Journal of Pharmaceutical & Biomedical Analysis
Vol. 6, Nos 6–8, pp. 555–564, 1988
Printed in Great Britain

0731–7085/88 $3.00 + 0.00

Design and application of an expert system for mobile phase optimisation in reversed-phase liquid chromatography*

ANTHONY F. FELL,†§ TIM P. BRIDGE‡ and M. HOWARD WILLIAMS‡

† *Department of Pharmaceutical Chemistry, University of Bradford, Bradford, BD7 1DP, W. Yorkshire, UK*
‡ *Department of Computer Science, Heriot-Watt University, 79 Grassmarket, Edinburgh EH1 2HJ, Scotland, UK*

Abstract: The selection of the optimum composition for the mobile phase in reversed-phase high-performance liquid chromatography (HPLC) is a complex task; conventional approaches require the expenditure of significant amounts of time by the analyst, particularly for complex mixtures of solutes of biological origin. Some of the existing strategies for the automated optimisation of mobile phase composition (e.g. Simplex), may fail if the elution order of the components changes; or they may require that standards be chromatographed in order to establish the retention behaviour of each component in a mixture (e.g. resolution mapping). These problems may be overcome if the retention behaviour of each individual solute can be established from the chromatogram of the mixture. In this regard, components can be tracked by exploiting the spectral information generated by a rapid scanning photodiode array detector. Unfortunately this information is often insufficiently detailed to allow an unambiguous model of retention behaviour to be constructed. The system developed by the Authors uses these spectral data as a basis for constructing one or more hypothetical retention models, each of which is refined or rejected as further information is obtained during the progress of the experiment. To improve the reliability of the retention models proposed by the system, the spectral data are utilised in a number of tests designed to assess the purity of each chromatographic peak. The information so generated may be used in conjunction with any previously acquired spectral data both to select an appropriate method for extracting spectra for each component from the matrix of (A, λ, t) data and to establish reliability parameters for the resultant spectra. The development and philosophy of the expert system developed for eluent optimisation in reversed-phase HPLC is discussed.

Keywords: *Liquid chromatography; photodiode array detection; expert system; eluent optimisation; peak homogeneity validation.*

* Presented at the "International Symposium on Pharmaceutical and Biomedical Analysis", September 1987, Barcelona, Spain.

§ To whom correspondence should be addressed.

Introduction

The problem of devising a suitable mobile phase for a separation by reversed-phase liquid chromatography, so as to produce the required separation of the solutes in an acceptable analysis time, is one that can require considerable expenditure of time and resources before a solution is identified. The conventional approach to such an optimisation requires the attention of a skilled chromatographer who has the expertise to select the constituents of the mobile phase from a combination of his knowledge of the analytes, the results of some initial experimentation and his experience of other similar optimisation problems.

To reduce the amount of time spent on the problem by the chromatographer a number of automated optimisation strategies have been developed [1, 2]. Unfortunately all of these techniques suffer from limitations in their applicability. They may require a large number of experiments to define the chromatographic response surface, as for instance the search strategy used by the Perkin–Elmer Solvent Optimisation System, (PESOS) or they may require that pure standards be available for chromatographing under a lesser number of conditions to produce a map of the response surface, as for instance in the technique of overlapping resolution mapping [3].

A system which avoids these drawbacks is the technique of Simplex optimisation [4] which assesses the chromatograms by means of a suitable response function [1], and predicts the optimum conditions by a gradient sensing search of the parameter space. However, Simplex optimisation may fail in cases in which the elution order of the components changes during the optimisation process.

What is required is a method of reproducing the skills of the expert analyst in an automated system. Such an approach can be provided by the use of an intelligent knowledge based or expert system [5]. Such a system is a computer program, or suite of programs, which encapsulates expertise derived from a human skill in such a form as to allow the information to be used to solve problems in a specific problem domain. A number of such systems have been created to address problems in the field of analytical chemistry [6].

An expert system for optimisation

The use of an "intelligent" expert system is the basis of the optimisation system currently under development by the Authors [7, 8]. This system attempts to determine the optimum composition of the mobile phase by means of an iterative regression modeling technique [9], with a Simplex optimisation routine [4] available as an alternative method for use in cases in which the iterative regression technique is not applicable. Some of the limitations of the iterative regression method will be dicussed later in this paper. The iterative regression and simplex methods form part of the overall framework of the optimisation system, as illustrated in Fig. 1. This system is organised on modular lines to facilitate the inclusion of novel or improved units at any stage.

