient programmes

If we consider the error in prediction at each step of the preceding example, the most critical step is the computation of the retention times from a pair of gradients (the one run-based calculation is only used to compute a rough estimate of the retention times of the first- and last-eluting peaks in order to obtain the second gradient profile). It has been found that the use of a second gradient differing in slope and solvent composition may lead to a threefold decrease in the error in prediction compared with a gradient programme that differs only in the slope. The actual problem is to appreciate which, and which kind of, difference is allowable, that is, should the main difference preferably concern the gradient slope or the solvent composition? To investigate this point, we designed five gradient programmes denoted A–E that differ in slope and/or composition. The calculation results from each pair of gradients were compared with isocratic data. Six isocratic runs were performed with volume fractions of organic modifier ranging from 0.65 to 0.90. All these gradients were chosen with a high slope in order to maximize the errors in prediction and thus to obtain significant differences between the different pairs of gradients.

The compounds used for this part of the study were eight phenolic antioxidants or preservatives of pharmaceutical and/or cosmetic interest (Table 3). These compounds were chosen as the

Table 3
List of phenolic solutes

Compound	Common name	Abbreviation
2-Butyl-4-hydroxyanisole		BHA
2-*tert.*-butylphenol		TBP
2-*tert.*-butyl-4-methylphenol		TBMP
2,6-Di-*tert.*-butyl-4-methylphenol	Butylated hydroxytolene	BHT
3,4,5-Trihydroxybenzoic acid propyl ester	Propyl gallate	PG
3,4,5-Trihydroxybenzoic acid octyl ester	Octyl gallate	OG
3,4,5-Trihydroxybenzoic acid dodecyl ester	Dodecyl gallate	DG
p-Hydroxybenzoic acid propyl ester		PHBP

resulting set of solutes exhibit a wide range of polarity, as seen from their isocratic retention times (see Table 4). Methanol was used as organic modifier and 1% of acetic acid was added to the aqueous fraction of the mobile phase to improve peak shape.

The results obtained with the Monte Carlo-based method of resolution show that of the ten possible pairs that could be constituted with the five gradients, only seven pairs of experiments led to acceptable values of S and $\ln k_w$. The incompatible pairs are AE, BE and CE, *i.e.*, gradients in which the greatest differences in initial mobile phase composition are encountered (10% methanol for A, B and C and 50% for E).

Calculation precision for S and ln k_w

For the seven remaining pairs of gradients, the best calculation precision for S and $\ln k_w$ is encountered with gradient pairs AC and AD, as depicted in Figs. 4 and 5. These two gradient pairs exhibit the maximum difference in slope. Gradient AC has the same initial and final compositions. The initial and final compositions of gradients A and D differ by 10% of the organic modifier.

As noted previously, if the gradient slopes are too similar, the set of two experiments describes nearly the same equations. Only a reasonable change in mobile phase composition may be allowed. An important difference in mobile phase composition will cause the two equations not to admit a single solution (the case of gradient A, B or C coupled with E), or a single solution with a low calculation precision (gradient pair DE). This can be explained as the parameter $\ln k_w$, *i.e.*, the logarithm of the hypothetical value of the capacity factor in the fully aqueous mobile phase, is extrapolated from the presumed value of the capacity factor in the initial mobile phase of the gradient elution.

Retention time prediction

The mean error and standard deviation (σ) of the error on retention time prediction for each isocratic mobile phase, presented in Table 5, were examined through cluster analysis to individualize similarities between gradient pairs. From the dendrogram presented in Fig. 6, it appears that the pairs of gradients can be associated in two main clusters: gradient pairs BC and CD in a first cluster and gradient pairs AB, AC, AD, BD and DE in a second cluster.

According to Table 5, gradient pairs BC and CD both lead to more inaccurate and less precise predictions. As expected, this cluster contains the gradient pairs with the most comparable conditions: gradients B and C share the same solvent composition and differ slightly in their slopes; gradients C and D share the same slope and differs by 10% in the initial and final organic modifier compositions.

