Journal of Chromatography, 485 (1989) 27–50
Elsevier Science Publishers B.V., Amsterdam — Printed in The Netherlands

CHROM. 22 055

REVIEW

SELECTION OF MOBILE PHASE PARAMETERS AND THEIR OPTIMIZATION IN REVERSED-PHASE LIQUID CHROMATOGRAPHY

HUGO A. H. BILLIET* and LEO DE GALAN
Delft University of Technology, Department of Analytical Chemistry, De Vries van Heystplantsoen 2, 2628 RZ Delft (The Netherlands)

CONTENTS

1. INTRODUCTION

To a large extent, the separation of sample components by reversed-phase liquid chromatography (RPLC) and reversed-phase ion-pair liquid chromatography (RP-IPLC) is governed by the composition of the mobile phase. Although basically a mixture of water and one or two organic solvents is to be designed, the exact composition is complicated by the strong influence of the type of organic solvents, their relative influence on the selectivity and the addition of other constituents such as buffers and ion-pairing reagents[1,2].

Through the use of small particles (5 or 3 μm), plate counts of 10 000 can be reached, which provide space for about 50 peaks in a chromatographic run of 20 min[3–5]. Hypothetically, then, a 50-component sample could be baseline separated in one run, provided that one can deliberately place each component in its own slot. Unfortunately, this is not the case. Herman *et al.*[3] have shown that for randomly distributed peaks, an unrealistically high peak capacity would be needed for the complete separation of as few as fifteen peaks. Solvent selectivity optimization with binary and ternary mixtures makes much more efficient use of the available separation space and reduces the required peak capacity to more realistic proportions (see Fig. 1). Assuming that the upper limit of peak capacities for columns in use today is around 50 in the isocratic mode, Fig. 1 shows that mixtures containing more than about seven

0021-9673/89/$03.50

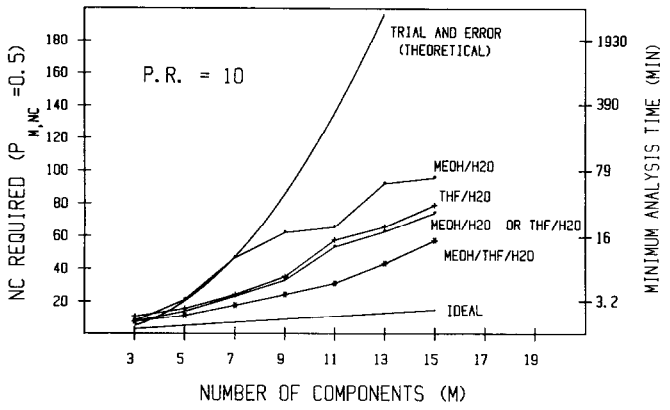

Fig. 1. Peak capacities needed for complete separation of an increasing number of solutes at the 50% probability level. The "ideal" curve refers to the situation where all the solutes are well separated, whereas the "trial and error" curve presents the random situation of no control at all[5]. The intermediate curves refer to the binary and ternary optimization in water–methanol–tetrahydrofuran. MEOH = Methanol. Peaks: 1 = benzyl alcohol; 2 = phenol; 3 = 3-phenylpropanol; 4 = 2,4-dimethylphenol; 5 = benzene; 6 = diethylphthalate. (From ref. 3 with permission.)

solutes cannot be expected to be fully separated without any selectivity optimization of the mobile phase. In contrast, in an "optimised" chromatogram, 13–15 compounds can be completely separated.

Organizing the solutes in a chromatogram by trial and error or "chromatographic intuition" can be a time-consuming experience. Attempts to develop a systematic, computer-assisted approach have been made in the past[1]. Because most useful variables (parameters) in RP- and IP-RPLC reside in the mobile phase, they are generally referred to as mobile phase optimization systems. Such an optimization consists of three distinct steps:

(i) The definition of the criterion that describes the goal of the analysis. Mostly, resolution-based criteria will be used that translate the aims of the analyst into a single numerical value.

(ii) The delineation of the parameter space, *i.e.*, the number and nature of the parameters to be considered and the range of variation. The success of all strategies depends critically on the correct selection of the parameters, the number constitutes the dimensionality and the parameter limits define the feasible area which can be used for optimization.

(iii) A logical procedure to locate the highest value of the criterion on the response surface (the variation of the criteron value over the parameter space). The procedure should search for the "global" optimum to distinguish it from secondary maxima in the often complex response surface.

It is the intention of this paper to review the optimization method developed in the authors' laboratory and to demonstrate its use in RP- and RP-IPLC.

2. CRITERIA

Given two chromatograms run under different conditions, it takes little experience for the analyst to select the better one at a glance. In contrast, it is

surprisingly difficult to formulate an algorithm that achieves the same result by computer logic. The reason is that we must strike a balance between a number of often conflicting aims, such as the separation of all constituents of a sample or a limited number of key components from their nearest neighbours (limited optimization); the best spreading of the peaks in the shortest possible analysis time, on a given column; or with the column length to be adjusted afterwards. In all instances, the selection should be suited to the goal of the analyst. These criteria can be used individually or, to combine more goals, sequentially[6]. Several mathematical expressions have been proposed, ranging from simple to complex, none being entirely satisfactory.

Generally, the primary goal of chromatographic selectivity optimization is to separate the solutes of interest from all the other (unimportant) components in the sample mixture. A direct approach is to calculate the resolution, R_s, between the peaks of interest and their nearest neighbours. The lowest value, $R_{s,\min}$, is the important one and should be at least equal to 1. This is the simplest criterion, but it suffers from severe limitations. For instance, the experimental measurement of the peak width is difficult for strongly overlapping peaks. Another factor not taken into account is the relative height or area of the two peaks. For non-Gaussian peaks, even the basic theoretical equation cannot be used[7].

The simple $R_{s,\min}$ criterion is not sufficient to characterize the extent of separation in a chromatogram containing more than two peaks. Taking the product of all the consecutive resolutions in a chromatogram makes it possible to condense the information on the separation quality into a single number. This product will be zero if any single pair of peaks is completely unresolved. The criterion also aims at an even spreading of the peaks over the total chromatogram, resulting in as much information as possible. To correct for changes in total analysis time, the criterion can be expressed as the fraction of the ideal separation, *i.e.*, equal spacing in R_s.

Many more criteria have been discussed in the literature, which means that the selection of the most suitable one is not straightforward. The sequential use of separation criteria is advised especially in the case of limited optimization, where we first look for satisfactory resolution, and thereafter for acceptable analysis time. Limited optimization can be seen as a special case of full optimization assigning weighting factors to the different solutes in the criteria. Table 1 gives an overview of suitable criteria for full and limited optimization.

Examining criteria often results in a heated debate as to which one is the best. There is no single best criterion: it is that criterion which translates the aims of the analyst in the most efficient way. A paper by Peeters *et al.*[8] describes an expert system to aid the analyst in choosing the best criterion for a particular separation problem and available types of columns.

3. PROCEDURE

Mathematical procedures for the efficient location of optimum conditions were known before their need arose in chromatography[9]. They can broadly be distinguished in three categories: grid-search methods, self-finding optimization procedures such as simplex and regression designs[10]. The authors have a clear preference for the last type.

The variation of the criterion over the parameter space (the response surface) often resembles a mountainous landscape with sharp peaks and deep valleys because

TABLE 1

OPTIMIZATION CRITERIA

Criterion	*Purpose*
$R_{s_{1,2}} = \frac{\sqrt{N}}{2} \cdot \frac{(k_2 - k_1)}{2(k_2+k_1+2)}$	Minimum resolution of worst resolved peak pair
$R_s = \prod_{i=1}^{n=1} \frac{(k_{i+1} - k_i)}{(k_{i+1} + k_i + 2)}$	Product of resolution; n is the number of peaks in the chromatogram
$r^* = \frac{\prod_{i=1}^{n=1} R_{s_{i+1,i}}}{\left[\sum_{i=0}^{n=1} R_{s_{i+1,i}}/(n-1)\right]^{n-1}}$	Product of resolution over all adjacent pairs divided by the average resolution
$r^* = \frac{\prod_{i=1}^{n} (R_{s_{i,i-1}} \quad R_{s_{i,i+1}})\ w_i/2}{\left[\sum_{i=0}^{n=1} R_{s_{i,i+1}}/p\right]^{p}}$	Product of R_s with weighting factors; n = the number of components in the sample, p = the number of peaks of interest; w_i = the weighting factor of the solute in the sample

minor shifts in the retention times of neighbouring peaks produce large changes in the resolution and consequently in the criterion value. In those cases a grid-search must be very finely spaced and, hence, requires a large number of chromatograms, whereas self-finding procedures run a severe risk of zooming in on a secondary maximum rather than the global optimum. Regresion designs, on the other hand, are free from these disadvantages, because they start from individual retention times that vary much more smoothly, so that a few well chosen chromatograms suffice to fit a simple model (linear or quadratic)[11].

