Journal of Chromatography, 631 (1993) 15–21
Elsevier Science Publishers B.V., Amsterdam

CHROMSYMP. 2632

Faster quantitative evaluation of high-performance liquid chromatography–ultraviolet diode-array data by multicomponent analysis

Jean-Louis Excoffier, Maureen Joseph, John J. Robinson and Terry L. Sheehan

Varian Chromatography Systems, 2700 Mitchell Drive, Walnut Creek, CA 94598 (USA)

ABSTRACT

Multicomponent analysis is a proven technique to analyze mixes of known UV-absorbing compounds in unknown amounts. While it is routinely used in spectrophotometers, its application to chromatography has so far been limited. This paper discusses an approach to obtain fast and accurate quantitation of unresolved chromatographic data, while providing for the detection of unexpected impurities. The applications of the technique to different conditions of chromatographic resolution are discussed, along with its potential for increasing sample throughput.

INTRODUCTION

Diode arrays are routinely used for qualitative analysis: spectra are used to confirm the identity of a peak, and spectral differences are exploited to assess peak purity. Quantitative analysis, however, still relies on the integration of unidimensional data, and therefore requires baseline resolution of all the compounds to be accurately quantitated.

Overlapping peaks are a fact of the chromatographer's life. Statistical studies [1,2] have shown that overlapping peaks are a likely occurrence in complex mixture, even with high efficiency separations. Even when baseline separation is possible, it may not be desirable, as costs increase with analysis time. Hence, there is an interest in multivariate techniques for the analysis of coelutions.

Different deconvolution techniques based on factor analysis have been proposed in the past to attempt to resolve an overlap of compounds with different UV spectra without knowledge of these spectra [3–5]. These techniques have several limitations: (i) approaching zero resolution, different spectra combine into a single factor; (ii) in the generic case of three or more peaks with low resolution, the problem has an infinity of solutions, all of which satisfy the constraints applied; (iii) not providing reference spectra limits these techniques to qualitative results.

On the other hand, multicomponent analysis (MCA) based on reference spectra works at any degree of resolution and provides accurate quantitative results validated by qualitative results. It also provides an optimum tool for the detection of unexpected impurities.

THEORETICAL

Multicomponent analysis

The basic hypothesis is that the unknown spectrum **U** results from the combination of known spectra **A**, **B**, **C** in "amounts" a, b, c, ...

$$\mathbf{U} = a\,\mathbf{A} + b\,\mathbf{B} + c\,\mathbf{C} + \ldots \quad (1)$$

This system of linear equations can be expressed in matrix notation (**S** is the matrix of the spectra, **X** is the solution vector a, b, c, ...):

$$\mathbf{U} = \mathbf{S} \cdot \mathbf{X} \quad (2)$$

Correspondence to: Jean-Louis Excoffier, Varian Chromatography Systems, 2700 Mitchell Drive, Walnut Creek, CA 94598, USA.

0021-9673/93/$06.00 © 1993 Elsevier Science Publishers B.V. All rights reserved

In the typical situation there are more wavelengths than compounds analyzed and the system is overdetermined. The matrix **S** is not square and cannot be inverted. The system is resolved by a simple "least squares" method. First, both sides of the equation are multiplied on the left by the transpose of the spectra matrix:

$$^{t}\mathbf{S} \cdot \mathbf{U} = {}^{t}\mathbf{S} \cdot \mathbf{S} \cdot \mathbf{X} \quad (3)$$

then the square matrix $^{t}\mathbf{S} \cdot \mathbf{S}$, which is the correlation matrix of the reference spectra can be inverted, provided the reference spectra are linearly independent. The solution is therefore given by:

$$\mathbf{X} = ({}^{t}\mathbf{S} \cdot \mathbf{S})^{-1} \cdot {}^{t}\mathbf{S} \cdot \mathbf{U} \quad (4)$$

Because the system is overdetermined, the validity of the decomposition can be assessed by comparing the original data set or target spectrum with the reconstructed spectrum obtained by combining the reference spectra in the calculated amounts. The mathematical aspects of this analysis have been presented in greater detail by Blackburn [6].