The starting point of the optimisation process is the selection of the column and choice of the organic modifiers to be investigated. In the current implementation these choices are made by the operator. However, rules can be continuously formulated and incorporated to allow the expert system to assist in these crucial stages.

The first experimental data required by the system are the results of a linear water–methanol gradient elution chromatogram. This is used to predict the eluting

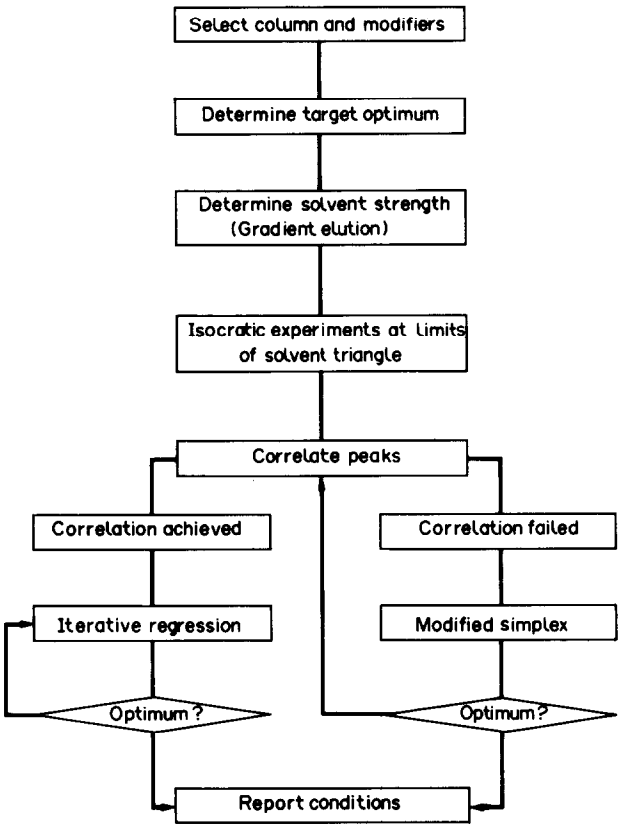

Figure 1
Outline structure of the expert system for HPLC optimisation.

power required to complete the analysis within the desired overall time [10], as indicated by the analyst. Once the eluting power has been determined, three isoeluotropic mobile phases are prepared using the modifiers selected during the initial stage. In the current system, these modifiers are normally restricted to methanol, acetonitrile and tetrahydrofuran. The compositions required to give approximately equal elution times for each mobile phase are determined by means of the transfer rules proposed by Schoenmakers [11].

The results of the three isocratic chromatograms are used as the starting point of the selectivity optimisation process, using either the iterative regression or the Simplex method, depending on the information which the system manages to extract from the chromatograms. As is explained below, the iterative regression technique requires that the retention time of each component be determined under each of the three sets of experimental conditions. If this information is not available due to overlap of the peaks or some other external factor, then the Simplex method is used. In all other cases the control of the optimisation is passed to the iterative regression.

Iterative regression optimisation

The iterative regression optimisation technique works by assuming that there is a linear relationship between the composition of the mobile phase and the logarithm of the

capacity factor of any given solute. If there is a variation in the selectivity between the mobile phases, then the relationship will be dependent upon the identity of the solute, that is, if the capacity factor k and the mobile phase composition Φ are related by the equation: $\ln(k) = m\,\Phi + c$, then the values of the constants m and c are dependent on the solute and the chromatographic conditions. This stage is illustrated in Fig. 2, a plot of log capacity factor versus mobile phase composition for four solutes.

The next stage in the construction of a model of the retention behaviour is to determine the separation of each pair of solutes with all possible mobile phases. This is

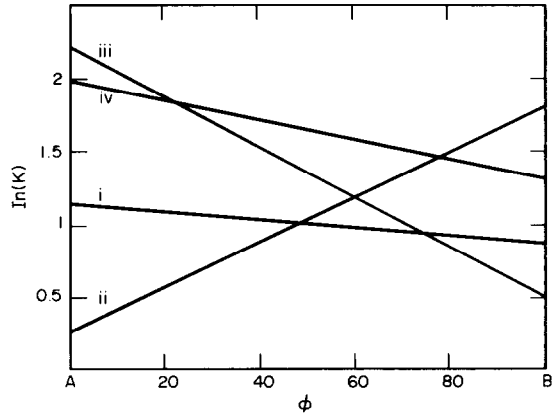

Figure 2
A plot of log capacity factor versus composition for a four component mixture. The solutes represented are: (i) phenol; (ii) diethylphthalate; (iii) *p*-hydroxybenzaldehyde; (iv) *n*-propyl-*p*-hydroxybenzoate. The mobile phase components are: (A) MeOH–water (60:40 v/v); (B) MeCN–water (48:52 v/v).

