Journal of Pharmaceutical & Biomedical Analysis
Vol. 11, No. 1, pp. 1–13, 1993
Printed in Great Britain

0731–7085/93 $24.00 + 0.00

Review

Strategies for peak-purity assessment in liquid chromatography

JOHN B. CASTLEDINE and ANTHONY F. FELL*

Pharmaceutical Chemistry, University of Bradford, Bradford BD7 1DP, UK

Abstract: One of the key requirements for the validation of chromatographic methods is to determine the purity of a chromatographic peak. Statistical modelling of the chromatographic process suggests that overlapping components are highly probable in a chromatogram. Hence extensive efforts have been directed at the development of sensitive, reliable and robust methods to assess peak purity. This is especially the case in the pharmaceutical industry, where liquid chromatography (LC) is widely utilized and the demands on method validation are justifiably high.

On-line multiwavelength absorptiometric detection is often used to generate the additional data required to facilitate peak-purity assessment in LC. This approach, using photodiode-array technology, is directly compatible with the aqueous-based reversed-phase LC solvents used extensively in drug analysis. Consequently, this work highlights many of the peak-purity algorithms, which may be applied using LC with diode-array detection. The relative merits of the individual techniques are discussed, and a rationale is developed for their application.

Keywords: *Absorbance ratio; chemometrics; diode-array detection; liquid chromatography; peak homogeneity; peak purity; principal components analysis.*

Introduction

To demonstrate that a particular chromatographic separation is appropriate for the sample matrix specified, it is necessary to validate the selectivity of that method. Thus, with the prominent use of liquid chromatography (LC) within the pharmaceutical industry and the consequences that could arise from incomplete drug characterization, it is not surprising that extensive effort has been directed at liquid chromatographic peak-purity assessment in recent years. Purity is itself not a quality that can be directly measured [1]. In the case of LC peak-purity determinations, it can only be assessed by demonstrating that interferents (i.e. impurities) are absent. Thus it is often preferable to express purity as determined relative to that for a suitably characterized reference sample. Such an assessment of chromatographic purity is closely related to the methodologies used both for the determination of peak identity and, where two or more overlapping peaks are distinguished, those techniques applied to facilitate the deconvolution and subsequent quantitation of the two (or more) components.

Recently Papas of the U.S. Food and Drug Administration highlighted 13 'ideal characteristics' that the 'completely automatic, accurate, and reliable integrator might be expected to possess' when encountering two or more partially overlapped peaks [2]. In addition to peak deconvolution and quantitation, many of these 'ideal characteristics', listed below, also apply to those methods used for the assessment of peak identity and peak purity. They should: (i) be able to determine the peak purity and number of components present; (ii) make no assumptions about peak shape; (iii) if based on a model, be able to provide confidence levels to the chromatographer; (iv) be able to adequately resolve and quantitate the components, regardless of the degree of overlap or resolution; (v) be insensitive to noise; (vi) be unaffected by the amount of peak tailing; (vii) assume unlimited computer speed, memory, and power; (viii) be unaffected by the weight per cent of the components present; (ix) be precise, accurate, and reliable; (x) be based on single-channel data; (xi) easily handle nonlinear and shifting baselines; (xii) require no prior knowledge of the components present nor require independent standard chromatograms; and (xiii) work either in real-time or within a reasonable time frame.

* Author to whom correspondence should be addressed.

Clearly no integrator or data processing algorithm exists that would fulfil all these criteria. However, the development of hyphenated techniques to generate multichannel data, in particular LC combined with a diode-array detector (DAD), has resulted in significant advances being made in this branch of chromatographic science.

The algorithms developed can be grouped according to various forms of classification. One of the most useful is to divide the methods into those which assess peak homogeneity and those which evaluate peak purity. Using LC–DAD, a homogeneous chromatographic peak is one that has similar absorptiometric properties throughout the peak. Such a peak is likely to consist of a single analyte although the possibility exists that a second analyte is also present, with both an identical retention time and similar band broadening properties. To identify this category of peak as impure, it is necessary to use techniques that compare the spectral characteristics of the analyte with those of a reference sample. Hence, while both peak-purity and peak-homogeneity algorithms are critically dependent upon differences in the spectral properties of the analyte and overlapping impurities, peak-homogeneity assessment also relies on chromatographic differences between the overlapping compounds.

A plethora of methodologies has been published to aid in the detection of simultaneously eluting LC solutes. This review aims to highlight the historical development of some of the more common techniques used primarily to examine either peak purity and/or peak homogeneity. Moreover, by incorporating descriptions of graphical, univariate and multivariate data handling approaches, a rationale is developed to identify the most useful algorithms for peak-purity assessment, taking account of the different requirements and various constraints that may apply in practice.

Methodologies Based on Single-Channel Data

Many of the algorithms used with multichannel LC detectors to assess either peak purity and/or peak homogeneity have their origin in approaches developed using single-channel data. Indeed, the results obtained using single variable-wavelength UV–vis detectors, either in the stopped-flow mode or by repetitive injections with detection at different wavelengths, gave an early demonstration of the potential of on-line spectroscopic measurement and have contributed to the development of LC–DAD technology and other hyphenated systems.

It is convenient to divide the methodologies based on single-channel data into three subsections. These approaches are: (1) statistical evaluation of the likelihood of analyte co-elution to assess peak purity: (2) curve resolution techniques for the determination of peak homogeneity; and (3) the use of stopped-flow, dual-detector or multiple-injection strategies to simulate multichannel data collection, with both the assessment of peak purity and peak homogeneity possible.

While category (3) has been superseded by technological advances, the limitations imposed by cost and, in a small number of cases by the lack of a suitable detection chromophore, have ensured that single-channel LC detection remains widespread. Consequently, the implications arising from the potential presence of interferents in such systems have resulted in the continuing development of curve-resolution approaches to peak deconvolution.

Statistical evaluation of LC analyte co-elution

Before considering the various practical approaches to the problem of peak purity, it is both interesting and informative to examine models of the situation based on statistical approaches. Much of the work published in this field has been presented by Giddings and Davis [3–6]. The equations generated, and the subsequent probabilities calculated, are based on the assumption that the component peaks of a complex mixture are spaced randomly within a chromatogram according to a Poissonian process, and furthermore, that the individual constituents are characterized by either a Gaussian or an exponentially-modified Gaussian distribution.

Davis and Giddings [4] found that 'Relative to the maximum peak content or peak capacity for closely spaced peaks, a random chromatogram will never contain more than about 37% of its potential maximum number of peaks and, worst of all from an analytical point of view, 18% of the potential single-component peaks . . . and a chromatogram must be approximately 95% vacant in order to provide a 90% probability that a given component of interest will appear as an isolated peak'. Later work on this subject [5] further developed this 'statis-

tical model of overlap' to enable the calculation of 'stand-alone probability values', expressing the likelihood that any given component would be adequately separated from its neighbours in a given chromatogram. The application of this theory to experimental GC chromatograms of complex mixtures resulted in calculated stand-alone probability values which were said to be discouragingly low [6]. Although the modelling of the chromatographic process in this way has its limitations, the work described above suggests that the potential for, and thus the occurrence of, overlapping components within a chromatogram is considerably more likely than may be generally realized.