APPLICATION OF MCA TO CHROMATOGRAPHY

In order to apply this method to HPLC with diode array detection, the concept of *area spectrum* was developed. An area spectrum is the result of the baseline-corrected integration of a region of a chromatogram at each wavelength, and is therefore a spectrum whose coordinates are expressed as peak area instead of peak height. An area spectrum can be thought of as the baseline-corrected spectrum of an elution fraction (Fig. 1a and b). This elution fraction might contain a single peak or a series of closely eluting peaks.

The area spectrum of a pure peak is identical in shape to the spectrum of the compound. Its magnitude represents an amount injected for a given detection system at a given flow-rate.

Spectrum analysis

Once area spectra have been collected for injections of known amounts of standards, the area spectrum corresponding to the overlapping peak of an unknown mix can be analyzed by MCA to provide quantitative results (Table I).

By computing the linear combination of the reference spectra with the multiplicative factors determined, one obtains a *reconstructed spectrum*, which ideally would be identical to the original or *target spectrum*. To assess the validity of the quantitative results, the reconstructed spectrum is com-

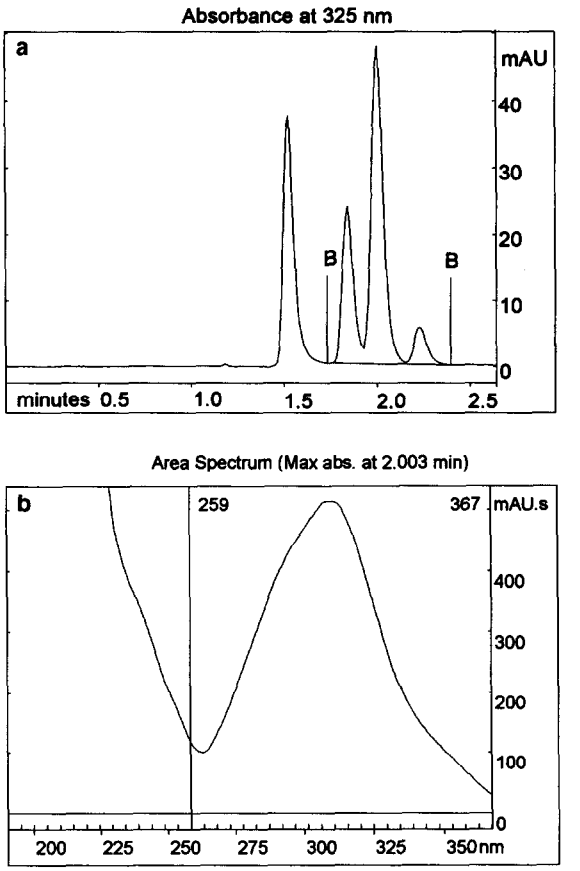

Fig. 1 (a) Integration of a chromatographic region to obtain an area spectrum, showing integration limits (B) and subtracted baseline. (b) Resulting area spectrum.

TABLE I

SPECTRUM ANALYSIS RESULTS OVER 259–367 nm

Name	Amount
MA	0.175 μg
OS	0.134 μg
OMC	0.119 μg
Similarity	1.00000
Dissimilarity	0.00192

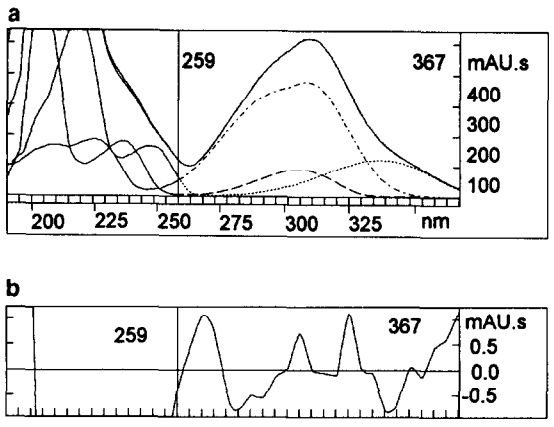

Fig. 2. (a) Spectral overlay. - · - · = OMC; – – – = MA; ----- = OS; ——— = sum and target. (b) Expanded difference between reconstructed sum and target.

pared to the target spectrum. This comparison can be carried out by graphical and numerical methods.

The graphical evaluation consists of an overlay of the target, the reconstructed sum, and the scaled component spectra (Fig. 2a). Because such a visual evaluation is subjective, small differences are hard to quantify. To make the difference between the target and the reconstructed sum easier to visualize, this difference is plotted separately on an expanded scale (Fig. 2b).

