Journal of Fluorescence, Vol. 14, No. 2, March 2004 (©2004)

Determination of Protolytic Constants by Trilinear Fluorescence Spectroscopy

Mikael Kubista,[1,2] Ismail Hassan Ismail,[1] Amin Forootan,[1] and Björn Sjögreen[1]

Received October 31, 2003; revised May 11, 2003; accepted May 11, 2003

Protolytic equilibria often have profound effects on chemical activity, since protolytic species usually behave quite differently. It is therefore important to characterize the protolytic properties of important chemicals. Here we present a new approach to study protolytic equilibria of fluorescent species that is extremely accurate and relies on minimum assumptions. We show that by measuring 2-dimensional excitation/emission scans of samples at different pH, the 3-dimensional experimental data set, $I(\lambda_{ex}, \lambda_{em}, C(\text{pH}))$, can be unambiguously decomposed into the spectral responses of the protolytic species present as well as their concentration. The approach is demonstrated on the protolytic equilibrium of fluorescein. Although the fluorescein monoanion cannot be obtained in pure form, the spectra and concentrations of both fluorescein species, as well as the protolytic constant, are determined with excellent accuracy. The proposed method is general and can be applied not only for studies of protolytic equilibria, but on any chemical equilibria and chemical reactions involving fluorescent species.

KEY WORDS: Trilinear spectroscopy; trilinear fluorescence; chemical equilibrium analysis; multidimensional data analysis; multidimensional fluorescence.

INTRODUCTION

By most techniques test samples are characterized by a single scalar measurement, such as its absorption of light of distinct energy, weight, density, melting temperature etc. Such measurements may be of use for simple quality control to, for example, find out if a particular sample behaves differently than expected, but they are not very useful for thorough characterizations. For such purposes spectroscopic methods are preferred. In a spectrum contributions from different species may be distinguished, which makes it possible to say something also about sample composition. Fluorescence spectra are particularly useful since they are inherently multi-dimensional [1]. For example, by recording emission spectra at different excitation wavelength it is possible to distinguish components that have different absorption spectra (and, hence, excitation spectra). Still, there is a limit to how well 2-dimensional spectra can be resolved. If no reference information is available, it is only possible to calculate ranges for the components' spectra [2]. This limitation is due to rotational ambiguity in the mathematical solution: any linear combination of the components' excitation spectra and the reciprocal linear combination of their emission spectra will fit the experimental data. The solutions can be narrowed by requiring non-negative spectral intensities, but there is still no way to find the truly correct solution. Furthermore, non-negative conditions may disqualify correct solution, which may have negative elements due to noise. In 1990 we showed that the rotational ambiguity is resolved if samples are analyzed in pairs [3]. The reason behind is that only one linear combination of the components' excitation and emission spectra will fit both 2-dimensional data sets. The mathematical

[1] Department of Chemistry and Biosciences, Chalmers University of Technology, Department of Clinical Chemistry, Göteborg University, and MultiD Analyses AB, Askim, Sweden.

[2] To whom correspondence should be addressed. E-mail: mikael.kubista@molbiotech.chalmers.se

1053-0509/04/0300-0139/0

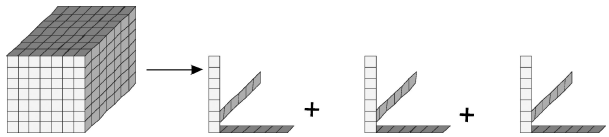

Fig. 1. Graphical illustration of trilinear decomposition.

procedure to find this solution is called Procrustes rotation [4,5]. Recently Procrustes rotation was generalized to larger number of samples [6], and found to be mathematically equivalent to trilinear decomposition and parallel factor analysis (PARAFAC) [7]. In trilinear decomposition we have a most powerful method to analyze trilinear fluorescence data. Without making any non-trivial assumptions or using reference information we can, from the measurement data only, calculate the excitation spectra, the emission spectra, as well as the concentrations of all species in all samples. This makes trilinear fluorescence spectroscopy most powerful to study independent test samples, and also to characterize chemical equilibria and chemical and photophysical reactions [5].

