Journal of Chromatography, 499 (1990) 499–522
Elsevier Science Publishers B.V., Amsterdam — Printed in The Netherlands

CHROM. 21 634

STRATEGY FOR PEAK TRACKING IN LIQUID CHROMATOGRAPHY ON THE BASIS OF A MULTIVARIATE ANALYSIS OF SPECTRAL DATA

JOOST K. STRĄSTERS, 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)
and
BERNARD G. M. VANDEGINSTE
Catholic University of Nijmegen, Department of Analytical Chemistry, Toernooiveld, 6525 ED Nijmegen (The Netherlands)

SUMMARY

Peak tracking, required for the interpretive optimization of the mobile phase in reversed-phase high-performance liquid chromatography, is performed by means of multiwavelength detection followed by a multivariate analysis of the resulting data. A combination of techniques (target factor analysis and iterative target transformation factor analysis) is necessary to circumvent problems caused by a limited resolution or lack of knowledge regarding spectral characteristics and to determine the number of relevant components. A procedure to perform the actual peak tracking, which combines data related to the spectral characteristics and relative peak areas of separated and coeluting components, is presented.

INTRODUCTION

A number of strategies developed to optimize the mobile phase composition in high-performance liquid chromatography (HPLC)[1–5] requires knowledge of the individual retention behaviour of all components in the sample, irrespective of whether the separation of all or only a part of the solutes is desired[6]. One of these so-called interpretive optimization strategies is the iterative regression design[7], which estimates the retention at various mobile phase compositions on the basis of a limited number of chromatograms of a given sample under various experimental conditions and a linear retention model. In order to derive this model, the location of the components in successive chromatograms must be known. The determination of the elution order in a set of chromatograms is refered to as peak tracking.

A simple method of peak tracking uses the relative areas of the peaks to derive the individual retention characteristics. Because of the expected peak overlap (one of the reasons to optimize the separation), some of the individual peak areas will be difficult to determine without assumptions on peak shape. Although an algorithm using combinations of areas to analyse coeluting peaks has been described[8], the result

0021-9673/90/$03.50

will depend heavily on the number of components assumed in the mixture, usually defined as the maximum number of peaks observed in any chromatogram. Otto *et al.*[9] circumvented problems related to the variation in area of peaks of the same component by applying fuzzy set theory and assuming a limited variation in the retention behaviour.

Drouen *et al.*[10] described the use of wavelength ratios, which should result in a characteristic level for every component. However, due to the variability of the spectra in different mobile phase systems and disturbances in the ratios caused by drifting or shifted baselines, a direct matching of levels is not always possible. Furthermore, in cases of severe overlap (resolution lower than 0.5) the individual levels of the components are no longer observed. The logical extension to multiwavelength detection by means of linear photodiode array detection (LPDA) and a subsequent comparison by eye of the resulting spectral data was described by the same authors[11]. Although individual spectra of the components are not always observed in peak clusters containing more than two components with severe chromatographic overlap, a simultaneous visual evaluation of a number of spectra recorded in a peak group can be used to estimate the number of components and select specific adsorption characteristics to aid in the peak tracking.

For a further automation of the method and an extension to more complex mixtures, we recently described the application of various multivariate analysis techniques to spectral data resulting from multiwavelength detection of peak clusters[12]: depending on the amount of information available, one or more of these techniques can be applied to assist in the determination of separate elution profiles and/or spectra of components in peak clusters. The multicomponent analysis applied to separate mixture spectra requires no resolution, but all components should be available in a library containing reference spectra measured in the same mobile phase. A selection of the relevant library spectra by means of a target factor analysis lowers the demands on the spectral quality of these spectra, but requires some chromatographic resolution. Various techniques related to self-modelling curve resolution such as the iterative target transformation factor analysis determine both spectral characteristics and elution profiles, but produce reliable results only when the chromatographic resolution is sufficiently large[13,14].

In this study we have used a combination of these techniques in a systematic approach to track peaks in various chromatograms of the same sample. The procedure is described by referring to two practical examples with special emphasis on a practical method to determine the number of relevant components.

THEORETICAL

The principal component analysis (PCA)[15]

The two multivariate methods applied in the peak-tracking procedure are the target factor analysis[15] and the iterative target transformation factor analysis[16]. Both methods use the results of an abstract analysis of the spectral data: since the spectra observed are a linear combination of contributions of individual components present in the cluster, the original data matrix, **D**, of n_s spectra over n_w wavelengths can be written as a product of the spectral characteristics of the components and the corresponding concentration profiles (Beer's law):

$$\mathbf{D} = \mathbf{S}\mathbf{C}^{\mathrm{T}} \tag{1}$$

S contains the n_c pure component spectra over n_w wavelengths and **C** contains the n_c individual elution profiles over n_s points. An analysis of the variance observed in the data by means of a principal component analysis[15] produces an alternative decomposition:

$$\mathbf{D} = \mathbf{R}\mathbf{V}^{\mathrm{T}} \tag{2}$$

When this description is derived from the covariance matrix $\mathbf{D}^{\mathrm{T}}\mathbf{D}$, the matrix **V** contains the n_s principal components (dimensions $n_s \times n_s$) which correspond to the eigenvectors of the covariance matrix. The matrix **R** (dimensions $n_w \times n_s$) is derived by a projection of **D** on **V**. When only the significant eigenvectors are considered (described in the next section) an equation with reduced matrices, **R'** (dimensions $n_w \times n_c$) and **V'** (dimensions $n_s \times n_c$), is derived which results in a reconstructed data matrix, **D'**. Apart from the experimental noise, **D'** is equal to **D**. The reduced matrices, **R'** and **V'**, can be transformed to their physically meaningful counterparts, **S** and **C**:

$$\mathbf{D}' = \mathbf{R}'\mathbf{V}'^{\mathrm{T}} \tag{3a}$$

$$= \mathbf{R}'\mathbf{T}\mathbf{T}^{-1}\mathbf{V}'^{\mathrm{T}} \tag{3b}$$

$$= \mathbf{S}\mathbf{C}^{\mathrm{T}} \tag{3c}$$

The matrix **T** is the transformation matrix (dimensions $n_c \times n_c$). Different methods indicated by the general descriptor "factor analysis" try to derive either **T** or $\mathbf{T}^{-1}$ depending on the available information on the system and/or by imposing a number of boundary conditions.

The target factor analysis[15,17]

The target factor analysis (TFA) selects the appropriate components from a library by an individual examination of the spectral characteristics: every reference spectrum is approximated in a least squares sense by means of the columns of **R** (the reference spectrum, $\boldsymbol{s}$, is projected onto the hyperspace defined by **R** and is compared with its projection, $\boldsymbol{s}'$):

$$\boldsymbol{s}' = \mathbf{R}(\mathbf{R}^{\mathrm{T}}\mathbf{R})^{-1}\mathbf{R}^{\mathrm{T}}\boldsymbol{s} \tag{4a}$$

$$= \mathbf{R}\boldsymbol{t} \tag{4b}$$

If the projection is valid, *i.e.*, if the spectrum resembles its projection closely and consequently is a component present in the examined cluster the required projection produces one of the columns of **T**, $\boldsymbol{t}$. Once all components present in the cluster have been identified, **T**, and hence $\mathbf{T}^{-1}$, is known and the corresponding elution profiles are calculated.

