ELSEVIER

Analytica Chimica Acta 302 (1995) 121–125

ANALYTICA CHIMICA ACTA

Quantitative spectral analysis of multicomponent equilibria

Mikael Kubista *, Robert Sjöback, Jan Nygren

Molecular Biotechnology Group, Department of Biochemistry and Biophysics, Chalmers University of Technology, S-413 90 Gothenburg, Sweden

Received 17 June 1994; revised manuscript received 30 August 1994

Abstract

A generalization is presented of a method for the determination of equilibrium constants for binary mixtures. It is shown that titrations involving several components can be quantitatively characterized in terms of component concentrations and spectral responses, even when none of the component spectra is known and the components cannot be obtained in pure form. The method is efficient even if the component spectra overlap extensively. It is applicable to all spectroscopic techniques that provide a linear response, opening up new possibilities to analyze complex chemical systems. The approach is illustrated by determining the absorption spectra and protolytic constants for the fluorescein protolytic equilibria.

Keywords: Multicomponent equilibria

1. Introduction

Spectroscopic methods are frequently used to analyze chemical equilibria. If the components involved can be obtained in pure form, or if their spectral responses do not overlap, such analysis is, in general, trivial. For many systems, particularly those with similar components, this is not the case, and these have been difficult to analyze. In 1971 Lawton and Sylvestre [1] introduced chemometric methods for spectral analysis. In their classical paper they showed that even two-component titrations cannot be analyzed unambiguously, and they provided means to obtain limiting solutions by assuming non-negative concentrations and signal responses. These criteria are difficult to apply to real data owing to noise, which can always be negative, and approaches to mitigate these criteria have been developed [2]. Attempts have also been made to extend their treatment to more components [3], but so far only with limited success. Still common to these methods is that they do not provide a unique solution based on physical criteria, but only limit the number of possible solutions.

Recently we developed a new method that provides a unique solution by requiring that the calculated concentrations obey an assumed equilibrium expression, and applied it successfully for two component equilibria [4]. In this work we extend this treatment to multicomponent mixtures and demonstrate its applicability by determining the absorption spectra of the four protolytic forms of fluorescein and the corresponding protolytic constants.

2. Theory

Spectra of the titration samples are digitized and arranged in a data matrix **A**, which is decomposed into

* Corresponding author.

SSDI 0003-2670(94)00433-1

an orthonormal basis set (Fig. 1) by Nipals or any equivalent method [5]:

$$\mathbf{A} = \mathbf{TP}' + \mathbf{E} \approx \mathbf{TP}' = \sum_{i=1}^{r} \mathbf{t}_i \mathbf{p}_i' \quad (1)$$

where the orthogonal target vectors $\mathbf{t}_i$ and orthonormal projection vectors $\mathbf{p}_i'$ are mathematical constructs that cannot be directly related to component spectra and concentrations. r is the number of independent spectroscopic components, which corresponds to the number of light-absorbing chemical species. It is determined by visual inspection of the $\mathbf{t}$ and $\mathbf{p}'$ vectors or by performing a chi^2-test [6]. $\mathbf{E}$ is an error matrix.

Assuming linear response the spectra in matrix $\mathbf{A}$ are linear combinations of the concentrations, $\mathbf{C}$, and spectral responses, $\mathbf{V}$, of the chemical components.

$$\mathbf{A} = \mathbf{CV} + \mathbf{E}' \approx \mathbf{CV} = \sum_{i=1}^{r} \mathbf{c}_i \mathbf{v}_i \quad (2)$$

From Eqs. 1 and 2 follows that there is a square matrix $\mathbf{R}$ ($r \times r$) that satisfies

$$\mathbf{T} = \mathbf{CR} \quad (3a)$$

$$\mathbf{P}' = \mathbf{R}^{-1}\mathbf{V} \quad (3b)$$

since $\mathbf{A} = \mathbf{CV} = \mathbf{C}(\mathbf{RR}^{-1})\mathbf{V} = (\mathbf{CR})(\mathbf{R}^{-1}\mathbf{V}) = \mathbf{TP}'$. If $\mathbf{R}$ can be determined the spectral responses ($\mathbf{V}$) and concentrations ($\mathbf{C}$) of the components can be calculated from the target ($\mathbf{T}$) and projection ($\mathbf{P}'$) matrices:

$$\mathbf{C} = \mathbf{TR}^{-1} \quad (4a)$$

$$\mathbf{V} = \mathbf{RP}' \quad (4b)$$

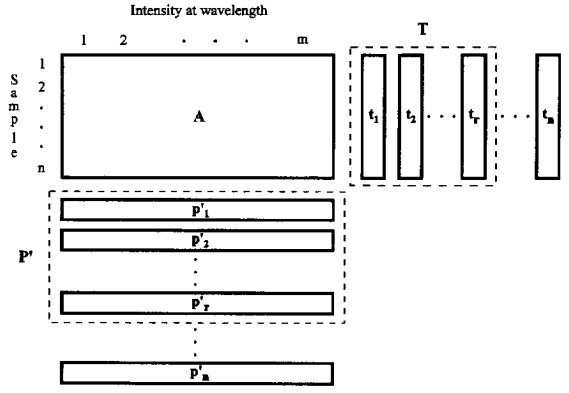

Fig. 1. The digitized spectra are arranged as rows in a data matrix $\mathbf{A}$, which is decomposed into a sum of products of orthonormal projection vectors ($\mathbf{p}'$) and orthogonal target vectors ($\mathbf{t}$).

Matrix $\mathbf{R}$ is determined as follows. Concentrations of the chemical species are calculated from the equilibrium expressions for various trial values of the equilibrium constants, and are fitted to the calculated target vectors according to Eq. 3a. The accuracy of this fit depends crucially on the trial values of the equilibrium constants, and best fit determines their values and the elements of matrix $\mathbf{R}$.

3. Experimental

Fluorescein was purchased from Sigma. It was found to contain no significant amount of contaminants and was used without further purification. Its concentration was determined spectroscopically, assuming a molar absorptivity of the dianion in 0.1 M NaOH of 76,900 $M^{-1}\,cm^{-1}$ [7] (unpublished results). Samples containing 14 μM fluorescein in 50 mM buffer at various pH (phosphate at pH $\geq$ 5 and citrate at pH $<$ 5) were prepared and absorption spectra were recorded on a CARY 2300 spectrometer. The optical pathlength was 1 cm and spectral data were collected at a resolution of five data points per nanometer.

4. Results and discussion

Fluorescein in aqueous solution is present as cation, FH_3^+, neutral species, FH_2, anion, FH^-, and dianion, F^{2-}. Their equilibria have been studied by absorption spectrometry previously [8,9], and the problem is acknowledged as very complex owing to extensive spectral overlap of the protolytic forms and the small difference between the successive protolytic constants. Other techniques, such as NMR spectrometry and fluorimetry, are not applicable owing to the tendency of fluorescein to aggregate at high concentrations [10] and to protolytic reactions in excited state (unpublished results).

Fig. 2 (top left) shows 17 absorption spectra of fluorescein recorded in the pH range 2–9.5. From inspection of the experimental spectra it is hard to guess even the number of protolytic species involved. The calculated four most significant projection vectors, $\mathbf{p}'$, are shown in Fig. 2 (bottom left). They all have clear spectral features, as compared to noise, evidencing the presence of four spectroscopically distinguishable