Trilinear data sets are sums of components' responses, which can be factorized into products of components' specific 1-dimensional contributions:

$$I(\ ,\ ,\ ) \propto \sum_{i=1}^{n} I_i(\ ,\ ,\ ) = \sum_{i=1}^{n} I_i(\ )I_i(\ )I_i(\ ) \quad (1)$$

Decomposition of trilinear data is unique and the components' specific responses are calculated without any modeling or making of any non-trivial assumptions. The principle of trilinear decomposition is illustrated graphically in Fig. 1.

Fluorescence is inherently a multidimensional technique and therefore highly suited for trilinear spectroscopic measurements. As we show here equilibrium constants can be determined with extreme accuracy by trilinear steady-state fluorescence spectroscopy even when the titration end-points are not reached.

MATERIALS AND METHODS

Fluorescein was purchased from Sigma and was used without further purification. Spectral analysis revealed the dye contained no significant amounts of contaminants. Its concentration was determined spectroscopically in 0.1 M NaOH assuming molar absorptivity of 76,900 $M^{-1}cm^{-1}$ for the fluorescein dianion [8]. A fluorescein stock solution (0.4 M) was prepared and split into two aliquots to which equal amounts of 5 mM phosphate buffer of either pH 5.37 or 8.30 was added. These two new stock solutions contained exactly the same amounts of fluorescein and of buffer, differing only in pH. The stock solutions were used as test samples with extreme pH. Additional test samples were prepared by mixing the stock solutions in appropriate ratios to obtain samples with reasonably evenly spaced pH values. For each test sample the pH was measured using two points calibrated pH meter. In analysis of data, the pH 5.37 sample was left out, since it contained about 10% of neutral fluorescein, which upon excitation is converted to negative fluorescein species [8].

Absorption spectra were measured on a CARY 5 spectrophotometer and fluorescence spectra were measured on a SPEX model FL1T1 spectrofluorometer at a resolution of five data points per nanometer. In no case did the absorption of the fluorescein samples exceed 0.06 making the necessary correction for the inner filter effect small [9].

Data analysis was performed using DATAN (www.multid.se) on a regular personal computer.

RESULTS

Figure 2 shows ex/em scans of fluorescein measured in the emission interval 470–650 nm using 390, 400, 410, 420, 430, 440, 450, and 460 nm excitations at different pH in the interval 5.66 < pH < 8.30. Maximum intensity is at 510 nm emission at all excitation wavelengths used and at all pH. With decreasing pH a shoulder appears at the high wavelength end of the emission spectrum and the intensity measured when using short excitation wavelength grows with decreasing pH relative to intensity measured at long excitation. Clearly, more than one fluorescent species contribute to the spectra and their relative contributions depend on pH and the excitation and emission wavelengths used.

Using very low total fluorescein concentration (200 nM) energy transfer between dyes and the inner filter effect are negligible [9]. Under such conditions the observed fluorescence equals the sum of the contributions from the fluorescein protolytic species present weighted with their concentrations:

$$I(\ _{ex},\ _{em}, c(\mathrm{pH})) \propto \sum_{i=1}^{n} I_i(\ _{ex},\ _{em}, \mathrm{pH})$$

$$= \sum_{i=1}^{n} I_i(\ _{ex},\ _{em}) c_i(\mathrm{pH}) \quad (2)$$

In general, the emission spectrum of a pure fluorescent species is independent of the excitation wavelength used and, vice versa, the excitation spectrum is independent of the wavelength of emission [10]. Hence, the

