Journal of Chromatography, 458 (1988) 335–353
Elsevier Science Publishers B.V., Amsterdam — Printed in The Netherlands

CHROMSYMP. 1531

STRATEGIES FOR AUTOMATED OPTIMISATION OF HIGH-PERFORMANCE LIQUID CHROMATOGRAPHIC SEPARATIONS INCORPORATING DIODE-ARRAY DETECTION

A. G. WRIGHT and A. F. FELL*
Department of Pharmaceutical Chemistry, University of Bradford, Bradford BD7 1DP (U.K.)
and
J. C. BERRIDGE
Pfizer Central Research, Sandwich, Kent CT13 9NJ (U.K.)

SUMMARY

An optimisation strategy based on the simplex lattice mixture design is automated by the development of a peak recognition algorithm which utilises multiwavelength detection data. The peak tracking routines are shown to deal with multiple peak overlap and extensive peak cross-over. The utility of these techniques is demonstrated for a model system of seven components. The efficiency of optimisation and quality of separation are compared, for the same test mixture, with an existing automated optimisation strategy, namely the sequential simplex procedure incorporating multichannel detection.

INTRODUCTION

High-performance liquid chromatography (HPLC) is well established as a powerful analytical tool for the separation and quantitation of mixtures. To obtain an adequate separation of all components of interest in an acceptable analysis time it is frequently necessary to adjust or optimise operating conditions. Mobile phase composition in particular is most frequently varied. For situations where a number of variables needs to be optimised simultaneously, there are formal strategies which may prove useful[1,2].

These formal optimisation strategies fall into two main categories: sequential experimental techniques and simultaneous experimental techniques. At least one strategy, however, has been developed[3] which combines both approaches.

Sequential procedures rely on search algorithms to traverse the specified response surface and locate a point of optimum or maximum response. The search is directed by the quality of preceding separations. The principal sequential method applied to HPLC separations is sequential simplex[4] or modified simplex[5]. This method has been widely used for the optimisation of factors such as: proportions of organic modifiers, flow-rate and temperature[6–9]. Sequential simplex has been successfully automated[6] and shown to be suitable for dealing with unknown samples[10].

Simultaneous experimental procedures collect data over the factor space

0021-9673/88/$03.50

according to a pre-defined scheme. These data are then used to fit an appropriate mathematical model. The use of such models allows chromatographic behaviour to be interpolated between the experimental points and optimum separation conditions to be predicted. A number of simultaneous procedures have been developed for use in HPLC and these include factorial design[11] and procedures based on simplex lattice mixture design[12–16]. In all these techniques the retention behaviour of individual solutes is modelled to enable a chromatogram to be predicted for any point on the response surface.

The hybrid technique developed by Schoenmakers *et al.*[13] is an iterative mixture design. A simple model is fitted to individual solute retention data collected from a simultaneous experimental scheme. The model is refined in a sequential, iterative process until a separation of pre-defined quality is located or the model can no longer be improved.

All three categories of optimisation strategy have both advantages and disadvantages. Table I describes some of the strengths and weaknesses for each approach, with reference to a particular technique. Of the three procedures only the sequential simplex procedure has been able so far to deal with unknown samples in a fully automated way[6,10]. Unfortunately, the existence of "local" optima may mean that an acceptable separation is not located. The two other procedures would address this problem but have not been automated for poorly characterised samples (*i.e.* where no reference standards are available) due to the need for solute recognition.

Peak recognition is a major problem which remains to be overcome. The failure of simultaneous and hybrid optimisation strategies to deal with unknown samples stems from the lack of peak recognition tools. This limitation has been widely appreciated by workers in the field of optimisation and a number of attempts have been made to overcome this difficulty. Drouen *et al.*[17] have attempted to utilise the

TABLE I

ADVANTAGES AND DISADVANTAGES OF THE MAJOR OPTIMIZATION APPROACHES

Each type is exemplified by a particular procedure.

Approach	*Advantages*	*Disadvantages*
Sequential (simplex)	No assumptions about retention behaviour	Requires many experiments (20–30)
	Can deal with unknown samples	May locate "local" optima
	Conceptually simple	Automation often limited to mixtures of two modifiers due to hardware
	Can be automated	
Simultaneous (simplex lattice mixture design)	Models whole response surface	Requires solute recognition
	Locates global and local optima	Accuracy dependent on model relationship selected
	Few experiments required (7–10)	Limited as to modifier
Hybrid (iterative regression analysis)	Models whole response surface	Requires solute recognition
	Locates global and local optima	Complex programming required when more than one variable considered
	Few experiments required	
	Models are accurate	

absorbance ratio method[18] for solute recognition. The ratio of absorbance at two detection wavelengths for a pure peak is both constant and characteristic. Drouen *et al.* have employed these ratios for peak recognition. Later work by these workers[19] utilised multichannel diode-array detectors to collect UV spectra for the components. Comparison of spectra allows peak recognition. Unfortunately, the spectral similarity between many components limits the utility of these approaches.

Closely overlapping components present extreme difficulties when it is necessary to obtain pure spectra for peak recognition. Recently powerful deconvolution techniques have been developed which can extract individual peak profiles and the component spectra from a multiple component absorbance–time–wavelength data matrix. Strasters *et al.*[20] have evaluated four of these chemometric approaches as a means of peak recognition. It was concluded that even these techniques have limitations: in most cases spectral differences and a degree of resolution are required if spectra are to be extracted. Once spectra are available recognition of solutes still suffers the problems of spectral similarity.

Other recognition techniques have also been employed: these mostly rely on peak area. Earlier work with the sequential simplex procedure[10] has shown that diode-array detection[21] offers significant potential for simple peak recognition strategies based on comparison of peak integrals determined at different detection wavelengths. Thus it was proposed to develop this strategy in order to automate an optimisation strategy based on the simplex lattice mixture design. This procedure was selected in preference to iterative mixture design as the computer programming required is less complex.

In the present studies an automated peak recognition and data handling program is developed for the optimisation of HPLC separations, on the basis of minimum selectivity, using data collected in a simplex lattice mixture design[16].

THEORY

Peak recognition has previously been attempted, in conjunction with a simplex lattice design optimisation strategy, by Issaq and McNitt[22]. They developed a computer program able to track solutes on the basis of percentage area for peaks at a single detection wavelength. The chromatographic separation that revealed the most peaks was used as a reference and all the peaks in other chromatograms were correlated to these data. Two-component overlap was dealt with by adding two reference components together and assessing the degree of fit with the chromatographic peak. Applicability of the routines was demonstrated by a hypothetical example.