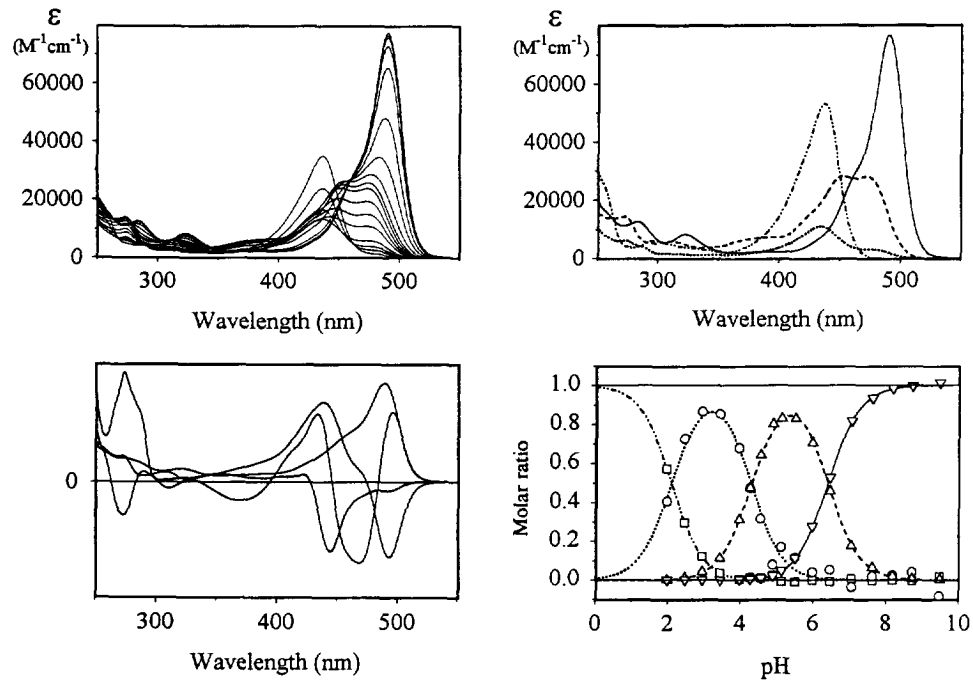

Fig. 2. *Top left*: absorption spectra (presented as molar absorptivities) of fluorescein recorded at pH 1.97, 2.47, 2.94, 3.43, 3.98, 4.27, 4.56, 4.89, 5.13, 5.51, 5.99, 6.47, 7.06, 7.62, 8.18, 8.72, 9.47. *Bottom left*: the four most significant projection vectors (significance decreases with increasing number sign-changes). Their singular values, which are the lengths of the corresponding target vectors, are 9.2×10^6, 3.6×10^6, 0.9×10^6 and 0.2×10^6. *Top right*: Calculated spectral responses of the four protolytic forms (cation: –·–·–; neutral species: ····; anion: – – –; dianion: ———). *Bottom right*: calculated concentrations using matrix **R** (cation: □; neutral species: ○; anion: △; dianion: ▽) compared with concentrations predicted by the calculated protolytic constants.

components. Their shapes, however, are clearly unphysical and cannot be directly related to the spectral responses of the four protolytic forms.

The concentrations of the four protolytic forms are given by:

$$\mathbf{c}_{\mathbf{FH_3}^+} = \frac{\mathbf{c}_{\mathbf{H}}^3+}{\mathbf{c}_{\mathbf{H}}^3+ + K_1\mathbf{c}_{\mathbf{H}}^2+ + K_1K_2\mathbf{c}_{\mathbf{H}}+ + K_1K_2K_3}\mathbf{c}_{\mathbf{tot}}$$

$$\mathbf{c}_{\mathbf{FH_2}} = \frac{K_1\mathbf{c}_{\mathbf{H}}^2+}{\mathbf{c}_{\mathbf{H}}^3+ + K_1\mathbf{c}_{\mathbf{H}}^2+ + K_1K_2\mathbf{c}_{\mathbf{H}}+ + K_1K_2K_3}\mathbf{c}_{\mathbf{tot}}$$

$$\mathbf{c}_{\mathbf{FH}^-} = \frac{K_1K_2\mathbf{c}_{\mathbf{H}}+}{\mathbf{c}_{\mathbf{H}}^3+ + K_1\mathbf{c}_{\mathbf{H}}^2+ + K_1K_2\mathbf{c}_{\mathbf{H}}+ + K_1K_2K_3}\mathbf{c}_{\mathbf{tot}}$$

$$\mathbf{c}_{\mathbf{F}^{2-}} = \frac{K_1K_2K_3}{\mathbf{c}_{\mathbf{H}}^3+ + K_1\mathbf{c}_{\mathbf{H}}^2+ + K_1K_2\mathbf{c}_{\mathbf{H}}+ + K_1K_2K_3}\mathbf{c}_{\mathbf{tot}} \qquad (5)$$

where K_1, K_2 and K_3 are the protolytic constants, and $\mathbf{c}_{\mathbf{H}}+$, $\mathbf{c}_{\mathbf{FH_3}^+}$, $\mathbf{c}_{\mathbf{FH_2}}$, $\mathbf{c}_{\mathbf{FH}^-}$, $\mathbf{c}_{\mathbf{F}^{2-}}$ and $\mathbf{c}_{\mathbf{tot}}$ are vectors having the concentrations of protons, fluorescein cation, neutral species, anion, dianion, and total fluorescein as elements. The vectors $\mathbf{c}_{\mathbf{H}}^2+$ and $\mathbf{c}_{\mathbf{H}}^3+$ have the square and cube of the proton concentration as elements. Different sets of trial concentration vectors (identified by a hat) were calculated for various trials values of the protolytic constants, and fitted to the target vectors:

$$\mathbf{t}_1 = \hat{r}_{11}\hat{\mathbf{c}}_{\mathbf{FH_3}^+} + \hat{r}_{21}\hat{\mathbf{c}}_{\mathbf{FH_2}} + \hat{r}_{31}\hat{\mathbf{c}}_{\mathbf{FH}^-} + \hat{r}_{41}\hat{\mathbf{c}}_{\mathbf{F}^{2-}}$$