Peak-homogeneity assessment using curve-resolution algorithms

As the demands on chromatography grow, with increasingly complex mixtures to be analysed ever more quickly, the interest in peak deconvolution, with the subsequent quantitation of the individual constituents within a peak, has also expanded. A substantial part of this work has focused on the use of single-channel detection data, using the chromatographic integrator with built-in curve-resolution algorithms. The accuracy of the various methods used has been reviewed [7].

The ability of chromatographic integrators to perform peak deconvolution requires the peak homogeneity to be evaluated from the single-channel detection data. This evaluation is generally based on double differentiation of the chromatographic signal, R, with respect to time, t [8, 9] to generate the second derivative, d^2R/dt^2 (Fig. 1). The resultant band-sharpening effect can be used to detect the presence of two or more components in an apparently single analyte peak, by enhancing the subtle changes in peak profile introduced by the underlying co-eluting species [10, 11].

Following the proposal to exploit differentiation in the time domain [8, 9], further developments have been reported by Berridge *et al.* [12, 13]. In isocratic LC the peak width increases with analysis time due to band broadening and other effects. Consequently the sensitivity of the derivative method will decrease as the chromatogram proceeds. Berridge proposed that a variable time-constant be used to successfully overcome this limitation [13]. In addition. Excoffier and Guiochon have reported the use of a pseudo-derivative function for enhanced peak sensing [14].

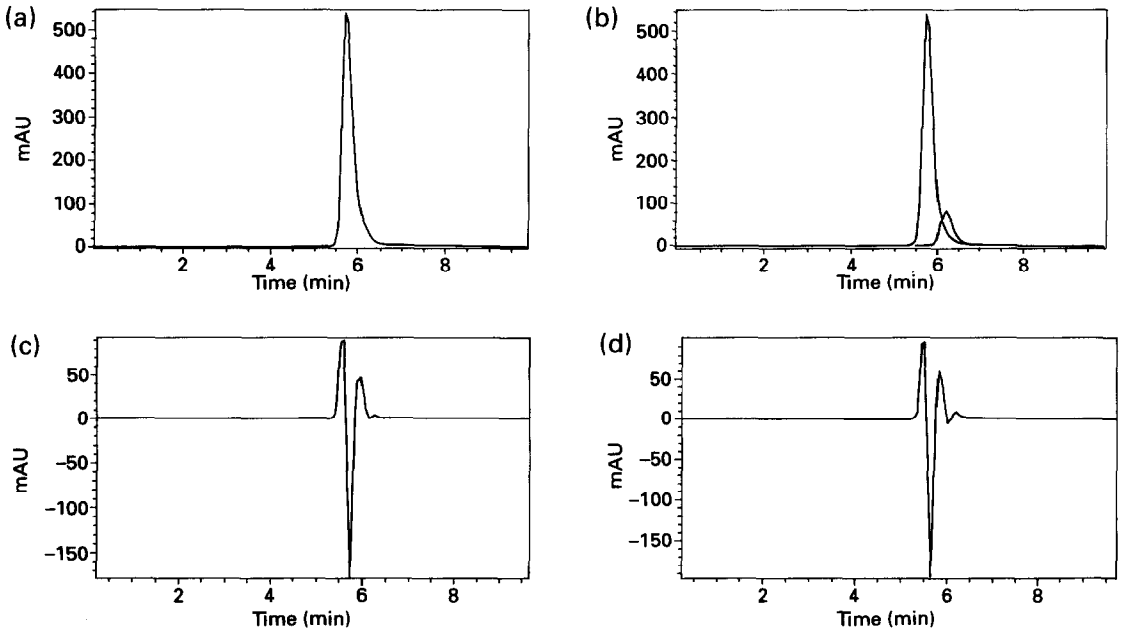

Figure 1
Peak-homogeneity assessment using the second-derivative transformation of the chromatographic profile in the time domain. The chromatogrpahic profiles of the individual components for a 'pure' and 'impure' chromatographic peak are shown in (a) and (b), respectively. The second-derivative transformations of these composite chromatographic profiles are shown in (c) and (d). The second negative peak in (d) indicates that the minor co-eluting component has been detected by this transformation.

Peak-purity assessment using strategies to simulate multichannel data collection

The use of spectroscopic data to aid in the quantification of co-eluting LC peaks was proposed by Skelly and Crummett in 1971 [15]. They suggested that duplicate injections with the detector set at two different wavelengths should enable the concentrations of the two partially-overlapped components to be determined after 'calculating the appropriate calibrations and solving simultaneous equations'. This was followed by a detailed theoretical analysis by Ostojic, describing the use of dual-wavelength data to generate *absorbance ratios* for the identification and resolution of overlapping chromatographic peaks [16]. Ostojic showed that for any pure eluting compound, the ratio of the absorbances detected at two different wavelengths should remain constant over the entire elution profile, provided that the Beer–Lambert law was obeyed and the experimental noise was negligible. Furthermore, the absorbance-ratio constant generated is characteristic of the compound of interest, for the specified wavelength-pair, thus facilitating solute identification in principle.

Subsequently, a practical application of dual-wavelength detection to assist in the quantitative analysis of acidic and neutral cannabinoids was described by Smith and Vaughan [17]. This work was expanded in a further publication which identified the absorbance-ratio (AR) technique as a simple way of assessing chromatographic peak purity [18]. In both cases the dual-wavelength detection was accomplished by connecting two variable-wavelength absorptiometric detectors in series. Similar work, using a variable-wavelength detector in series with a fixed (254 nm) wavelength detector was presented by Krstulovic *et al.* [19, 20].

A comprehensive practical examination of the absorbance-ratio technique was published by Yost *et al.* of the Perkin–Elmer Corporation in 1977 [21] who outlined the two main strategies available for determining the presence of a detectable interferent which elutes simultaneously with the compound of interest. When spectral data on the 'pure' compound of interest are available, a comparison of these data with the spectral data obtained from the chromatographic peak allows the purity of the peak to be assessed. When such reference data are not obtainable, it is possible to acquire spectra at various points through the chromatographic peak, in order to assess the homogeneity or heterogeneity of the peak. In both cases it is necessary to know the confidence limits of the ratios generated. From the experimental data presented it was concluded that relative standard deviations in the absorbance ratios greater than 2% could be taken as a reasonable indication that the peaks were not identical or homogeneous.

The problem of wavelength selection was also considered by Yost *et al.* To overcome the possibility that both the compound of interest and the potential interferent may yield similar AR values at the wavelengths selected, it was suggested that three, or more, well separated wavelengths should be used to allow several AR values to be determined throughout each chromatographic peak. As this necessitated the use of the relatively time-consuming stopped-flow techniques, single absorbance ratios continued to be used until the development of the multiwavelength detector in 1979. The limitations of a single wavelength-pair in uniquely identifying pharmaceutical compounds may be observed from work presented by Baker and co-workers [22] using two detectors in series.

