Anal. Chem. **1998,** *70,* 4841−4846

Unambiguous Characterization of a Single Test Sample by Fluorescence Spectroscopy and Solvent Extraction without Use of Standards

Jan Nygren, Abdalla Elbergali, and Mikael Kubista*

Department of Biochemistry and Biophysics, Chalmers University of Technology, 413 90 Göteborg, Sweden

We show that a test sample containing several components can be analyzed unambiguously by a two-dimensional spectroscopic technique, such as fluorescence, if it can be split into two aliquots of unequal component concentrations. Standard spectra are not needed, nor is it necessary to make assumptions about regions without spectral overlap or about spectral shapes. The requirements are that the spectral response is linear and that the two-dimensional contribution from each component can be factorized into a product of its one-dimensional responses: $I(\lambda_1,\lambda_2) = \sum_{i=1}^{r} c_i I_i(\lambda_1) I_i(\lambda_2)$. In fluorescence spectroscopy this is the normal situation for noninteracting species. The approach is illustrated on a test sample containing three components, which is split into aliquots by solvent extraction. Even though the components' spectral responses overlap extensively, all spectral features were correctly reproduced in the analysis.

A common task in analytical chemistry is to identify the components in test samples and determine their concentrations. If the components have distinct responses this may be done spectroscopically. But if their spectra overlap, their contributions cannot be distinguished in a single measurement. Using self-modeling curve resolution (SMCR),[1] Lawton and Sylvestre showed that the components' spectral profiles and concentrations could be estimated within a range by assuming that all responses are nonnegative. This range is usually rather large, and additional constraints are needed to obtain an acceptable solution. For example, one may apply physical constraints[2,3] or assume particular spectral shapes[4,5] or maximum dissimilarity of all spectra.[6] Methods not based on SMCR have also been developed. Some use physical constraints and require that the components are in chemical equilibrium,[7,8] while other make assumptions about spectral shapes[9] or concentration profiles.[10] Still another approach is based on the Procrustes rotation,[11] which mathematically is equivalent to the general rank annihilation method (GRAM).[12] These methods require two sets of spectra, to which the components shall contribute with the same intensity distributions but with different relative magnitudes. These could be fluorescence,[13] liquid chromatography diode array-UV (LC-DA-UV),[14] or NMR[15,16] spectra. Fluorescence spectra may either be emission spectra recorded at two excitation wavelengths or excitation spectra recorded at two emission wavelengths. Alternatively, a pair of samples can be analyzed by an appropriate two-dimensional (2-D) spectroscopic technique, which, as demonstrated here, opens the possibility to unambiguously characterize a single test sample. The test sample is split into aliquots by a standard extraction procedure, where after the aliquots are characterized by fluorescence excitation/emission spectroscopy. This generates two sets of 2-D spectra that can be analyzed without a priori assumptions by the Procrustes rotation.

MATERIALS AND METHODS

Fluorescence Measurements. *n*-Heptane was purchased from Merck, and POPOP, anthracene, and diphenylanthracene (DPA) were from Sigma. The dyes were used without purification. In all samples, the components' concentrations were sufficiently low to make energy transfer and quenching negligible. Sample absorption was also low, making the correction for the inner-filter effect negligible.[17] Fluorescence spectra were measured on a Spex Fluorolog τ2 spectrofluorometer and were digitized with 5 data points/nm. Emission spectra were recorded between 385 and 500 nm (0.2-nm intervals) at excitation wavelengths between 300 and 380 nm (2-nm intervals), and excitation spectra were recorded between 300 and 380 nm (0.2-nm intervals) at emission wavelengths between 385 and 500 nm (2-nm intervals).