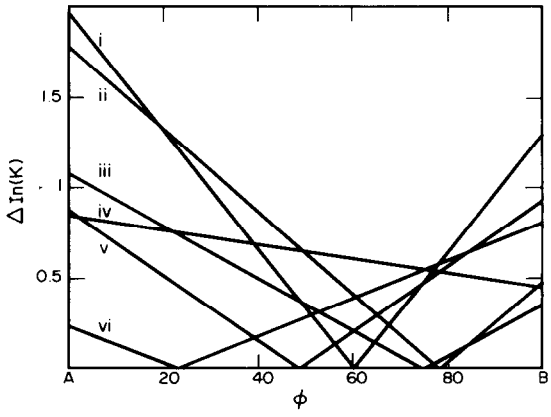

Figure 3
A plot of delta ln k v mobile phase composition derived for each pair of solutes from the data in Fig. 2. The indicated optimum lies at 100% B. The region of achievable resolution at each value of ϕ is bounded by the axes and the lowest line of the plot. A value of 0 for $\Delta\ln(k)$ indicates that the two solutes coelute in that mobile phase. The solute pairs associated with each line are: (i) diethylphthalate and *p*-hydroxybenzaldehyde; (ii) diethylphthalate and *n*-propyl-*p*-hydroxybenzoate; (iii) phenol and *p*-hydroxybenzaldehyde; (iv) phenol and *n*-propyl-*p*-hydroxybenzoate; (v) phenol and diethylphthalate; (vi) *p*-hydroxybenzaldehyde and *n*-propyl-*p*-hydroxybenzoate. The mobile phases are as in Fig. 2.

achieved by considering the separation between the curves of the $\ln(k)$ plot. Figure 3 is a plot of $\Delta\ln(k)$ versus Φ, derived from Fig. 2.

The indicated optimum composition is that at which the separation of the *worst resolved pair* of solutes is maximised. In Fig. 3 this composition is indicated to be 100% component B. An experiment is next carried out using the mobile phase indicated by this plot and the resultant chromatogram examined. If, as is likely during the early stages of the optimisation, the resolution achieved does not match the prediction, then the actual retention behaviour is used to produce a refined model of the response surface and a new mobile phase composition selected. The cycle of experimentation and refinement is repeated until the predicted and actual results agree, at which time the best separation available using the column and modifiers selected will have been obtained.

The example illustrated in Fig. 2 is for two organic modifiers only. The current system considers the use of three modifiers in a direct extension of the modeling process which produces planes of response rather than the lines of Fig. 3. A contour plot of such a response surface is given in Fig. 4. In principle the technique may be extended to consider other factors such as pH or the concentration of ion-pairing reagent present.

Selection of optimisation strategy

As discussed above, the expert system is provided with a choice of optimisation strategies, which may be deployed during the search for the optimum mobile phase composition. The strategy to be deployed at any stage is selected by the expert system on the basis of the spectral information available.

The preferred optimisation strategy is the iterative regression technique, since this requires fewer experimental separations than is the case for a Simplex search. It does, however, rely on the correct assignment of spectral profiles to individual components in the mixture. This assignment of spectra to components takes place in the module labelled "Correlate Peaks" in the diagrammatic representation of the expert system, Fig. 1. This module first examines each peak for homogeneity and then extracts the appropriate spectral profile(s). For a pure peak this is achieved by simply recording the spectra from the peak maximum. For a composite peak, however, the spectral information must be

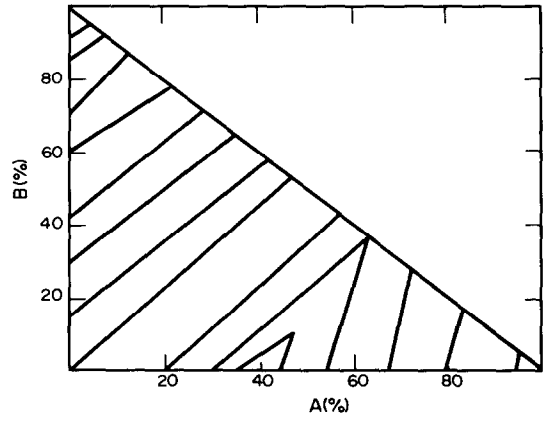

Figure 4
Contour plot of minimum resolution for a four component mixture, chromatographed in three different mobile phases. Resolution contours are based on a scale of arbitrary units. Mobile phase components are: (A) MeOH–water (60:40) v/v); (B) MeCN–water (48:52 v/v); and (C) THF–water (42:58 v/v).

deconvoluted from the total data set. To this end the system employs iterative target transformation factor analysis (ITTFA) [12] to determine the probable number of components contributing to a peak, and to extract their spectral profiles.