The requirements for this analysis to work are the following: the spectral characteristics for all components should be (slightly) different and some resolution

(approximately $R = 0.1$) should be present in order to derive an hyperplane with the correct dimensionality. A slight change in spectral characteristics, for instance caused by a change in the mobile phase composition, is acceptable as long as the correct components are selected from the library[12].

The iterative target transformation factor analysis (ITT-FA)[13,14,16]

Like other forms of self-modelling curve resolution[18,19], the ITT-FA derives both elution profiles and spectra by imposing a number of boundary conditions on the derived solution, *i.e.*, non-negativity of concentrations and a restriction to unimodal profiles. However, as the name indicates, no specific peak model is assumed.

The ITT-FA starts with a first estimate of the elution profile, which is a spike at the location derived by means of an orthogonal abstract rotation (the varimax rotation). This estimate, $\boldsymbol{e}_1$, is projected onto the hyperplane described by $\mathbf{V}$, and the resulting projection, $\boldsymbol{e}'_1$, is corrected for negative concentrations and secondary maxima. The resulting profile, $\boldsymbol{e}_2$, is again projected and corrected until no significant improvement in subsequent projections is observed:

$$\boldsymbol{e}'_i = \mathbf{V}(\mathbf{V}^T\mathbf{V})^{-1}\mathbf{V}^T\boldsymbol{e}_i \tag{5a}$$

$$= \mathbf{V}\boldsymbol{t}_i \tag{5b}$$

$$\boldsymbol{e}_{i+1} = (\boldsymbol{e}'_i)_{\text{corrected}} \tag{6}$$

The final transformation found to derive the profile, $\boldsymbol{t}_i$, is one of the rows of $\mathbf{T}^{-1}$ in eqn. (3b). Once all profiles in a peak cluster have been determined in this way, the transformation matrix, $\mathbf{T}$, is constructed and the corresponding spectra are calculated (eqn. 3c).

Similar to TFA, the spectra should show some dissimilarity in order to derive the correct hyperplane used for the derivation of the elution profiles. However, the amount of resolution required for a correct analysis is significantly greater in comparison with TFA (approximately $R = 0.4$ for components present in equal concentrations). The quality of the resulting spectra, used in a subsequent comparison between various chromatograms of the same sample, will also depend on the spectral similarity of the components and the relative quantities. In a previous publication[13] we described a selection criterion to decide on the reliability of the derived spectra based on the observed values of these parameters (as opposed to the real, unknown values).

The number of relevant components

A specific problem confronting the analyst in practical optimization problems is how to determine the number of components is an unknown mixture that can be detected by UV spectroscopy. Often, one assumes this number to be equal to the maximum number of peaks in any chromatogram of the sample. However, this number is a good guess at best. A similar problem must be dealt with in the case of separate peak clusters: the individual elution profiles (TFA) or spectra (ITT-FA) can be calculated only once all spectra or profiles have been determined. Consequently, a knowledge of the number of (relevant) components is essential to obtain the desired information.

A much better estimate as compared to the observed number of peak maxima can be derived by a multivariate treatment of the spectral data. As indicated before, the number of principal components (PCs) derived is originally equal to the number of columns, n_s, in the data matrix, *i.e.*, the number of spectra recorded. However, because each PC describes the maximum amount of variation in the data which is not described by the previous ones, only the first n_c PCs are related to the relevant information with regard to spectra and elution profiles. The remaining PCs are required only to describe the exprimental noise and can be removed without changing the information content of the resulting data matrix, $\mathbf{D}'$ (eqn. 3a). A number of methods to determine the number of significant PCs has been described in the literature[15,20–22] and can be roughly divided into two depending on the knowledge of the experimental error.

When the experimental error is known, one can reconstruct the data matrix $\mathbf{D}'_{n_c}$ using the reduced matrices $\mathbf{R}'_{n_c}$ and $\mathbf{V}'_{n_c}$; n_c indicates the number of columns and hence the assumed dimensionality of the data, *i.e.*, the number of components. $\mathbf{D}'_{n_c}$ is compared with $\mathbf{D}$ and when the difference falls just within the experimental error the correct number of components is used. Adding more components to the description corresponds to overfitting of the data and will not result in a significant improvement of the quality. When using the reduced matrices, $\mathbf{R}'$ and $\mathbf{V}'$, part of the experimental error is removed from the data. Malinowski and Howery[15] defined this removed error by the term "extracted error" (XE):

$$\mathrm{XE}^2_{n_c} = \sum_{i=1}^{n_s} \sum_{j=1}^{n_w} [d_{ij} - (d'_{ij})_{n_c}]^2 / n_s n_w \tag{7}$$

This error can be used to estimate the real error, RE, in the data:

$$\mathrm{RE}^2_{n_c} = \mathrm{XE}^2_{n_c} n_s / (n_s - n_c) \tag{8}$$

Several short-cuts to determine the value of RE exist, using the eigenvalues λ related to the eigenvectors of the covariance matrix, $\mathbf{Z}$ (the PCs). Since the last $n_s - n_c$ eigenvalues are related only to the experimental error, this error can be estimated from:

$$\mathrm{RE}^2_{n_c} = \sum_{i=n_c+1}^{n_s} \lambda_i / [n_w (n_s - n_c)] \tag{9}$$

The real error is estimated for different values of n_c and the value of n_c where the error drops below the known experimental error is taken as the number of components. A similar measure of uncertainty, the chi-squared criterion of Bartlett, is also based on the difference between the original and the reconstructed data matrix (see ref. 15, p. 76)

$$\chi^2_{n_c,\mathrm{exptl.}} = \sum_{i=1}^{n_s} \sum_{j=1}^{n_w} [d_{ij} - (d'_{ij})_{n_c}]^2 / \sigma^2 \tag{10}$$

where the known experimental error, σ, is assumed to be constant over the full wavelength range, which is acceptable to a first approximation. This value is compared with the theoretical value:

$$\chi^2_{n_c,\text{th.}} = (n_w - n_c)(n_s - n_c) \tag{11}$$

The value of n_c where the difference between $\chi_{n_c,\text{exptl.}}$ and $\chi_{n_c,\text{th.}}$ is minimal is taken as the correct dimensionality of the data.

When the experimental error is not known, one has to evaluate the amount of variance added to the data by the inclusion of an additional PC as compared to the remaining variance in the non-significant PCs. One of the functions used for this evaluation is the indicator function, IND, described by Malinowski[20]:

$$\text{IND}_{n_c} = \text{RE}_{n_c}/(n_s - n_c)^2 \tag{12}$$

This value should reach a minimum when the correct number of factors is used in the reproduction. However, this minimum is observed clearly only when the real (true) factors underlying the data are significantly different, *i.e.*, the angles between axes representing these factors in the data space are sufficiently large to distinguish differences between spectra or elution profiles from deviations caused by the experimental error. A computationally more complicated but also more fundamental approach is the so-called cross-validation[21]: the principal component analysis is repeated a number of times with part of the data omitted from the data matrix. This data section is predicted on the basis of models derived from the PCs by applying varying dimensionalities. This is repeated for all sections of the data matrix (preferably all individual elements). The resulting estimates are compared to the true values and the decrease in the remaining error observed with an increase in the number of factors used for the prediction is evaluated to derive the correct number of components.