$$\mathbf{t}_2 = \hat{r}_{12}\hat{\mathbf{c}}_{\mathbf{FH_3}^+} + \hat{r}_{22}\hat{\mathbf{c}}_{\mathbf{FH_2}} + \hat{r}_{32}\hat{\mathbf{c}}_{\mathbf{FH}^-} + \hat{r}_{42}\hat{\mathbf{c}}_{\mathbf{F}^{2-}}$$

$$\mathbf{t}_3 = \hat{r}_{13}\hat{\mathbf{c}}_{\mathbf{FH_3}^+} + \hat{r}_{23}\hat{\mathbf{c}}_{\mathbf{FH_2}} + \hat{r}_{33}\hat{\mathbf{c}}_{\mathbf{FH}^-} + \hat{r}_{43}\hat{\mathbf{c}}_{\mathbf{F}^{2-}}$$

$$\mathbf{t}_4 = \hat{r}_{14}\hat{\mathbf{c}}_{\mathbf{FH_3}^+} + \hat{r}_{24}\hat{\mathbf{c}}_{\mathbf{FH_2}} + \hat{r}_{34}\hat{\mathbf{c}}_{\mathbf{FH}^-} + \hat{r}_{44}\hat{\mathbf{c}}_{\mathbf{F}^{2-}} \qquad (6)$$

Each fit defines the elements of a trial matrix $\hat{\mathbf{R}}$. The value of the protolytic constants were varied to minimize the sum of square residuals:

$$\chi^2 = \sum_{i=1}^{4}\sum_{j=1}^{17}\left(T_{ij} - \sum_{k=1}^{4}\hat{r}_{ik}\hat{c}_{kj}\right)^2 \qquad (7)$$

This was performed by grid search and a single local minimum with $\chi^2 = 0.018$ was found for

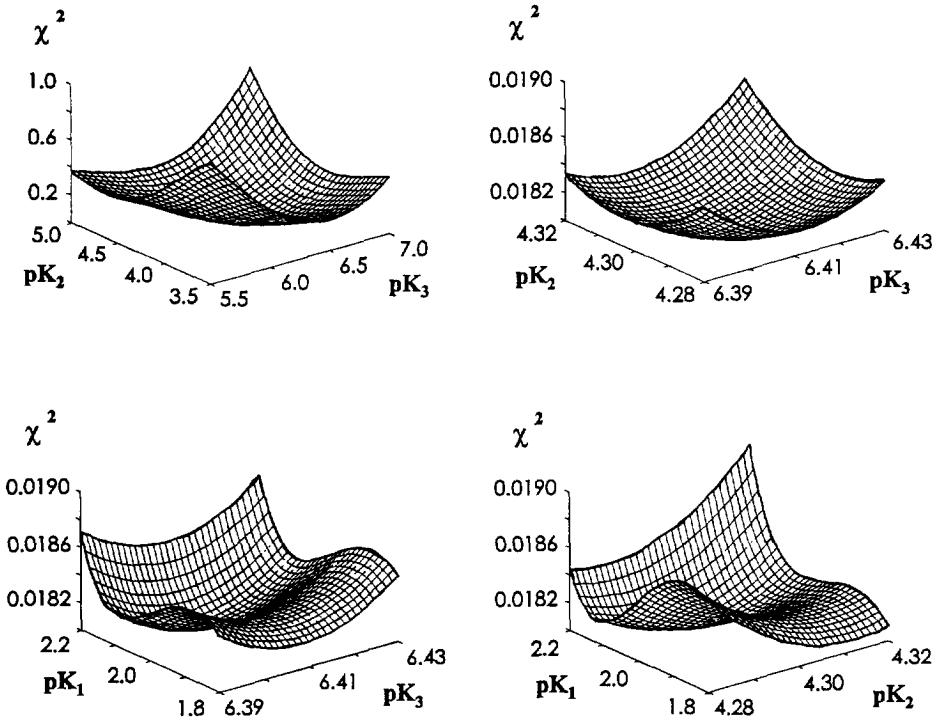

Fig. 3. Subsections of the χ^2 surface. *Top left*: dependence of χ^2 on pK_2 and pK_3 for $pK_1 = 2.1$. *Top right*: enlargement of the χ^2 dependence on pK_2 and pK_3 for $pK_1 = 2.1$. *Bottom left*: χ^2 dependence on pK_1 and pK_3 for $pK_2 = 4.30$. *Bottom right*: χ^2 dependence on pK_1 and pK_2 for $pK_3 = 6.41$. In the pK_2, pK_3 subsurface a single global minimum is found. In the pK_1, pK_3 and pK_1, pK_2 subsurfaces a single local minimum is observed. Lower χ^2 values are obtained for $pK_1 < 1.75$, but this region of the χ^2 surface can be ruled out since it predicts negative spectral features for the cationic species.