The Delft iterative procedure can be split up into several distinct steps in the case of uncharged solutes and reversed-phase conditions. In this section, only a short enumeration of the steps will be given. The starting binary methanol–water composition can be derived from a standard gradient scan (0–100% methanol in 15 min)[12]. Isoeluotropic binary acetonitrile (ACN)–water and tetrahydrofuran (THF)–water compositions can be calculated using transfer rules[12,13]. Chromatograms taken at those three binary compositions constitute the start of the iterative procedure. The above-mentioned simple linear model is used to describe the solute retention behaviour over the ternary compostion. It is then simple to calculate resolution values and, hence, the optimization criterion over the parameter space. It is true that we rely on the accuracy of the model, but its prediction can be easily verified and the result fed back into the procedure. In this way, the model is updated and a decision about new experiments to be performed can be made.

There are, however, two problems associated with this approach. First, the smooth behaviour does not refer to the retention time itself, but to the natural logarithm of the capacity factor. The calculation of several hundred antilogarithms makes high demands on the computing time, so that with the current microcomputers regression desings are restricted to one or two parameter optimizations. Also, the need

to model the capacity factors of all solutes in the sample makes it mandatory to locate corresponding solutes in different chromatograms, a procedure known as peak tracking.

The procedure can be stopped not only if experiments have already been run at the suggested compositions, but also if the predicted optimum is the same as it was before, or if it can be established in the calculation step that no further improvement may be expected from an additional iteration cycle. The philosophy of the design is to locate the global optimum using a minimum number of experiments and making use of the insight gained from all experimental data.

4. THE PARAMETER SPACE

A general statement can be: choose those parameters which have the largest influence on selectivity. No optimization criterion or procedure will resolve an overlapping peak pair; only the chromatographic phase system will do that. The success of the method depends heavily on the pre-knowledge of the sample constituents and on the correct selection of the parameter space formed by the type and number of parameters and their range. The number constitutes the dimensionality of the parameter space and the limits define the feasible area which can be used for optimization.

Some variables at our disposal are discontinuous, such as the column length and the type of stationary phase. For reasons of practicality, continuously variable parameters are to be prefered. Examples are the flow-rate and the column temperature, but above all the composition of the mobile phase. In RP-LC with uncharged solutes, the mobile phase is a mixture of water and one or more organic solvents such as methanol, ACN and THF. The type and concentration of the organic modifiers are of prime importance in RP–LC optimization. An accepted method for estimating the isocratic retention of any solute molecule in the three common binary eluents methanol–water, ACN–water and THF–water is first to pre-analyse the sample using an exploratory methanol–water linear gradient scan. From measured net retention times under gradient conditions, capacity factors in methanol water binary isocratic eluents can be estimated spanning an adequate capacity factor range. The corresponding ACN–water and THF–water concentrations, predicted to yield chromatograms exhibiting the same range of capacity factors, can be calculated using simple transfer rule equations.

For ternary optimization, essentially a one-parameter optimization problem, the three starting binary compositions form the basis of the phase selection diagram (see Fig. 2). If all three modifiers are to be used, a two-parameter optimization problem arises as the mixing ratio of two of the three ternary combinations can be changed independently. Basically, the procedure remains the same.

5. DISTINGUISHING IONIC STATES

Whereas it is relatively simple to select an appropriate parameter space for mixtures containing uncharged solutes, this task becomes considerably more difficult when differently charged solutes are present in a sample. The eventual selection of the parameter space must be dictated by the following considerations. First, changes in

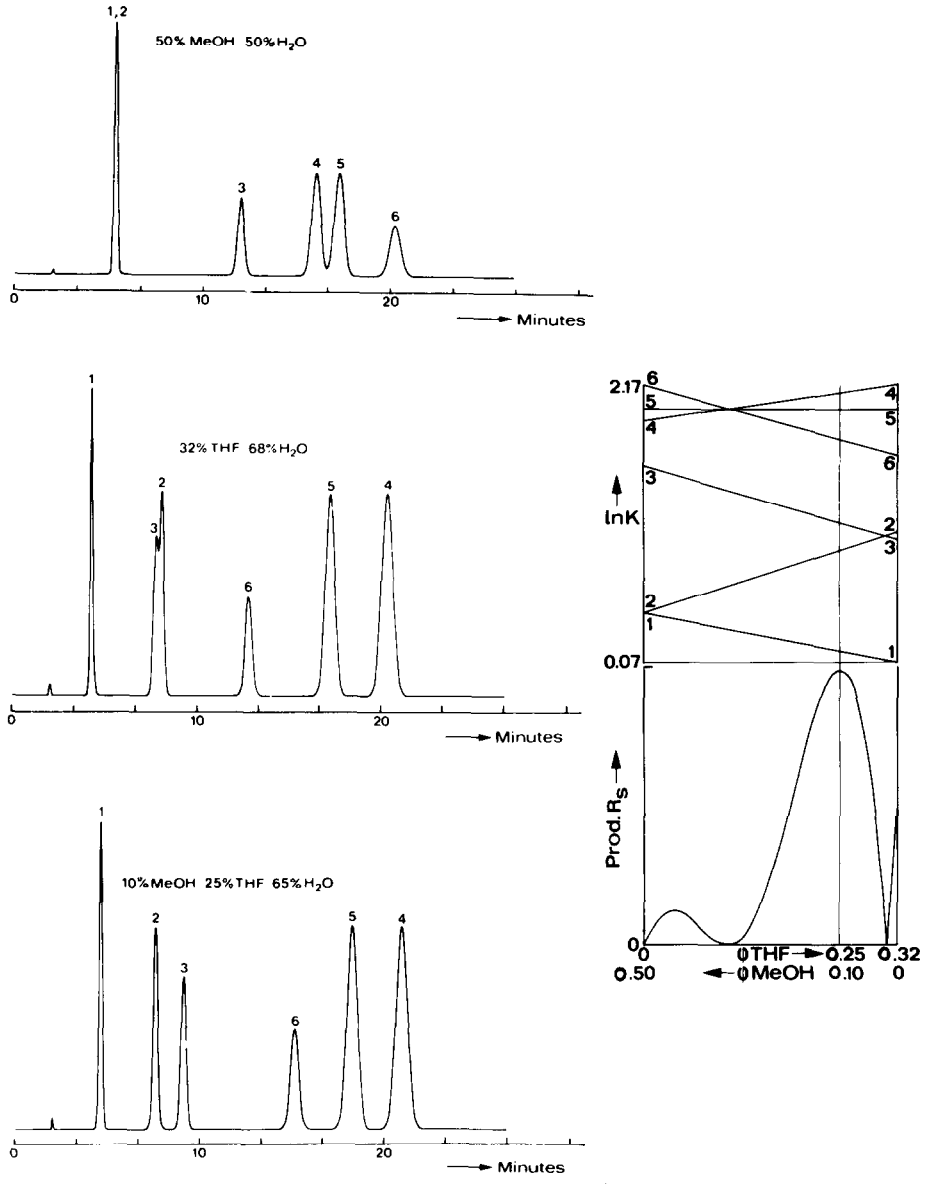

Fig. 2. Chromatograms of a sample containing six aromatic solutes in two isoeluotropic binary mobile phases. Also given is the chromatogram taken at the optimum ternary composition given by the solvent selection diagram (right). The solvent selection diagram shows the response surface of the resolution product calculated by using the linear interpolation model between the starting binaries. MeOH = Methanol. (From ref. 26 with permission.)

selectivity for charged solutes can be accomplished by a number of mobile phase parameters such as pH, ion-pair type and concentration, ionic strength in addition to organic modifier type and concentration. Obviously, it is impractical to attempt to optimize all these parameters at once for a given mixture. With the exception of the

simplex method, most of the known optimization procedures at present can involve only the simultaneous optimization of two or at most three parameters. Second, the question arises of which parameters are the more relevant, which parameters are to be selected for the formation of a parameter space to solve a particular optimization problem. The retention movement for uncharged molecules with the content of organic modifier in the mobile phase is largely predictable. This is much more difficult for retention changes of differently charged solutes in a mixture under the influence of pH, ion-pair concentration and organic modifier content.