The second cluster may be divided into smaller structures with an increasing inaccuracy and poorer precision from the centre of the cluster towards its boundaries. (i) The inner cluster is constituted by the gradient pairs AB and AD, which lead to the best compromise between accuracy and precision. As shown previously, these gradient pairs also induce the best calculation precision for the parameters S and $\ln k_w$. (ii) The next cluster is surrounded by a larger one corresponding to the gradient pair BD, that leads to more accurate but also less precise retention time prediction than the previous two. This gradient pair is similar to AD, except that the difference in gradient slope is less important. (iii) The following cluster includes the gradient AC, which is the gradient pair with the same initial and final compositions that exhibits the maximum difference in slope. (iv) The outer cluster involves the gradient pair DE, which differs only in the initial and final compositions and yields a large underestimate of the solute retention times associated with a poor precision.

From this analysis, the most accurate results are obtained with gradients differing at least in their slope, and also by a limited difference in their solvent compositions. When comparing two pairs of gradients with the same difference in slopes, the pair of gradients that also differ in a limited difference in composition yield the best results. The most inaccurate results were obtained with gradients with similar slopes or too large differences in solvent compositions.

Table 4
Retention times of phenolic compounds under gradient and isocratic elution (average column dead time 1.12 min)

Compound[a]	A	B	C	D	E	$\phi = 0.65$	$\phi = 0.70$	$\phi = 0.75$	$\phi = 0.80$	$\phi = 0.85$	$\phi = 0.90$
PG	15.39	18.75	20.94	17.30	4.04	2.22	2.03	1.88	1.76	2.04	1.90
PHBP	19.16	24.66	28.74	24.73	9.03	3.81	3.05	2.60	2.26	2.30	1.90
BHA	20.63	27.05	32.65	27.93	11.98	5.66	4.19	3.33	2.70	2.53	2.05
TBP	21.45	28.31	34.57	29.67	13.78	7.40	5.26	4.00	3.09	2.72	2.17
TBMP	22.10	29.77	36.66	31.70	15.94	10.37	6.94	4.98	3.65	2.74	2.17
OG	22.36	30.28	37.58	32.64	17.00	14.26	9.36	5.75	3.76	2.84	2.35
BHT	25.56	34.78	43.47	40.35	23.26	–	31.62	18.02	10.18	6.12	3.77
DG	25.78	35.09	44.07	42.55	23.92	–	–	–	13.18	6.59	3.95
Composition											
Initial (%)	10.0	10.0	10.0	20.0	50.0						
Final (%)	90.0	90.0	90.0	80.0	90.0						
Duration (min)	20.0	30.0	40.0	30.0	20.0						
Slope (% min^{-1})	4.0	2.67	2.0	2.0	2.0						

[a] See Table 3 for compound names.