All methods described in this section have one important characteristic in common: they are all devised to determine the number of statistically significant factors or components. Unfortunately, this also includes deviations in the normally distributed experimental error caused by drifting baselines, non-linearities in the absorbance characteristics of the diode array, insignificantly small contributions of impurities, etc. What we are really interested in is the number of relevant components, *i.e.*, those components exhibiting significant UV absorption and physical meaningful chromatographic behaviour. When the ITT-FA is used to analyse the above disturbances, a unimodal is possibly forced upon non-unimodal profiles, resulting in large errors in the subsequent inverse transformation (eqn. 3b) and consequently in erroneous spectral characteristics. Incorrect profiles can influence other spectra derived for the cluster under examination, including those related to physical significant solutes, or give rise to apparent components without any physical significance.

In order to avoid these kinds of deviations, we propose an alternative procedure. The ITT-FA is repeated a number of times using different dimensionalities, n_c. Each analysis, the resulting profiles and spectra are corrected for negative absorbances and concentrations in case the procedure was stopped before physically meaningful elements were derived. The calculated and corrected profiles and spectra are used to recalculate the data matrix:

$$\mathbf{D}'_{n_c} = \mathbf{S}'_{n_c}\mathbf{C}'^{\mathrm{T}}_{n_c} \tag{13}$$

By a comparison of this reproduced data matrix with the original one, the extracted error is calculated (eqn. 7). When XE reaches a minimum, the cluster is described by the n_c most reliable profiles and spectra which can be derived given the aforementioned boundary conditions. Although the neglect of small impurities, drifting baseline, etc., will result in a larger extracted error than would be expected on the basis of the experimental error, this deviation is less serious than the one resulting from the use of forced unimodal profiles and is a better indication of the number of relevant components in a cluster.

EXPERIMENTAL

Chromatographic conditions

The full peak-tracking procedure will be described on the basis of two examples. The first example consists of two chromatograms of a mixture of thirteen components listed in Table I. The mixture was eluted on a Novapak C_{18} column (Millipore Waters, Milford, MA, U.S.A.), 10 cm × 8 mm I.D. and particle size 5 μm, using 1.5 ml/min of tetrahydrofuran (THF)–water (40:60) and methanol–water (60:40) as mobile phases, respectively (Fig. 1).

The second example was part of a study on full and limited optimization[6] and consists of three chromatograms of a mixture of nine components (Table I). The mixture was eluted on a 20 cm × 4.6 mm I.D. column packed with 5-μm ODS-Hypersil using 1.0 ml/min of THF–water (31:69), methanol–water (50:50) and acetonitrile–water (38:62) as mobile phases, respectively (Fig. 2).

Retention times of the nine-component mixture observed during the subsequent steps of the optimization procedure can be found in ref. 6. The solutes were obtained from Fluka (Buchs, Switzerland) and E. Merck (Darmstadt, F.R.G.) and used in suitable concentrations (0.1–0.2 mg/ml).

TABLE I

THE COMPOSITION OF THE TWO MIXTURES USED IN THE EXAMPLES

Mixture A	*Mixture B*
1 Acetanilide	1 Benzyl alcohol
2 Methylparaben	2 Dimethyl phthalate
3 Benzaldehyde	3 Phenol
4 Acetophenone	4 Benzonitrile
5 Cinnamyl alcohol	5 *p*-Cresol
6 Nitrobenzene	6 Diethyl phthalate
7 Methyl benzoate	7 3,4-Dimethylphenol
8 Anisole	8 Benzene
9 Diethyl phthalate	9 2,4-Dimethylphenol
10 Methyl salicylate	
11 Ethyl benzoate	
12 1-Nitronaphthalene	
13 Benzophenone	

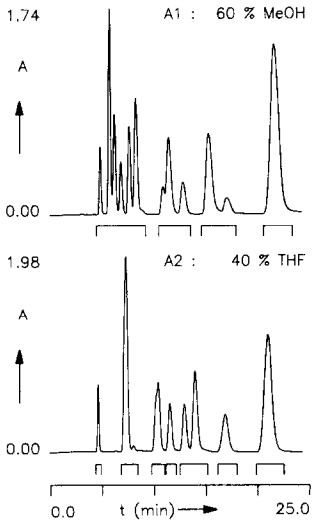

Fig. 1. Two chromatograms of a thirteen-component mixture (composition in Table I). Experimental conditions in the text. The peak clusters selected for the multivariate analysis are indicated by the brackets below the chromatograms. MeOH = methanol.

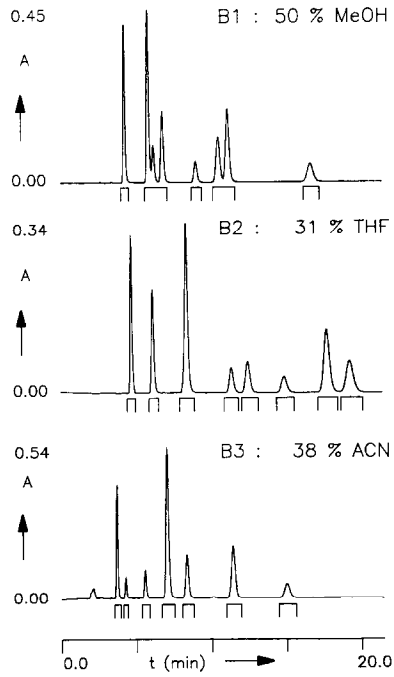

Fig. 2. Three chromatograms of a nine-component mixture (composition in Table I). ACN = Acetonitrile Other details as in Fig. 1.

Instrumentation

The chromatographic system consisted of an HP 1090 chromatograph, equipped with an HP 1040A linear photodiode array detector (Hewlett-Packard, Waldbronn, F.R.G.). The detector was connected to an HP-85 desktop computer, equipped with an HP-IB IEEE-488 interface and RS-232C serial interface. The data were temporarily stored on $5\frac{1}{4}$-in. flexible disks using an HP82910M disk-drive.

The data files collected by means of the detector were transferred to an Olivetti M24 personal computer (Olivetti, Ivrea, Italy) by means of serial interfaces on both computers. Further analysis of the data was performed on an IBM-PC (IBM, U.S.A.), equipped with two $5\frac{1}{4}$-in. disk-drives, an expansion unit with two 10-Mbyte hard disks, the Intel 8087 mathematical coprocessor and an HP-7470A graphics plotter with serial interface.

The spectra were recorded between 210 and 400 nm. The first 10 min of the chromatograms were sampled with approximately 30 spectra per minute. The remainder was analysed using 20 spectra per minute, since broadening of the peaks reduced the requirements on the sampling frequency.

Software

All software used in the procedure was developed using the Turbo Pascal compiler version 4.0 (Borland International, CA, U.S.A.). The algorithm for the TFA was adapted from the description by Malinowski and Howery[15]; the ITT-FA was adapted from the description by Vandeginste *et al.*[16]. Correlation coefficients were calculated according to Reid and Wong[23]. The eigenvectors and eigenvalues of the covariance matrix of the data matrix were determined by the HQRII algorithm[24].