Prior to the development of multiwavelength chromatographic detectors, stopped-flow conditions were also used to collect data at several wavelengths for use with multivariate factor analysis. The matrix manipulation technique of *principal components analysis* (PCA), first proposed for use in educational psychology [23], has been applied in many scientific disciplines including spectroscopy. Lawton and Sylvestre [24] investigated self-modelling spectroscopic curve-resolution using PCA, and were the first to recognize that it could also be utilized in chromatography. The ability of this form of factor analysis to calculate the number of significant components in a spectral data matrix, obtained from GC–MS and stopped-flow LC systems, was subsequently demonstrated by Halket [25, 26].

The Assessment of Peak Purity and Peak Homogeneity using Multichannel Absorptiometric Detection in LC

In 1973 Santini and co-workers reviewed the relatively new and fast growing technique of rapid-scanning spectroscopy in 'Rapid Scan-

ning Spectroscopy: Prelude to a New Era in Analytical Spectroscopy' [27]. As the title indicated, this indeed proved to be the case, since the further development of rapid-scanning spectroscopic detectors significantly influenced both spectroscopy and LC detection. The historical evolution of the LC diode-array detector (DAD) has been expertly reviewed by several authors, at various stages in the development of this technology [8, 28–30]. Further insight into the design concept and implementation of a DAD system was published by George and Maute of Hewlett–Packard in Germany [31]. This paper discussed the primary needs of the analyst, the goals to be achieved in hardware and software, and some potential applications of such a detection system.

As previously described, the principles of many peak-purity assessment algorithms were conceived prior to the development of DAD technology. It is however apparent from the literature that the commercial availability of LC–DAD systems in *ca* 1979 resulted in an upsurge of interest in the determination of LC peak purity, leading to the extensive development and refinement of the algorithms used. Furthermore, routine use of LC–DAD also enabled chromatograms to be monitored at several wavelengths, resulting in the detection of impurities which had indeed been resolved by chromatography, but had remained previously undetected due to large differences in the wavelength requirements for the chromophores of the parent compound and the various related product(s) [32].

The peak-purity techniques devised may be usefully categorized as either univariate or multivariate data handling methods. The univariate techniques adopt simple mathematical approaches, often incorporating data from reference and sample chromatograms, and thus may evaluate peak purity and/or peak homogeneity. In general, the multivariate data handling approaches tend to be computer and/or detector intensive methods, utilizing the information available from a single spectral data matrix and consequently assessing peak homogeneity. These powerful chemometric techniques have been primarily developed for chromatographic peak-deconvolution, and the software developed for peak-homogeneity assessment may not take full advantage of the current potential of chemometrics when used in conjunction with LC–DAD.

Information on the general application of multivariate mathematical and statistical methods to chromatographic measurements is readily available in the section on 'Chemometrics' published bi-annually in *Analytical Chemistry* [33].

Univariate statistical techniques

The use of *absorbance ratios* to characterize chromatographic peaks, widely used prior to LC–DAD, has remained a prominent method for peak-purity assessment. It is interesting to note that Bylina *et al.* [34], working to develop a rapid-scanning spectrophotometric LC detector using cathode-ray tubes, also significantly contributed to the initial development of AR methodology. This work, which was published prior to the theoretical discussions of Ostojic [16], was the first to develop the idea of constantly monitoring the AR with respect to elution time. For a homogeneous peak the AR plot generated is, in theory, a square-wave signal. Thus, any deviations from this shape suggests the presence of a co-eluting impurity [Fig. 2(a)]. Using a mixture of fluorene and carbazole, Bylina *et al.* demonstrated that this method could detect co-elution under conditions which gave rise to an apparently single chromatographic peak [34].

With the advent of LC–DAD several workers critically examined the use of AR plots. The performance of the AR technique has been compared with that of generating the second derivative of the elution profile (d^2A/dt^2) [30]. Variations in DAD baseline offset were shown to cause deviations from the square-wave plot for homogeneous peaks [35, 36]. In addition, the latter report also assessed the significant influence of peak overlap upon the absorbance ratiogram of two adjacent peaks.

The development of LC–DAD also facilitated the further investigation of AR wavelength selection, a problem originally discussed by Yost and co-workers [21]. In 1980 White demonstrated, using two variable-wavelength detectors in series, that at least two ratios were necessary to individually characterize a series of barbiturates [37]. Later work using a multi-wavelength detector incorporated data at additional wavelengths to enhance the characterization of barbiturates [38]. A similar approach using up to seven absorbance ratios, generated from wavelengths spaced throughout the spectral region in which the analytes

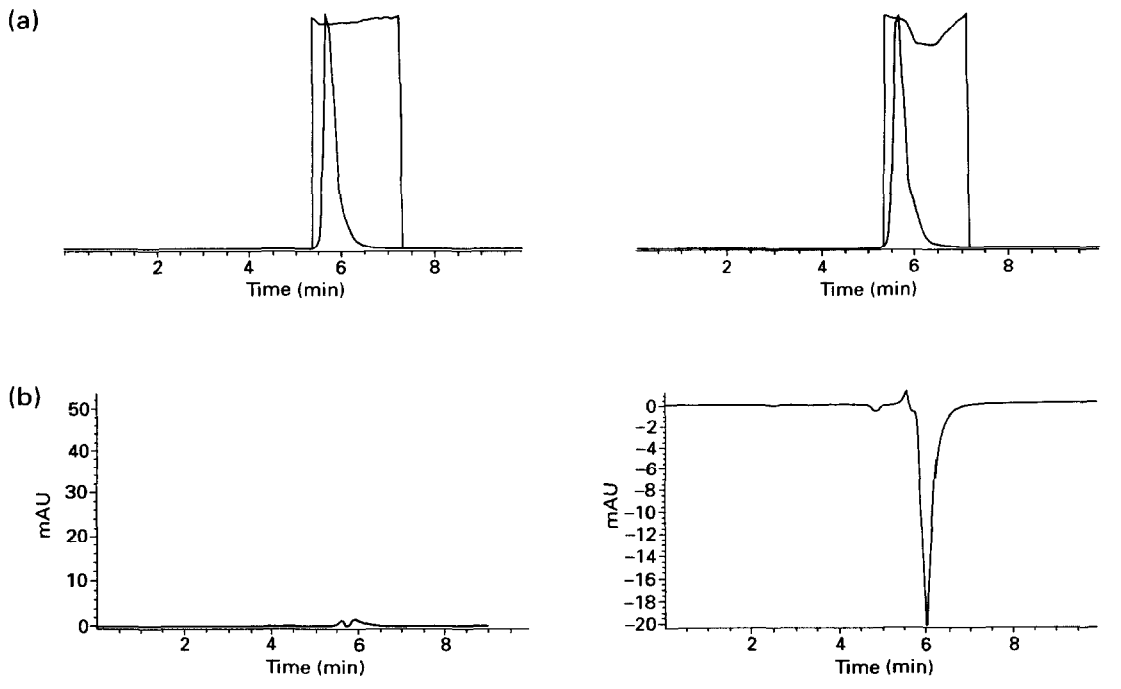

Figure 2
(a) Absorbance ratio plots for the chromatographic profiles presented in Fig. 1. (b) Spectral suppression plots for the chromatographic profiles presented in Fig. 1.

absorb, was subsequently used to examine dye components in forensic analysis [39].