The presence or absence of various ionization classes can be derived by observing the retention behaviour of the sample compounds in a series of four water–methanol gradients (gradients from 100% water to 100% methanol to ensure complete elution of all the compounds). The first is a gradient at pH 7.5 and the second is a similar gradient at pH 2.5. Between these two runs, the retention of neutral compounds, strong acids and strong bases will remain the same. Weak acids, however, will be fully ionized and, hence elute early at pH 7.5, whereas they are neutralized and more retained at low pH (2.5). The reverse is true for weak bases that are ionized at pH 2.5 but are neutral at pH 7.5 (see Fig. 3). The presence of weak acids and bases can thus be concluded from

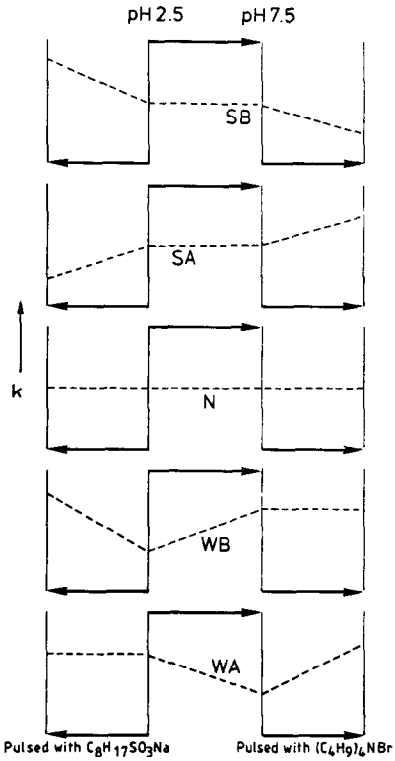

Fig. 3. Idealized classification of sample constituents according to their ionic state. The retention behaviour of (from top to bottom) strong bases, strong acids, neutral compounds, weak bases and weak acids is shown when they are subjected to water–methanol gradients at low (2.5) and high (7.5) pH with or without addition of an anionic (octane sulphonate) or a cationic (quaternary ammonium salt) ion-pairing reagent. The various retention shifts allow solute classification.

retention shifts observed between these two gradient runs.

To distinguish between neutral compounds, strong acids and strong bases, two more gradient chromatograms are run. One is at pH 2.5 with a pulse anionic ion-pair reagent injected just before the sample injection. The reagent will complex with strong bases and increase their retention, and will also speed up the hydrophobic anionic species. A final gradient at pH 7.5 with a pulse injection of a cationic pairing reagent will reveal the presence of strong acids (retarded) and hydrophobic cationic species (advanced). Fundamental to the success of this scanning strategy is the accurate tracking of the peaks in the four gradients. This can generally be achieved by using the information on peak heights, peak area, spectra and/or combinations.

What remains to be done is to select the appropriate optimization parameters once the ionic state of the sample has been determined. Several relatively simple cases have been worked out (see Table 2), but more complex combinations are still to be considered. A rule-based expert system is under development to translate the results of the gradient scans into the requirements for the parameter space.

6. THE IMPORTANCE OF THE SOLUTE NUMBER

Although chromatographic theory predicts that any separation can be achieved provided that one waits long enough, there are practical constraints. With increasing retention, the peak becomes broader and ultimately vanishes in the baseline noise. The peak capacity indicates the number of available peak positions, but we cannot spot the solutes at arbitrary positions. This "limitation" of chromatographic technology is demonstrated in Fig. 4 based on a computer simulation of many separations in either a binary methanol–water or a ternary mobile phase containing variable proportions of water–methanol and THF[3].

Each curve refers to a particular number of solutes and relates the probability of achieving a complete separation ($R_s = 1.0$) depending on the peak capacity, N_c. Clearly, some samples containing seven "easy" solutes can be separated at relatively low peak capacity (<20), whereas other samples with seven "difficult-to-separate" compounds require much larger peak capacities (>40). It is also clear from this exercise that a binary eluent (methanol–water) requires a higher peak capacity (is less selective) to achieve the same probability for a given number of solutes than a ternary

TABLE 2

PREFERRED OPTIMIZATION PARAMETERS

Sample contains[a]					*Suitable parameters*		
N	*SA*	*SB*	*WA*	*WB*	*Modifier (%)*	*pH*	*Ion-pair reagent*
×					+		
	×				+		Cationic
×		×					Anionic
	×		×			+	Cationic
×			×	×		+	Either

[a] N = neutral; SA = strong acid; SB = strong base; WA = weak acid; WB = weak base.

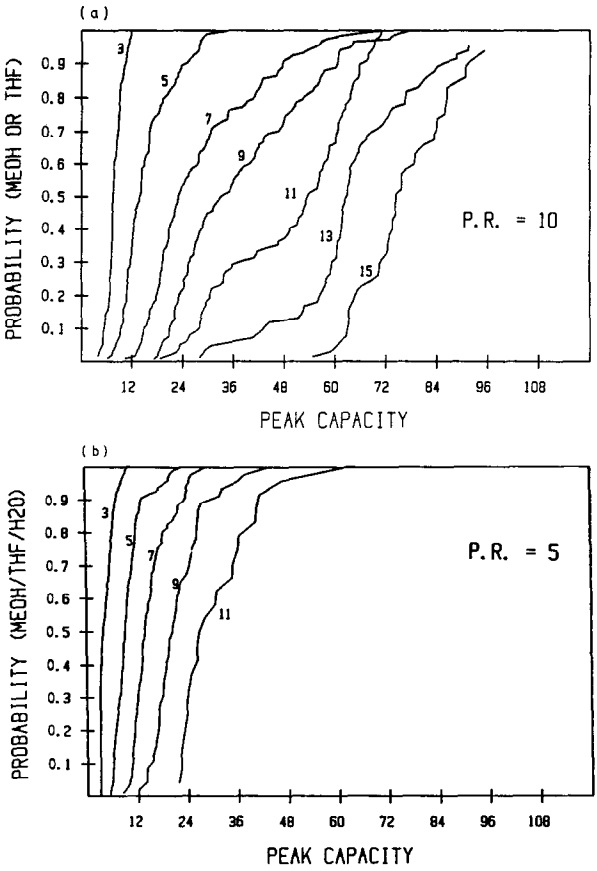

Fig. 4. Probability of successful separation. Curves referring to a certain number of solutes in a sample relate the probability of separation of all components to the peak capacity after eqn. 1 for a water–methanol or a water–tetrahydrofuran mobile phase. (b) Same curve but now showing that ternary mobile phases need less peak capacity to achieve the same probability. These curves can be used to design the starting binary methanol–water mobile phase. (From ref. 3 with permission.)

eluent (methanol–water–THF). Apparently, the addition of one more organic modifier increases the ability to "manipulate" the retention and thus the selectivity of the chromatographic system. However, in all instances, even for a small number of solutes in a sample, the necessary peak capacity is significantly larger than the number of solutes.