The numerical evaluation of the comparison is provided by two complementary parameters: similarity and dissimilarity. The similarity is also known as the correlation coefficient and is the traditional mathematical approach to comparing two sets of data. However, when the spectra are very similar, the correlation coefficient gets very close to one, is a quadratic measure of the difference between the spectra and is, therefore, not a very human-friendly tool. From the observation that the correlation coefficient is the cosine of the angle between the two vectors stems the idea of using the sine of the angle, which can be called dissimilarity. This parameter has an interesting property: since the sine function is a linear measure of small angles, the dissimilarity provides a linear measurement of the impurity level when comparing to a pure reference.

In a mathematical form, A and B being two spectra of coordinates $\mathbf{A}_i$ and $\mathbf{B}_i$, the parameters can be expressed as:

$$\text{Sim}\ (A, B) = \frac{\sum \mathbf{A}_i \mathbf{B}_i}{\sqrt{\sum \mathbf{A}_i^2} \sqrt{\sum \mathbf{B}_i^2}} \quad (5)$$

$$\text{Dissim}\ (A, B) = \sqrt{1 - \text{Sim}^2\ (A, B)} \quad (6)$$

Identical spectra have a similarity of 1 and a dissimilarity of 0. Spectra with a dissimilarity of 0.001 have a similarity of 0.999999. A dissimilarity of 0.002 correspond to a double level of the same impurity, while the corresponding similarity value would be 0.999996.

The dissimilarity parameter can be used to empirically or theoretically establish an acceptable error level based on the noise level. An empirical approach will take into account all experimental errors, instead of basing predictions on a model of the error.

Chromatogram analysis

At each point of the chromatographic region integrated, the spectrum can be decomposed on the reference spectra. This provides an absorbance value for each component and a residual error. As a result, the individual peak profiles and the sum chromatogram can be reconstructed and overlaid to the original chromatogram (Fig. 3a). Basic parameters such as retention time, peak height, peak width, and resolution can be derived from the peak profiles (Table II).

Error analysis

The residual error between the reconstructed spectrum and target spectrum in each point of the

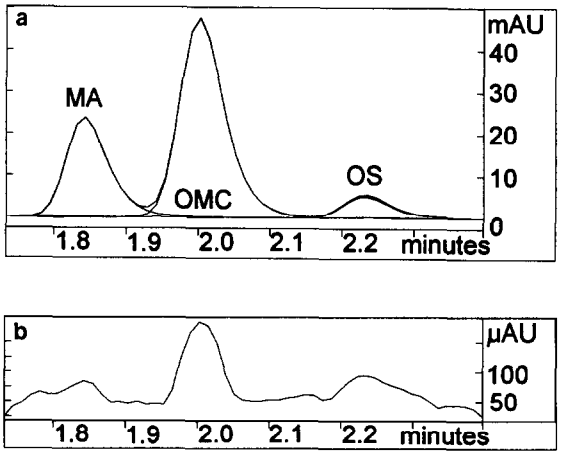

Fig. 3. (a) Reconstructed peak profiles. (b) Error chromatogram.

TABLE II

PLOT ANALYSIS RESULTS AT 325 nm

Name	t_R (min)	Peak height mAU	Peak width (s)	R_s
MA	1.842	23.86	4.1	–
OMC	2.002	47.88	4.2	1.37
OS	2.231	5.08	4.6	1.85

chromatogram can be reduced to a root mean square value:

$$\mathrm{RMS} = \sqrt{\frac{\sum_{i=1}^{i=n} (\mathbf{t}_i - \mathbf{r}_i)^2}{n}} \quad (7)$$

where $\mathbf{t}_i$ and $\mathbf{r}_i$ are the coordinates of the target and reconstructed spectra at channel i. Plotting this residual error for the whole region analyzed will indicate the location of spectral discrepancies caused by unexpected impurities (Fig. 3b). It is equivalent to plotting the modulus of the result of spectral suppression of the component spectra from the original chromatogram [7]. This error plot detects any spectral characteristics that cannot be explained by the component spectra and is therefore an optimum tool for the detection of unexpected impurities.