In cases where the spectral characteristics of both co-eluting species are known, the optimum AR wavelength pair can be readily determined by computing the ratio of the individual spectra, to locate the points of maximum and minimum ratio, which correspond to the combination of wavelengths which permit spectral discrimination with maximum sensitivity [30]. Using this optimum wavelength pair, as little as 1% w/w of noscapine could be detected co-eluting with papaverine. Furthermore, in this example it was found to be advantageous to display the results as a spectral suppression plot rather than as a ratiogram, as noted below.

Spectral suppression plots were described by Carter *et al.* [40] and Fell *et al.* [30], who established the theoretical basis. If the Beer–Lambert law is obeyed, i.e. the absorbance of a solution of pure analyte is directly proportional to the concentration of the analyte at a given wavelength, for any pure analyte the ratio of absorbances at any two wavelengths, 1 and 2, is constant and independent of concentration, and may be represented thus:

$$A_1/A_2 = K_{1,2} \qquad (1)$$

where A_1 and A_2 are the absorbances at each wavelength and $K_{1,2}$ is the AR constant.

By rearrangement of equation (1) the basis for eliminating the chromatographic response due to this pure, known analyte becomes:

$$A_1 - K_{1,2} \cdot A_2 = 0. \qquad (2)$$

Plotting the above expression as a function of chromatographic elution time yields a bipolar spectral suppression chromatogram where the response due to the known analyte is suppressed. Consequently, any deviations (positive or negative) from zero indicate the presence of an impurity [Fig. 2(b)]. Using mixtures of spectrally dissimilar compounds, e.g. carbamazepine and acridone [40], as little as 0.1% w/w of the co-eluting impurity could be detected when using an appropriate wavelength pair. While a number of workers have supported the use of several wavelength pairs in cases where the spectral characteristics of the potential co-eluting species are unknown, others have created and investigated the use of other univariate algorithms which incorporate spectral information from more than two wavelengths.

In 1981 Poile and Conlon of the Perkin–Elmer Corporation published an alternative

but closely related method to the absorbance ratio method [41]. This algorithm, called the *Absorbance Index*, computes the quotient, at a series of wavelengths, of the two spectra obtained at the inflection points on the leading and trailing edges of the chromatographic peak. In theory, for a homogeneous peak the resultant graph of this quotient against wavelength should be a straight line parallel to the wavelength axis. Deviations from such a straight line plot are indicative of spectral differences within the chromatographic peak and thus suggest the presence of an impurity. To facilitate the use of this algorithm in a numerical single-figure format, rather than using a graphical approach, Poile and Conlon proposed the use of a discriminator (D) to measure the degree of variation in the series of calculated quotient (Q) values, generated from a minimum of nine equally spaced wavelengths. The discriminator (D) may be defined as the quotient of the largest to the smallest Q values. For a homogeneous peak, under ideal conditions, D would be equal to unity. However, to accommodate the spectral differences resulting from instrumental and environmental variations (generally described as 'noise'), a value for D of between 1 and 1.5 was suggested as a practical limit for spectral discrimination in cases where the signal:noise ratio was greater than 10:1.

Other DAD manufacturers have also devised individual peak purity algorithms, notably Varian Chromatographic Systems, USA, and Hewlett–Packard, Germany. The *purity parameter* (PUP) developed for use with the Varian Polychrom 9060 detector was described in detail by Alfredson and Sheehan [42]. The PUP has been defined as the absorbance-weighted mean wavelentth of a spectrum at a specific elution time and is analogous to the statistical calculation of the moment of a distribution over a wavelength range from λ_0 to λ_n nm:

$$\mathrm{PUP}(A) = \frac{\sum_{i=0}^{i=n} A_i^2 \lambda_i}{\sum_{i=0}^{i=n} A_i^2}.$$

In practice, the calculation is performed on a discrete set of wavelengths and may be restricted to a subset of the full wavelength range. As with absorbance ratios, the PUP values obtained for each spectrum are wavelength (range) dependent and characteristic of the analyte; thus they can be used to assess both peak homogeneity and/or peak purity. A comparison of the use of absorbance ratios and PUP values for the discrimination of LC solutes has been presented by White [43]. Both techniques were shown to be capable of detecting 1% w/w of metoprolol in mixtures with atenolol. Further work, assessing the chromatographic peak purity of the benzodiazepines temazepam and lormetazepam based on PUP values, has been reported by Chan and Carr [44]. The PUP algorithm has also been used, in conjunction with other techniques, for the qualitative analysis of dyes [45]. In all the above publications it has been shown that it is necessary to optimize the wavelength range used when calculating the PUP to give maximum spectral discrimination. Although the technique has been found to be robust and sensitive in practice (Lincoln and Fell, unpublished data), the need for optimization may limit its usefulness and general applicability.

Various ratio techniques have been incorporated in the Hewlett–Packard HPLC Chemstation software. Using the *Color View* suite of programs [46], ratiograms can be constructed which normalize each spectrum collected for a LC peak (i.e. the spectra are scaled so that each spectrum has the same maximum absorbance value within the spectral range of interest); then points of equal absorbance are displayed graphically using a colour scale. Homogeneous peaks are indicated by a series of straight coloured bands, parallel to the time axis. These data may also be displayed as ratio plots, at defined wavelengths, and are analogous to the use of the AR plots discussed above. A further indicator of peak homogeneity is given by ratioing the maximum absorbance from each spectrum with the absorbance signal (i.e. LC peak trace) at a defined wavelength.

The combination of multiple absorbance ratios to give a single-figure assessment of peak purity has been examined by Marr *et al.* [47, 48]. *Multiple absorbance ratio correlation* (MARC), as proposed by Marr *et al.*, is based on the correlation between five-point absorbance ratio vectors at points distributed throughout the chromatographic peak [48]. This technique may be internally referenced by comparing the vectors obtained with that of the apical spectrum, thus assessing peak homogeneity (IMARC). Alternatively if a reference

spectrum is available as an external source, this may be used in the comparison, allowing both peak homogeneity and peak purity to be appraised (EMARC). By incorporating data at several wavelengths, the MARC algorithm was shown to be less sensitive to wavelength choice than the single AR [48].

The multiple absorbance ratio approach has also been used to successfully determine the composition of complex mixtures, and thus can indicate the presence or otherwise of any additional 'unknown' solutes in overlapping peak profiles. *Multiple spectral suppression* [49, 50] is similar in concept to the spectral suppression technique. Using matrix mathematics, the original equations can be expanded to incorporate data at additional wavelengths and hence several components can be suppressed simultaneously. Seaton *et al.* [50] have suggested that up to four known components can be suppressed using this technique.