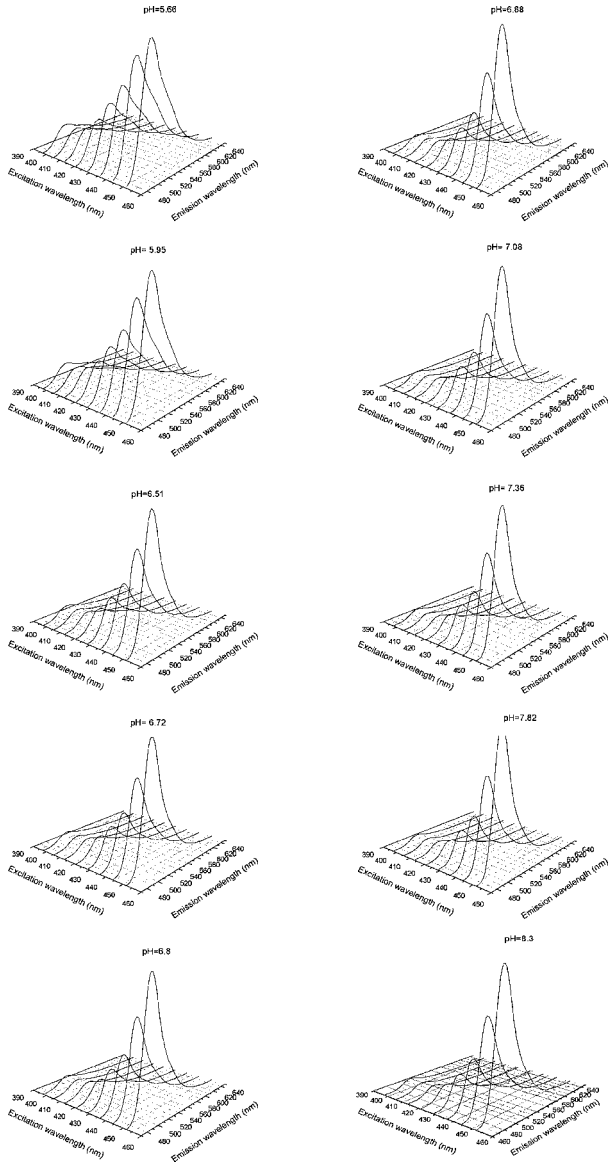

Fig. 2. Emission spectra of fluorescein at pH = 5.66, 5.95, 6.51, 6.72, 6.80, 6.88, 7.08, 7.36, 7.82, and 8.30 measured between 470–650 nm with five data points collected per using 390, 400, 410, 420, 430, 440, 450, and 460 nm excitation.

excitation/emission matrix can be factorized:

$$I(\ _{\mathrm{ex}},\ _{\mathrm{em}}, c(\mathrm{pH})) \propto \sum_{i=1}^{n} I_i(\ _{\mathrm{ex}},\ _{\mathrm{em}}, \mathrm{pH})$$

$$= \sum_{i=1}^{n} I_i(\ _{\mathrm{ex}}) I_i(\ _{\mathrm{em}}) c_i(\mathrm{pH}) \qquad (3)$$

From Eq. (3) follows that measurements of excitation/emission scans at different pH make up a trilinear data set that can be decomposed into $I_i(\ _{\mathrm{ex}})$, $I_i(\ _{\mathrm{em}})$, and $c_i(\mathrm{pH})$, which are the excitation intensities, emission intensities, and concentration dependence on pH for the i:th protolytic species.

Before proceeding with trilinear decomposition we must estimate the number of protolytic species, n, that contribute to the spectra in the studied pH interval. This is done by principal component analysis, PCA [11]. Conventional PCA is for 2-dimensional data arrays and decomposes a data matrix **A** into a product of two orthogonal matrices.

$$\mathbf{A} = \mathbf{TP}' + \mathbf{E} = \sum_{i=1}^{q} \mathbf{t}_i \mathbf{p}_i' + \mathbf{E} \qquad (4)$$

T and **P'** are referred to as target and projection matrices and they contain sets of q orthogonal target and projection vectors. **E** is the residual matrix. For noise free data when **TP'** exactly matches **A** and **E** is zero then q equals the number of independent species, n, that are present. Hence, n can be determined from the number of principal components required to reproduce the data. However, experimental data contain noise and there is always a noise residual. Noise can be accounted for by using statistical indicators to estimate when the residual matrix **E** contains only noise [12,13].