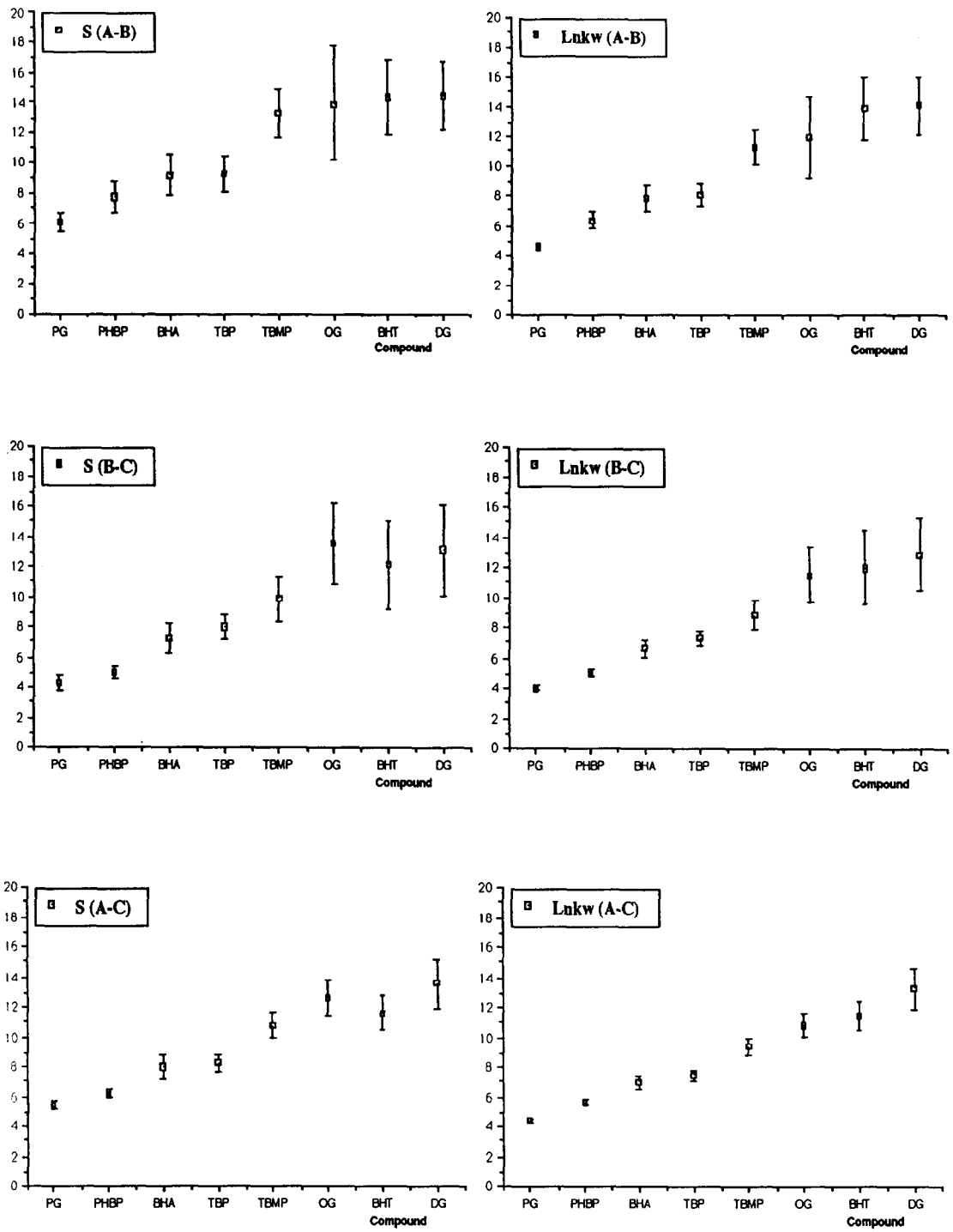

Fig. 4. Plots of S and $\ln k_w$ with calculated standard errors. Gradients of various initial and final composition. For abbreviations, see Table 3.