Curve fitting, to estimate resolutions and concentrations, was performed on the basis of bigaussian profiles and a non-linear procedure described by Bevington[25]. This procedure was preferred to a direct calculation of peak width and areas based on the calculated profiles, because of the limited sampling frequency over the chromatographic peaks, and the inherent uncertainty in the derived results.

RESULTS AND DISCUSSION

In order to illustrate the procedure currently under investigation, two mixtures were eluted under varying chromatographic conditions (Table I and Figs. 1 and 2). In the following discussion the thirteen-component mixture will be referred to as mixture A, and the corresponding chromatograms as A1 (60% methanol) and A2 (40% THF). The nine-component mixture is referred to as mixture B, and the chromatograms will be identified by B1 (50% methanol), B2 (32% THF) and B3 (38% acetonitrile). Although the composition of both mixtures was accurately known, the strategy as such assumes no prior knowledge of the number of components or their identities, hence chromatograms of "unknown" mixtures can be analysed in the same way.

An overview of the full procedure is given in Fig. 3. In the following sections the separate steps of the strategy will be discussed in more detail.

Processing of separate clusters

The first step is the partition of every chromatogram in peak clusters. These clusters are for instance selected on the basis of a critical absorbance threshold. This

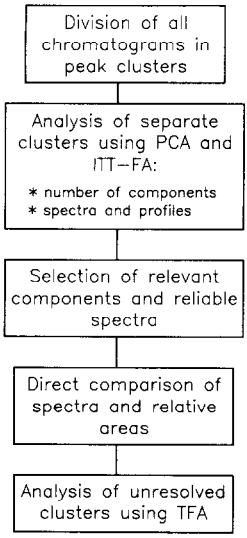

Fig. 3. Overview of the peak tracking procedure.

level is tested for at all wavelengths, thus ensuring detection of all components in the mixture exhibiting significant absorption at any wavelength. Once all regions with significant UV activity have been selected, the corresponding mixture spectra are subjected to the principal component analysis, resulting in the abstract matrices **R** and **V**.

Fig. 4 illustrates the procedure applied to the separate peak clusters (step 2 in Fig. 3). After the principal component analysis, the maximum number of components,

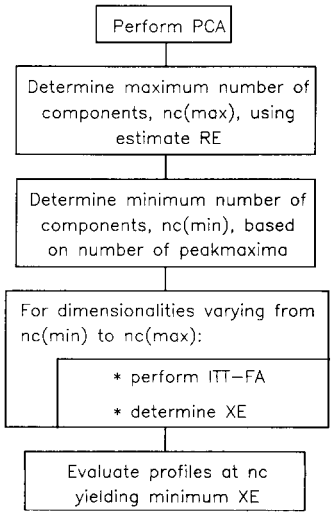

Fig. 4. The analysis of an individual peak cluster.

$n_{c_{max}}$, present in the cluster is determined on the basis of the real error (RE, eqn. 9) determined for different dimensionalities, n_c, and compared with the known experimental error of 0.15 ma.u. The minimum number of components, $n_{c_{min}}$, is defined by the number of observed peak maxima in the cluster. Of course, $n_{c_{max}}$ should be larger than or equal to $n_{c_{min}}$, otherwise one is confronted with split peaks or components with high spectral similarity and a correct analysis of the cluster by means of the ITT-FA will not be possible. The next step consists of a repeated execution of the ITT-FA varying the number of components from $n_{c_{min}}$ to $n_{c_{max}}$. The extracted error, XE, is calculated for every dimensionality as the difference between the elements of the original and of the reconstructed data matrix (eqn. 7). The dimensionality yielding the smallest XE is taken to be the true number of components, n_c.

The last step in Fig. 4 consists of an evaluation of the calculated profiles: by a non-linear fitting procedure a bigaussian profile is used to estimate the area and peak width for every profile determined by means of ITT-FA. The area is based on profiles corresponding with normalized spectra (normalization to the norm, *i.e.*, the sum of squared absorbances equals 1), and is used to distinguish between minor components (less than 1% of the mixture) and major components. The minor components are ignored in the remainder of the analysis and the relative peak areas of the other components are calculated. The peak width is used to determine the resolution from neighbouring components, for the subsequent selection of reliable spectra.

The above procedure is illustrated in Fig. 5 for the third cluster in chromatogram A1. The number of components based on the value of RE was four. Since the number of peak maxima observed was two, the ITT-FA was performed with two, three and four components, resulting in XE values of 0.9, 2.6 and 6.9. As a comparison, the ITT-FA was also performed with an assumed dimensionality of 1, resulting in XE = 310.3. The latter value is extremely high, as expected, since two peaks are clearly observed in the original chromatogram. The spectrum which corresponds with this profile will be equal to the first column of **R**, hence it will be a mixture spectrum containing elements of the spectra of both components. When three components are assumed to be present, a small profile is observed fronting the first real component, which is characterized by a rather "noisy" spectrum. This deviation is probably caused by a shift in the baseline or a small impurity in the front of the peak. Although the profile derived for this disturbance is not correct, as expressed by the increased value of XE, the confusion caused by a profile like this will be minimal since the peak area derived is negligible. However, if four components are assumed to be present, a large additional peak is observed, probably corresponding with non-linearities in the observed absorptions as indicated by the similarity of spectra 1 and 4. Since this fourth profile is in fact bimodal, the error introduced by the ITT-FA will be more noticeable, as indicated by the value of XE. Obviously, these kind of disturbances will cause confusion with regard to the correct number of components in the mixture and must be avoided. By selecting the dimensionality corresponding with the minimum value of XE, the best profiles and spectra are derived to define the cluster. Consequently the number of components in the cluster was taken to be two.

The full results for all chromatograms in the two examples are summarized in Tables II (mixture A) and III (mixture B). In most cases the number of components based on the value of RE, χ^2 and/or IND is too high because of the disturbances mentioned in the Theoretical section. Especially the IND function is much too

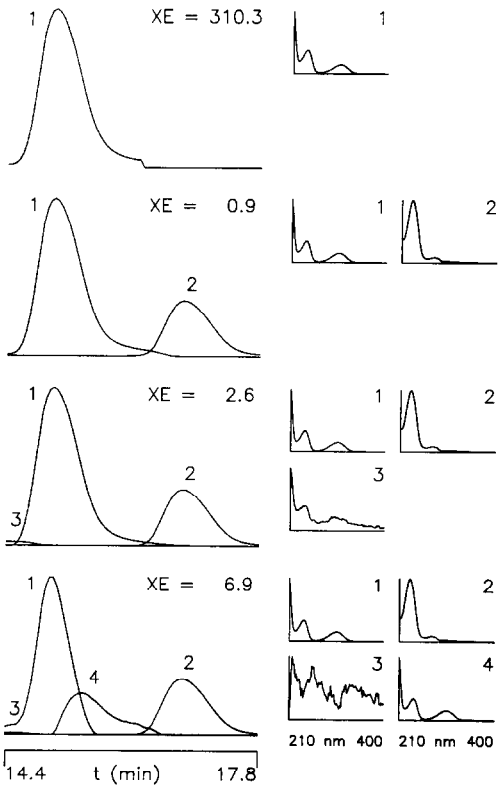

Fig. 5. The results of the ITT-FA for the third peak cluster in chromatogram A1. The analysis is performed for different dimensionalities of the cluster, resulting in the individual elution profiles and the corresponding spectra. The value of the extracted error, XE, is indicated for each analysis.