3-Dimensional data can be analyzed by PCA by slicing the 3-dimensional data set into 2-dimensional arrays that are laminated into a large 2-dimensional matrix (Fig. 3). Fluorescence data are typically sliced along the sample dimension and joined along the wavelength dimension with least data points. Figure 4 shows the result of PCA for the laminated fluorescein data in Fig. 2. Data are reconstituted with the two most significant pairs of principal components as predicted by the statistical indicators [12]. This is also in agreement with previous studies of fluorescein, which show that the fluorescein monoanion and dianion dominate at pH > 5.5 [8]. The agreement between measured and reconstructed data was excellent for all samples as exemplified for the samples with extreme pH (pH = 5.66 and pH = 8.30) in Fig. 4. The agreement did not improve significantly by adding a third pair of principal components, from which we conclude that two protolytic species contribute to the fluorescence of fluorescein in the range $5.66 \leq \mathrm{pH} \leq 8.30$.

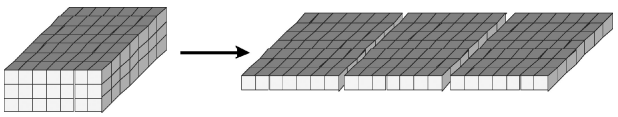

Fig. 3. Slicing and lamination of trilinear data for PCA.

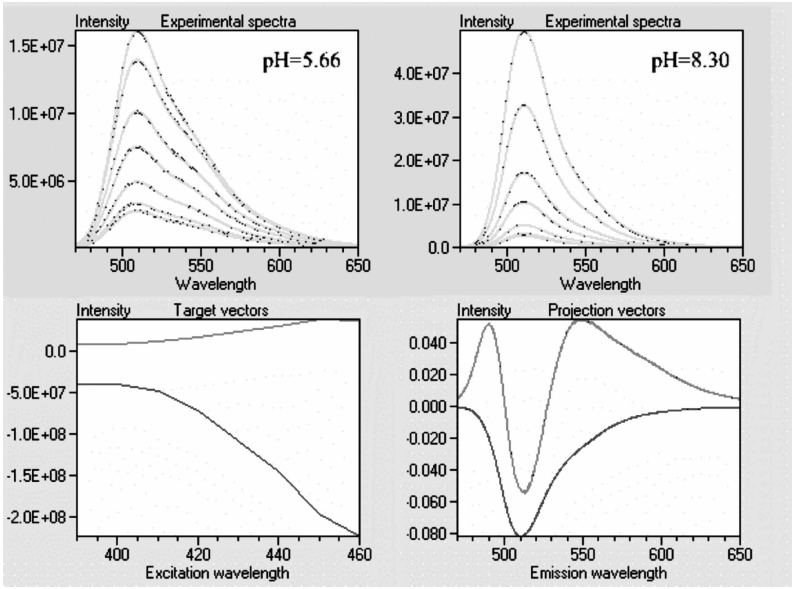

Fig. 4. PCA of the fluorescein excitation/emission pH titration in Fig. 2. Two pairs of principal components are used. Measured (purple) and reconstructed (green) spectra are overlaid in the top left (pH = 5.66) and the top right (pH = 8.30) figures. The difference between measured and reconstructed spectra is hardly distinguishable. Bottom graphs show the two most significant pairs of principal components. The target vectors (left) are orthogonal and correspond to the sample at pH 5.66. The projection vectors (right) are orthonormal and they are common to all samples.