Application to purity analysis

It is worth mentioning that this technique can be used to analyze the spectral homogeneity of a peak independently of any reference. In this case, every spectrum in the peak is compared to the area spectrum, *i.e.*, the average spectrum of the peak. The error plot would reveal the magnitude and location of the spectral discrepancy. An RMS threshold can be derived from the noise level, and used as an objective purity criterion. The threshold can be determined empirically or based on the error analysis theory developed by Malinovski [8–10].

Other sources of error

Since MCA is based on linear mathematical techniques, any non-linearity in the spectra results in spectral inhomogeneity. Besides the well-known scattered light effect, typically observed at absorbances greater than 1.5 AU, non-linearity can also result from scan rate and optical bandwidth effects.

The scanning effect, due to concentration changes during the scanning of the spectrum, can be avoided by a correction in the scanning software as advocated by Keller *et al.* [11] and implemented in the instrumentation used in this study since 1985. As reported by Dose and Guiochon [12], the diode-array hardware typically averages the intensity across the optical bandwidth. This average intensity does not follow the Beer-Lambert law, since the average absorbance is not the logarithm of the average intensity. As a result, regions of fine structure or of high slope/absorbance ratio will show small non-linearities at relatively low absorbances (100 mAU). The effect of these non-linearities may exceed the noise level and create an error pattern. However, since this effect is concentration dependent, the pattern will be centered on the peak. An isolated peak in the error plot, not lined up with one of the component peaks, can only be interpreted as an impurity.

EXPERIMENTAL

Mixes of four sunscreen components were analyzed under various resolution conditions: baseline resolution, partial resolution, and no resolution.

Instrumentation

All experiments were performed on a Varian LC Star Workstation, composed of a 9010 ternary gradient pump, 9100 AutoSampler, PolyChrom 9065 diode-array detector, equipped with Rev C LC Star Workstation Software, and Rev E PolyView Spectral Processing Software (Varian Chromatography Systems, Walnut Creek, CA, USA). The column used for the separation was a Varian MicroPak SP-C8-IP-5, 5 μm, 15 cm × 4.0 mm I.D. (Varian, Sunnyvale, CA, USA).

Chemicals

The sunscreen agents, menthyl-*o*-aminobenzoate (MA), 2-ethylhexyl *p*-methoxycinnamate (OMC), and 2-ethylhexyl salicylate (OS), were supplied by Haarman & Reimer (Springfield, NJ, USA). The fourth sunscreen agent 2-hydroxy-4-methoxybenzophenone (OXY, oxybenzone) was supplied by Sig-

ma (St. Louis, MO, USA). All reagents were dissolved in methanol. All solvents were supplied by Fisher Scientific (Fair Lawn, NJ, USA). The methanol was optima grade and the water was HPLC grade.

Chromatographic procedures

Baseline resolution. A complete separation of the four sunscreen components was achieved using a mobile phase gradient of methanol–water (75:25) to methanol–water (95:5) in 8 min at a flow-rate of 1 ml/min. Acetic acid was added to the mobile phase at 0.3%. The 9065 Polychrom diode-array detector was set to display the chromatogram at 306 nm. Data was acquired over the 190–369 nm wavelength range.

A complete separation was also obtained isocratically on the same column with a mobile phase of methanol–water (85:15) at 1 ml/min. A PolyView–MCA library was built to quantitate the sunscreen components based on injections of the individual components using the same mobile phase conditions as the analysis. The library amounts (on column) were: OXY = 406.8 ng, MA = 293.2 ng, OMC = 278.8 ng, and OS = 440.4 ng.

Partial resolution. In order to reduce the resolution, the mobile phase was adjusted to 100% methanol and run isocratically at both 1 ml/min and 2 ml/min. A library was built using the same standards as above, but injected individually using the 100% methanol mobile phase.

Zero resolution. In this case the sunscreen mixture was reduced to three components (OMC, MA, OS) and a 50 ft. length of coiled stainless-steel tubing was inserted in place of the analytical column. The mobile phase was 100% methanol. Mixtures of the sunscreens were analyzed over the flow-rate range 1–4 ml/min. The mixture injected contained 538.5 ng OMC, 448.5 ng MA, and 349.5 ng OS. To build the library for use with the PolyView–MCA software standard solutions of the sunscreen components were injected individually. These levels were OS = 862 ng, MA = 1.913 ug, and OMC = 6.344 ug. A new library was built since the mobile phase contained no acetic acid, and spectral shifts could occur due to the change in mobile phase.