Of the possible ways that can be used to describe the size of the separation space in chromatography, the system peak capacity is most useful. It is defined as the maximum number of Gaussian peaks that may be placed within a specified interval such that the resolution between all successive peak profiles is constant. In isocratic liquid chromatography, the peak capacity, N_c, of the interval bounded by the capacity factors of the last- and first-eluting peaks, k_ω and k_α is given by

$$N_c \approx 1 + \frac{1}{4} \cdot \frac{\sqrt{N}}{R_s} \cdot \ln\left(\frac{1 + k_\omega}{1 + k_\alpha}\right) \quad (1)$$

where N is the plate count of the column and R_s can be set to the desired value (*e.g.*, 1 or 1.5). Apparently, the peak capacity is determined by the quality of the column and by the capacity factor range. The latter is largely determined by the composition of the sample (the polarity range). If the sample components cover a broad polarity range, they elute over a correspondingly large capacity factor range. Conversely, if the solutes are very similar, they tend to elute within a narrow capacity range, and there is little one can do about it. Although it is generally possible to position the first peak at any value of k_α, it is not simultaneously possible to retard the last peak to an arbitrarily large value of k_ω. It is therefore necessary to relate the total number of components in the sample and their polarity range to the result of eqn. 1.

The necessary data are obtained as follows. Fig. 5a represents a water–methanol gradient. There are several means of estimating the total number of solutes from this chromatogram. Simply counting the number of peaks and shoulders provides a lower

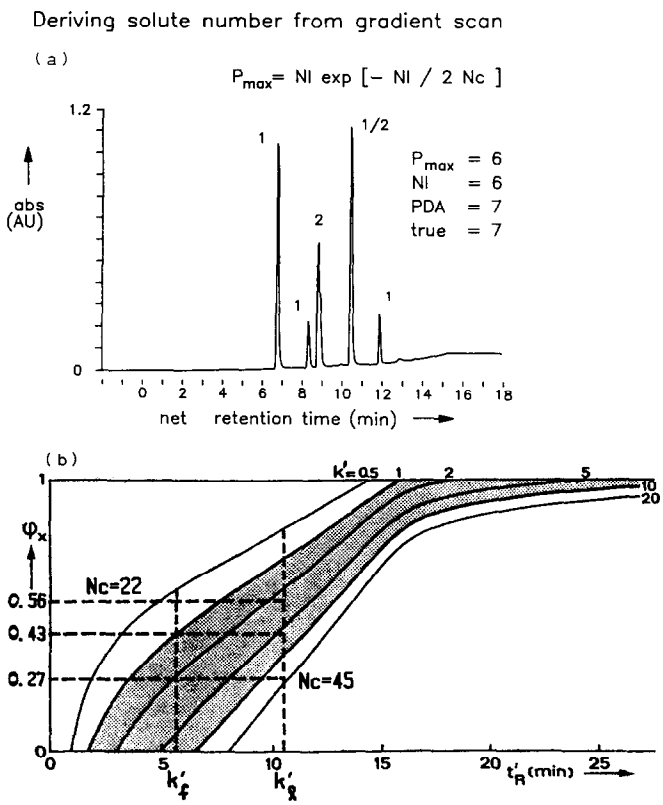

Fig. 5. (a) Chromatogram of seven solutes taken with a linear water–methanol gradient in 15 min. The number of solutes is estimated from P_{max}, the number of peak maxima; NI, a statistical correction for possible peak overlap[5]; PDA, principal component analysis of photodiode-array spectra collected over each peak. (b) Curves relating the isocratic composition, $\psi_{methanol}$, to gradient net retention time, t_R, calculated for various isocratic capacity factors. The dashed lines refer to the above sample eluting from the gradient between 6 and 11 min and the peak capacities obtained in various methanol fractions calculated with eqn. 1. (From ref. 12 with permission.)

estimate of, in this instance, six solutes. The statistical Davis–Giddings model can be used to account for peak overlaps in such a small number of solutes over the observed time frame[5]. In the present example, the theory predicts no overlap and, hence it also estimates the presence of six solutes. Another method is to examine the UV spectra taken across each peak by a photodiode-array detector[14]. Analysis shows variation in the spectra for the peak eluting at 10.6 min, which is thus suspected to be impure. A statistical technique known as principal component analysis indicates the presence of two solutes adding up to a total of seven. Actually, all three estimates are sufficiently accurate for our purposes.

The real value of the gradient scan emerges from Fig. 5b. These curves relate the gradient net retention times of each solute to capacity factors that would be observed with isocratic water–methanol binary compositions[4,12]. For example, if the sample is eluted from the same column with water–methanol (57:43), then the first solute (gradient net retention time 6 min) would elute with a capacity factor $k_\alpha = 1$, whereas the last solute would elute with a capacity factor $k_\omega = 6$. Substituting these values in eqn. 1 (with $R_s = 1$ and $N = 8000$) yields a peak capacity $N_c = 30$ under these chromatographic conditions and for this particular sample.

Apparently, suppose we had chosen a stronger eluent containing 56% methanol, all solutes elute earlier ($k_\alpha = 0.6$ and $k_\omega = 3$) so that the peak capacity is reduced to $N_c = 22$. Conversely, we can raise the peak capacity to a value as high as $N_c = 45$ by lowering the methanol content to 27%, albeit at the cost of a much longer analysis time, as the last solute now elutes at $k = 20$, which is impractically high.

All three values for the peak capacity ($N_c = 30$, 22 and 45) are much higher than the total number of solutes present in the sample (seven). Complete separation would seem easy but, unfortunately, this is not so. As can be seen from Fig. 4, a peak capacity of $N_c = 22$ (56% methanol) yields only a 30% change of having the mixture completely separated in a binary water–methanol system. At $N_c = 30$ (43% methanol), the probability is raised to 60%. A peak capacity of $N_c = 45$ is needed to raise the probability to an acceptable level of 90%, which is achieved with 27% methanol at the cost of a very long analysis time. In a ternary mixture of water–methanol–THF similar probabilities are found, but now for a larger number of solutes (eleven instead of seven).

The following chain of decisions results from the above discussion:

(i) Estimate the number of solutes in the sample from the gradient run.

(ii) Establish the peak capacity needed to achieve a certain separation probability (selected by the analyst).

(iii) Convert the peak capacity N_c to combinations of isocratic capacity factors k_α and k_ω through eqn. 1 (k_α or k_ω must be chosen by the analyst).

(iv) Use the gradient net retention times (Fig. 5a) to derive the isocratic eluent composition that yields the necessary capacity factor range (Fig. 5b).

It should be noted that we need only a single experiment, with a water–methanol standard gradient, to formulate the starting methanol–water binary eluent that should effect a separation of the sample with a certain probability and/or, in the case of failure, forms the basis of the optimization procedure including other modifiers such as ACN and THF.

7. CHANGE TO OTHER MODIFIERS

Whatever probability level is chosen, the particular sample may defeat the statistics and still remain incompletely separated in the isocratic methanol–water eluent selected. At this point and, as discussed earlier, other organic modifiers such as ACN and THF can add more selectivity to the mobile phase. The question to be answered is how we can design and optimize such complex mobile phases.

On changing from one modifier to another, we wish to retain the peak capacity, *i.e.*, the range of capacity factors over which the sample elutes. Such a transfer is provided by the isoeluotropic curves in Fig. 6. These were derived from a representative data set and adapted recently[15]. They represent the fraction of ACN and THF that will give roughly the same retention as a certain binary methanol–water fraction. For example, the curve for THF predicts that a mixture of 50% THF yields the same retention time as 70% methanol for an average solute. Obviously, and fortunately, the prediction cannot be exactly true for all solutes owing to the "selective effect", otherwise there would be no point in changing the organic modifier. A problem can arise when the predicted isoeluotropic compositions result in retention behaviour that deviates strongly from the expected one. Herman *et al.*[15] and others[12,13,16] have revised the existing transfer rules and included a correction procedure with one additional measurement.

An example is shown in Fig. 7, where we have aimed for 75% probability. Indeed, the retention ranges in the three isocratic solvents are very similar and the separation seems satisfactory. Unfortunately, a principal component analysis on the collected spectra reveals the presence of nine solutes, whereas none of the chromatograms in Fig. 7 results in nine fully separated peaks. For example, in 30.8% THF the third peak actually contains two solutes. If the same two solutes also coincide in the other two chromatograms, there is little prospect of improvement. However, if different solute pairs coelute in the three chromatograms, then a carefully selected mixture of two binaries might accomplish the desired separation. To make the decision we must once more recognize corresponding solutes in the three chromatograms.