Two questions remain to be answered as a result of this work: (a) how to deal with two or more components of very similar area and (b) how to deal with the case where only partial resolution of peaks is observed in the reference chromatogram itself. The peak recognition algorithm developed by Wright *et al.*[10] aimed to address the peak similarity problem by using multiwavelength detection. The problem of overlap in the reference chromatogram did not arise in this earlier work as only well resolved chromatograms were studied.

Clearly the need for well separated peaks in the reference chromatogram imposes

an impossible constraint on the simplex lattice based optimisation procedures. In order to track peaks using the existing algorithm, every peak in each chromatogram has to be well resolved. If this is the case then optimisation may be unnecessary. Thus, it was proposed to extend the existing procedure to cover both complete and partial peak overlap.

Both procedures described[10,22] require the reference data to be extracted from one of the separations, a major limitation. However, the basic assumption of any area-based recognition approach is that area (or integral) does not change significantly with mobile phase composition. If this fundamental assumption holds then the reference data for a particular solute can, in principle, be taken from any of the chromatograms where the solute is fully resolved from neighbouring peaks. In the proposed procedure complete resolution of all components is not required in a single chromatogram; it is simply necessary that complete resolution of each solute from its neighbours occurs in at least one of the seven chromatograms resulting from the mixture design.

Algorithms for automated optimisation

The programs needed for automated simplex lattice mixture design optimisation can be summarised in six steps: (i) integration of all peaks in each chromatogram at each detection wavelength; (ii) extraction of all apparently resolved peaks; (iii) reduction of resolved peak array to remove any duplicates and any composite peaks, to arrive at a representative set of reference data; (iv) correlation of chromatographic peaks in each separation with the reference data; (v) fitting of the mathematical model to retention data; (vi) calculation of optimum separation conditions.

Routines for steps v and vi are straightforward and will not be discussed further.

Integration of all chromatograms involves the simultaneous handling of several data sets, one for each detection wavelength selected. Absorbance data are encoded as data strings and stored on magnetic disk. Also included in the coded data are peak event markers which signify start of peak, apex, valley, and end of peak, indicating the appropriate points where integration is started and finished. The integration routine employs Simpson's rule for the determination of area under a chromatographic peak.

The peak event codes also offer a convenient way of selecting pure (or apparently pure) peaks for the initial reference array. If a chromatographic peak is recognised by a start code, and the end code occurs without an intervening valley code then the peak is assumed to be pure. If a valley code is observed but the height above baseline corresponds to less than 5 mAU then the peak is also regarded as pure.

Routines for reduction of the array of "pure" peaks to a representative set of reference data form the core of an automated procedure. The principles are best described schematically (Fig. 1). It is necessary to remove any multiple occurrences of the same component so that each solute is represented only once in the reference array. A complication is the presence of composite peaks in the array (Fig. 2). Extra peaks occur which cannot be deleted by comparison with the other solutes present. Combinations of the other peaks (two together, or three together etc.) have then to be compared. These comparisons between peaks require a function which can assess similarity to allow objective decisions as to whether or not a peak should be included in the reference array.

The peak recognition algorithm developed previously[10] provides an assessment

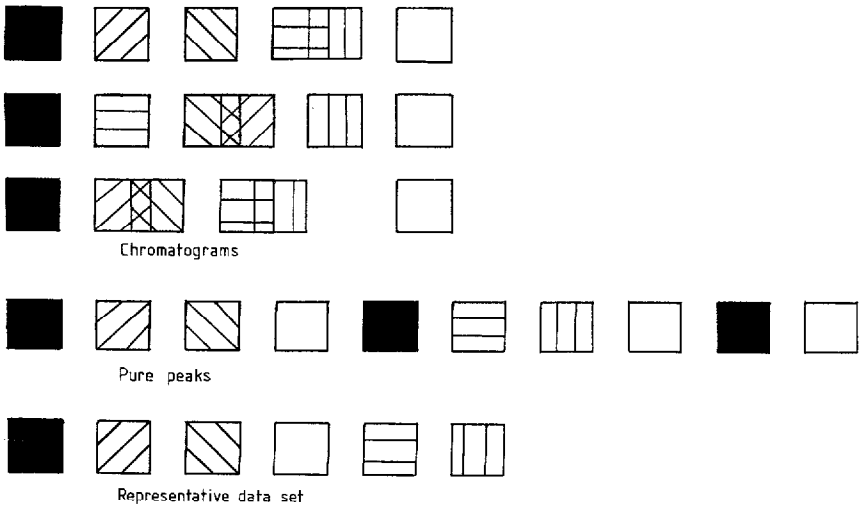

Fig. 1. Schematic diagram representing the basic principle behind the proposed program for extracting reference data from the simplex lattice design. Chromatographic peaks are represented by the squares. Any square exhibiting only one pattern is assumed to be pure. Removal of any pattern occurring more than once leads to a representative data set — a reference archive.

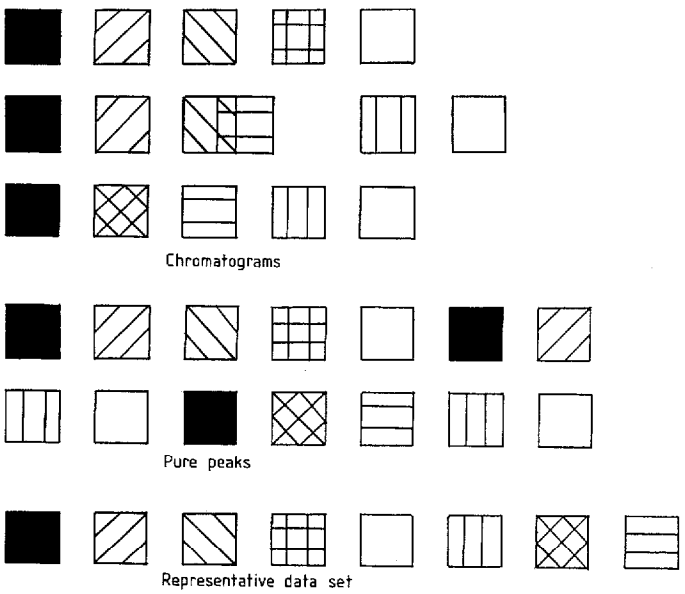

Fig. 2. Similar data showing the problem associated with complete overlap. Completely fused peaks appear to be pure but cannot be removed from the array of pure peaks. In this instance the representative data set consists of eight peaks instead of six.