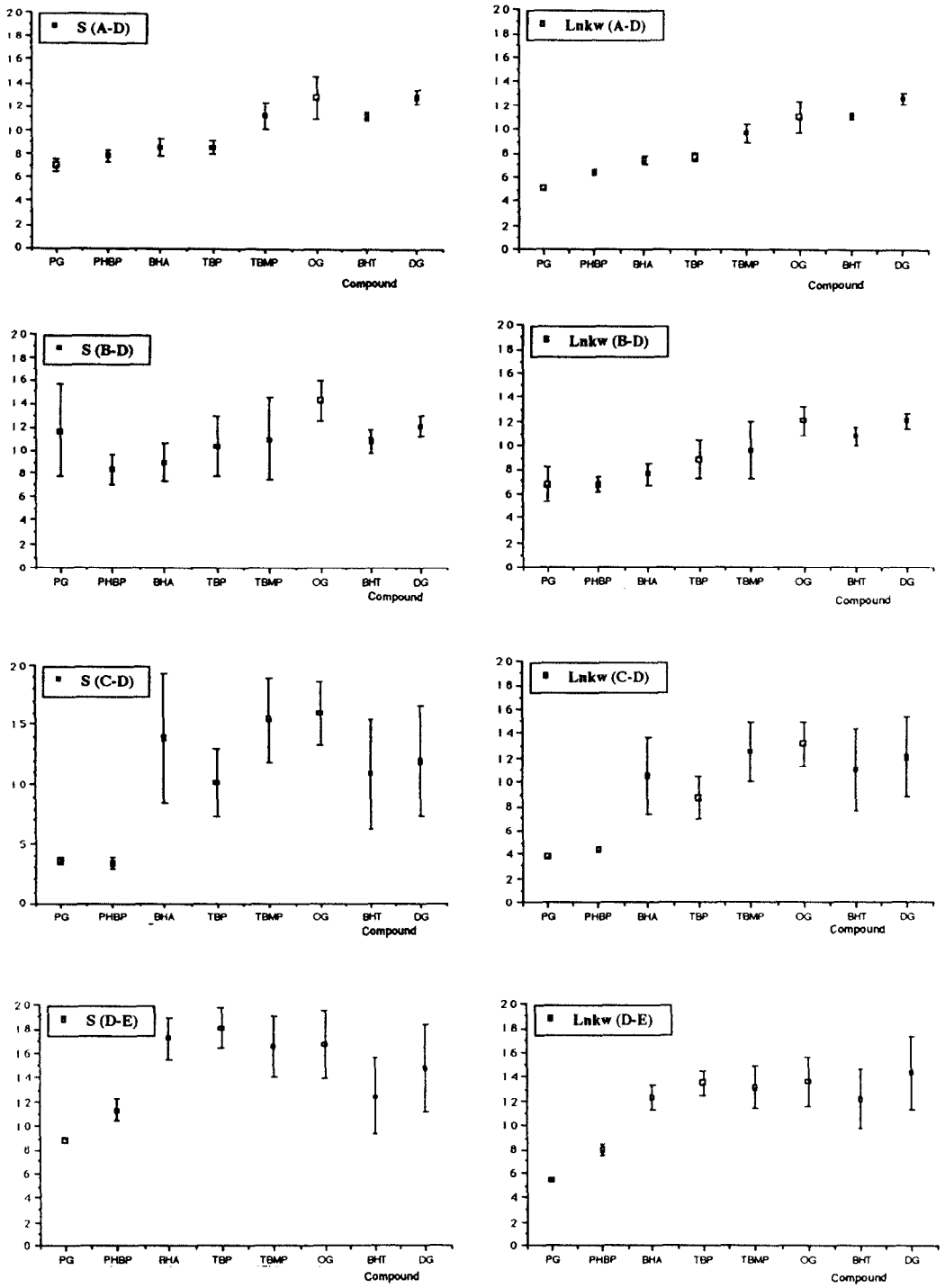

Fig. 5. Plots of S and $\ln k_w$ with calculated standard errors. Gradients of various initial and final composition. For abbreviations, see Table 3.

Table 5
Percentage error in isocratic retention times prediction using the Monte Carlo-based calculation of the parameters S and $\ln k_w$ from a pair of gradients