Indeed, we have now encountered several instances where peak recognition is

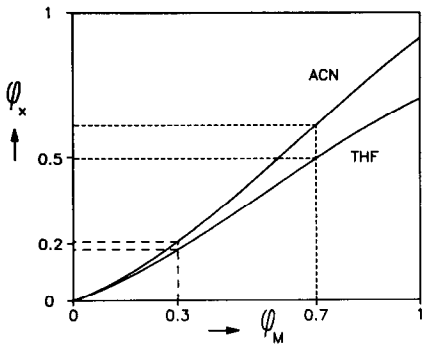

Fig. 6. Transfer curves. The two curves show the volume fractions φ_X, of acetonitrile and tetrahydrofuran that yield the same average retention time as a certain volume fraction of methanol. (From ref. 2 with permission.)

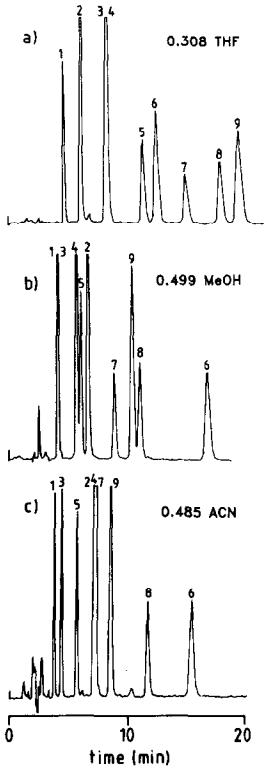

Fig. 7. Isocratic chromatograms taken with the initial binaries (49.9% methanol, 30.8% tetrahydrofuran, 48.5% acetonitrile) derived from the gradient scan and selected to yield a 75% probability of separation success. The sample contains nine aromatic solutes. Note that in all three chromatograms at least one solute pair is still unresolved. (From ref. 25 with permission.)

needed. Other optimization methods such as Sentinel and the critical band method also need some kind of peak recognition[17,18]. One unequivocal, although lengthy, procedure is to inject each sample component separately. This is only feasible in the rare situaton that all solutes are known and available.

8. METHODS CURRENTLY AVAILABLE FOR PEAK TRACKING

The simplest form of peak recognition is applied in one-dimensional detection systems using a single wavelength or refractive index detection and considering peak area as being specific for the different solutes. This method can be applied when changes in experimental conditions do not influence the detection characteristics too much. Apart from the fact that within one chromatogram several peaks can show almost the same area, and that experimental conditions do have an influence on the detection (spectral) properties peak overlap will cause severe problems as the individual peak areas will be more difficult to determine. A possible solution can be found in curve-fitting procedures with predetermined peak models or in evaluating different combinations of individual peak areas observed in other chromatograms. Both methods depend heavily on the assumed number of compounds: when severe

overlap is observed in all chromatograms of a given sample, more elaborate procedures are required. Problems related to the inherent uncertainty in the observed peak areas caused by overlap and/or experimental conditions have been approached by the "fuzzy-set" theory[19].

It is possible to increase the specificity of detection considerably by using more wavelengths and/or different, more specific, detectors. The most extreme example is the use of a variable-wavelength detector tuned to a certain wavelength that is specific for one solute and does not detect the others. This detection is unique and that particular peak is tracked unambiguously.

The use of two detectors at different wavelengths will increase the specificity considerably. The absorbance ratio over a pure peak is a constant value independent of the concentration and only dependent on the molar absorptivities at the wavelengths of interest.

This so-caled ratio method can be used for peak tracking because the value of the ratio is a unique quantity related only to spectral qualities of the solute(s). However, this method also suffers from serious drawbacks. For instance, it is difficult to estimate the ratio in the case of severe peak overlap or in the case of a baseline with a different drift at both wavelengths. Furthermore, small changes in spectra due to different experimental conditions (change of organic modifier during the optimization procedure) can strongly influence the ratio, causing additional confusion in the tracking procedure[20].

An obvious extension of this method is to use a multiwavelength detector such as the linear photodiode-array detector which records complete spectra of the eluting compounds in a very short time. For fully separated peaks the characteristic spectra can be determined easily and compared with the spectra recorded in other chromatograms. If the spectra of the solutes in the sample differ enough, a visual comparison can be sufficient to recognize the peaks. A number of methods exist to perform this comparison in a more objective way; usually the correlation coefficient is applied, which is equivalent to the sum of squared differences of normalized and averaged spectra. Alternative methods based on absolute differences, root of the mean squared differences and the fuzzy-set theory have also been described[21]. All methods imply or require some form of normalization to eliminate concentration effects.

The more powerful methods for peak tracking are based on multivariate analysis of mixture spectra. It involves the following steps (illustrated in Fig. 8)[14].

The chromatogram is divided into mutually well separated peaks (peak clusters) and the number of solutes hidden under each peak is determined by principal component analysis (PCA). A PCA is a multivariate treatment of the data matrix containing the spectra recorded over the elution profile. This results in a list in which a selection is to be made in significant principal components and those describing the experimental noise. The most common problem related to the estimation of the number is that too many solutes are proposed. Several methods to determine the number of significant principal components have been described[22]. The knowledge of the experimental error is an important factor.

All spectraly pure peaks contain only one solute and thus yield a reliable spectrum for that solute. When the peak contains two or more overlapping solute bands, a multivariate analysis technique known as Iterative Target Transformation Factor Analysis (ITTFA) can be used to resolve the total peak profile in individual

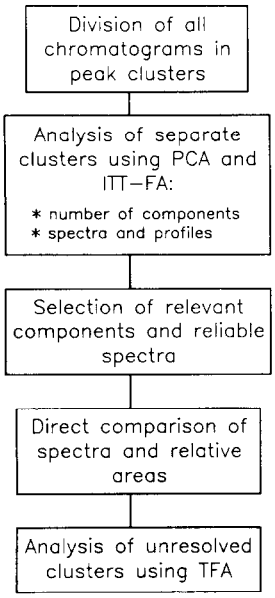

Fig. 8. Overview of the peak tracking procedure. Details are given in the text.

profiles for all solutes. First, the peak locations of the components (number being derived by PCA) are determined by application of a varimax rotation[23]. The first estimate of each elution profile consists of a single delta function at one peak location. In an iterative way, this "spike" is adjusted for negative concentrations and secondary maxima until it encounters preset conditions. This procedure is repeated for every component (peak location) in the cluster. It is necessary that all profiles have been determined before reliable UV spectra can be calculated for the cluster.

Step 3 in Fig. 8 refers to an evaluation of the calculated profiles and spectra. As the peak tracking is primarily based on a comparison of spectra, it is essential to estimate the equality of a calculated spectrum, *i.e.*, its similarity with the unknown true calculated spectrum. The resolution and spectral dissimilarity between the two neighbouring solutes are the important factors here. In general, when two components are resolved with a resolution better than $R_s \geqslant 0.4$, not too disimilar (mutual correlation coefficient >0.5) and present in approximately equal concentrations, the calculated spectra will have a high quality and will be sufficiently accurate for reliable comparison with spectra observed in other chromatograms[21].

In the next step, all reliable spectra in two chromatograms are compared and any combinaton yielding a correlation coefficient larger than 0.99 is considered to be a match. If, as a result, all reliable spectra have been matched and thus have been exhausted, we would be left only with unreliable spectra in the two chromatograms and their matching would be very difficult. Generally, however, we achieve only a partial match and in either chromatogram some reliable spectra remain unmatched.

Unidentified peaks in one chromatogram are now searched for the presence of an unmatched reliable spectrum from the other chromatogram using another multivariate technique called Target Factor Analysis (TFA). This method needs test

spectra available from earlier chromatograms, or from a library. The test of whether the mixture spectra in a cluster contain a particular test spectrum is based on the following principle: from a PCA analysis carried out on the data matrix one obtains a number of abstract spectra with which each of the measured mixture spectra can be reconstructed by linear combinations. It is therefore plausible that also the pure spectra constituting the mixture spectra can be reconstructed with the same set of abstract spectra. The test consists of two steps. A multicomponent analysis is caried out with the abstract spectra on the spectrum being tested. This results in that linear combination of the abstract spectra with which the tested spectrum is best reconstructed. By taking that linear combination, one thereafter recalculates the target, which is called the estimated target. The test fails when tested and estimated target differ to a larger extent than is expected from noise level considerations[24].