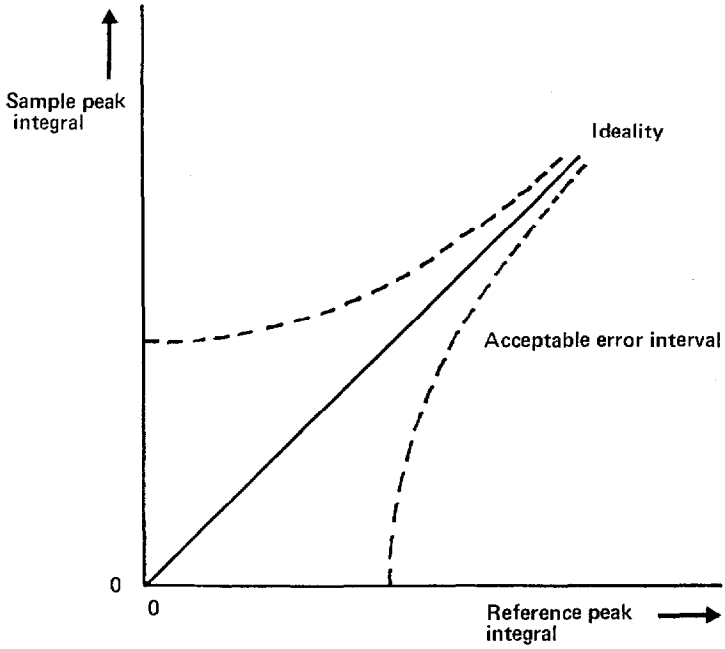

Fig. 3. Sample peak integral as a function of reference peak integral to illustrate the increase in acceptable error which should be included, as reference and sample peak integrals decrease.

of similarity between two peaks but has a major limitation, in that no provision is made for the larger degree of error associated with small relative to larger peaks (Fig. 3). The significance of the degree of difference between two chromatographic integrals is dependent upon the size of the integrals. A 30% difference when each of the integrals is less than 2% of the total chromatographic integral, is clearly less significant than a 30% difference when the integrals are greater than 10% of the total integral. The original algorithm[10] made no allowance for this and so a new algorithm was developed.

The peak percentage areas for a selected reference peak are summated over all the detection wavelengths used:

$$I_R^{tot} = \sum_{i=1}^{n} I_R^{\lambda_i} \tag{1}$$

where I_R^{tot} is total percentage area summated over all wavelengths, $I_R^{\lambda_i}$ is percentage area at wavelength λ_i, and n is the number of wavelengths used. The total percentage area over all detection wavelengths indicates the contribution the reference peak makes to the total chromatographic integral over the same detection channels. The size of this contribution should dictate the degree of difference which can be tolerated between the sample and reference peaks before it must be concluded that they arise from different solutes. The value of I_R^{tot} dictates the value of this maximum acceptable difference

TABLE II

MAXIMUM DIFFERENCES BETWEEN SAMPLE AND REFERENCE PEAK INTEGRALS WHICH MAY BE TOLERATED BEFORE THE PEAKS ARE REGARDED AS DIFFERENT

n is the number of detection wavelengths used.

Range for I_R^{tot}	*Maximum value of D as the fraction of* I_R^{tot}
$< 1.67\ n$	0.475
1.67–2.50 n	0.450
2.50–3.33 n	0.400
3.33–4.00 n	0.325
4.00–4.67 n	0.225
$> 4.67\ n$	0.100

(Table II). The actual difference at each wavelength is determined by the absolute difference in peak area for reference peak and sample peak:

$$D^{\lambda_i} = |I_R^{\lambda_i} - I_2^{\lambda_i}| \tag{2}$$

where: $I_2^{\lambda_i}$ is percentage area for a sample peak at wavelength λ_i. The total difference (D) is defined as:

$$\mathrm{D} = \sum_{i=1}^{n} D^{\lambda_i} \tag{3}$$

For the reduction of the reference array it is not necessary to calculate a numerical value for an assessment of similarity. It is sufficient to compare D with the maximum value allowed (Table II). If D exceeds the acceptable maximum then the solutes are regarded as different.

The computer program proposed in this work deals firstly with the reference array of pure (or apparently pure) peaks. The first peak is taken as a reference and values for D are determined with every other peak in the array. Where the difference does not exceed the maximum value allowed (Table II) the subsequent peak is deleted from the array as being a duplicate of the reference peak. Where the value of D exceeds the maximum allowable level the peak is assumed to be due to another component. Repeating the routine for the second and subsequent peaks in the reference array deletes any further duplicate peaks.

Once duplicate peaks have been removed, any extra sets of data, over and above the representative set, are assumed to be attributable to composite peaks. A similar approach to the one above is used to delete these additional data. The percentage areas of two or more of the remaining peaks can be added together and used as a reference for comparison with the other peaks.

The peaks of the reference array are sorted into ascending order on the basis of percentage area for the first detection wavelength selected. It is then assumed that a composite peak will be larger than any of the individual peaks overlapped to produce it. Therefore, if two peaks in the ordered reference array are combined together (*i.e.* 1 and 2), it follows that only larger peaks (*i.e.* 3 onwards) need be considered as

potential composite peaks. Calculation of the maximum allowable differences and assessment of similarity follows the same approach as before, except that the percentage areas for the combined peaks are used as reference data. The current software can consider two and three component overlap. Any peak which may be concluded to arise from a two or three peak combination is deleted from the reference array. After completion of this second process only one set of data should be retained for each solute (a representative set of reference data *i.e.* peak integrals at each detection wavelength).

The availability of a complete and unique set of reference data means that peak recognition in the other chromatograms is possible. The current recognition algorithm does not determine a numerical value, it simply allows a decision as to the inclusion or exclusion of a peak in the reference array. For recognition of peaks in chromatograms it is necessary to calculate a numerical assessment of similarity between reference and chromatographic peaks. To achieve this the recognition algorithm was amended.

To calculate a specific similarity assessment value it is still necessary to determine both I_R^{tot} for the reference peak, and D between the reference and chromatographic peaks. As discussed above the error associated with small percentage areas is less significant than that for larger percentage areas. Thus, I_R^{tot} is used to select a weighting factor to determine how much of the absolute difference should be regarded as significant (Table III).

The numerical value for measuring the similarity is calculated by taking the ratio of significant difference to I_R^{tot}. A value close to zero indicates a good fit. These calculations are not limited to single reference peaks and may be applied to comparisons of two and three reference peak combinations with chromatographic peaks.