Gradient	$\phi = 0.65$		$\phi = 0.70$		$\phi = 0.75$		$\phi = 0.80$		$\phi = 0.85$		$\phi = 0.90$		$\bar{m}$	σ		
	$\bar{m}$	σ	$\bar{m}$	σ	$\bar{m}$	σ	$\bar{m}$	σ	$\bar{m}$	σ	$\bar{m}$	σ			$\bar{m}$	σ
AB	36.6		27.4		16.8		8.4		−4.0		−6.7		13.6	17.2		
		10.0		11.8		9.9		9.2		10.5		10.4			10.3	0.9
AC	40.4		28.8		21.8		13.9		2.8		1.7		18.3	15.1		
		20.3		22.7		18.0		14.9		7.9		11.6			15.9	5.5
AD	31.2		18.6		10.6		3.9		−5.3		−5.8		8.9	14.4		
		11.3		15.2		10.6		8.6		12.3		14.3			12.1	2.4
BC	59.9		47.6		43.7		34.8		23.2		23.6		38.8	14.4		
		44.5		47.2		40.3		34.7		18.8		22.3			34.6	11.8
BD	22.8		13.5		7.1		2.2		−5.3		−5.5		5.8	11.1		
		21.2		20.2		16.3		12.9		17.6		17.5			17.6	2.9
CD	72.2		62.0		59.1		48.4		33.5		31.8		51.2	16.2		
		58.3		56.1		48.6		43.7		23.3		22.7			42.1	15.7
DE	−9.9		−22.3		−28.2		−25.4		−25.6		−21.1		−22.1	6.5		
		21.1		16.7		18.7		22.6		30.3		31.2			23.4	6.0
$\bar{m}$	36.2		25.1		18.7		12.3		2.7		2.6					
σ	26.5		26.8		27.9		23.9		19.8		18.6					
$\bar{m}$		27.6		28.9		24.0		20.7		15.1		16.4				
σ		19.5		18.3		16.4		14.8		5.8		5.2				

$\bar{m}$ = mean of the percentage error; σ = corresponding standard deviation.

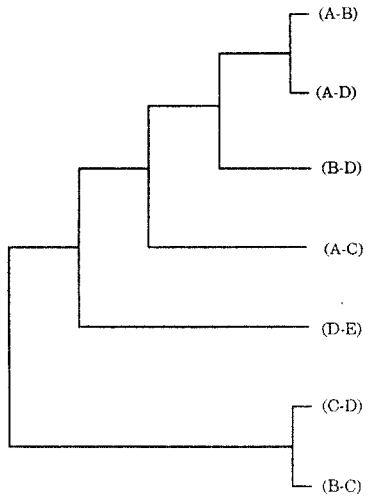

Fig. 6. Cluster analysis of results in Table 5.

5. Conclusions

A schematic presentation of our iterative algorithm is depicted in Fig. 1. This optimization strategy appears flexible enough to make use of all the experiments realized during the mobile phase development. The computer algorithm accepts any case of solute elution, during or after the gradient rise for linear or multi-linear gradients. The initial experiments must be selected with care to ensure the maximum accuracy from the beginning of the development process. As with previously reported algorithms, the computer routine discussed in this study is able to compute solute retention from two gradients differing in their slopes. However, from our examples, an improvement in the accuracy of retention time prediction is observed when the two gradients also vary by a limited difference in solvent composition. In this case, this improvement may be attributed to the greater independence between the parameters of the two gradients leading to a more pertinent solution. This may be explained by the meaning of the linear model, which, in turn, is a tangent to the quadratic relationship that links the solute capacity factor and the volume fraction of the organic modifier. Consequently, for the same solute, the linear model will lead to different values of S and $\ln k_w$ depending on the investigated solvent compositions. In the case of gradient experiments, this difference will also exist depending on the starting and ending mobile phases. Computing the common solution of two gradient runs with slight differences in solvent composition led to averaged values of S and $\ln k_w$. Nevertheless, as a limit, if the two experiments are carried out with a large difference in solvent composition, the two tangents to the quadratic relationship will concern two distinct regions of the quadratic curve. Consequently, the values of the slope and the intercept will be different for the two tangents. In this case, the two equations that describe the two gradients runs will not admit a common solution. In this step, the accuracy will depend heavily on the non-ideal process generated by the solvent delivery system and the gradient delay. The iterative computation of the linear model allows one to minimize the error in prediction generated by a partial account of this parameter. Hence, if some additional experiments are needed, the retention model can be extended to a more accurate quadratic form. From its computation, which uses the Nedler and Mead simplex controlled by an error-reduction criterion, this model allows a further improvement in retention time prediction. This permits a low prediction error in the case of difficult developments that require several chromatographic runs.