The use of correlation coefficients in a data reduction and search system for digital absorbance spectra was proposed by Reid and Wong [51] and they were subsequently used for the computer retrieval of infrared spectra by Tanabe and Saeki [52]. Frank *et al.* [53] have reported the use of correlation coefficients, calculated from the absorbance valuies at six or more wavelengths obtained from a LC–DAD system, for the assessment of peak-homogeneity. The results published show this method to be more reliable than the use of an absorbance ratio. In addition, the presence of 2% w/w of a spectrally similar impurity (correlated with each other, $r = 0.9997$), eluting with a resolution of only 0.37 from the main LC peak, could be detected.

More recently correlation coefficients have been used for LC peak-identification [54] and peak-purity assessment [55, 56]. In the latter case the *multiple peak-area correlation technique* (MPACT) was described (Fig. 3). This

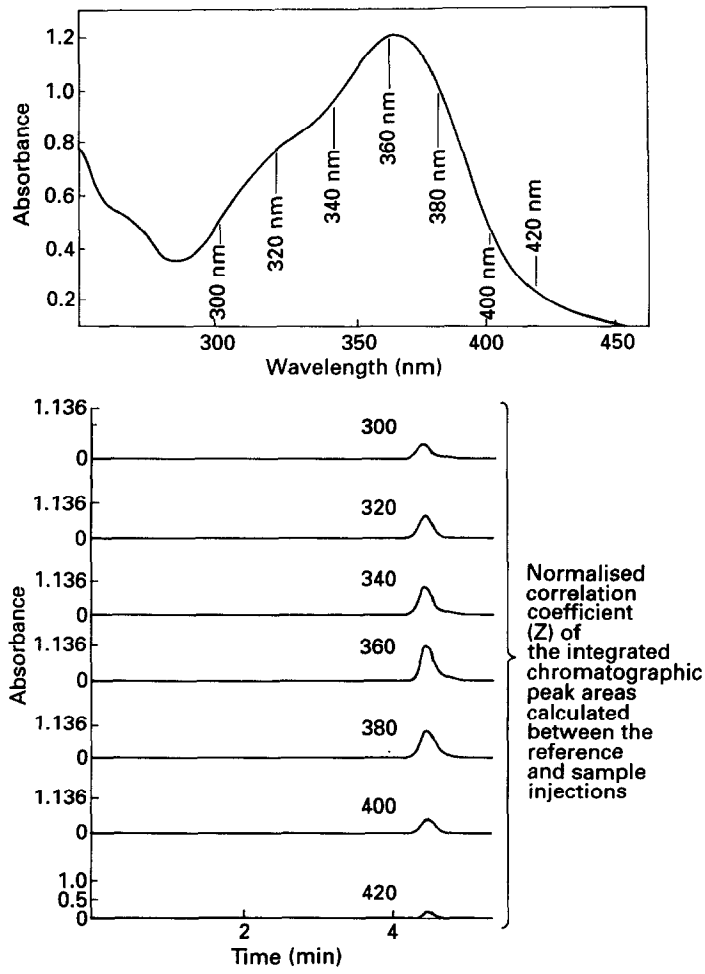

Figure 3
The multiple peak-area correlation technique (adapted from ref. 55).

technique determines the correlation between reference and sample peak-area data, calculated from the absorbance values at seven detection wavelengths, equally spaced throughout the absorption spectrum. The use of peak-areas minimizes the effects of variation in the signal:noise ratio of the data across the chromatographic peak. Furthermore, MPACT is not susceptible to the potential effects of spectral skewing, may be used with gradient elution and gives a single-figure purity-parameter suitable for routine statistical analysis. Using this technique, 1% w/w of a potentially related compound of sulphasalazine in a mixture with sulphasalazine could be detected, regardless of the chromatographic resolution between the two compounds [56].

The ability of multi wavelength LC detectors to generate large amounts of data necessitates a reduction of the data set prior to interrogation for an efficient application of peak-purity algorithms. The absorbance ratio methodology and many of the related techniques described above, bring about data reduction in the wavelength domain. Other peak-purity techniques have been developed which, in general, utilize all the spectral information, but with a non-continuous use of the time data.

The simplest form of this latter category is the visual comparison of the normalized absorption spectra collected at the apex and at the up- and down-slope inflection points of the chromatographic peak [57]. This simple *spectral overlay* method has remained popular, often being used in conjunction with other complementary peak-purity techniques, e.g. AR plots [58, 59]. Variations of this technique have been presented by George and Maute [31] and Zech *et al.* [60]. To highlight spectral differences, George and Maute generated a difference spectrum by subtracting the down-slope spectrum from that collected at the up-slope inflection point. In the work of Zech *et al.*, sample spectra obtained from the chromatogram were visually compared with reference spectra to assess peak purity.

Since absorption spectra characteristically lack the fine structure necessary for distinguishing between similar analytes, post-run processing of the spectral data to generate the second- and fourth-derivative spectra (with respect to wavelength) to enhance spectral differences has been examined, as has the $\log_{10}(A)$ transformation [61, 62]. The use of the second-derivative transformation in the wavelength domain has subsequently been used, for example, to enhance spectral differences in the characterization of phloroglucinols [63].

Another property of derivative spectra was employed in an early and novel approach to peak-purity determination using multiwavelength detection [64, 65]. Generation of the first derivative, $dA/d\lambda$, results in a zero signal at the λ_{max} of the spectrum. Consequently, Milano and co-workers showed that by plotting the first derivative of the spectrum (calculated at the λ_{max} of the main analyte for each point in the chromatogram) with respect to time, the signal due to the main analyte could be suppressed, so that any deviations from the baseline were indicative of co-eluting impurities.

An alternative approach using the data generated by the LC–DAD system is the examination of peak symmetry, and retention time shifts. These methods, which assess peak homogeneity, are based on the assumption that any coeluting species will distort the chromatographic peak profile of the main analyte at certain detection wavelengths. Consequently, if there is an impurity with a different spectrum, by monitoring at defined wavelength a change in either the *apparent* retention time, or the peak symmetry value will be detected. Although both these parameters are highly sensitive to changes in the chromatographic performance, Wright *et al.* [66] successfully incorporated the use of apparent retention time shifts into sequential simplex optimization software for the determination of peak overlap during method development.

Similarly, pseudo three-dimensional and isometric plots of the data collected by the LC–DAD system have been used to visually examine peak shape and thus to assess peak homogeneity [67, 68]. To enhance the visual discrimination of peak overlap, Fasanmade *et al.* [69] examined the use of pseudo three-dimensional derivative spectrochromatograms and the resulting isometric (contour) plots. The increased number of characteristic features afforded by the second-derivative transformation was found to give a significant improvement, although as with all the graphical approaches outlined above, operator interpretation is required, thus limiting the application in a standardized fashion.

Multivariate statistics applied to LC–DAD data matrices

Multivariate statistics can be defined as 'a collection of powerful mathematical tools that can be applied to chemical analysis when more than one measurement is acquired for each sample' [70]. In general, the various methods adapted and/or developed for peak deconvolution using LC–DAD 'spectrochromatograms' have the same starting point. The data matrix is divided into submatrices, each of which represent one set of overlapping elution profiles or an apparently single analyte peak. A principal components analysis (PCA) is then performed on each submatrix to determine the number of spectrally distinct and partially resolved analytes present.