Peak recognition in chromatograms may be achieved on the basis of these similarity values. Peak similarity is calculated for each individual reference peak with every chromatographic peak in the seven chromatograms. The three best fitting reference peaks are stored in descending order of fit. Similarity values are calculated between chromatographic peaks and combinations both of two reference peaks, and of three reference peaks. The three best fitting two-peak combinations and the two best fitting three-peak combinations are recorded.

The program for identifying the peaks in the chromatograms contains three sets of routines. Which set is used depends on the number of peaks detected. Where all

TABLE III

DETERMINATION OF THE FINITE DIFFERENCE BETWEEN THE REFERENCE AND SAMPLE INTEGRALS WHICH MAY BE REGARDED AS SIGNIFICANT

Range for I_R^{tot}	*Weighting factor*	*Significant absolute difference (D)*
$< n$	0.2	0.2
1–2 n	0.3	0.3
2–3 n	0.4	0.4
3–4 n	0.5	0.5
4–5 n	0.6	0.6
5–6 n	0.8	0.8
>6 n	1.0	1.0

expected peaks are detected in a given chromatogram, only similarity values with single reference peaks need be considered. When one less peak occurs in the chromatogram, this means that two-peak overlap must be considered. The absence of two or more peaks leads to both two and three peak overlap being examined.

The first case is where all peaks are detected. The best similarity value for each chromatographic peak (with single reference peaks) is studied. A value of 0.05 or below is assumed to correspond to a correctly assigned peak. All peaks in the chromatogram are considered in this way and as many peak identities as possible are assigned. Any remaining non-assigned peaks are then studied. The best fitting reference peak is provisionally assigned in each case. The chromatogram is then examined to check that each reference peak has only been assigned once. If this is the case then elution order has been assigned and the process can be halted for this chromatogram. If peaks are assigned more than once then any provisionally identified peak takes its second best identity and the checking procedure is repeated. A third best identity is available if necessary. Eventually an elution order can be assigned even for chromatograms where there is partial overlap of several components. A fit value of less than 0.05 should be very specific for one component. Two solutes would have to have almost identical absorptivity–concentration combinations at the detection wavelengths for both to be fitted by one reference peak.

The second case to be considered is that for one peak less than expected in the sample chromatogram. Single reference peaks are assigned to as many peaks as possible in the same way as before. The two routines diverge at this point. The chromatographic peak with the worst single peak fit is located and two-peak combination data are studied. If the similarity for the best fitting two-peak combination is less than 0.15 then the two peaks are provisionally assigned. A check is made to ensure that these peaks have not already been positively assigned to "single" components. Where these peaks are unused the identity is then positively assigned. If either or both peaks have been used then the next best fitting pair of peaks is considered. The process is repeated provided that the similarity value is less than 0.20. When a two-peak combination cannot be assigned the second worst fitted peak from the single reference peak data is located and these routines repeated. A check that all reference peaks are assigned only once is carried out. A process analogous to the one used with single reference peaks is applied to arrive at the correct elution order.

The third case is for two fewer peaks than expected. As for the previous case, single reference peak fits are used to assign identity to as many peaks as possible. Once this has been done, the worst single peak fit is located and the three-peak combination data examined. The process is analogous to that for two-peak combinations. However, only one peak in the chromatogram is considered and the second to worst fitted peak is not evaluated. If a three-peak combination does not satisfactorily describe the worst fitted peak then it is assumed that the discrepancy is due to two instances of two-peak overlap. Thus, the two worst fitted peaks are studied to see which two combinations fit best. The process is identical with that used for the first part of the routine for fitting two overlapped peaks as described above. Elution order is established by ensuring that each peak is used only once.

These proposed algorithms should enable: peak recognition; the determination of elution orders; and the assignation of the correct retention time to each solute from every chromatogram. When these data are supplied to a routine for fitting to the

special cubic equation[12–14], retention time can be predicted at any point on the response surface for any of the solutes provided that the model employed is adequate[23]. These retention models allow optimisation of the separation.

In the present work a model system exhibiting peak overlap and multiple peak cross-over is used to exemplify the new procedure. Such a sample also provides a very severe test of the sequential simplex procedure since multiple elution orders mean there are many local maxima.

EXPERIMENTAL

Apparatus and materials

Two systems were used for this work. Automated simplex lattice mixture design employed a Hewlett-Packard (Wokingham, U.K.) 1040A diode-array detector, an HP85B computer controller, an LDC (Stone, U.K.) Constametric 3000 pump and a Rheodyne (Alltech, Carnforth, U.K.) 7010 injection valve fitted with a 20-μl loop. Separations were performed at room temperature (*ca.* 21°C) on a 125 × 4.6 mm I.D. Partisphere C_{18} 5-μm column (Whatman). The flow-rate used 1.25 ml/min.

Automated sequential simplex procedures employed a Hewlett-Packard 1090A liquid chromatograph comprised of an HP 1040A diode-array detector, a DR5 ternary pumping system and a 100-position autosampler. The system was controlled by an HP85B computer. Separations were performed at controlled temperature (*ca.* 27°C) also using a Partisphere column. The flow-rate used was 1.25 ml/min.

Mobile phases consisted of different mixtures of methanol, acetonitrile, tetrahydrofuran (THF) (Rathburn Chemicals, Peebles, U.K.) and distilled water. A seven component model system consisting of benzyl alcohol (BA), *p*-cresol (PC), propyl *p*-hydroxybenzoate (PHB), butyl *p*-hydroxybenzoate (BHB), diethyl phthalate (DP), toluene (TOL) and benzophenone (BP) was used throughout.

Software

Programs for the sequential simplex optimisation procedure were written in BASIC as a series of "hook" programs on the HP85B. The quality of separation was assessed using a chromatographic response function (CRF)[10]. The maximum number of injections was limited to 25.

Programs for the simplex lattice mixture design were written in BASIC on the HP85B as separate programs called from a short linking program.

RESULTS AND DISCUSSION

In this study the aim was to demonstrate the applicability and feasibility of an automated simplex lattice mixture design based on simple peak recognition procedures. The model system was selected to provide an exacting test of the proposed procedure.

Techniques based on the simplex lattice mixture design in HPLC require appropriate isoeluotropic eluents of water with each of the following: methanol, acetonitrile and THF. In this instance methanol–water (60:40, v/v) producing an analysis time of *ca.* 8 min was selected by an iterative procedure. A single gradient scan would also allow the appropriate composition to be predicted[24,25], although it does

suffer some limitations. The methanol–water composition found was used to predict the acetonitrile–water and THF–water eluents of the same solvent strength using the rules developed by Schoenmakers *et al.*[19]. These compositions were predicted to be 46:54 (v/v) and 40:60 (v/v), respectively. A "fine tuning" adjustment of composition to obtain three nominally isoeluotropic eluents was required, resulting in the following being selected: methanol–water (60:40, v/v), acetonitrile–water (47:53, v/v), and THF–water (42.5:57.5, v/v). Recently Haddad and Sekulic[26] have published an iterative procedure for the fine tuning of eluent composition which may be more efficient than the process used here.