References

[1] A. Tchapla, *Analusis*, 20, No. 7 (1992) 71.
[2] P.J. Schoenmakers, H.A.H. Billiet and L. de Galan, *J. Chromatogr.*, 185 (1979) 179.
[3] L.R. Snyder, *J. Chromatogr.*, 13 (1964) 415.
[4] P.J. Schoenmakers, H.A.H. Billiet and L. de Galan, *J. Chromatogr.*, 205 (1981) 13.
[5] M.A. Quarry, R.L. Grob and L.R. Snyder, *Anal. Chem.*, 58 (1986) 907.
[6] S. Heinisch, P. Rivière and J.L. Rocca, *Chromatographia*, 32 (1991) 559.
[7] S. Heinisch and J.L. Rocca, *Analusis*, 18 (1990) 83.
[8] S. Heinish, J.L. Rocca and M. Kolosky, *Chromatographia*, 29 (1990) 482.
[9] L.R. Snyder, J.W. Dolan and J.R. Gant, *J. Chromatogr.*, 165 (1979) 3.
[10] J.W. Dolan, J.R. Gant and L.R. Snyder, *J. Chromatogr.*, 165 (1979) 32.

[11] M.A. Quarry, R.L. Grob, L.R. Snyder, J.W. Dolan and M.P. Rigney, *J. Chromatogr.*, 384 (1987) 163.
[12] H. Poppe, J. Paanakker and M. Bronckorst, *J. Chromatogr.*, 204 (1981) 77.
[13] M.A. Quarry, R.L. Grob and L.R. Snyder, *J. Chromatogr.*, 285 (1984) 1.
[14] M.A. Quarry, R.L. Grob and L.R. Snyder, *J. Chromatogr.*, 285 (1984) 19.
[15] J.P. Foley, J.A. Crow, B.A. Thomas and M. Zamora, *J. Chromatogr.*, 478 (1989) 287.
[16] D.P. Herman, H.A.H. Billet and L. de Galan, *J. Chromatogr.*, 463 (1989) 1.
[17] E.D. Katz, K. Ogan and R.P.W. Scott, *J. Chromatogr.*, 352 (1986) 67.
[18] R. Rosset, M. Caude and A. Jardy, *Chromatographies en Phases Liquide et Supercritique*, Masson, Paris, 1991, pp. 327–346.
[19] R.M. McCormick and B.L. Karger, *Anal. Chem.*, 52 (1980) 2249.
[20] J.D. Baty and S. Sharp, *J. Chromatogr.*, 432 (1988) 13.
[21] W.H. Press, B.P. Flannery, S.A. Teukolsky and W.T. Vetterling, *Numerical Recipes in Pascal*, Cambridge University Press, Cambridge, 1989, pp. 274–279; 212–253; 326–330.
[22] J.A. Nedler and R. Mead, *Comput. J.*, 7 (1965) 308.
[23] S.A. Tomellini, R.A. Hartwick and H.B. Woodruff, *Anal. Chem.*, 57 (1985) 811.
[24] C. Ortigosa, *Lyon Pharm.*, 40 (1989) 315.
[25] M.W. Routh, P.A. Swartz and M.B. Denton, *Anal. Chem.*, 49 (1977) 1422.
[26] S.N. Deming and S.L. Morgan, *Anal. Chem.*, 45 (1973) 278A.
[27] F.H. Walters and S.N. Deming, *Anal. Chim. Acta*, 167 (1985) 361.
[28] H. Maurer and K. Pfleger, *J. Chromatogr.*, 222 (1981) 409.
[29] E. Marengo, C. Baiocchi, M.C. Gennaro and P. Bertolo, *Chromatographia*, 27 (1989) 19.
[30] N. Lundell, *Ph.D. Thesis*, Uppsala University, Uppsala, 1992.