In comparison with the previous technique, TFA needs significantly less resolution and it can recognize a solute even when it is resolved by only $R_s = 0.1$. For all the solutes in a cluster there should be a "reliable" spectrum available.

To conclude the peak tracking procedure, we can remark that the number of analytically significant components is best estimated on the basis of the results of the PCA combined with ITTFA for solutes showing a certain concentration (UV activity).

The combination of TFA and ITTFA eliminates the limitations of each method separately; unknown spectra of moderately coeluting components are determined by means of the ITTFA and clusters with too little resolution to perform the ITTFA are resolved by means of TFA using reliable spectra derived from other chromatograms.

When components with very similar spectra are present in a mixture, additional information such as the relative areas of the peaks must be used in order to perform an unambiguous identification.

The major assumption is that spectral characteristics of the components do not change much with varying experimental conditions. This is perhaps a safe assumption in modifier optimization, but questionable in pH optimization.

9. PRACTICAL EXAMPLE

The optimization and peak tracking procedure will be described on the basis of a nine-component mixture. The stationary phase was 5-μm ODS-Hypersil (Shandon Southern Products, slurry-packed into a 20 cm $\times$ 4.6 mm I.D. column. The nine components were benzyl alcohol, dimethyl phthalate, phenol, benzonitrile, *p*-cresol, diethyl phthalate, 3,4-dimethylphenol, benzene and 2,4-dimethylphenol.

9.1. Determination of the optimization search area

First, a water-to-methanol gradient scan is carried out in 15 min (see Fig. 9). The first peak is eluted at 10.21 min and the last at 13.96 min. Once the number of solutes is known (nine solutes were used and found by performing a PCA analysis), the peak capacity needed to solve the separation problem at a preselected probability level can be determined from the curve in Fig. 4b[25]. We target the separation of the nine solutes at a probability level of 0.75, which results in a required peak capacity $N_c = 26$. In this particular instance, the starting binary methanol–water composition resulting in a probability of 0.738 can be derived from Fig. 5b and yields a composition of 50:50. The chromatogram taken at this composition shows eight peaks (see Fig. 7b), so the

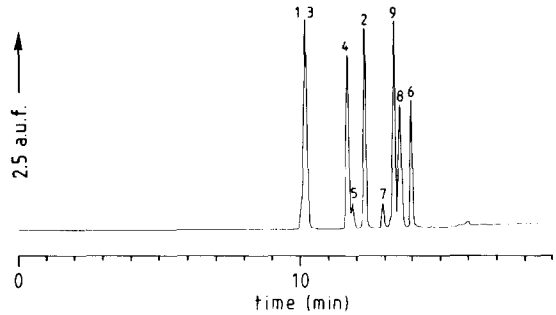

Fig. 9. Gradient elution chromatogram of the nine-solute mixture from Fig. 7. The net retention time of the first- and last-eluting peaks from this chromatogram serve to calculate the initial starting binaries for Fig. 7. (From ref. 25.)

isoeluotropic THF–water and ACN–water eluents are tried in order to investigate the possibilities of ternary optimization. In order to maintain the same probability level with the ternary mobile phase mixtures, the binaries should have the same peak capacity. The empirical transfer rules described in ref. 15 are used to calculate the corresponding ACN–water and THF–water compositions. Chromatograms taken at these compositions (see Fig. 7a and c) again do not show the expected nine peaks, so ternary optimization is to be considered. After having defined the suitable vector space (the three starting binaries), an optimization criterion must be selected according to the goal of the analyst. The goal of the optimization for this sample is satisfactory resolution for the critical pair ($R_{s,\min}$).

Before the solvent selection diagram can be constructed, peak tracking in the three starting chromatograms is needed.

9.2. Peak tracking procedure

Although in the present sample the composition is known, the peak tracking strategy as such assumes no prior information. An overview of the procedure was given in Fig. 8.

The first step is the partition of every chromatogram into peak clusters (Fig. 10). These clusters are selected on the basis of a critical absorbance threshold. This level is tested for at all wavelengths, thus ensuring detection of all components in the sample exhibiting significant absorption at any wavelength. Once all regions with significant absorption have been selected, the corresponding mixture spectra are subjected to the PCA analysis, resulting in an estimate of the number of components. Different criteria to select the significant eigenvectors have been proposed[22]. In the case of the nine-component mixture, Table 3 shows the result of the PCA and the ITTFA on the clusters. In comparing the different criteria, it is obvious that an estimation of the peak area (and related concentration) can be very helpful in judging the correct number of solutes present in the clusters.

The following distinctions can be made: a cluster is considered fully resolved if only one component is present or when all components are sufficiently resolved to fulfil the above requirement of spectral quality and reliable concentration (peak area); and a cluster is unresolved if one or more of the solutes involved are uncertain with

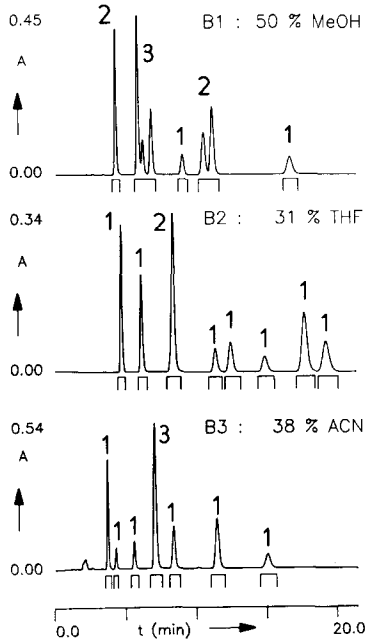

Fig. 10. The same three chromatograms as in Fig. 7 redrawn to show the division into clusters for the peak tracking procedure. The peak clusters selected for the multivariate analysis are indicated by the brackets below the chromatogram.

respect to spectra and peak area. These clusters are analysed in more detail in the next steps of the procedure.

The reliable spectra selected on the basis of the above evaluation are stored in a library (not to be confused with an "external library"). From Table 3, we know how many solutes should be present in the different clusters of the three chromatograms show in Fig. 10. All single-component peaks result in a reliable spectrum. This is the case for the clusters 3, 4 and 5 in B1, clusters 1, 2, 4, 5, 6, 7 and 8 in B2 and clusters 1, 2, 3, 5, 6 and 7 in B3. These spectra are already part of the three libraries. Additional spectral information is needed from the clusters 1 and 2 in B1, cluster 3 in B2 and cluster 4 in B3.

The chromatographic resolution in the first cluster of chromatogram B1 is below 0.4; however, owing to its large abundance the spectrum of the first component is tentatively used in the spectral comparison in the next step of the procedure. The resolution between the first and second component in the third cluster of chromatogram B2 is 0.48 but, because the concentration of the first component seems very low, its spectral characteristics are not included in the library. The same holds for the last component (impurity?) in the cluster.

The resolutions between the three components in the fourth cluster in chromatogram B3 are estimated to be 0.25 and 0.32. The spectral quality and the relative peak areas are not included in the library.

Apparently, we are dealing with a nine-component mixture which produces three libraries: B1 and B2 both containing seven reliable and one tentative spectrum,

TABLE 3

ESTIMATED DIMENSIONALITIES OF THE CLUSTERS IN THE THREE CHROMATOGRAMS OF THE NINE-COMPONENT MIXTURE, BASED ON THE KNOWN EXPERIMENTAL ERROR (RE), THE χ^2 TEST, THE INDICATOR FUNCTION OF MALINOWSKI (IND) AND THE RESULTS OF THE ITT FA (ITT)

In addition, the number of significant components selected after examination of the concentrations (*N*-est) and the actual number of components based on knowledge of the mixture, (*N*-act) are listed. The three chromatograms are identified as B1, recorded in 50% methanol, B2, recorded in 32% THF, and B3, recorded in 38% ACN. For explanantion of the different criteria, see ref. 14.