(1) Lawton, W. H.; Sylvestre, E. A. *Technometrics* **1971**, *13*, 617−633.
(2) Sun, Y.-P.; Sears, D. F., Jr.; Saltiel, J. *J. Am. Chem. Soc.* **1989**, *111*, 706−711.
(3) Saltiel, J.; Sears, D. F., Jr.; Choi, J.-O.; Sun, Y.-P.; Eaker, D. W. *J. Phys. Chem.* **1994**, *98*, 35−46.
(4) Sun, Y.-P.; Bennet, G.; Johnston, K. P.; Fox, M. A. *Anal. Chem.* **1992**, *64*, 1763−1768.
(5) Steinbock, O.; Neuman, B.; Cage, B.; Saltiel, J.; Müller, S. C.; Dalal, N. S. *Anal. Chem.* **1997**, *69*, 3708−3713.
(6) Meister, A. *Anal. Chim. Acta* **1985**, *57*, 971−985.
(7) Kubista, M.; Sjöback, R.; Nygren, J. *Anal. Chim. Acta* **1995**, *302*, 121−125.
(8) Nygren, J.; Andrade, M. A.; Kubista, M. *Anal. Chem.* **1996**, *68*, 1706−1710.
(9) Neal, S. L.; Davidson, E. R.; Warner, I. M. *Anal. Chem.* **1990**, *62*, 658−664.
(10) Meader, M.; Zuberbuhler, A. D. *Anal. Chim. Acta* **1986**, *181*, 287−291.
(11) Kubista, M. *Chemom. Intell. Lab. Syst.* **1990**, *7*, 273−279.
(12) Sanches, E.; Kowalski, B. R. *Anal. Chem.* **1986**, *58*, 496−499.
(13) Scarmino, I.; Kubista, M. *Anal. Chem.* **1993**, *65*, 409−416. Information about the DATAN program is available on the Internet: http://www.bcbp.chalmers.se/mbg/chemo/chemo.html.
(14) Sanches, E.; Ramos, L. S.; Kowalski, B. R. *J. Chromatogr.* **1987**, *385*, 151−164.
(15) Schulze, D.; Stilbs, P. *J. Magn. Reson.* **1993**, *105*, 54−58.
(16) Antalek, B.; Windig, W. *J. Am. Chem. Soc.* **1996**, *118*, 10331−10332.
(17) Kubista, M.; Sjöback, R.; Eriksson, S.; Albinsson, B. *Analyst* **1994**, *119*, 417−419.

10.1021/ac980381n CCC: $15.00
Published on Web 10/22/1998

Chemometric Analysis. Sample pairs were analyzed by measuring either emission spectra at different excitation wavelengths or excitation spectra at different emission wavelengths. The two analyses are independent but equivalent. The results differ in which spectral region is determined as a continuous spectrum and which is calculated at discrete intensities. The region scanned in the experiments is the one for which continuous spectra are determined.

The spectra recorded on the first sample were arranged as rows in a matrix $\mathbf{I^A}$, and spectra recorded on the second sample were arranged as rows in a matrix $\mathbf{I^B}$. This gave the equations

$$\mathbf{I^A} = \mathbf{X}\mathbf{C^A}\mathbf{M}$$

$$\mathbf{I^B} = \mathbf{X}\mathbf{C^B}\mathbf{M} \qquad (1)$$

where $\mathbf{X}$ and $\mathbf{M}$ are normalized excitation and emission profiles of the components, and $\mathbf{C^A}$ and $\mathbf{C^B}$ are diagonal matrices containing components' concentrations in the two samples. Renormalizing $\mathbf{X}$ the equations can be written

$$\mathbf{I^A} = \mathbf{X^A}\mathbf{M}$$

$$\mathbf{I^B} = \mathbf{X^B}\mathbf{M} = \mathbf{X^A}\mathbf{D}\mathbf{M} \qquad (2)$$

$\mathbf{D}$ is a diagonal matrix containing the ratios between the components' concentrations, $c_i^{\mathrm{B}}/c_i^{\mathrm{A}}$, in the two samples. The $\mathbf{X^A}$, $\mathbf{D}$, and $\mathbf{M}$ matrices are not known beforehand but are calculated from the experimental spectra. $\mathbf{I^A}$ and $\mathbf{I^B}$ are laminated and decomposed into principal components using, for example, the NIPALS routine:[18]

$$\begin{bmatrix}\mathbf{I}^{\mathrm{A}}\\ \mathbf{I}^{\mathrm{B}}\end{bmatrix} = \mathbf{T}\mathbf{P}' = \begin{bmatrix}\mathbf{T}^{\mathrm{A}}\\ \mathbf{T}^{\mathrm{B}}\end{bmatrix}\mathbf{P}' \qquad (3)$$

The r most significant principal components, where r is the number of spectroscopically distinguishable components in the samples, are sufficient to adequately describe the $\mathbf{I^A}$ and $\mathbf{I^B}$ matrices. The other $n-r$ principal components contain noise and are discarded. The upper half of the $\mathbf{T}$ matrix, $\mathbf{T^A}$, corresponds to $\mathbf{I^A}$, and the lower part, $\mathbf{T^B}$, corresponds to $\mathbf{I^B}$.