The spectra extracted from one chromatogram are next matched with those from any previous chromatograms, to enable the changes in retention time for each component to be assessed. If this correlation procedure succeeds, then the optimisation is directed by the iterative regression routine. If, on the other hand, the correlation procedure fails, then the expert system will select a modified Simplex strategy to continue the optimisation. The system will repeat the correlation step after each new experiment under the Simplex regime, and if sufficient additional data have been acquired to allow a successful correlation to be achieved, the optimisation is redirected to the control of the iterative regression routine.

Determination of retention behaviour

Whilst the technique of iterative regression optimisation is capable of locating the optimum conditions for a given separation after as few as five or six experiments, its success depends entirely on the correct determination of the retention time for each component in each experiment. If pure standards of each solute are available, then this information may be obtained by chromatographing each standard under each condition. This has two significant disadvantages. Firstly, such an approach requires a large number of individual chromatograms and secondly, the constituents of the mixture must be known in advance in order to allow selection of the relevant standards, which may not be available, particularly in the case of samples of biomedical origin.

To avoid this problem the system uses the ultraviolet absorption spectra of the eluting peaks as a label for tracking components between chromatograms. These spectra are obtained on-line by means of a photodiode array detector [13]. Whilst the tracking of components by their UV spectra is straightforward for compounds with significantly different spectral profiles which elute in isolation, the problem assumes significance if two or more components have similar spectra or where co-elution occurs. At this stage there is no easy answer for the first problem case, that is of indistinguishable spectra, but mathematical tools are available to assist in the extraction of spectra from compound peaks [14]. The development of rules to select an appropriate method for deconvoluting the spectral information contained in the spectrochromatogram of an heterogeneous peak will allow the system to operate with a greater degree of independence than is currently the case.

The matrix-based deconvolution methods, whilst offering a significant advance in the extent of usable information which can be extracted from a chromatogram, exact a penalty in terms of the computer power required to run them. To reduce the time spent on analysing the data from each chromatogram the system is provided with a number of tools for the rapid assessment of peak homogeneity. The tests currently available to the system are: spectral difference; first derivative chromatogram; second derivative chromatogram; absorbance ratio (time domain); variation of apparent retention time with detector wavelength; absorbance ratio (wavelength domain); ratio of derivative chromatograms; multiple absorbance ratio.

The intention underlying all these approaches is to give a quick assessment of the homogeneity of the target peak as a filter before any heavy-weight multivariate statistical tools are employed, so as to reduce the total amount of computer time required in examining each chromatogram.

In all cases the tests are applied to a peak or peak group which has been extracted from the total spectrochromatogram. The operation of each of these tests is discussed below.

Spectral difference. This compares the profiles of spectra taken from the leading and trailing edges of the peak. The routine extracts spectra from the data matrix at times 1/3 and 2/3 of the apparent peak width and normalises them at the wavelength of maximum absorption in the first spectrum. The square root of the sum of squares (RSS) difference between the two profiles is calculated and this value is reported. For a pure peak this difference will be due only to noise in the detector, but for an heterogeneous peak containing two or more components with dissimilar spectral profiles a larger difference will be determined.

First derivative chromatogram. The derivative of a chromatographic profile may be expected to be more sensitive than the original curve to the existence of a minor component [15]. To make use of this property, the total absorbance chromatogram is differentiated with respect to time by a simple difference mechanism:

$$\frac{dA}{dt}(i) = (A_{(i+1)} - A_{(i-1)})/2.$$

This method of differentiation was selected to reduce the loss of data inherent in any such averaging differentiation, whilst retaining the advantage of a high speed of operation. The number of zero crossings in the derivative profile is determined. A facility is included to allow crossings produced by noise at low signal amplitudes to be discounted. For a pure peak, one zero crossing is expected, whereas an impure peak would be expected to give rise to a profile exhibiting an increased number of zero crossings.