Cluster	*RE*	χ^2	*IND*	*ITT*	*N-est*	*N-act*
B1-1	3	2	6	3	2	2
B1-2	3	3	9	3	3	3
B1-3	1	1	3	1	1	1
B1-4	3	3	7	2	2	2
B1-5	2	1	3	1	1	1
B1-total	12	10	28	10	9	9
B2-1	2	2	5	1	1	1
B2-2	1	1	4	1	1	1
B2-3	3	3	6	3	3	2
B2-4	2	1	3	2	1	1
B2-5	2	1	5	1	1	1
B2-6	1	1	2	1	1	1
B2-7	2	2	4	1	1	1
B2-8	2	2	3	1	1	1
B2-total	15	13	32	11	10	9
B3-1	1	1	4	1	1	1
B3-2	2	1	3	1	1	1
B3-3	1	1	3	1	1	1
B3-4	3	3	7	3	3	3
B3-5	2	1	3	1	1	1
B3-6	2	2	4	1	1	1
B3-7	1	1	3	1	1	1
B3-total	12	10	27	9	9	9

and B3 with six reliable spectra. Fig. 11a shows the first results in a schematic way. In the next step of the procedure the library spectra from the three chromatograms are compared (see Fig. 11b). When the spectra are different enough, a close match expressed by a correlation coefficient is sufficient for identification purposes. However, when two or more components have very similar characteristics, an additional source of information is required. This is supplied by the relative peak areas.

First the library derived from chromatogram B2 (containing most spectra) is compared with the library B1, the correlation coefficient being the match factor. Also, the tentative components are recognized as both spectra and areas are matched with solutes in B1 and consequently are now considered as "defined". When comparing the spectrum of the second component in chromatogram B2 with the spectra in library B1, two spectra with high correlation are observed (see Table IV). By examining the relative areas of both components and comparison with the relative area of component B2-2, component B1-5 is clearly matched with the component in question. As all components in library B2 are matched with spectra in library B1, no spectra remain to be used in the next step of the analysis.

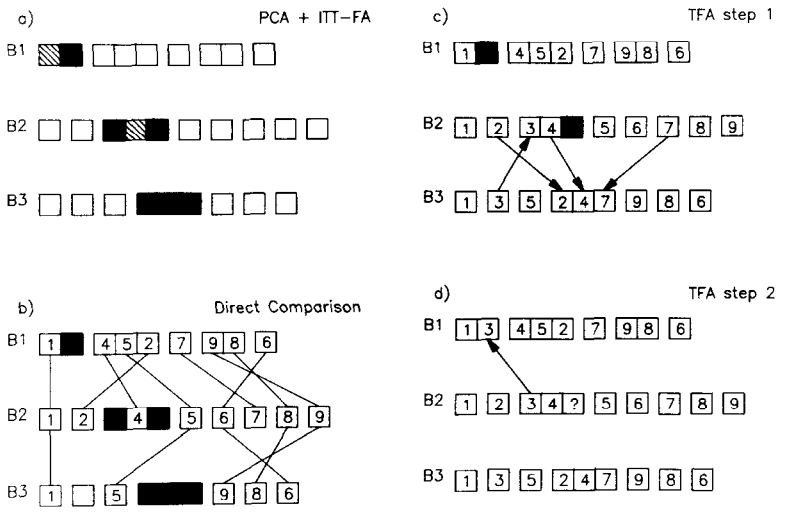

Fig. 11. The results of the four steps in the peak tracking procedure for the three chromatograms of the nine-component mixture in Fig. 10. (a) Conclusions of the evaluation of the spectra; spectra are defined as: reliable (white boxes), tentative (hatched boxes) or unreliable (black boxes). (b) Components recognized by a direct comparison, indicated by a line connecting matched solutes. (c) Components recognized by means of TFA using spectral data from chromatograms B2 and B3. (d) Component 3 recogized by means of TFA using the updated spectral library of chromatogram B2.

The next comparison involves libraries B2 and B3. Five components are identified unambiguously, indicated by the lines and numbers in Fig. 11b. Despite high correlations between the spectra of components 2, 4 and 7 with the spectra in library B3, no direct match is observed because of large differences in the relative areas. As a consequence, a new library B2a is developed containing the spectra of these compounds. Simultaneously, a library B3a is created containing the spectrum of the

TABLE 4

ILLUSTRATION OF THE COMPARISON OF THE SPECTRA AND RELATIVE AREAS RESULTING FROM THE ANALYSIS OF CHROMATOGRAMS B1, B2 AND B3

ρ is the correlation coefficient of spectrum B2-2 (library B2) with the spectra in libraries B1 and B3. When the correlation exceeds a value of 0.995, the relative areas are compared with the relative area of component B2-2 (13.8%). An unambiguous identification is indicated by an asterisk.

Component library B1	ρ	*Area (%)*	*Component library B3*	ρ	*Area (%)*
B1-1	0.83245		B3-1	0.82032	
B1-3	0.90942		B3-2	0.87021	
B1-4	0.90383		B3-3	0.90818	
B1-5	0.99976	13.9*	B3-7	0.92077	
B1-6	0.91285		B3-8	0.76801	
B1-7	0.91648		B3-9	0.99942	7.9
B1-8	0.76731				
B1-9	0.99983	7.9			

second component of library B3, as no direct match was observed between this spectrum and the spectra in library B2.

The unresolved clusters remaining after the PCA and the ITTFA are analysed using the just created libraries in a target factor analysis (TFA). As the requirements with respect to the resolution are considerably reduced, clusters wich could not be analysed by means of the ITTFA can be resolved provided the spectra of the components are known.

The example requires two steps, as illustrated in Fig. 11c and d. First the unresolved cluster in chromatogram B3 is analysed with the reliable, but as yet unidentified, spectra from chromatogram B2. The three components (2, 4 and 7) contained in library B2a are all located in the fourth cluster in chromatogram B3. As it was the only remaining cluster, this result was to be expected, but the principle can be applied when more unresolved clusters are present in the chromatogram. The only remaining unidentified spectrum in B3 is found in the unresolved cluster (3) in chromatogram B2. We are left with chromatogram B1. In the preceding step all of the known reliable spectra from B2 could be matched. From B3 we have found a new, ninth component (3). Not surprisingly, in a TFA this component is found to be located in the unresolved cluster (1) in chromatogram B1 (see Fig. 11d). As a high correlation is observed, the nine components are now recognized in all three chromatograms and only the impurity observed in the third cluster in the second chromatogram remains unidentified. This "unknown" solute will be ignored in the procedure.

Tables containing retention times are now available to be used in the next step of the optimization strategy.

9.3. Continuation of the optimization procedure

The next step in the procedure is to set up the solvent selection diagram. The logarithms of the solute capacity factors measured in the three binaries are connected by straight lines over the ternary eluent composition (see Fig. 12). These plots are used

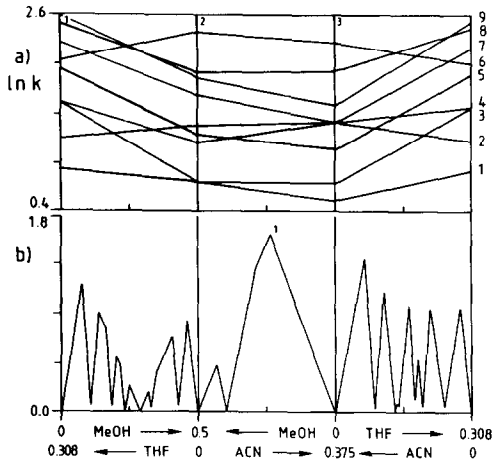

Fig. 12. Solvent-selection diagram constructed from the chromatograms shown in Fig. 7 and after the peak recognition procedure (see Fig. 11). (a) Plots of ln k; (b) response surface for the minimum resolution criterion $R_{s,min}$. (From ref. 25 with permission.)

for the calculation of the minimum resolution criterion, $R_{s,min}$ (lower part of Fig. 12). Clearly, the separation selectivity is strongly influenced by the ternary eluent composition (strong variation in elution order) and a maximum value of $R_{s,min} = 1.64$ is predicted for a ternary mobile phase containing 24.4% methanol and 19.2% ACN. After measuring the predicted composition, the retention data are entered into the model and a new optimum is predicted (Fig. 13). This is repeated until no more improvement is gained. The final optimum is found at 15% methanol and 26.3% ACN. The chromatogram verifying this optimum is shown in Fig. 14.