RESULTS AND DISCUSSION

Baseline resolution

The gradient conditions led to a baseline separation in 10 min (Fig. 4). Results achieved with MCA are presented in Table III. The interpolating baseline correction removes any baseline effects due to the gradient profile. The standard for the first compound (OXY) was found to contain an impurity interfering with the other three and therefore was not used in studies at lower resolutions.

Partial resolution

The plot analysis (Fig. 3a) shows the reconstructed peak profiles and a residual error or less

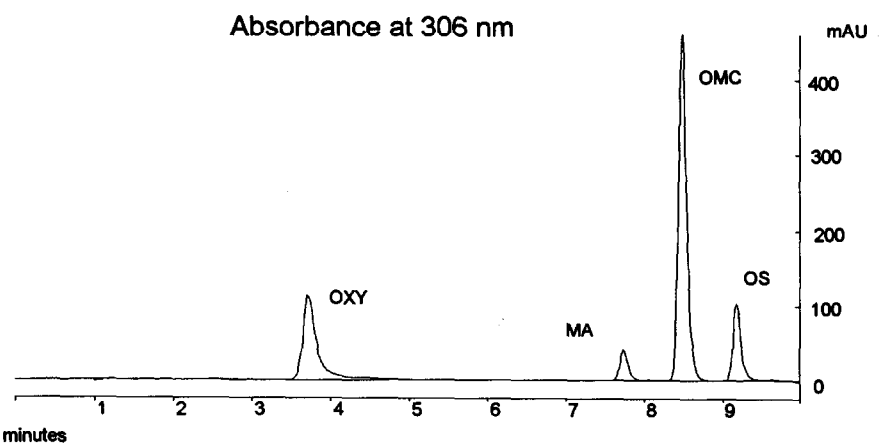

Fig. 4. Baseline separation in 10 min (gradient).

TABLE III

QUANTITATION RESULTS FOR BASELINE SEPARATION ($n = 6$)

Compound	Resolution	Accuracy (%)	Precision (%)	k'
2-Hydroxy-4-methoxybenzophenone	9.20	101	1.34	2.36
Menthyl-*o*-aminobenzoate	16.90	98.1	0.87	5.98
2-Ethylhexyl *p*-methoxycinnamate	4.13	97.1	0.27	6.66
2-Ethylhexyl salicylate	4.00	99.5	2.7	7.29

TABLE IV

QUANTITATION RESULTS WITH LIMITED RESOLUTION ($n = 6$)

Compound	1 ml/min				2 ml/min			
	Resolution	Accuracy (%)	Precision (%)	k'	Resolution	Accuracy (%)	Precision (%)	k'
Menthyl-*o*-aminobenzoate	8.00	102	1.42	0.79	2.83	100	0.21	0.60
2-Ethylhexyl *p*-methoxycinnamate	1.05	101	0.58	0.92	0.69	101	0.26	0.69
2-Ethylhexyl salicylate	1.99	101	0.84	1.17	1.62	95.4	0.84	0.89

TABLE V

FIA QUANTITATION WITH ZERO RESOLUTION ($n = 6$)

Compound	4 ml/min			3 ml/min		
	Resolution	Accuracy (%)	Precision (%)	Resolution	Accuracy (%)	Precision (%)
Methyl-*o*-aminobenzoate	0.00	100.6	0.4	0.00	99.6	0.4
2-Ethylhexyl *p*-methoxycinnamate	0.00	98.0	0.5	0.00	98.0	0.4
2-Ethylhexyl salicylate	0.00	102.3	0.8	0.00	96.0	0.6