sensitive to disturbances in the chromatograhic process, even in the case of pure peaks. By comparison, the repeated execution of ITT-FA is much better, although it still overestimates the total number of solutes in the sample. If we impose a threshold of 1% for the relative peak area of a component to be considered as significant, the numbers reduce to N_{est} which are equal to the true numbers, N_{act}, in the three chromatograms. Indeed, only two discrepancies remain: in cluster 2 of chromatogram A1 and cluster 3 of chromatogram B2 an additional component is detected. The relative areas of the peaks are equal to 1.5 and 2.0% respectively of those of the total mixtures and are probably due to minor impurities or disturbances in the chromatographic process. When these solutes are also detected in chromatograms A2, B1 and B3, they will be regarded as impurities in the mixtures. However, if they are not detected in the other chromatogram(s), either because we are dealing with an artefact, or because of close coelution with a component present in excess, they are ignored because of the low concentrations. Since only thirteen components are detected in chromatogram A2 and nine components in both chromatograms B1 and B3, these two solutes will not be detected in the other chromatograms. The point is stressed here because in practical experiments it is almost impossible to avoid these kinds of disturbances, and automated procedures should be equipped to handle them.

TABLE II

ESTIMATED DIMENSIONALITIES OF THE CLUSTERS IN THE TWO CHROMATOGRAMS OF THE THIRTEEN-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 two chromatograms are identified as A1, recorded in 60% methanol, and A2, recorded in 42% THF (Fig. 1).

Cluster	*RE*	χ^2	*IND*	*ITT*	N_{est}	N_{act}
A1-1	10	10	21	8	6	6
A1-2	5	5	10	4	4	3
A1-3	4	4	9	2	2	2
A1-4	4	4	12	3	2	2
A1 total	23	23	52	17	14	13
A2-1	3	2	7	2	1	1
A2-2	6	5	11	5	4	4
A2-3	3	3	7	2	2	2
A2-4	2	2	5	1	1	1
A2-5	4	4	13	3	3	3
A2-6	2	2	8	1	1	1
A2-7	3	3	13	1	1	1
A2 total	23	21	64	15	13	13

The selection of reliable spectra

Step 3 in Fig. 3 refers to an evaluation of the calculated spectra and profiles. The results of the ITT-FA are dependent on the extent of chromatographic resolution, the spectral similarity between the components and the relative concentrations of the components[13]. Since the peak tracking is based primarily on a comparison of spectra, the quality of the calculated spectra, *i.e.*, the similarity between the true and calculated spectrum expressed by the correlation coefficient, ρ, is essential. Based on simulations of two-component clusters, described by bigaussian profiles, relationships to derive this quality by evaluating the observed chromatographic resolution, observed spectral similarity of both components and observed relative concentrations were developed. Especially in the case of low resolution, these observed values will deviate from the true ones due to reconstruction errors in the ITT-FA. In general, when two components are reasonably resolved (resolution > 0.4), not too dissimilar ($\rho_{12} > 0.5$) and present in approximately equal concentrations, the calculated spectra will have a high quality ($\rho > 0.999$), sufficient for reliable results in comparison with spectra observed in other chromatograms. However, at lower resolution or in the case of large differences in relative concentrations, this quality will not be obtained for one or both calculated spectra, in which case they will not qualify for use in a direct comparison. Although the relationships applied were derived for two-component clusters, an extension to larger clusters by means of a pairwise evaluation of subsequent peaks seems justified to a first approximation[26].

However, application of the above procedure in practical situations sometimes

TABLE III

ESTIMATED DIMENSIONALITIES OF THE CLUSTERS IN THE THREE CHROMATOGRAMS OF THE NINE-COMPONENT MIXTURE

The three chromatograms are identified as B1, recorded in 50% methanol, B2, recorded in 32% THF and B3, recorded in 38% acetonitrile (Fig. 2). Other details as in Table II.

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

leads to acceptance of components not quite fulfilling the demands imposed on the spectral quality, *i.e.*, in the case of closely coeluting components (resolution $R < 0.4$) with high spectral similarity ($\rho_{12} > 0.9$) and increased tailing not very well described by the bigaussian model used in the derivation of the relationships used in the evaluation. In order to make the selection procedure more flexible, spectra derived from closely coeluting solutes will remain suspect and candidates for improvement. Furthermore, the selection procedure is directed only at the quality of the spectra and consequently may result in incorrect relative areas in the case of components with similar spectra but different concentrations at low resolutions. If the relative areas are required in order to derive an unambiguous identification, the conclusions with regard to these clusters based on the results of the ITT-FA must be regarded with caution.

In view of the previous discussion, we make the following distinctions.

A cluster is considered fully resolved when only one component is present or when all components are sufficiently resolved with respect to the concentrations (minimum resolution 0.4) to rely on both spectral data and related concentrations.

A cluster is unresolved when one or more of the components involved are uncertain with respect to the spectra and/or relative areas. These clusters are analysed in more detail in the next steps of the procedure.

A component is fully defined when both its spectral characteristics and its relative area are known.

A component is tentative when some assumptions are available with respect to its spectrum, but due to the close proximity of other undefined or tentative components the exact characteristics are not yet defined.

A component is undefined when, due to its low concentration or close proximity to other components, its spectral characteristics are unknown.

The spectra selected on the basis of the above evaluation, are stored in a library identified by the code of the chromatogram. For instance, chromatogram A1 will produce library A1, which will be compared with library A2 derived from chromatogram A2.

The result of the selection procedure for the first example is illustrated in Fig. 6a. The required spectral quality was set at 0.995, a practical compromise between an high correlation, ensuring acceptance of only correct spectra, and a larger deviation from the ideal value of 1 in order to include most components in the library. Since some changes in the spectra are often observed due to differences in the mobile phase, extreme high demands on the quality of the spectra will be impractical. Obviously, the resolution in the second cluster of chromatogram A2 is too low to rely on the derived

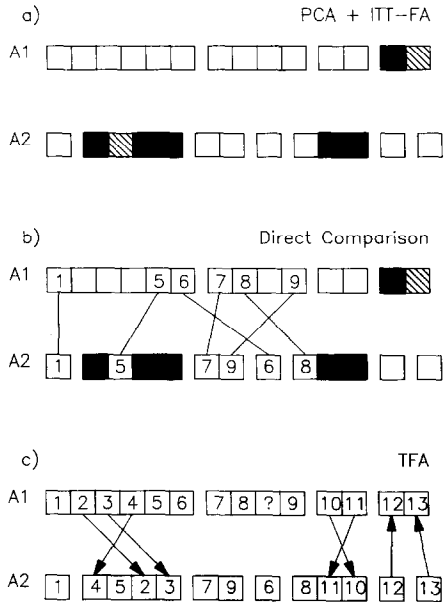

Fig. 6. The results of the three steps in the peak tracking procedure for the two chromatograms of the thirteen-component mixture, A1 and A2. Numbers refer to Table I. The unknown component is indicated by "?". (a) The conclusions of the evaluation of the spectra. Spectra are defined as: reliable (□), tentative (▤) or unreliable (■). (b) Components recognized by a direct comparison, indicated by a line connecting matched solutes. (c) Components recognized by means of TFA. Arrows point from the targets to the clusters.