$K_1 = 7.9 \times 10^{-3}$, $K_2 = 5.0 \times 10^{-5}$, $K_3 = 3.9 \times 10^{-7}$. This corresponds to $pK_1 = 2.1 \pm 0.1$, $pK_2 = 4.30 \pm 0.02$ and $pK_3 = 6.41 \pm 0.02$, where the uncertainties are estimated from inspection of the subsections of the χ^2 surface shown in Fig. 3. Narrow minima are observed for pK_2 and pK_3, whereas for pK_1 the minimum is more shallow. This implies a higher accuracy in the determination of pK_2 and pK_3 than of pK_1, which is a consequence of the latter falling on the edge of the pH range investigated. In fact the χ^2 minimum is only local. Lower χ^2 is obtained with $pK_1 < 1.75$ (Fig. 3 bottom). These values of pK_1, however, can be excluded since they predict a spectrum of the cationic species with negative features thus corresponding to unphysical solutions. The accuracy of the analysis was checked by systematically removing one or two spectra from the data set. Analysis of these subsets gave in all cases protolytic constants which were within the specified uncertainty limits. The values of the protolytic constants we have determined are significantly different from those determined previously by single-point analysis of absorption data ($pK_1 = 1.95$, $pK_2 = 5.05$ and $pK_3 = 7.00$ [8]; $pK_1 = 2.2$, $pK_2 = 4.4$ and $pK_3 = 6.7$ [9]), but agrees with the previous determination of pK_3 by chemometric methods ($pK_3 = 6.44$ [4]). This most certainly reflects the much higher accuracy in the present method of analysis which utilizes all spectral information.

The **R** matrix determined for the optimum set of protolytic constants was used to calculate the spectra of the protolytic forms (Fig. 2, top right). These have a higher signal-to-noise ratio than the recorded spectra owing to the noise reduction achieved by using only the four most significant target and projection vectors in the analysis.

Appearance of maxima, side-peaks and shoulders in the calculated spectra was independent of the values of the protolytic constants within the specified uncertainty limits. When determined from three independent titration series the spectral shapes were practically the same

and the intensities of the 13 peaks and shoulders larger than 3000 M^{-1} cm^{-1} varied by less than 10%.

Matrix **R** was also used to calculate concentrations of the protolytic species (Eq. 4a). These are compared with the concentrations predicted from the equilibrium constants in Fig. 2 (bottom right). This comparison is a self-consistency check of the analysis. Stochastic variations between the concentrations calculated these two ways is a strong indication of successful analysis. Changing the protolytic constants by even small amounts outside the specified uncertainty limits results in systematic variations: data points at pH values below the erroneous pK_a either falling constantly below the lines, and those at pH values above the erroneous pK_a falling constantly above the lines, or vice versa.

Acknowledgements

This work was supported by the Swedish Research Council for Engineering Sciences, the Swedish Natural Science Research Council, and by Magn. Bergvalls Stiftelse.

References

[1] W. Lawton and E. Sylvestre, Technometrics, 13 (1971) 617.
[2] S. Kawata, H. Komeda, K. Sasaki and S. Minami, Appl. Spectrosc., 39 (1985) 610.
[3] N. Ohta, Anal. Chem., 45 (1973) 553; K. Sasaki, S. Kawata and S. Minami, Appl. Optics, 22 (1983) 3599; A. Meister, Anal. Chim. Acta, 161 (1984) 149; K. Sasaki, S. Kawata and S. Minami, Appl. Optics, 23 (1984) 1955; O. Borgen and B.R. Kowalski, Anal. Chim. Acta, 174 (1985) 1; Y. Sun, D.F. Sears and J. Saltiel, Anal. Chem., 59 (1987) 2515.
[4] M. Kubista, R. Sjöback and B. Albinsson, Anal. Chem., 65 (1993) 994.
[5] R. Fisher and W. MacKenzie, J. Agric. Sci., 13 (1923) 311; H. Wold, in F. Daved (Ed.), Research Papers in Statistics, Wiley, New York, 1966, p. 441; K.V. Mardia et al., Multivariate Analysis, Academic Press, London, 1979.
[6] I. Scarminio and M. Kubista, Anal. Chem., 65 (1993) 409.
[7] M. Imamura, Bull. Chem. Soc. Jpn., 31 (1958) 962.
[8] V. Zanker and W. Peter, Chem. Ber., 91 (1958) 572.
[9] L. Lindquist, Arkiv Kemi, 16 (1960) 79.
[10] G. Weber and F.W.J. Teale, Trans. Faraday Soc., 54 (1958) 640.