The principal components and the spectra of the sample components are related by a rotation.

$$\mathbf{X^A} = \mathbf{T^A}\mathbf{R}^{-1}$$

$$\mathbf{X^B} = \mathbf{T^B}\mathbf{R}^{-1}$$

$$\mathbf{M} = \mathbf{R}\mathbf{P}' \qquad (4)$$

where $\mathbf{R}$ is a square $r \times r$ rotation matrix. The $\mathbf{T^A}$ and $\mathbf{T^B}$ matrices are related through multiplication by a matrix $\mathbf{Q}$:

$$\mathbf{T^B} = \mathbf{T^A}\mathbf{Q} \qquad (5)$$

which is calculated by the least-squares method:

$$\mathbf{Q} = ((\mathbf{T^A})'\mathbf{T^A})^{-1}(\mathbf{T^A})'\mathbf{T^B} \qquad (6)$$

$\mathbf{Q}$ is called the "Procrustes rotation" of $\mathbf{T^A}$ relative to $\mathbf{T^B}$, after the highwayman Procrustes in the Greek tale who provided travelers with lodging and during the night either cut their legs or extended them to make them fit exactly into his bed. Matrices $\mathbf{D}$ and $\mathbf{R}$ are the eigenvalues and eigenvectors, respectively, of $\mathbf{Q}$. They are calculated by solving the eigenvalue equation:

$$\mathbf{Q}'\mathbf{R}' = \mathbf{R}'\mathbf{D} \qquad (7)$$

Matrix $\mathbf{R}$ is finally normalized.[13]

Assessing the Goodness of Fit. Calculated spectra were compared to those recorded separately of the pure components by calculating a correlation coefficient:

$$\mathrm{Corrcoef} = \sum_{j=1}^{j=m} I_{\mathrm{calc}}(\lambda_j)\cdot I_{\mathrm{true}}(\lambda_j) \Big/ \left(\left(\sum_{j=1}^{j=m} I_{\mathrm{calc}}(\lambda_j)\cdot I_{\mathrm{calc}}(\lambda_j)\right)\left(\sum_{j=1}^{j=m} I_{\mathrm{true}}(\lambda_j)\cdot I_{\mathrm{true}}(\lambda_j)\right)\right)^{1/2} \qquad (8)$$

where I_{calc} and I_{true} are mean-centered vectors containing the calculated and the pure spectra, respectively. For identical spectra Corrcoef equals 1. This is not even expected for a perfect match owing to experimental noise.

RESULTS

The approach is demonstrated on a sample containing the dyes POPOP, anthracene, and DPA at about the same concentrations in *n*-heptane. The *n*-heptane was saturated with 85% ethanol in water to minimize spectral shifts caused by subsequent extractions; 85% ethanol in water was used instead of pure ethanol because the latter is miscible with *n*-heptane. The sample was characterized by two independent fluorescence measurements. Fluorescence emission spectra were recorded at 41 excitation wavelengths (Figure 1, top left), and excitation spectra were recorded at 58 emission wavelengths (Figure 2, top left). A 7-mL aliquot of the sample was then extracted with 5 mL of 85% ethanol in water, and new sets of emission (Figure 1, top right) and excitation (Figure 2, top right) spectra were recorded.

This yielded a total of four data sets, two sets of excitation spectra and two sets of emission spectra, which were analyzed separately. The two spectra of the same kind were arranged in two matrices, $\mathbf{I^A}$ and $\mathbf{I^B}$, and processed by the DATAN program.[13] First the number of components was determined by statistical tests,[19] and then the components' spectral responses were calculated by Procrustes rotation analysis (eqs 3−7). The bottom right panel in Figure 1 compares the calculated components' emission profiles ($\mathbf{M}$) with those recorded separately on the pure compounds in *n*-heptane saturated with 85% ethanol. The correspondence, as assessed by the calculated correlation coefficients (Table 1), is excellent. Many of the coefficients are of the same magnitude as a typical correlation coefficient (0.9997) between spectra measured on two samples containing the same pure

(18) Wold, H. *Research Papers in Statistics*; Wiley: New York, 1966; pp 411−444.

(19) Elbergali, A.; Nygren, J.; Kubista, M. *Anal. Chim. Acta*, in press.