This is followed by one of two general approaches to transform either the abstract spectra or abstract chromatograms (as represented by the eigenvectors generated using PCA) into pure spectra, or into single-analyte chromatographic profiles. The first approach, usually termed *curve resolution*, is based on the extrapolation of mixed spectra (or mixed chromatographic profiles) into pure spectra (or single-analyte chromatograms). Further details of this method are described by Sharaf and Kowalski [71, 72]. The second approach is based on an iterative process to construct the elution profiles followed by the least-squares fitting of these estimated elution profiles to the spectrochromatogram, thus deriving the individual spectra. Methods using this technique include; iterative target testing factor analysis (ITTFA) [73–75], evolving factor analysis (EFA) [76] and fixed size moving window evolving factor analysis [77]. While the above methods do not require prior knowledge about the constituents of multicomponent peaks, variations of the techniques have been developed to improve the quantification of known components in unknown mixtures, e.g. rank annihilation by evolving factor analysis [78].

From the brief discussion above it is apparent that the assessment of peak homogeneity using such multivariate techniques depends predominantly on the application of PCA. With model data (in the absence of noise), the number of analytes present is equal to the number of eigenvalues generated. In LC–DAD, the eigenvectors corresponding to the least significant eigenvalues merely reconstruct the noise in the system. A consequence of this is that the data submatrix can be recreated with reduced noise, prior to the application of peak deconvolution algorithms. The number of eigenvectors (corresponding to the highest eigenvalues) required to reproduce all spectra accurately is termed the rank. The determination of rank, and hence the homogeneity or heterogeneity of the chromatographic peak, is a non-trivial problem.

Many criteria have been developed for the determination of rank. In an examination of the methods available in 1976, Duewer and coworkers concluded that the concurrent use of two or more of these methods was required to give a good indication of the rank of a data matrix [79]. Since then a variety of techniques have been proposed, with the aim of improving the reliability of factor analysis. This may be attributed, in part, to the increasing number of potential applications for the multivariate analysis of chemical data.

In 1977, Malinowski published his 'Theory of Error in Factor Analysis' [80], in which he established that the eigenvalues resulting from PCA can be grouped into two sets: a primary set containing the true factors, together with a mixture of error, and a secondary set which consists of pure error. Moreover, three types of errors were shown to exist: real error (RE): imbedded error (IE) and extracted error (XE). Theoretically, the real error (RE) is the difference between the pure data, free from error, and the raw experimental data which contain experimental error. The extracted error (XE) is a measure of the error removed by the factor analysis, while the imbedded error (IE) represents the amount of error which remains in the data regenerated from the primary eigenvectors.

Using this approach Malinowski demonstrated that these parameters could be used to determine the rank of a data matrix [80–82]. In cases where the experimental error in the data matrix can be modelled, the RE parameter (also termed the residual standard deviation) can be used. Initially, the RE value is calculated by incorporating only data from the most significant eigenvalue extracted. If the RE value is equivalent to the known standard deviation of the system noise then the rank is one. If the RE value is greater than the experimental noise then the next largest eigenvalue is included in the calculation. This is repeated until the RE value gives a good estimate of the system noise.

In cases where prior knowledge of the error is not obtainable, the IE values extracted by the interpolation process outlined above can be used. Since RE $\geq$IE, and assuming that the errors are distributed randomly and uniformly throughout the data matrix, then as increasingly more primary eigenvectors are included in the reconstruction process, the IE value should decrease. Furthermore, as secondary eigenvectors are included in the reproduction, the IE value will begin to increase. Consequently, the point at which the IE function reaches a minimum should indicate the rank. In practice, non-uniformity of the error distribution, exaggerated by the PCA technique, limits the application of this technique.

An empirical algorithm, the factor indicator function (IND), based on the error functions, has been shown to be more robust in certain cases of the non-uniformity of the error distribution [81]. However, as the IND function is not generally applicable and the theoretical basis of the algorithm is uncertain, other methodologies have subsequently been developed by Malinowski. Most notable of these techniques is the use of reduced eigenvalues [83] and their subsequent inclusion into a statistical *F*-test to test for the equivalence of suspected secondary eigenvalues [84]. These techniques also rely on the data matrices containing uniformly distributed errors, and should therefore be used cautiously.

While the above approaches use information obtained from the eigenvalues, other workers in this area have examined the use of eigenvectors for the determination of rank. Shrager and Hendler [85] proposed the use of an autocorrelation function of eigenvectors to identify those eigenvectors with a high amount of noise. These eigenvectors were considered to be non-significant, i.e. those corresponding to secondary eigenvalues. Subsequently, Rossi and Warner used the frequency distribution of Fourier-transformed eigenvectors as an alternative method of identifying eigenvectors with a high amount of noise [86]. Recently the multivariate technique of canonical correlation has been successfully applied to rank estimations, using the emission–excitation matrices of fluorescent mixtures as experimental data [87]. Further examination of this technique is required before its full potential will be known.

An alternative approach to rank determination is the use of cross-validation, first applied to chemical data by Wold in 1978 [88]. Cross-validation may be performed by either deleting a fraction of the data points or a fraction of the sample from the data matrix, followed by a PCA of the reduced data set. This process is repeated until all of the data fractions have been omitted in turn. The optimal number of principal components (PCs) required for reproducing the deleted data points or samples is determined by minimizing the predictive residual error sum of squares (PRESS). This number of significant PCs is used as an estimate of the rank of the data matrix.

Of all the methods described cross-validation is probably the most reliable. However the other techniques described above are, in general, easier to implement, as they utilize information calculated during PCA. Consequently, the concurrent use of several of these algorithms, and/or those methods described by Duewer *et al.* [79], continues to be one of the best practical approaches to determining the true rank of a data matrix and hence for the assessment of peak homogeneity using LC–DAD with multivariate statistics.

Conclusions

This review of peak-purity and peak-homogeneity assessment techniques clearly shows that no single algorithm is capable of fulfilling all the requirements of the analyst. It is however possible to group the various methods into several categories, thus allowing a rationale to be developed to identify the most useful algorithms for different requirements and under various constraints (Fig. 4). Although the preferred method(s) in each group are highlighted, more dependable results may be obtained by the concurrent use of several complementary techniques, as illustrated by rank determination in PCA. Moreover, using this approach a collection of several univariate techniques has been incorporated into an expert system to improve the reliability of operator-independent peak-homogeneity evaluations [88–91].