Second derivative chromatogram. The first derivative chromatogram is further differentiated by the same method, and the number of zero crossings again determined. Two such crossings are expected for an homogeneous peak; an increase in the number of zero crossings is an indicator of heterogeneity. Both derivative tests check to ensure that the available data are larger than the differentiation band width. If differentiation is not possible, due to a lack of available data, i.e. the peak width is less than twice the width of the differentiation window, then a value of -1 zero crossings is returned as a warning "flag" to the system.

Absorbance ratio (time domain). The ratio of absorbances measured at two wavelengths should be invariant with time for an homogeneous peak [16]. This test locates the time at which maximum absorbance occurs (t_{max}) and identifies the wavelength of maximum absorbance (λ_{max}) and the wavelength which gives half this absorbance value at this time. The ratio of the signals at these two wavelengths is then calculated as a function of time across the peak. The RSS deviation from the value of the ratio at t_{max} is determined and reported. Ideally this should be zero, or for a real system, not exceeding the variation in signal due to detector noise.

Variation of apparent retention time with detector wavelength. For a pure peak the measured retention time will be independent of the wavelength used to monitor the

elution. For a composite peak, however, the apparent retention time may vary according to the detector wavelength employed [17]. The extent of this variation is governed by the difference in elution times of the contributing components, the difference, or similarity, of the individual spectral profiles and the relative amounts of the components. The value of t_{max} is determined for each detector channel and the difference between the largest and smallest values reported. A minimum value of absorbance must also be selected, to avoid reporting noise signals for wavelengths at which the solute does not absorb. For a pure peak any variation would be due solely to noise in the detector.

Absorbance ratio (wavelength domain). Just as the ratio of two chromatograms recorded at different wavelengths should be invariant with time, so the ratio of two spectra recorded at different times should be invariant with wavelenth. If the spectra are normalised before the ratio is determined, then the ratio should, of course, be unity. As in the other tests, a variation from this value is indicative of heterogeneity.

Ratio of first derivative chromatograms. The idea of absorbance ratio evaluation may be extended into the derivative domain. This test normalises the chromatograms on each channel with respect to λ_{max} at t_{max} and then calculates the time derivative for each chromatogram. The ratio of these chromatograms to that at λ_{max} (which should be unity) is then calculated and the RSS deviation computed and reported.

Multiple absorbance ratio. This routine is identical to the derivative ratio routine, except that the ratio is calculated on the original, undifferentiated, data. Once more a deviation from unity is a measure of the degree of heterogeneity in the peak under investigation.

The sensitivity of each of the above tests will depend on a number of factors, such as the signal-to-noise ratio, the relative amplitudes of the contributing solutes and the degree of similarity between the component spectra. Since these factors cannot be readily determined for an experimental chromatogram, rules are currently being developed to allow the interpretation of the results of the tests, which may be expected to disagree with each other, in the light of the known signal parameters (peak amplitude, peak width, known detector noise level etc.).

Experimental

The solutes employed, 3-hydroxybenzaldehyde, *n*-propyl-*p*-hydroxybenzoate, diethyl phthalate and phenol (BDH, Poole, UK) were used without further purification. HPLC-grade methanol, acetonitrile and tetrahydrofuran (Rathburn Chemicals, Walkerburn, UK) were used as received. All eluents were filtered through Millipore® 0.45 μm filters in all-glass apparatus and degassed under reduced pressure in an ultrasonic bath for 10 min.

Eluent delivery was achieved by the use of two PU4010 pumps under the control of a PU4850 chromatography data station (Pye Unicam Ltd, Cambridge, UK). The 100 × 5 mm column was packed with 5-μm Hypersil ODS. Solute detection was achieved using a Pye-Unicam PU4021 diode array detector.

The data used to illustrate the process of iterative regression optimisation were taken from an optimisation of a four component mixture, containing the solutes listed above. The final chromatogram produced by the system is illustrated in Fig. 5. The lack of

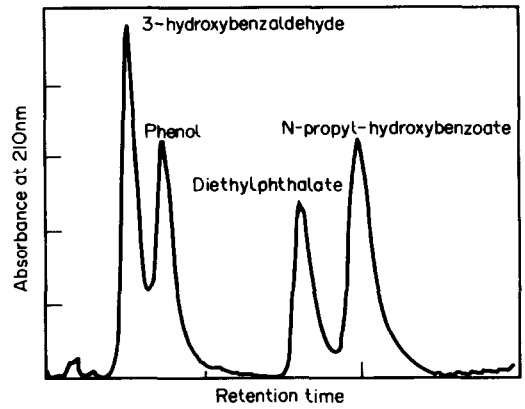

Figure 5
Chromatogram of a four-component mixture using the mobile phase composition predicted by the expert system.

complete baseline-resolution is probably a consequence of the operator-specified analysis time. A longer analysis time would allow the expert system a greater degree of freedom, and may result in improved resolution.