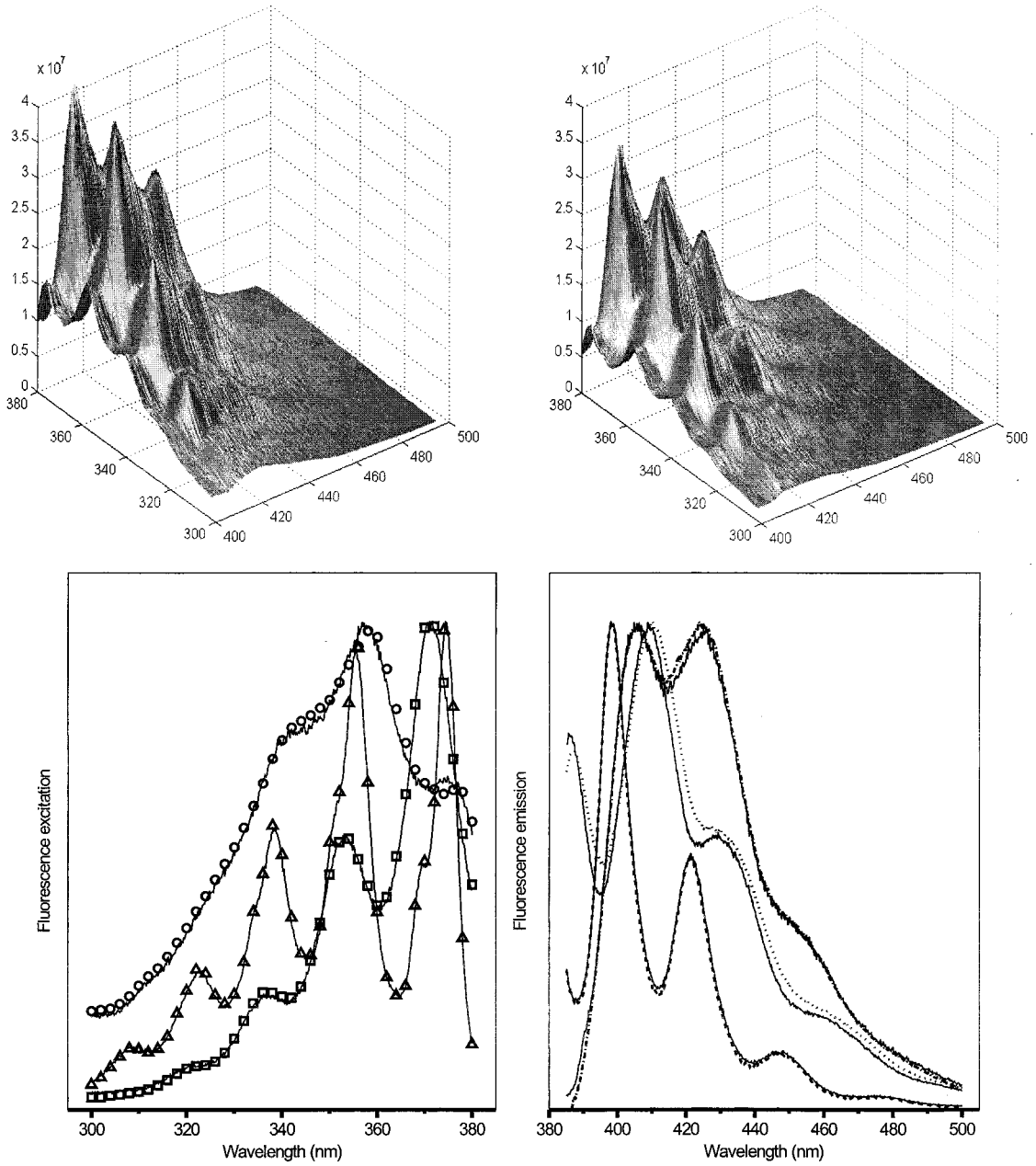

Figure 1. Top: fluorescence emission spectra recorded on the test sample before (left) and after (right) extraction with ethanol. Bottom: (left) calculated excitation intensities of anthracene (△), DPA (□), and POPOP (○), and excitation spectra of the pure components recorded separately (lines); (right) calculated emission spectra of anthracene (- - -), DPA (— · —) and POPOP (···) compared to emission spectra of the pure compounds (lines).

compound. All significant spectral features are reproduced in the calculated spectra, from which the compounds can readily be identified. The signal-to-noise ratio in the calculated spectra is significantly higher than in the measured spectra. This is a valuable feature of the analysis procedure and is a result of the noise reduction achieved when a large number of experimental spectra are reproduced by a small number of significant principal components. It is equivalent to the signal-to-noise improvement achieved when averaging repeated scans and shall not be compared with model-dependent noise-reducing procedures such as smoothing.