Chromatograms were run for the seven mobile phases required by the simplex lattice mixture design[12–14]. Data were acquired simultaneously at 240, 260 and 280 nm. Each set of chromatographic data was stored on magnetic disk for subsequent integration and processing.

The number of peaks detected in the chromatograms ranged from five to seven. In no chromatogram were all peaks completely resolved (Fig. 4) and so the sample was a good candidate for optimisation. Applying the suite of programs described above allowed retention data to be determined for each solute from every one of the seven chromatograms. These data enabled a special cubic function[12–14] to be fitted. This function describes retention behaviour for every point on the response surface. Glajch *et al.*[27] found that the most accurate models were those where the logarithm of capacity factor was modelled. Therefore, the special cubic function fitted to these data was:

$$\ln k' = a_1x_1 + a_2x_2 + a_3x_3 + a_{12}x_1x_2 + a_{13}x_1x_3 + a_{23}x_2x_3 + a_{123}x_1x_2x_3 \quad (4)$$

where x_1, x_2 and x_3 are proportions of isoeluotropic eluents 1, 2 and 3 respectively, a_1–a_{123} are coefficients calculated from the seven experimental points of the simplex lattice design.

The optimisation criterion selected for use with this automated procedure was selectivity (α) for the least well separated pair of adjacent peaks (*i.e.* the minimum α). Laub and Purnell[28] first used selectivity as an optimisation criterion, but it was Weyland *et al.*[16] who applied it to a simplex lattice design based procedure to generate minimum alpha plots. The selectivity between the least well separated pair of adjacent peaks was plotted against mobile phase composition. The highest point on the minimum alpha plot (MAP) corresponded to the best separation for the worst separated pair of peaks. All other peaks were better separated.

The use of minimum α as an optimisation criterion has been criticised since the separation of other peaks may suffer by concentrating on just the worst pair. However, in this work an adequate separation of all peaks was the only consideration. Thus, if the worst peak pair could be separated it follows that all other peaks should also be resolved. With this criterion, the only parameter which had to be measured or predicted was the column dead or void time, t_0. Knox[29] proposed a simple relationship which allowed t_0 to be estimated from the column dimensions:

$$t_0 = \frac{\pi d_c^2}{4} \varepsilon_0 \cdot L \cdot \frac{1}{F} \quad (5)$$

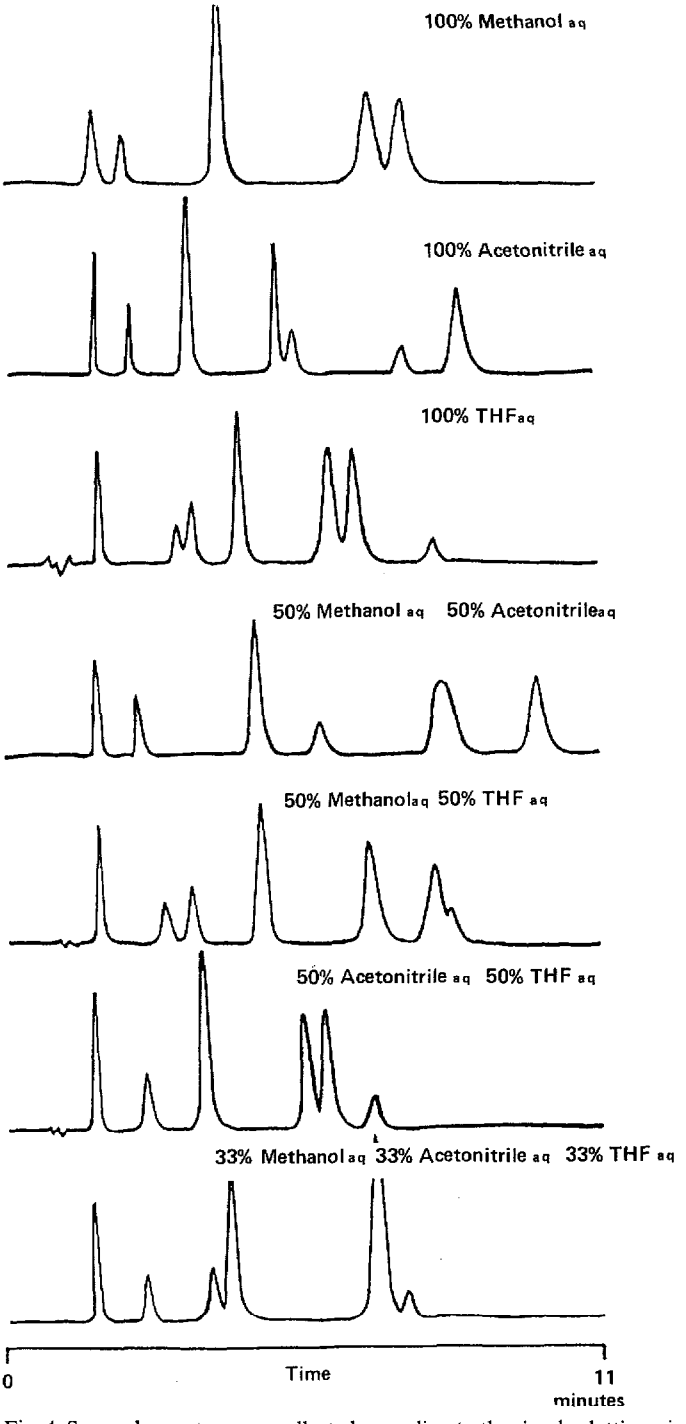

Fig. 4. Seven chromatograms collected according to the simplex lattice mixture design described (aq means aqueous).