10. CONCLUSIONS

The Delft iterative optimization design offers a rapid selection of mobile phase compositions depending on the nature and number of the species present in the sample. The method starts with a water–methanol gradient (or several gradients with controlled pH if ionic species are present). These scans reveal the presence of the different compounds and provide a prediction of a suitable water–methanol eluent to achieve complete separation at a specified probability level. The gradient scans taken at different pHs offer the analyst the information needed to decide on the selectivity parameters and their range of variation.

The transfer to isoeluotropic ACN– and THF–water mixtures forms the basis of the solvent selection diagram. The retention behaviour over the ternary mobile phase composition is assumed to be linear. This linear model can be refined as soon as a new set of experimental data becomes available. Any optimization criteron can be used depending on the desired separation. The number of parameters to be optimized simultaneously is limited to three for practical reasons such as the number of initial experiments and the computer time needed for the calculations.

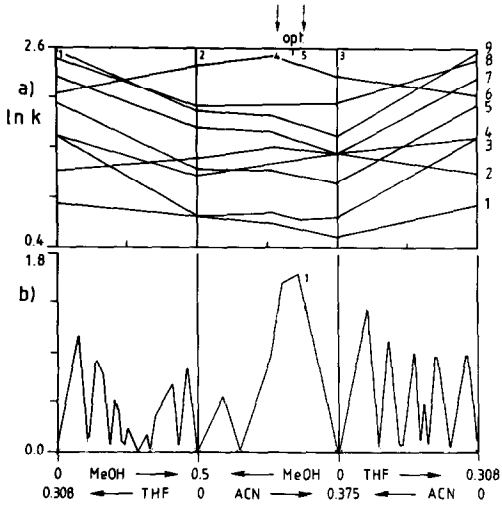

Fig. 13. Optimization of the nine-solute mixture in Fig. 12. (a) After two more chromatograms taken at points 4 and 5 and connecting the retention data by straight lines, an optimum is predicted in the ternary methanol–acetonitrile parameter space. (b) Optimization criterion $R_{s,min}$. (From ref. 25 with permission.)

As the procedure uses a model based on the retention behaviour of individual components, peak tracking during the course of the optimization process is inevitable. Multivariate handling of spectral data resulting from a linear photodiode-array detector is capable of unambiguous tracking of the solutes in the chromatograms. The major assumption is that the spectral characteristics of the components do not change much under varying conditions. The method has proved to be fast and reliable for many different samples.

11. SUMMARY

An iterative procedure for optimizing the mobile phase composition in reversedphase liquid chromatography and reversed-phase ion-pair liquid chromatography is reviewed. The method consists of a gradient-scouting technique to delineate the parameter space for the case of reversed-phase systems. In ion-pairing chromatography, the choice of the parameters and their range depend very much on the sample composition. Initial chromatographic experiments are described to determine the presence of strong and weak acids or basis and neutral components in the sample.

In addition to the type of solutes present in the sample, their number also has a large influence on the chances of success in achieving a complete separation of all components in the sample. When the number of components has been determined and a desired probability of success has been imposed, the result from the gradient run predicts the suitable binary compositions of the three starting mobile phases containing on the one hand water and on the other hand either methanol, acetonitrile or tetrahydrofuran. The assumption that the logarithm of the capacity factor changes linearly with the mixing rate of these three binary mixtures makes it possible to construct a phase-selection diagram. A resolution-based criterion can then be calculated and the predicted optimum located in the ternary mobile phase composi-

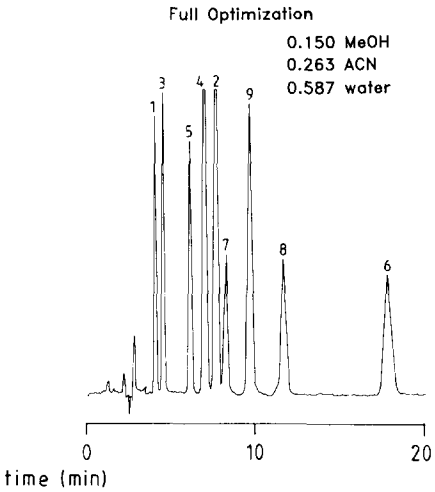

Fig. 14. Final chromatogram of the nine-solute mixture in Fig. 7. The optimum ternary composition for complete separation is derived from Fig. 13. (From ref. 25 with permission.)

tion. After recording a chromatogram under the predicted optimum conditions, the retention data are used to refine the linear retention model and a new optimum is predicted. This procedure is repeated until no further improvement is to be expected.

The described procedure needs the location of the different solutes in the recorded chromatogram. If the (unknown) components absorb UV radiation, such peak tracking can be performed using multivariate analysis of data obtained with a linear photodiode-array detector.

The entire procedure is illustrated with an example of the separation of nine components.

REFERENCES

1 P. J. Schoenmakers, *Optimization of Chromatographic Selectivity (Journal of Chromatography Library,* Vol. 35), Elsevier, Amsterdam, 1986.
2 J. C. Berridge, *Techniques for the Automated Optimization of HPLC Separations,* Wiley, Chichester, 1985.
3 D. P. Herman, H. A. H. Billiet and L. de Galan, *Anal. Chem.*, 58 (1986) 2999.
4 L. de Galan, D. P. Herman and H. A. H. Billiet, *Chromatographia,* 24 (1987) 108.
5 J. M. Davis and J. C. Giddings, *Anal. Chem.*, 53 (1983) 418.
6 A. Bartha, H. A. H. Billiet and L. de Galan, *J. Liq. Chromatogr.*, 12 (1989) 173.
7 J. K. Strasters, A. Bartha, H. A. H. Billiet and L. de Galan, *J. Liq. Chromatogr.*, 11 (1988) 1827.
8 A. Peeters, L. Buydens, D. L. Massart and P. J. Schoenmakers, *Chromatographia,* 26 (1988) 101.
9 S. N. Deming and S. L. Morgan, *Experimental Design: a Chemometric Approach,* Elsevier, Amsterdam, 1987.
10 L. de Galan and H. A. H. Billiet, *Adv. Chromatogr.*, 25 (1986) 63.
11 P. J. Schoenmakers and Th. Blaffert, *J. Chromatogr.*, 384 (1987) 117.
12 P. J. Schoenmakers, H. A. H. Billiet and L. de Galan, *J. Chromatogr.*, 205 (1981) 13.
13 H. B. Patel and T. M. Jefferies, *J. Chromatogr.*, 389 (1987) 21.
14 J. K. Strasters, H. A. H. Billiet, L. de Galan and B. G. M. Vandeginste, *J. Chromatogr.*, 499 (1990) 499.
15 D. P. Herman, H. A. H. Billiet and L. de Galan, *J. Chromatogr.*, 463 (1989) 1.
16 P. R. Haddad and S. Sekulic, *J. Chromatogr.*, 392 (1987) 65.
17 J. L. Glajch and J. J. Kirkland, *Anal. Chem.*, 55 (1983) 319A.
18 H. Colin, A. Krstulovic, G. Guiochon and J. P. Bounine, *Chromatographia,* 17 (1983) 209.
19 M. Oto, W. Wegscheider and E. Lankmayer, *Anal. Chem.*, 60 (1988) 517.
20 A. C. J. H. Drouen, H. A. H. Billiet and L. de Galan, *Anal. Chem.*, 56 (1984) 971.
21 J. K. Strasters, H. A. H. Billiet, L. de Galan, B. G. M. Vandeginste and G. Kateman, *Anal. Chem.*, 60 (1988) 2745.
22 E. R. Malinowsky, D. G. Howery, *Factor Analysis in Chemistry*, Wiley, New York, 1980.
23 B. G. M. Vandeginste, W. Derks and G. Kateman, *Anal. Chim. Acta,* 173 (1985) 253.
24 B. G. M. Vandeginste, G. Kateman, J. K. Strasters, H. A. H. Billiet and L. de Galan, *Chromatographia,* 24 (1987) 127.
25 A. Bartha, H. A. H. Billiet and L. de Galan, *J. Chromatogr.*, 458 (1988) 371.
26 P. J. Schoenmakers, A. C. J. H. Drouen, H. A. H. Billiet and L. de Galan, *Chromatographia,* 15 (1982) 688.