The left bottom panel in Figure 1 shows calculated excitation intensities (**C**) of the components. The excitation spectra of the pure compounds measured separately are shown for comparison. Here also the correspondence is excellent. The ratios between the concentrations of the components in the two samples are calculated by the Procrustes rotation analysis as the *d*-values ($d_i = c_i^A / c_i^B$). These were 0.47, 0.83, and 0.98 for POPOP, DPA, and anthracene, respectively, reflecting that 53%, 17%, and 2% of the three dyes had been removed by the ethanol extraction. Figure 2 shows the independent analysis of the excitation data sets that were recorded on the same two samples. Here also the determined excitation spectra and emission intensities are in excellent agreement with those measured separately on the pure compounds (Table 1). The concentration ratios calculated by this analysis were 0.48, 0.83, and 0.98, which are essentially identical to those determined by the analysis of the emission data sets above.

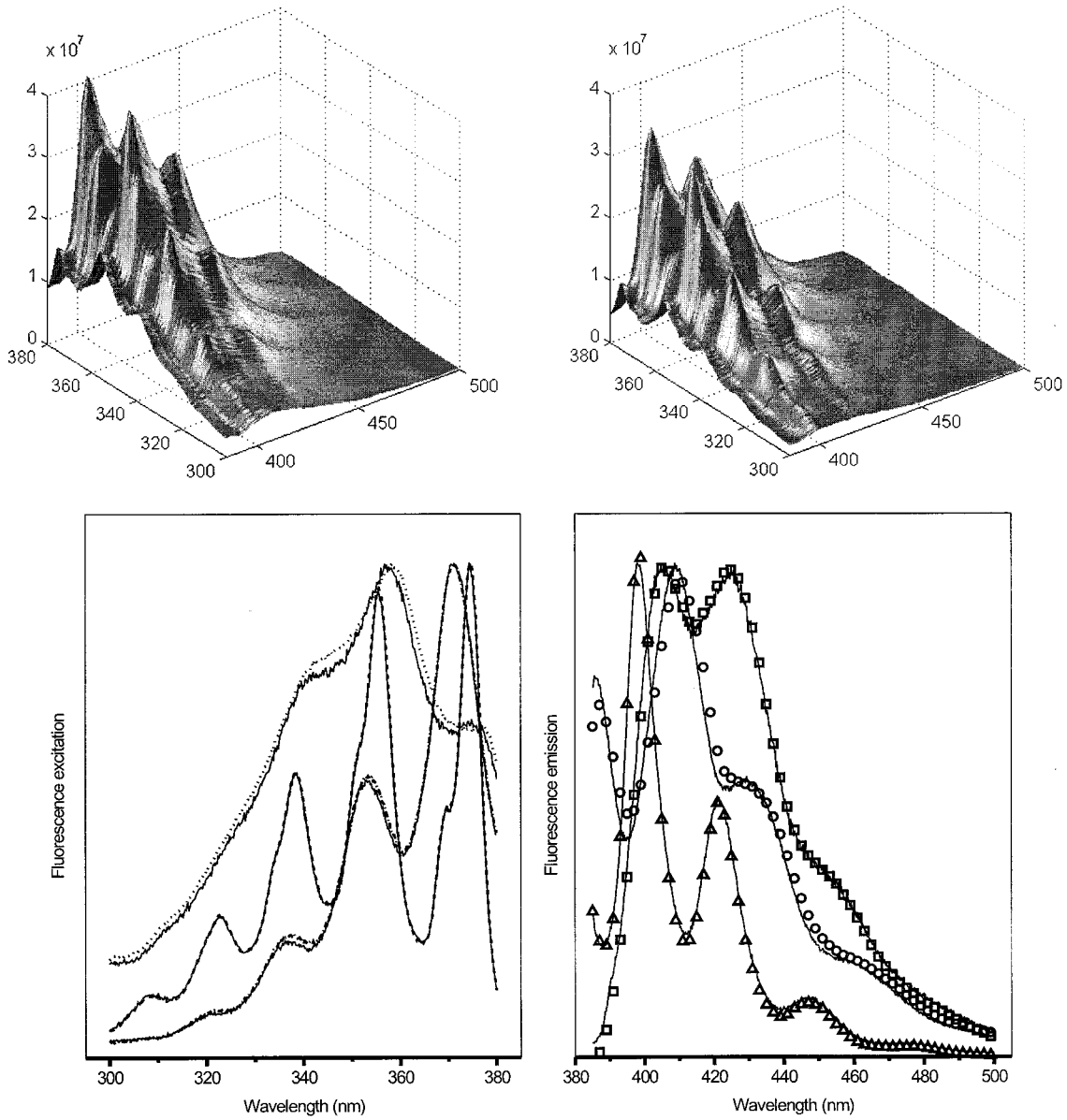

Figure 2. Top: fluorescence excitation spectra recorded on the test sample before (left) and after (right) extraction with ethanol. Bottom: (left) calculated excitation spectra of anthracene (- - -), DPA (— · —), and POPOP (···) compared to excitation spectra of the pure compounds (lines); (right) calculated emission intensities of anthracene (△), DPA (□), and POPOP (○) compared to the emission spectra of the pure compounds (lines).

The extraction procedure was repeated once by adding 18 mL of 85% ethanol in water to the sample and again recording sets of excitation and emission spectra. Both analyses gave the *d*-values 0.18, 0.51, and 0.88 for POPOP, DPA, and anthracene, reflecting that 82%, 49%, and 12% of the compounds had been removed this time. Again the calculated spectra were in excellent agreement with those recorded of the pure compounds (Table 1). The correlation coefficients for the POPOP spectra and the emission spectrum of DPA were somewhat lower than in the previous analysis, while for the anthracene spectra and the excitation spectrum of DPA, they were slightly higher.