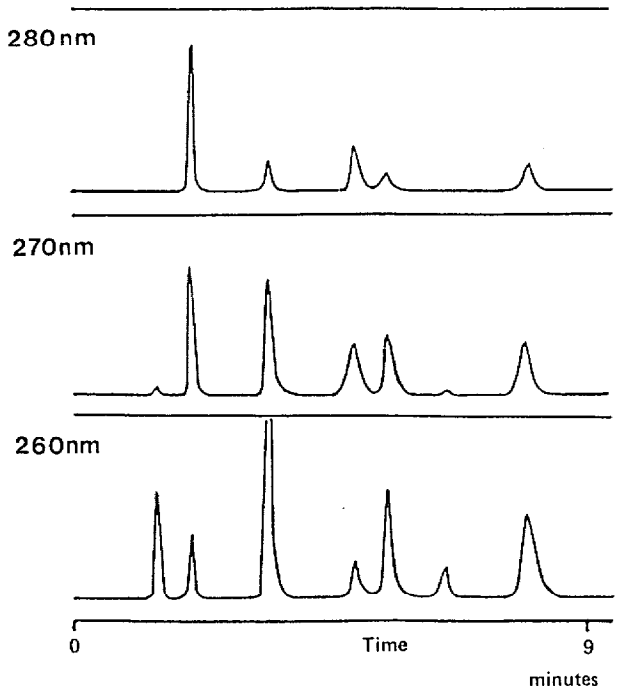

Fig. 5. The predicted optimum separation at different detection wavelengths. Mobile phase: 60% (v/v) aqueous methanol–47% (v/v) aqueous acetonitrile (25:75, v/v). Minimum α: 1.15.

where d_c is the column diameter, ε_0 is total porosity (taken as 0.75 for silica based packings), L is column length, and F is the flow-rate. These variables were known therefore allowing an estimate of t_0 (*ca.* 1.25 min for this column and these conditions).

The recognition algorithm would also allow many different criteria to be used, including those where specific peaks are of particular interest. Weighting factors may be applied to the resolution or separation of these components and a recognition procedure would enable this without the injection of standards.

The final optimisation program in the suite predicts an optimum by grid search of the predicted response surface. The grid uses 5% steps in eluent composition and involves more than 200 points. Optimum separation conditions of 47% (v/v) aq. acetonitrile–60% (v/v) aq. methanol (75:25, v/v) were predicted. The minimum selectivity expected for this point was 1.20. A chromatogram run with these separation conditions produced baseline resolution of all components and a minimum selectivity of 1.15 (Fig. 5). The error in selectivity is due to the lack of fit of the special cubic function and possible errors in the estimated value of t_0.

Absolute identity of the component peaks in each chromatogram was established by examination of the elution orders assigned and injection of individual reference standards for two sets of separation conditions (the methanol–water binary eluent and the optimum separation conditions) (Table IV). The complexity of retention behaviour was immediately apparent.

The programs developed were shown to deal with situations where two-

TABLE IV

ELUTION ORDER ENCOUNTERED FOR THE SEVEN EXPERIMENTS OF THE SIMPLEX LATTICE MIXTURE DESIGN AND THE OPTIMUM, TOGETHER WITH THE THREE CHROMATOGRAMS DESCRIBED BY SNEE TO ASSESS THE DEGREE OF FIT OF THE SPECIAL CUBIC MODEL

Two-letter codes together signify total overlap.

Mobile phase composition			*Peak*						
Methanol (%)	*Acetonitrile (%)*	*THF (%)*	*1*	*2*	*3*	*4*	*5*	*6*	*7*
100	0	0	BA	PC	PHB/DP	BHB/TOL	BP	–	–
0	100	0	BA	PC	PHB	BHB	DP	TOL	BP
0	0	100	BA	PC	DP	PHB	BP	BHB	TOL
50	50	0	BA	PC	PHB	DP	BHB/TOL	BP	–
50	0	50	BA	PC	DP	PHB	BP	BHB	TOL
0	50	50	BA	PC	PHB/DP	BHB	BP	TOL	–
33	33	33	BA	PC	PHB	DP	BHB/BP	TOL	–
25	75	0*	BA	PC	PHB	DP	BHB	TOL	BP
67	16	16	BA	PC	DP	PHB	BHB/TOL/BP	–	–
16	67	16	BA	PC	PHB	DP	BHB	TOL	BP
16	16	67	BA	PC	DP	PHB	BP	BHB	TOL

* Optimum separation conditions with aqueous isoeluotropic organic phases (see text).

component overlap occurs, and where two-component overlap occurs twice in the same chromatogram. Three further chromatograms were run (eluent compositions in Table IV) as potential checks of degree of fit for the special cubic model as outlined by Snee[14]. While these data were not considered in this study they were included for future extensions of the method. Elution orders were correctly assigned for these chromatograms and an instance of three-component overlap was satisfactorily identified. The limited computer memory available on the instrument used for these studies means that more complicated overlap situations cannot be dealt with.

The elution orders encountered over the response surface were plotted (Fig. 6). In all, seven different elution orders were predicted and each had a corresponding local maximum. A contour MAP (Fig. 7) was generated to establish which elution order or orders were capable of yielding an adequate separation. From this plot it was concluded that two of the local maxima would yield adequate separations although the optimisation program had in fact correctly identified the global optimum.

This seven-component test mixture was also optimised using the automated sequential procedure[10]. The use of a ternary solvent delivery system meant that methanol, acetonitrile, THF and water could not all be included in the same procedure unless isoeluotropic binary eluents were used. The isoeluotropic compositions employed were those from the simplex lattice design although column temperature was more than 5°C higher than with the simplex lattice procedure. As a result analysis time was reduced from 8 to 5 min. Thus, minimum α values for separations resulting from the two procedures were taken as the criterion for comparison as a 5°C increase in temperature was not expected to significantly alter selectivity. Comparisons of