pure component spectra (the observed resolutions were 0.33, 0.40 and 0.26 and consequently the real resolutions are approximately 0.25, 0.34 and 0.10; ref. 13). Only the spectrum of the second component is included in the library as tentative, because of an higher concentration and consequently a more reliable spectrum. The same holds for the last two components in the fifth cluster (observed resolution 0.28, real resolution 0.15). The last cluster of chromatogram A1 contains two components with resolution 0.29 (real resolution approximately 0.17). Since the concentration (UV activity) of the second component is much higher than that of the first one, its spectrum is tentatively used in the subsequent comparison. As a result, the first example contains thirteen (perhaps fourteen) components (Table II), and produces two spectral libraries: A1, containing twelve reliable and one tentative spectrum, and A2 containing seven reliable and one tentative spectrum. In chromatogram A1 one unresolved cluster remains to be analysed in the next steps of the procedure since it contained one unreliable spectrum. In chromatogram A2 two unresolved clusters remain: cluster 2 (three unreliable spectra) and cluster 5 (two unreliable spectra).

The same procedure is followed for the second example (Fig. 7a): the chromatographic resolution in the first cluster of chromatogram B1 is very low (observed 0.32, estimated 0.23), however, due to the large difference in concentration, the spectrum of the first component is tentatively used in the spectral comparison. The resolution between the first and second components in the third cluster of chromatogram B2 is 0.48 (estimated real resolution 0.44) but, due to the fact that the concentration of the first component seems very low (observed fraction in the cluster 0.13, estimated true fraction 0.09), its spectral characteristics are not included in the library. The same holds true for the last component (impurity) in the cluster. The spectral characteristics of the central component are probably sufficiently accurate, but the component is regarded as tentative because of the overall uncertainty in the cluster. Although a fraction of 0.13 cannot be considered negligible, the ITT-FA does not perform well when used to analyse a cluster with widely varying concentrations[13]. The identity of the component(s) present in low concentration must be determined by other methods (step 5 in Fig. 3). Finally, the resolutions between the three components in the fourth cluster of chromatogram B3 are estimated to be 0.25 and 0.32. Despite an high spectral similarity, the spectral characteristics and relative areas cannot be relied upon, and are not included in the library. Apparently we are dealing with a nine- (perhaps ten) component mixture (Table III) which produces three spectral libraries: B1 and B2, both with seven reliable and one tentative spectrum and B3, with six reliable spectra. In chromatogram B1 the first cluster is unresolved (one spectrum was unreliable), in chromatogram B2 the third cluster is unresolved (two unreliable spectra) and in chromatogram B3 the fourth cluster must be examined in the next steps of the procedure (three unreliable spectra).

The direct comparison of spectra

The fourth step in the peak tracking procedure consists of a direct comparison of the spectra (both the reliable and tentative ones) derived from the chromatograms. When all components differ sufficiently in their spectral characteristics a close match between the spectra, expressed by a correlation coefficient, is adequate for identification purposes. However, when two or more components have very similar spectral characteristics, for instance in the case of homologues, an additional source of

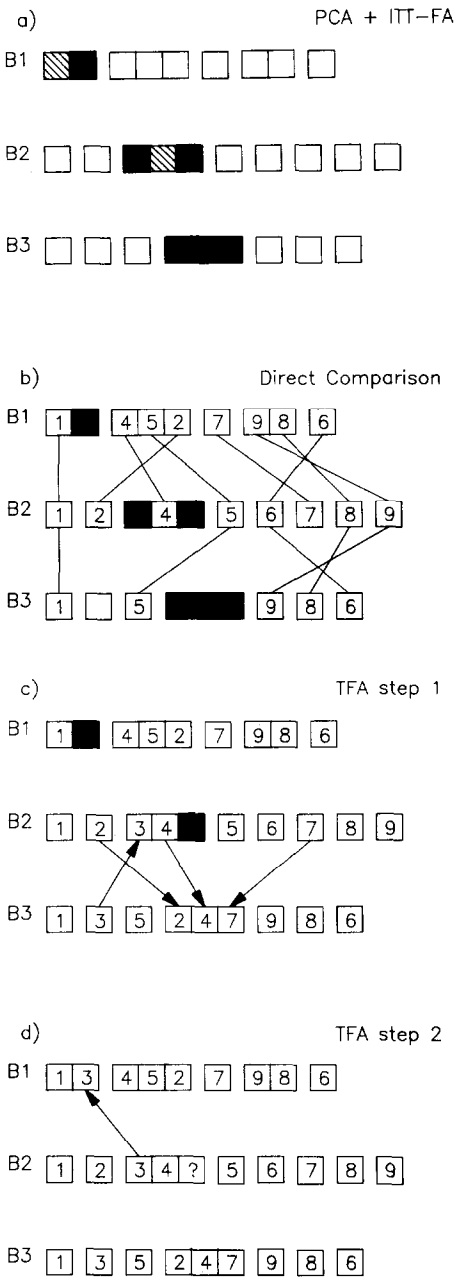

Fig. 7. The results of the four steps in the peak tracking procedure for the three chromatograms of the nine-component mixture, B1, B2 and B3. Numbers refer to Table I. The unknown component is indicated by "?". (a) The conclusions of the evaluation of the spectra. Spectra are defined as: reliable (□), tentative (▤) or unreliable (■). (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. Arrows point from targets to clusters. (d) Component recognized by means of TFA using the updated spectral library of chromatogram B2.

information is required. This is supplied by the relative areas of the peaks. In the rare situation that two different components in the sample have identical spectra as well as equal areas, it will be impossible to distinguish between them and one has to work with two or more hypotheses to describe the retention behaviour of the mixture. As a first assumption, the hypothesis without cross-over of peaks seems the most reasonable one, since homologues will usually react in a similar fashion to changes in the specificity of the solvent. This assumption is equivalent to the assumption of limited variation in retention[12].

During the comparison of the spectra (and areas) of two chromatograms, two additional libraries are produced containing those spectra which are not observed in the other chromatogram. Again this is decided on the basis of a critical value of the correlation coefficient, for instance $\rho > 0.995$, and an assumed variance in the relative area, for instance 10%. Due to the variable nature of the chromatographic process, the spectral comparison is used as a first selection criterion, and the relative area is used to check the validity of the identification. When a solute from a series of spectrally similar components is not yet included in a library, this will be detected by the comparison of relative areas as well (provided the solutes do not have both similar spectra and equal area). This is illustrated in Table IV: when comparing the spectrum of the second component of chromatogram B2 with the spectra in library B1, two spectra with high correlation are observed. By examining the relative areas of the two components and comparison with the relative area of component B2-2, component B1-5 is clearly matched with the component in question. Since this component is missing in library B3 due to close overlap with other components, only a considerable similarity with one of the other solutes is observed. This is detected by a large difference in the relative areas (13.8 *vs.* 7.9%), and consequently component B2-2 is not yet found in chromatogram B3, so that its spectral characteristics are retained for use in the next step of the procedure.