The importance of the data set size was tested by analyzing subsets of the original data. One analysis was made using only every fourth spectrum. This reduced the average correlation coefficient for the spectral profiles (**V**) by 0.03% and that for the calculated intensities (**C**) by 0.06% (Table 1). In a second analysis, only every fifth data point was used. This unexpectedly, and most likely accidentally, increased the average correlation coefficient for **V** by 0.007% while that for **C** decreased by 0.01% (Table 1). Finally, an analysis was performed where both the number of spectra (4-fold) and the number of data points per spectrum (5-fold) were reduced. This decreased the average correlation coefficients by 0.02% for **V** and by 0.06% for **C** (Table 1).

DISCUSSION

Previously, Procrustes rotation has been used to analyze series of samples by two spectroscopic measurements.[11,13] Here we demonstrate that a pair of samples, generated from a single sample, can be analyzed by a series of spectroscopic measurements. The procedure requires that two samples have the components at different relative concentrations. Further, the spectroscopic signal shall be proportional to the components'

Table 1. Correlation Coefficients

	all data		reduced no. of spectra		reduced no. of points		reduced no. of spectra and points	
	first extraction	second extraction	first extraction	second extraction	first extraction	second extraction	first extraction	second extraction
				Correlation of Calculated **V** Matrices				
emission								
POPOP	0.995 45	0.994 91	0.995 33	0.994 70	0.995 53	0.994 91	0.995 47	0.994 83
DPA	0.999 46	0.999 02	0.998 96	0.998 73	0.999 50	0.999 23	0.999 42	0.999 29
anthracene	0.999 73	0.999 66	0.998 98	0.999 31	0.999 88	0.999 71	0.999 67	0.999 28
excitation								
POPOP	0.998 39	0.998 03	0.998 52	0.997 74	0.998 49	0.998 07	0.997 78	0.997 48
DPA	0.999 71	0.999 81	0.999 02	0.999 78	0.999 75	0.999 76	0.999 47	0.999 70
anthracene	0.999 77	0.999 75	0.999 56	0.999 58	0.999 80	0.999 84	0.999 69	0.999 66
av value	0.999 64		0.998 35		0.998 71		0.998 48	
				Correlation of Calculated **C** Matrices				
emission								
POPOP	0.995 01	0.994 64	0.992 35	0.991 32	0.994 79	0.994 44	0.992 34	0.991 08
DPA	0.999 33	0.999 23	0.999 80	0.999 51	0.999 49	0.999 06	0.999 57	0.999 38
anthracene	0.999 60	0.999 83	0.998 67	0.999 84	0.999 67	0.999 74	0.999 85	0.999 87
excitation								
POPOP	0.998 16	0.998 19	0.997 65	0.997 58	0.997 84	0.998 26	0.997 23	0.997 71
DPA	0.999 81	0.999 84	0.999 78	0.999 95	0.999 59	0.999 78	0.999 92	0.999 93
anthracene	0.999 88	0.999 88	0.999 86	0.999 85	0.999 42	0.999 86	0.999 70	0.999 86
av value	0.998 62		0.998 01		0.998 50		0.998 04	

concentrations, and the 2-D response from each component shall be a product of its one-dimensional responses.