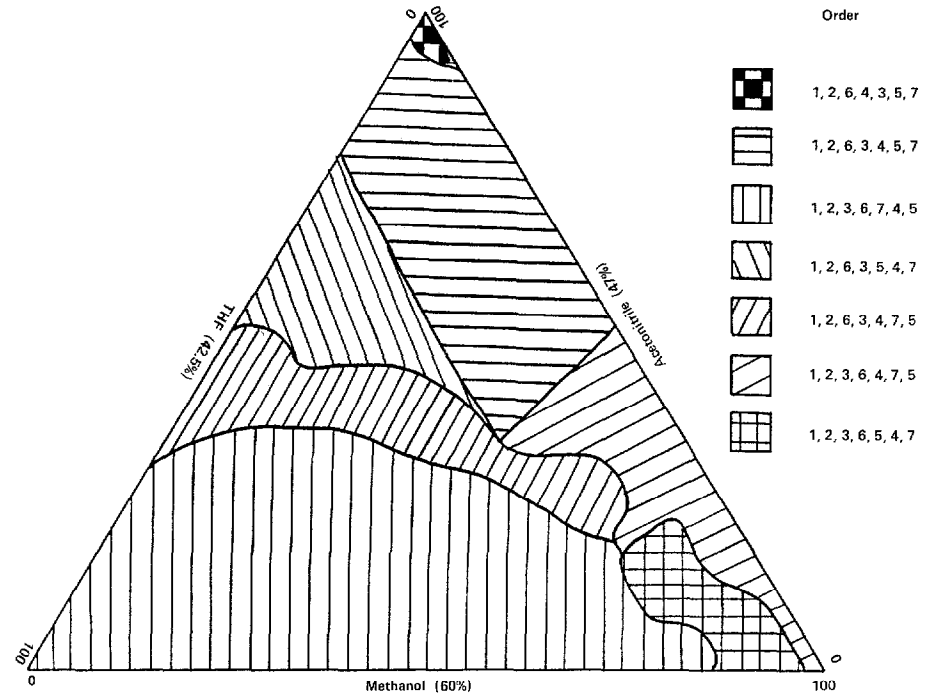

Fig. 6. Plot of elution orders encountered over the entire response surface. Peaks: 1 = benzyl alcohol, 2 = *p*-cresol, 3 = diethyl phthalate, 4 = butyl *p*-hydroxybenzoate, 5 = toluene, 6 = propyl *p*-hydroxybenzoate, 7 = benzophenone.

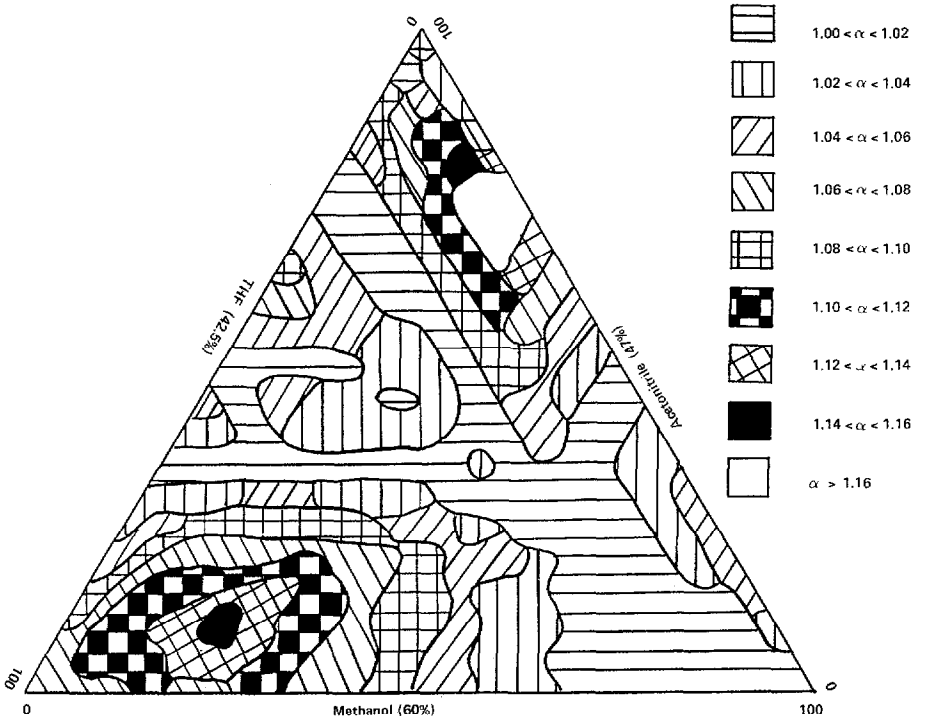

Fig. 7. Contour plot of minimum α values over the response surface.

resolution were not suitable criteria as decreasing the average capacity factor for two peaks leads to a decrease in their separation[2].

The CRF used for the sequential simplex optimisation study was:

$$\text{CRF} = \sum_{i=1}^{n-1} R_s + n - |t_A - t_L| \quad (6)$$

where R_s is the resolution between adjacent peaks (limited to a maximum value of 1), n is the number of peaks detected, t_A is the required retention time for the last peak, and t_L is the actual retention time for the last peak. The final term was only included if it exceeded 1 min. The time term was of little importance to the CRF as the use of iso-eluotropic eluents was designed to constrain analysis time.

On carrying out the sequential simplex optimisation procedure for the separation, the stop criterion[10] halted the search after 20 injections as the CRF had attained values within 5% of the maximum CRF on three occasions. The best separation encountered was selected for comparison with the simplex lattice mixture design optimum. Baseline resolution of all components was not achieved and the minimum α value was 1.10. The eluent composition producing this separation was: 60% methanol–47% (v/v) aq. acetonitrile–42.5 (v/v) aq. THF (7.0:61.1:31.9, v/v/v) (Fig. 8).

The complexity of retention behaviour for this test mixture meant that the sequential simplex procedure was unable to fully optimise the separation. A number of pairs of peaks in the sample proved difficult to resolve. As the separation of one pair of peaks was improved the separation of another pair deteriorated.

Since both simultaneous and sequential experimental optimisation strategies were applied to the same test sample, it was possible to compare them. In this case multiple peak cross-over meant that the simultaneous procedure, capable of modelling the entire response surface, located a better separation than the sequential procedure, which relied on a search algorithm. The problems associated with local optima and the sequential simplex procedure were highlighted by the complex nature of the response surface.

The simplex lattice mixture design requires the selection of appropriate isoeluotropic binary eluents of water with methanol, acetonitrile, and THF. In these studies an iterative procedure was employed, necessitating two chromatograms for each of the three eluents. These six chromatograms required about 2 h to complete. Once isoeluotropic eluents were located seven mixed eluents were prepared and the chromatograms run, taking about 3 h. Programs for predicting optimum separation conditions required about 80 min, while to run the final chromatogram required 30 min. Thus, an adequate separation for this mixture was located within a working day (*ca.* 7 h).

The sequential simplex procedure required 20 injections taking a total time of around 6.5 h. Additionally, the selection of the isoeluotropic binary eluents still has to be considered. As the chromatograph had only three solvent reservoirs, it was necessary to employ isoeluotropic binary eluents if the full range of selectivities (*i.e.* due to methanol, acetonitrile, and THF) were to be exploited. Therefore, the selection of isoeluotropic eluents, as in the earlier procedure, required about 2 h. Thus, the sequential simplex procedure required 9 h for operation (although it was highly automated), but could achieve an adequate separation of the test sample (Fig. 8), even

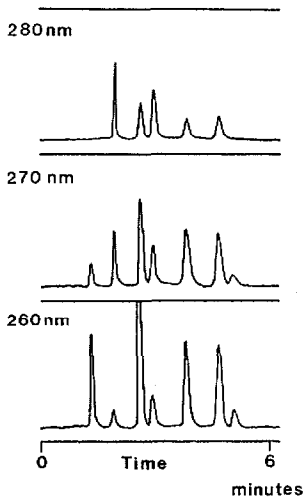

Fig. 8. Best separation located by the sequential simplex optimisation procedure. Mobile phase: 60% (v/v) aqueous methanol–47% (v/v) aqueous acetonitrile–42.5% (v/v) aqueous THF (7.0:61.1:31.9, v/v/v). Minimum α: 1.10.

though this was not the global optimum revealed by the contour plot (Fig. 7).