All the practical techniques outlined above rely on differences in detector response between the parent compound and the co-eluting impurity. In general, the differences are those between the absorption spectra of the two compounds. The UV–vis spectra characteristically lack fine structure and hence, despite

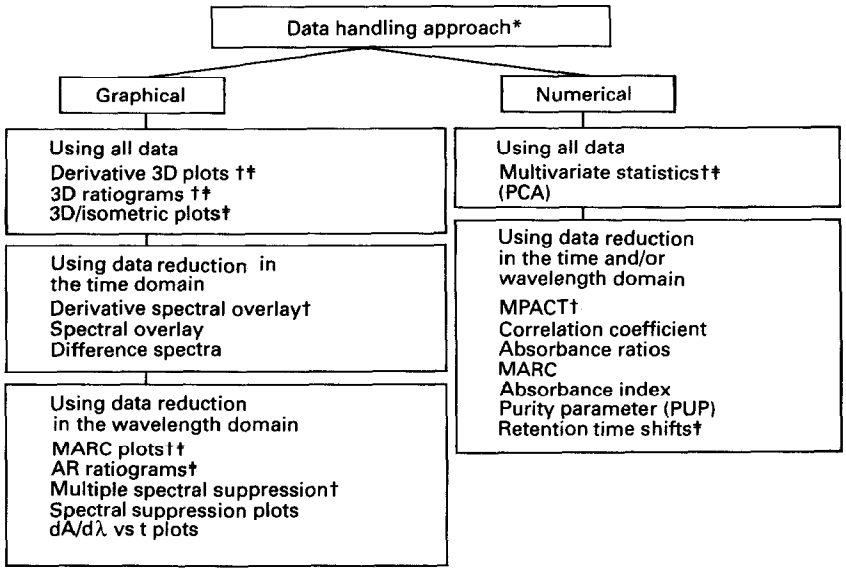

Figure 4
Peak-purity and/or peak-homogeneity assessment methods. A rationale for use with LC–DAD data. *See text for a description of the alternative approaches based on single-channel data or statistical modelling. †The methods of choice, as highlighted by the description of their relative merits in the text. ‡These methods assess only peak homogeneity, whereas the other methods may be used to assess both peak homogeneity and peak purity when a suitable reference standard is available.

the advances made in peak-purity assessment techniques, the detection of structurally similar impurities eluting simultaneously with the parent compound is a continuing problem. While enhanced spectral discrimination of LC analytes containing a chromophore has been achieved through the use of post-column technology to generate pH-shifted spectra [63, 92, 93], the use of this technique generates additional noise. The increased noise is such that the practical application of post-column technology for enhanced peak-purity determinations has not been found to be viable [94].

The availability of mass spectrometry as a reliable and robust LC detector should, when used with multivariate statistical analysis, overcome the problem of spectral discrimination and thus allow the assessment of LC peak purity and peak homogeneity to be made with a higher degree of confidence.

Acknowledgements — Thanks are extended to Bertil Sellberg and Rolf Modin at Kabi Pharmacia Therapeutics AB for stimulating discussions. One of the authors (J.B.C.) would like to thank Kabi Pharmacia Therapeutics AB (Uppsala, Sweden) for kindly providing the studentship for this research.

References

[1] J. Van Rompay, *J. Pharm. Biomed. Anal.* **4**, 725–732 (1984).
[2] A.N. Papas, *CRC Crit. Revs Anal. Chem.* **20**, 359–404 (1989).
[3] J.C. Giddings, *Anal. Chem.* **39**, 1027–1028 (1967).
[4] J.M. Davis and J.C. Giddings, *Anal. Chem.* **55**, 418–424 (1983).
[5] J.M. Davis and J.C. Giddings, *Anal. Chem.* **57**, 2168–2177 (1985).
[6] J.M. Davis and J.C. Giddings, *Anal. Chem.* **57**, 2178–2182 (1985).
[7] A.N. Papas and T.P. Tougas, *Anal. Chem.* **62**, 234–239 (1990).
[8] A.F. Fell, *Anal. Proc.* **17**, 266–271 (1980).
[9] A.F. Fell, *Anal. Proc.* **17**, 512–519 (1980).
[10] J.W. Ashley Jr and C.N. Reilley, *Anal. Chem.* **37**, 626–630 (1965).
[11] G.C. Allen and R.F. McMeeking, *Anal. Chim. Acta* **103**, 73–108 (1978).
[12] J.C. Berridge, *Chromatographia* **16**, 172–174 (1982).
[13] J.C. Berridge and K.S. Andrews, *Analyst* **109**, 287–289 (1984).
[14] J.L. Excoffier and G. Guiochon, *Chromatographia* **15**, 543–545 (1982).
[15] N.E. Skelly and W.B. Crummett, *J. Chromatogr.* **55**, 309–318 (1971).
[16] N. Ostojic, *Anal. Chem.* **46**, 1653–1659 (1974).
[17] R.N. Smith and C.G. Vaughan, *J. Chromatogr.* **129**, 347–354 (1976).
[18] R.N. Smith and M. Zetlein, *J. Chromatogr.* **130**, 314–317 (1977).
[19] A.M. Krstulovic, D.M. Rosie and P.R. Brown, *Anal. Chem.* **48**, 1383–1386 (1976).
[20] A.M. Krstulovic, P.R. Brown and D.M. Rosie, *Anal. Chem.* **49**, 2237–2241 (1977).
[21] R. Yost, J. Stoven and W. MacLean, *J. Chromatogr.* **134**, 73–82 (1977).
[22] J.K. Baker, R.E. Skelton and C.Y. Ma, *J. Chromatogr.* **168**, 417–427 (1979).
[23] H. Hotelling, *J. Educ. Psychol.* **24**, 417–441 (1933).
[24] W.H. Lawton and E.A. Sylvestre, *Technometrics* **13**, 617–633 (1971).