$$I^{A}(\lambda_1, \lambda_2) = \sum_{i=1}^{r} c_i^{A} I_i(\lambda_1) \cdot I_i(\lambda_2)$$

$$I^{B}(\lambda_1, \lambda_2) = \sum_{i=1}^{r} c_i^{B} I_i(\lambda_1) \cdot I_i(\lambda_2)$$

$$c_i^{B}/c_i^{A} \neq c_j^{B}/c_j^{A} \quad (i \neq j; \quad i, j = 1...r) \tag{9}$$

This is the usual case in, for example, fluorescence. The signal is recorded as a function of excitation and emission wavelengths, and the contribution from each component is proportional to its contribution and is a product of its normalized excitation and emission spectra. Note that the 2-D spectra cannot be analyzed separately owing to rotation ambiguity. It is only when they are analyzed as a pair that the ambiguity is resolved.

A test sample can be split into two aliquots of sufficiently different component concentrations by various means. Here we use standard solvent extraction. Other possibilities are distillation, adsorption, precipitation, chromatographic separation, filtration, etc. As discussed elsewhere, Procrustes analysis is very sensitive to spectral shapes and even small shifts between the two samples may have a deleterious effect.[13] Solvent extraction may shift the components' spectral responses owing to some admixture of the two solvents. This can be minimized by saturating the sample with the solvent before the extraction, which can be done by mixing the sample with a small amount of solvent to achieve saturation with minimum loss of sample components. The extraction is then made with a much larger volume of solvent. As seen in our example, such precaution minimizes the spectral shifts to such a degree that they do not interfere with the analysis.

When fluorescence is used, either the excitation or the emission wavelength can be scanned, while the other is varied discretely. Scanning the excitation wavelength has the advantage that excitation spectra usually are identical to absorption spectra, which may simplify the identification of the components. Determining separately the excitation and emission spectra has the advantage that the results can be compared in a self-consistent check. In our example, the determined emission spectra in Figure 1 (bottom right) and the calculated emission intensities in Figure 2 (bottom right) from the two independent analyzes are in excellent agreement. This is also the case for the excitation spectra in Figure 2 (bottom left) and the excitation intensities in Figure 1 (bottom left). These correspondences are strong indications of successful analysis.

In our analyses of the test sample, we collect data points every 0.2 nm in the scanned spectra and every 2 nm in the discretely varied region. This gives $576 \times 41 = 23\,616$ data points for the emission scan and $401 \times 58 = 23\,258$ data points for the excitation scan. These large data sets are evidently sufficient to determine the components' spectral responses with excellent precision despite the extensive spectral overlap between the components' responses. Decreasing the number of spectra collected (n) or the number of data points collected per spectrum (m) has little effect on the accuracy of the calculated spectra. Not even decreasing both n and m is very detrimental to the analysis. It seems that decreasing n is somewhat more deleterious than decreasing m, and the effect on **C** seems somewhat larger than the effect on **V**, which may be due to $m \gg n$. The differences are, however, minor and probably not statistically valid. Previously it has been shown that changing n mainly effects the accuracy in **V** and changing m effects mainly **C**.[13] The precision in the relative concentrations (**D**) could not be assessed in our analysis, since the concentrations were not determined independently, but it is expected to depend on both n and m.[20]

(20) Booksh, K.; Kowalski, B. R. *J. Chemom.* **1994**, *8*, 45−63.

CONCLUSIONS

We show that a single test sample can be split into two aliquots of sufficiently different components' concentration to allow their spectral responses to be determined unambiguously by an appropriate 2-D spectroscopic technique such as fluorescence. The analysis can be done by the DATa ANalysis (DATAN) program, which is available on the Internet: http://www.bcbp.chalmers.se/mbg/chemo/chemo.html.

Received for review April 3, 1998. Accepted August 26, 1998.

AC980381N