Implementation

The expert system has been implemented on an IBM PS/2 Model 50 microcomputer, which provides a usable memory of 640 Kbytes, with a 20Mbyte hard disk. The logic processing is performed by a program written in microProlog [18]. As this language is not suited to the large amount of numerical processing required in the interpretation of the spectrochromatograms, an extension to the language is used [19] to allow the prolog to execute programs in other languages such as Pascal [20]. A modular design has been adopted, so that changes and extensions to the system may be implemented without forcing modifications to previously implemented sections.

Conclusions

The system including the tests and logical modules described above has been implemented and demonstrated for the optimisation of the separation of a four-component mixture. This system is still under development and will receive additional modules to increase its "intelligence" and thereby reduce the need for operator intervention. In particular rules to guide the selection and use of an appropriate deconvolution technique will be established and will form a key part of the enhanced system. The system, as implemented, combines elements of logic processing with numerical processing routines, and serves to demonstrate the potential advantages to be obtained from such a synergistic combination.

Acknowledgements — The Authors acknowledge receipt of a Post-Doctoral Assistantship from the Science and Engineering Research Council for one of us (TPB) and the generous loan of equipment by Pye Unicam Ltd, Cambridge.

References

[1] J. C. Berridge, *Techniques for the Automated Optimisation of HPLC Separations*. Wiley, Chichester (1986).
[2] P. J. Schoenmakers, *Optimisation of Chromatographic Selectivity*. Elsevier, Amsterdam and New York (1986).
[3] J. L. Glajch, J. J. Kirkland and J. J. Squire, *J. Chromatogr.* **199**, 57–79 (1980).
[4] J. C. Berridge, *Analyst* **109**, 291–293 (1984).
[5] F. Hayes-Roth, D. A. Waterman and D. B. Lenat, *Building Expert Systems*. Addison-Wesley, Reading, MA (1983).
[6] T. P. Bridge, M. H. Williams and A. F. Fell, *Chem. Brit.* **23**, 1085–1088 (1987).
[7] T. P. Bridge, M. H. Williams and A. F. Fell, *Anal. Proc.* **25**, 43–44 (1987).
[8] T. P. Bridge, M. H. Williams and A. F. Fell, *Chromatographia* **24**, 691–695 (1987).
[9] A. C. J. H. Drouen, *Computerized Optimization and Solute Recognition in Liquid Chromatographic Separations*. PhD Thesis, Technical University Delft (1985).
[10] L. R. Snyder, J. W. Dolan and J. R. Grant. *J. Chromatogr.* **165**, 3–30 (1979).
[11] P. J. Schoenmakers, H. A. H. Billiet and L. de Galan, *J. Chromatogr.* **205**, 13–30 (1981).
[12] G. G. R. Seaton and A. F. Fell, *Chromatographia* **24**, 208–216 (1987).
[13] A. F. Fell, H. P. Scott, R. Gill and A. C. Moffat, *Chromatographia* **16**, 69–78 (1982).
[14] E. R. Malinowsky and D. G. Howery, *Factor Analysis in Chemistry*. Wiley, New York (1981).
[15] T. P. Bridge, A. F. Fell and R. H. Wardman, *J. Soc. Dyers and Colourists* **103**, 17–27 (1987).
[16] A. F. Fell, H. P. Scott, R. Gill and A. C. Moffatt, *J. Chromatogr.* **282**, 123–140 (1983).
[17] A. G. Wright, A. F. Fell and J. C. Berridge, *Chromatographia* **24**, 533–540 (1987).
[18] K. L. Clark and F. G. McCabe, *micro-PROLOG: Programming in Logic*. Prentice-Hall International, Englewood Cliffs, NJ (1984).
[19] K. T. Milne, M. H. Williams, B. J. Clark and A. F. Fell, *Anal. Proc.* **23**, 157–160 (1986).
[20] K. Jensen and N. Wirth, *PASCAL User Manual and Report*. Springer, Berlin (1975).

[*Received for review 10 December 1987*]