[25] J.M. Halket, *J. Chromatogr.* **186**, 443–455 (1979).
[26] J.M. Halket, *J. High Resol. Chromatogr. — Chromatogr. Comm.* **2**, 197–198 (1979).
[27] R.E. Santini, M.J. Milano and H.L. Parduc, *Anal. Chem.* **45**, 915A–927A (1973).
[28] M.J. Milano, S. Lam, M. Savonis, D.B. Pautler, J.W. Pav and E. Grushka, *J. Chromatogr.* **149**, 599–614 (1978).
[29] A.F. Fell, B.J. Clark and H.P. Scott, *J. Pharm. Biomed. Anal.* **1**, 557–572 (1983).
[30] A.F. Fell, H.P. Scott, R. Gill and A.C. Moffat, *J. Chromatogr.* **282**, 123–140 (1983).
[31] S.A. George and A. Maute, *Chromatographia* **15**, 419–425 (1982).
[32] B.F.H. Drenth, R.T. Ghijsen and R.A. De Zeeuw, *J. Chromatogr.* **238**, 113–120 (1982).
[33] S.D. Brown, *Anal. Chem.* **62**, 84R–101R (1990).
[34] A. Bylina, D. Sybilska, Z.R. Grabowski and J. Koszewski, *J. Chromatogr.* **83**, 357–362 (1973).
[35] P.A. Webb, D. Ball and T. Thornton, *J. Chromatogr. Sci.* **21**, 447–453 (1983).
[36] A.C.J.H. Drouen, H.A.H. Billiet and L. De Galan, *Anal. Chem.* **56**, 971–978 (1984).
[37] P.C. White, *J. Chromatogr.* **200**, 271–276 (1980).
[38] P.C. White and T. Catterick, *J. Chromatogr.* **280**, 376–381 (1983).
[39] P.C. White and T. Catterick, *J. Chromatogr.* **402**, 135–147 (1987).
[40] G.T. Carter, R.E. Schiesswohl, H. Burke and R. Yang, *J. Pharm. Sci.* **71**, 317–321 (1982).
[41] A.F. Poile and R.D. Conlon, *J. Chromatogr.* **204**, 149–152 (1981).
[42] T. Alfredson and T. Sheehan, *Am. Lab.* **17**, 40–54 (1985).
[43] P.C. White, *Analyst* **113**, 1625–1629 (1988).
[44] H.K. Chan and G.P. Carr, *J. Pharm. Biomed. Anal.* **8**, 271–277 (1990).
[45] P.C. White and T. Catterick, *Analyst* **115**, 919–923 (1990).
[46] Hewlett-Packard, HPLC Chemstation: Using Color View, Part No. 79994-90011, 1st edn, March 1987.
[47] J.G.D. Marr, Ph.D. Thesis, University of Bradford, UK (1988).
[48] J.G.D. Marr, G.G.R. Seaton, B.J. Clark and A.F. Fell, *J. Chromatogr.* **506**, 289–301 (1990).
[49] J.G.D. Marr, P. Horvath, B.J. Clark and A.F. Fell, *Anal. Proc.* **23**, 254–256 (1986).
[50] G.G.R. Seaton, J.G.D. Marr, B.J. Clark and A.F. Fell, *Anal. Proc.* **23**, 424–426 (1986).
[51] J.C. Reid and E.C. Wong, *Appl. Spectrosc.* **20**, 320–325 (1966).
[52] K. Tanabe and S. Saéki, *Anal. Chem.* **47**, 118–122 (1975).
[53] J. Frank Jr, A. Braat and J.A. Duine, *Anal. Biochem.* **162**, 65–73 (1987).
[54] H.P. Sievert, S. Wu, R. Chloupek and W.S. Hancock, *J. Chromatogr.* **499**, 221–234 (1990).
[55] J.B. Castledine, A.F. Fell, R. Modin and B. Sellberg, *Anal. Proc.* **29**, 100–104 (1992).
[56] J.B. Castledine, A.F. Fell, R. Modin and B. Sellberg, *J. Chromatogr.* **592**, 27–36 (1992).
[57] J.C. Miller, S.A. George and B.G. Willis, *Science* **218**, 241–246 (1982).
[58] F. Nyberg, C. Pernow, U. Moberg and R.B. Erikson, *J. Chromatogr.* **359**, 541–551 (1986).
[59] E.M. Kirk and A.F. Fell, *Clin. Chem.* **35**, 1288–1292 (1989).
[60] K. Zech, R. Huber and H. Elgass, *J. Chromatogr.* **282**, 161–167 (1983).
[61] A.F. Fell, H.P. Scott, R. Gill and A.C. Moffat, *Chromatographia* **16**, 69–78 (1982).
[62] A.F. Fell, H.P. Scott, R. Gill and A.C. Moffat, *J. Chromatogr.* **273**, 3–17 (1983).
[63] A.F. Fell, T.Z. Woldemariam, P.A. Linley, G. Jian, M.D. Luque De Castro and M. Valcárcel, *Anal. Chim. Acta* **234**, 89–95 (1990).
[64] M.J. Milano, S. Lam and E. Grushka, *J. Chromatogr.* **125**, 315–326 (1976).
[65] M.J. Milano and E. Grushka, *J. Chromatogr.* **133**, 352–354 (1977).
[66] A.G. Wright, A.F. Fell and J.C. Berridge, *Chromatographia* **24**, 533–540 (1987).
[67] B.J. Clark, A.F. Fell, H.P. Scott and D. Westerlund, *J. Chromatogr.* **286**, 261–273 (1984).
[68] A.F. Fell, B.J. Clark and H.P. Scott, *J. Chromatogr.* **297**, 203–214 (1984).
[69] A.A. Fasanmade, A.F. Fell and H.P. Scott, *Anal. Chim. Acta* **187**, 233–240 (1986).
[70] K.R. Beebe and B.R. Kowalski, *Anal. Chem.* **59**, 1007A–1017A (1987).
[71] M.A. Sharaf and B.R. Kowalski, *Anal. Chem.* **53**, 518–522 (1981).
[72] M.A. Sharaf and B.R. Kowalski, *Anal. Chem.* **54**, 1291–1296 (1982).
[73] P.J. Gemperline, *J. Chem. Inf. Comput. Sci.* **24**, 206–212 (1984).
[74] B.G.M. Vandeginste, W. Derks and G. Yateman, *Anal. Chim. Acta* **173**, 253–264 (1985).
[75] G.G.R. Seaton and A.F. Fell, *Chromatographia* **24**, 208–216 (1987).
[76] M. Maeder, *Anal. Chem.* **59**, 527–530 (1987).
[77] H.R. Keller and D.L. Massart, *Anal. Chim. Acta* **246**, 379–390 (1991).
[78] H. Gamp, M. Maeder, C.J. Meyer and A.D. Zuberbuehler, *Anal. Chim. Acta* **193**, 287–293 (1987).
[79] D.L. Duewer, B.R. Kowalski and J.L. Fasching, *Anal. Chem.* **48**, 2002–2010 (1976).
[80] E.R. Malinowski, *Anal. Chem.* **49**, 606–612 (1977).
[81] E.R. Malinowski, *Anal. Chem.* **49**, 612–617 (1977).
[82] E.R. Malinowski, *Anal. Chim. Acta* **103**, 339–354 (1978).
[83] E.R. Malinowski, *J. Chemom.* **1**, 33–40 (1987).
[84] E.R. Malinowski, *J. Chemom.* **3**, 49–60 (1988).
[85] R.I. Shager and R.W. Hendler, *Anal. Chem.* **54**, 1147–1152 (1982).
[86] T.M. Rossi and I.M. Warner, *Anal. Chem.* **58**, 810–815 (1986).
[87] X.M. Tu, D.S. Burdick, D.W. Millican and L.B. McGown, *Anal. Chem.* **61**, 2219–2224 (1989).
[88] S. Wold, *Technometrics* **20**, 397–405 (1978).
[89] A.F. Fell, T.P. Bridge and M.H. Williams, *J. Pharm. Biomed. Anal.* **6**, 555–564 (1988).
[90] T.P. Bridge, M.H. Williams and A.F. Fell, *J. Chromatogr.* **465**, 59–67 (1989).
[91] T.P. Bridge, M.H. Williams and A.F. Fell, *Anal. Chim. Acta* **223**, 175–182 (1989).
[92] K. Hostettmann, B. Domon, D. Schaufelberger and M. Hostettmann, *J. Chromatogr.* **283**, 137–147 (1984).
[93] A.F. Fell, J.B. Castledine, B. Sellberg, R. Modin and R. Weinberger, *J. Chromatogr.* **535**, 33–39 (1990).
[94] J.B. Castledine, A.F. Fell, R. Modin and B. Sellberg, *J. Chromatogr.* **626**, 127–134 (1992).

[Received for review 22 July 1992; revised manuscript received 25 August 1992]