The sequential simplex procedure did have a major advantage, in that its operation was fully automated once the isoeluotropic eluents were prepared. Simplex lattice mixture design data had to be collected in an interactive process, although with suitable chromatographic hardware its automation would be straightforward. As with sequential simplex the only information required at the outset of optimisation were the column parameters (diameter, length etc.).

CONCLUSIONS

A peak recognition algorithm based on percentage area at a number of detection wavelengths has been developed. This algorithm enabled the simplex lattice mixture design to be automated for a test mixture where individual standards were not available. Separation optimisation was achieved by combining the simplex lattice design with a minimum alpha (selectivity) criterion. Baseline resolution of a seven component test mixture was achieved despite there being seven different elution orders across the response surface.

Automated sequential simplex optimisation was applied to the same sample and was found to be unable to locate the global optimum. A comparison of the two procedures revealed that not only was simplex lattice mixture design able to obtain an adequate separation but it did so in substantially fewer chromatograms and took 20% less time than sequential simplex. With an automated instrument the time reduction could conceivably be greater. Thus, automated simultaneous optimisation procedures incorporating peak recognition algorithms now provide a reasonable alternative to the sequential simplex procedure for unknown samples. Samples producing complex response surfaces can be optimised by the simplex lattice design, more efficiently and more reliably than by the sequential simplex procedure. Simultaneous procedures are

more generally applicable than sequential procedures, provided that appropriate mathematical models can be fitted to the chromatographic data.

The suite of programs currently used was written primarily to demonstrate the feasibility of this approach. A number of problems are associated with the routines: (i) programs take a significant time to run on the HP 85B; (ii) each peak must be detected in at least one of the chromatograms; (iii) the minimum number of peaks which can be dealt with is only two less than the maximum.

These difficulties may all be addressed. A chromatograph with a multiple solvent delivery system would increase the rate of data acquisition, while more efficient programming and a faster computer/program language would speed up data handling. The detection of all peaks in one chromatogram may not be necessary if solvochromic effects are insignificant. Cases of multiple peak overlap may be dealt with if more computer memory were available.

These algorithms described have been successfully applied to the test mixture without any knowledge of the number of components or their UV spectra. However, the optimisation routine is dependent upon the accuracy of the model and if retention behaviour can not be accurately described by the special cubic function adopted, then errors will result. The possibility of using a procedure such as the iterative mixture design[3] remains to be assessed. The basic suite of programs may be elaborated and extended to form a useful and generally applicable automated optimisation procedure. This will form the basis of further work.

ACKNOWLEDGEMENT

A.G.W. is grateful to Pfizer Central Research, U.K. for providing the studentship for this work.

REFERENCES

1 J. C. Berridge, *Techniques for the automated optimisation of HPLC Separations*, Wiley-Interscience, New York, 1985.
2 P. J. Schoenmakers, *Optimization of Chromatographic Selectivity*, Elsevier, Amsterdam, 1986.
3 P. J. Schoenmakers, A. C. J. H. Drouen, H. A. H. Billiet and L. de Galan, *Chromatographia*, 15 (1982) 688.
4 W. Spendley, G. R. Hext and F. R. Himsworth, *Technometrics*, 4 (1962) 441.
5 J. A. Nelder and R. Mead, *Comput. J.*, 7 (1965) 308.
6 J. C. Berridge, *J. Chromatogr.*, 244 (1982) 1.
7 J. C. Berridge and E. G. Morrissey, *J. Chromatogr.*, 316 (1984) 69.
8 J. C. Berridge, *Analyst*, 109 (1984) 291.
9 A. S. Kester and R. E. Thompson, *J. Chromatogr.*, 310 (1984) 372.
10 A. G. Wright, A. F. Fell and J. C. Berridge, *Chromatographia*, 24 (1987) 533.
11 M. Otto and W. Wegscheider, *J. Chromatogr.*, 258 (1983) 11.
12 H. Scheffe, *J. Royal Stat. Soc. B*, 20 (1958) 344.
13 J. W. Gorman and J. E. Hinman, *Technometrics*, 4 (1962) 463.
14 R. D. Snee, *Chemtech*, 9 (1979) 702.
15 J. L. Glajch, J. J. Kirkland, K. M. Squire and J. M. Minor, *J. Chromatogr.*, 199 (1980) 57.
16 J. W. Weyland, C. H. P. Bruins and D. A. Doornbos, *J. Chromatogr. Sci.*, 22 (1984) 31.
17 A. C. J. H. Drouen, H. A. H. Billiet and L. de Galan, *Anal. Chem.*, 56 (1984) 971.
18 R. Yost, J. Stoveken and W. MacLean, *J. Chromatogr.*, 134 (1977) 73.
19 A. C. J. H. Drouen, H. A. H. Billiet and L. de Galan, *Anal. Chem.*, 57 (1985) 962.

20 J. K. Strasters, H. A. H. Billiet, L. de Galan, B. G. M. Vandeginste and G. Kateman, *J. Chromatogr.*, 385 (1987) 181.
21 A. F. Fell, B. J. Clark and H. P. Scott, *J. Chromatogr.*, 316 (1984) 423.
22 H. J. Issaq and K. L. McNitt, *J. Liq. Chromatogr.*, 5 (1982) 1771.
23 P. J. Schoenmakers and T. Blaffert, *J. Chromatogr.*, 384 (1987) 117.
24 L. R. Snyder, J. W. Dolan and J. R. Gant, *J. Chromatogr.*, 165 (1979) 3.
25 P. J. Schoenmakers, H. A. H. Billiet and L. de Galan, *J. Chromatogr.*, 205 (1981) 13.
26 P. R. Haddad and S. Sekulic, *J. Chromatogr.*, 392 (1987) 65.
27 J. L. Glajch, J. J. Kirkland and L. R. Snyder, *J. Chromatogr.*, 238 (1982) 269.
28 R. J. Laub and J. H. Purnell, *J. Chromatogr.*, 161 (1978) 49.
29 J. H. Knox, *J. Chromatogr. Sci.*, 15 (1977) 352.