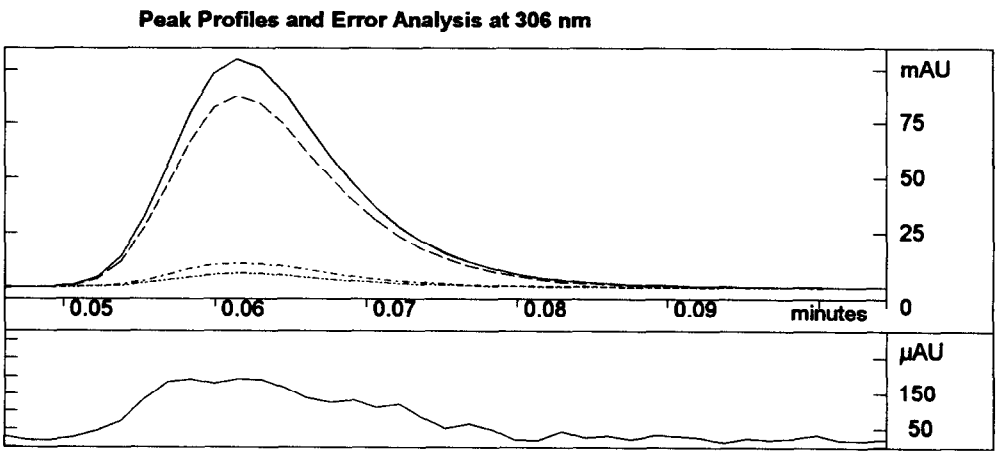

Fig. 5. Reconstructed peak profiles at zero resolution. – – – = OMC; – · – · – = OS; – – – – – = MA; ——— = sum and target.

TABLE VI

SAMPLE THROUGHPUT *VS.* TYPE OF SEPARATION

Separation	Run time (min)	Equilibration time (min)	Report time (min)	Total analysis time (min)	Samples per hour
Baseline (gradient)	10	10	1.5	21.5	2.8
Baseline (isocratic)	16	0	1.5	17.5	5.2
Incomplete	1.5	0	1.5	3	20
FIA	0.1	0	1.5	1.6	37.5

than 150 μAU (Fig. 3b). Table IV summarizes the results for flow-rates of 1 and 2 ml/min. It is worth noting that both sets of results were obtained based on the same reference area spectra acquired at 2 ml/min, which confirms the assumption that area spectra are inversely proportional to the flow-rate.

Zero resolution

The flow injection simulation led as expected to zero resolution. The plot analysis reveals the peak profiles of the different compounds. MCA analysis performed as well in these conditions (Fig. 5). Table V summarizes the results as flow-rates of 3 and 4 ml/min. In all cases, accuracy and precision are within the range expected from experimental errors. A quick estimate of the analysis cycle time achieved under different conditions (adding an isocratic baseline separation) shows that MCA can conservatively provide a 5-fold improvement in sample throughput (Table VI).

Poor baseline correction would have an adverse effect on this technique. A practical limitation in trying to speed up a separation is the interference between the void peak and the least retained peaks.

CONCLUSIONS

MCA can automate fast quantitation of unresolved chromatographic data, while providing peak identity confirmation and purity evaluation. The validity of the quantitative results is supported by error analysis features that are used to establish objective fit criteria. A fused group can be analyzed as long as it is baseline-resolved from other groups (or the void peak) and composed of 2 to 6 known components which have different enouhg spectra. When working with partial separations, a 5-fold increase in sample throughput over complete separations required by common quantitation methods can be achieved. The application to flow injection analysis (FIA) must be limited to mixtures of 2 to 6 compounds. More complex samples need some degree of separation.

REFERENCES

1 J. M. Davis and J. C. Giddings, *Anal. Chem.*, 55 (1983) 418.
2 G. Guiochon, D. P. Herman and M. F. Gonnord, *Anal. Chem.*, 56 (1984) 995–1003.
3 W. H. Lawton and E. A. Sylvester, *Technometrics*, 13 (1971) 617–633.
4 P. J. Gemperline, *Anal. Chem.*, 58 (1986) 2656–2663.
5 D. W. Osten and B. R. Kowalski, *Anal. Chem.*, 56 (1984) 991–995.
6 A. J. Blackburn, *Anal. Chem.*, 37 (1965) 1000–1003.
7 A. F. Fell, H. P. Scott, R. Gill and A. C. Moffat, *J. Chromatogr.*, 282 (1983) 123–140.
8 E. R. Malinovski and D. G. Howery, *Factor Analysis in Chemistry*, Wiley, New York, 1980.
9 E. R. Malinovski, *Anal. Chem.*, 49 (1977) 606–612.
10 E. R. Malinovski, *Anal. Chem.*, 49 (1977) 612–617.
11 H. R. Keller, D. L. Massart, P. Kiechle and F. Erni, *Anal. Chim. Acta*, 256 (1992) 125–131.
12 E. V. Dose and G. Guiochon, *Anal. Chem.*, 61 (1989) 2571–2579.