When processing more than two chromatograms, one starts by selecting the chromatogram with the maximum number of reliable spectra. Next, the contents of the

TABLE IV

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			

library related to this chromatogram are compared to all other libraries collected during the previous step in the procedure.

The results for both examples are presented in Figs. 6b and 7b. Example A is based on a comparison of spectral characteristics only, resulting in a recognition of five components. The tentative spectrum (corresponding with component 13) is not matched with any of the spectra in library A2. Consequently a library A1a containing the spectra of solutes 2–4, 10, 11 and 13 (tentative) of chromatogram A1 is created. At the same time a library A2a with solutes 12 and 13 of chromatogram A2 is constructed.

Example B is analysed in two steps. First the library derived from chromatogram B2 is compared with library B1. The tentative components are recognized since both spectra and areas are matched with solutes in B1 and consequently are now considered as "defined". Since all components in library B2 are matched with the spectra in library B1, no spectra remain to be used in the next step of the analysis. The next comparison involves libraries B2 and B3. Five components are identified unambiguously, indicated by the lines and numbers in Fig. 7b. Despite high correlation 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 second component of library B3, since no direct match was observed between this spectrum and the spectra in library B2.

Analysis of unresolved clusters

The unresolved clusters remaining after the PCA and ITT-FA are analysed using the libraries created in the previous step in a TFA. Since the requirements with respect to the resolution are considerably reduced, clusters which cannot be analysed by means of the ITT-FA can be resolved, provided the spectra of the components are known.

The procedure continues as follows: for every unresolved cluster one determines which pure component spectra ("defined components") are available from the cluster itself because of a reliable performance of the ITT-FA for the corresponding section of the cluster, for instance component 8 when analysing the fifth cluster in chromatogram A2. These components are required to determine the complete transformation matrix **T** and have often been identified in the previous step of the procedure. The spectra are added to the library containing the unidentified spectra from the other chromatogram and the TFA is performed. On the basis of the results of the projection (eqn. 4), the relevant spectra are selected from the library. The inverse transformation is applied to derive the corresponding elution profiles. However, before the profiles, and hence retention times (required for the optimization procedure!), can be determined, spectra of all components in the cluster should be available. When only minor components in a cluster are not yet identified, the retention times can be estimated by performing the TFA assuming a reduced dimensionality. In the case of severe spectral similarity within the mixture, the relative areas of the profiles can now be determined and used as an additional indication of the identity.

An example of this approach is given in Table V: the library A1a is used in a target test on the unresolved clusters of chromatogram A2, starting with the second cluster. First the library is extended by including the spectrum of component 5 as determined by means of ITT-FA in step 2 of the procedure. After projecting the spectra, the components 2–4 (and obviously component 5) are selected as solutes

TABLE V

RESULTS OF THE TFA WITH FIVE SPECTRA SELECTED FROM LIBRARY A1, STORED IN LIBRARY A1a, ON THE SECOND AND FOURTH CLUSTERS OF CHROMATOGRAM A2

The success of the test is expressed by means of the correlation coefficients, ρ_2 and ρ_4, between the target spectra and their projections on the hyperplanes related to the clusters. The conclusion expresses the number of the cluster in which the components are situated in chromatogram A2.

Target spectrum	*Cluster A2-2* ρ_2	*Cluster A2-4* ρ_4	*Tested spectrum present in cluster*
A1-2	0.99623	0.42776	2
A1-3	0.99722	0.55250	2
A1-4	0.99962	0.70628	2
A1-10	0.83222	0.99818	4
A1-11	0.92815	0.99974	4

present in cluster 2 on the basis of the high correlation between the target spectra and their projections. Since all major components have been identified in this cluster, the inverse transformation is performed and the cluster is fully resolved. Similarly, the components 10 and 11 (in combination with component 8) are identified in cluster 5 and used to determine the individual profiles. In this way, all unresolved clusters are analysed and the locations of the missing components are determined.

When more than two chromatograms are involved, the procedure is repeated a number of times: when a new component has been identified on the basis of TFA, its spectrum is removed from the library used in the TFA and added to the library corresponding with the chromatogram in which the compound has been identified. This adjusted library is then used in a direct comparison with the other chromatograms involving only those components not yet recognized in a previous step. New components not identified in a chromatogram by a direct comparison are added to the library used in the TFA of unresolved clusters in that chromatogram, etc. These steps are repeated until all components are identified or the contents of the libraries remain constant. This is summarized in Fig. 8.

The results for example A are illustrated in Fig. 6c. By a combination of the spectral data in both chromatograms all components in chromatogram A2 are identified, and only one impurity in chromatogram A1 remains unrecognized. Example B requires two steps, illustrated in Fig. 7c 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 (2,4,7) contained in library B2a are all located in the fourth cluster of chromatogram B2. Since it was the only remaining cluster, this result was expected, but the principle can be applied when more unresolved clusters are present in the chromatogram (example A); *vise versa*, the only remaining unidentified spectrum from B3 is found in the unresolved cluster of chromatogram B2 (correlation 0.99924). We are left with chromatogram B1. In the preceding step all of the known reliable spectra from B2 were 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 of chromatogram B1 (Fig. 7d). Since an high correlation is observed, nine components are now recognized in all three chromatograms, and only

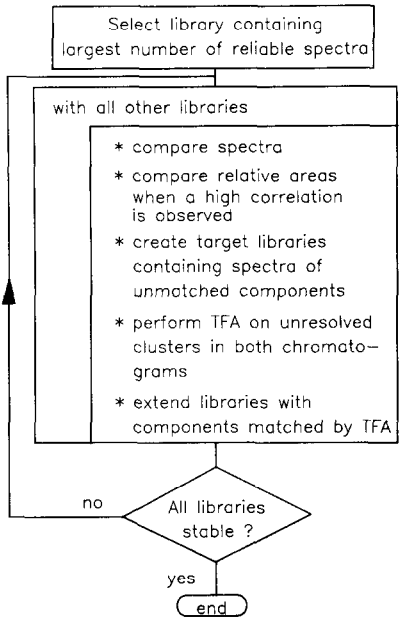

Fig. 8. The comparison and combination of spectral data derived from the separate chromatograms.

the impurity observed in the third cluster of the second chromatogram remains unidentified.

The conclusions with respect to the peak tracking in the two examples are summarized in Figs. 9 and 10. Tables with retention times corresponding to the individual solutes are now available for use in an interpretive optimization strategy[6].

Limitations

The procedure described in this section has a number of limitations. First and foremost it is assumed that the spectral characteristics of the components do not change too much with varying experimental circumstances. It is clear that none of the methods described in the introduction can be applied if the absorption characteristics of the components change, for instance because of dissociation or association with varying pH. If reversed-phase chromatography is restricted to a change in organic modifier such as methanol, tetrahydrofuran and acetonitrile, such strong conformational changes do not occur. Although shifts and other changes in the spectra have been reported[10,12], the shape of the spectra remains fairly constant throughout, and an unambiguous identification is usually possible, although slight distortions are observed in the profiles derived by means of TFA. However, when other modes of liquid chromatography are involved (for instance ion-pairing chromatography) these assumptions no longer hold and peak tracking is limited to those components which have stable spectral characteristics. Increased uncertainty will result in the case of coelution of components with unstable spectra.

Also in the case where the spectra of the components do not change too much, a number of limitations exists. When a particular component is coeluted closely with

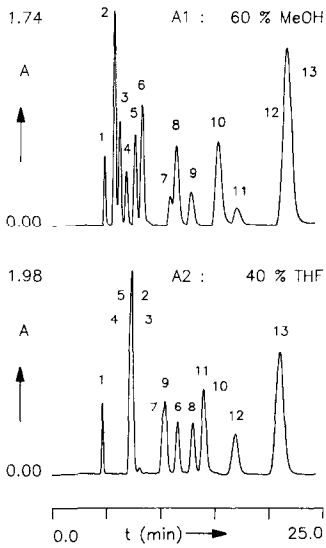

Fig. 9. The results of the peak tracking procedure for the two chromatograms of the thirteen-component mixture. Numbers refer to Table I. MeOH = methanol.

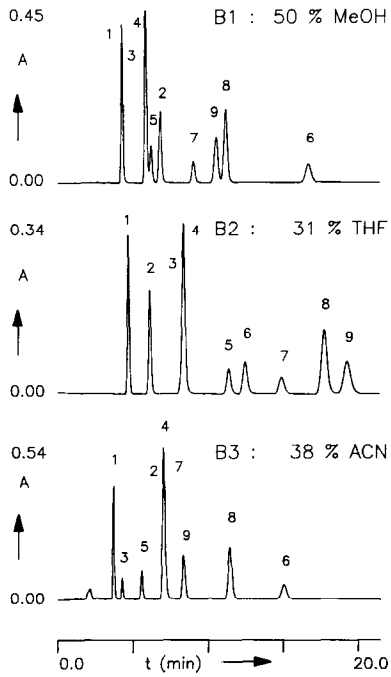

Fig. 10. The results of the peak tracking procedure for the three chromatograms of the nine-component mixture. Numbers refer to Table I. MeOH is methanol.

other components in all available chromatograms, no reliable spectrum is obtained for tracking that component. Here two situations may be observed: if the same components are coeluted together in all chromatograms, the changes of finding a suitable optimum are limited. On the other hand, a given component may be closely coeluted with different components in subsequent chromatograms. For example, between chromatograms B1 and B2 only, it would have been impossible to find the spectral characteristics, and hence the exact retention times, of component 3. Now, as long as only one component remains unknown, its uniqueness automatically allows approximate tracking. When two or more components are involved, only a limited peak tracking might be possible, resulting in several alternatives with respect to the retention behaviour of a number of components. Often it will be possible to choose one of the alternatives on the basis of subsequent chromatograms. Otherwise an increase in peak capacity (decrease in elution strength) will often result in the additional resolution required for reliable spectral data.

CONCLUSIONS

The procedure described is capable of unambiguous peak tracking in chromatograms of unknown mixtures based on a multivariate handling of spectral data resulting from multiwavelength detection. The number of analytically significant components is best estimated on the basis of the results of the ITT-FA, combined with an examination of the responses involved (UV activity).

The combination of TFA and ITT-FA eliminates the limitations of each method separately: unknown spectra of moderately coeluting components are determined by means of the ITT-FA. Clusters with too little resolution for the ITT-FA are resolved by means of TFA, using 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. Clearly, by limiting the comparison to the areas of spectrally similar components, the number of possible matches is greatly reduced, and hence the accuracy of the analysis is increased.

The major assumption is that the spectral characteristics of the components do not change much with varying experimental conditions. However, the probability of erroneous identifications is limited and the procedure issues a warning when not all components have been tracked in all chromatograms. The influence of the remaining components on the quality of the separation can be determined on the basis of a measure of concentration. Especially if procedures are to be implemented in automated systems, this last point is of great importance: a procedure should be able to indicate success or failure, rather than issue an answer at any cost.

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, 1986.
3 J. L. Glajch, J. J. Kirkland and J. M. Minor, *J. Liq. Chromatogr.*, 10 (1987) 1727.
4 L. R. Snyder and M. A. Quarry, *J. Liq. Chromatogr.*, 10 (1987) 1789.

5 S. N. Deming, J. G. Bower and K. D. Bower, *Adv. Chromatogr. (N.Y.)*, 24 (1984) 35.
6 A. Bartha, H. A. H. Billiet and L. de Galan, *J. Chromatogr.*, 458 (1988) 371.
7 A. C. J. H. Drouen, H. A. H. Billiet, P. J. Schoenmakers and L. de Galan, *Chromatographia*, 16 (1982) y48.
8 H. J. Issaq and K. McNitt, *J. Liq. Chromatogr.*, 5 (1982) 1771.
9 M. Otto, W. Wegscheider and E. P. Lankmayr, *Anal. Chem.*, 60 (1988) 517.
10 A. C. J. H. Drouen, H. A. H. Billiet and L. de Galan, *Anal. Chem.*, 56 (1984) 971.
11 A. C. J. H. Drouen, H. A. H. Billiet and L. de Galan, *Anal. Chem.*, 57 (1985) 962.
12 J. K. Strasters, H. A. H. Billiet, L. de Galan, B. G. M. Vandeginste and G. Kateman, *J. Chromatogr.*, 385 (1987) 530.
13 J. K. Strasters, H. A. H. Billiet, L. de Galan, B. G. M. Vandeginste and G. Kateman, *Anal. Chem.*, 60 (1988) 2745.
14 G. G. R. Seaton and A. F. Fell, *Chromatographia*, 24 (1987) 208.
15 E. R. Malinowski and D. G. Howery, *Factor Analysis in Chemistry*, Wiley, New York, 1980.
16 B. G. M. Vandeginste, W. Derks and G. Kateman, *Anal. Chim. Acta*, 173 (1985) 253.
17 M. McCue and E. R. Malinowski, *Appl. Spectrosc.*, 37 (1983) 463.
18 W. H. Lawton and E. A. Sylvestre, *Technometrics*, 13 (1971) 617.
19 M. Maeder and A. Zilian, *Chemometr. Intell. Lab. Syst.*, 3 (1988) 205.
20 E. R. Malinowski, *Anal. Chem.*, 49 (1977) 612.
21 H. T. Eastment and W. J. Krzanowsky, *Technometrics*, 24 (1982) 73.
22 E. R. Malinowski, *J. Chemometrics*, 3 (1988) 49.
23 J. C. Reid and E. C. Wong, *Appl. Spectrosc.* 20 (1966) 220.
24 Y. Beppu and I. Ninomiya, *Computers Chem.*, 6 (1982) 87.
25 P. R. Bevington, *Data Reduction and Error Analysis for the Physical Sciences*, McGraw-Hill, New York, 1989, Ch. 11.
26 J. K. Strasters, H. A. H. Billiet, L. de Galan, B. G. M. Vandeginste and G. Kateman, *Abstracts 12th International Symposium on Column Liquid Chromatography, Washington, 1988*, Tu-L-13.