Principal components are mathematical constructs. They are defined as the sets of orthogonal vectors that best reproduce the experimental data. The orthogonality criterion requires that all principal components but the first pair have negative features, which makes them very different from normal spectra. In fact, the principal components are linear combinations of the excitation and emission spectra of the chemical species present. To determine the linear combinations we need at least two excitation/emission data sets [3]. In Fig. 2 we have 10 sets. Assuming presence of two protolytic species the data were decomposed into the components' specific responses by trilinear decomposition using DATAN (www.multid.se.). Measured spectra and spectra reconstructed from the calculated component specific responses are compared in the top left graph in Fig. 5. The agreement is as good as the reproduction from the principal components (Fig. 4), but this time the base vectors have physical meaning. The emission spectrum and excitation intensities determined for the fluorescein dianion is indistinguishable from those measured in 1 M NaOH, where the dianion is the only species present [8]. The fluorescein monoanion cannot be obtained in pure form and direct comparison of calculated and measured spectra is not possible. But the results are in excellent agreement with previous determinations of the fluorescein monoanion spectra [8]. Plotted on top of each other, the calculated (Fig. 5) and measured [8] spectra of the fluorescein monoanion totally overlap (not shown). When normalized to maximum intensity of 1, the sum of squared differences:

$$\mathrm{SSD} = \frac{1}{m}\sum_{i=1}^{m}(I(\ \mathrm{em})_{\mathrm{calc}} - I(\ \mathrm{em})_{\mathrm{meas}})^2 \qquad (5)$$

between the spectra was 10^{-5}. The square root, which is 0.003, shall be compared with maximum intensity. Hence, the error in the determination of spectral profiles in this particular case is of the order of 3 per mille of signal maximum. This is of the order of the noise level in the measured spectra.

Only the shapes, not the magnitudes of the components' specific responses, can be determined by trilinear decomposition due to scaling ambiguity. In Fig. 5 the responses are presented with spectra normalized to unit area and all weights are collected in the calculated concentrations. In Fig. 6 the results are renormalized to show concentrations as molar ratios and the weights are divided equally between the spectral responses. Inspecting the calculated molar ratios, the two species are found to have equal concentration at pH = 6.41. Although no

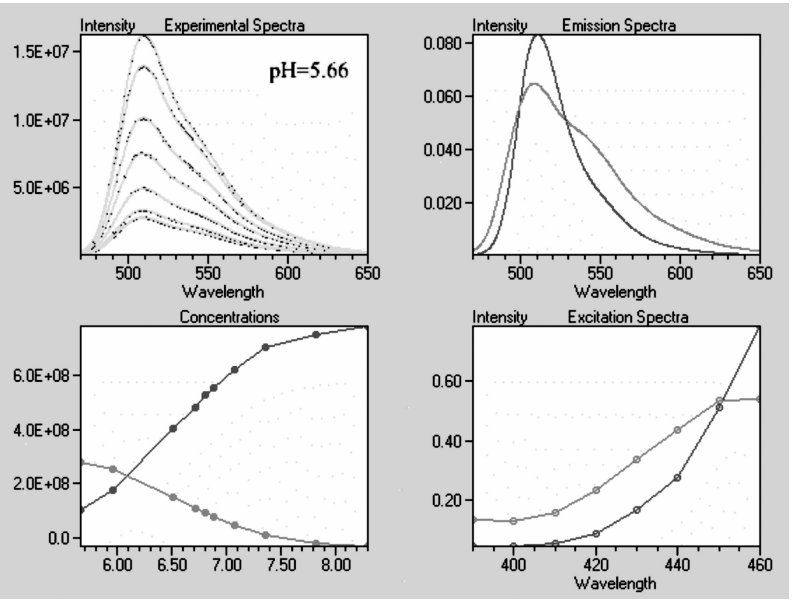

Fig. 5. Trilinear decomposition of the fluorescein excitation/emission pH titration in Fig. 2. Top left graph compares measured (purple) and reconstructed (green) spectra at pH = 5.66. Calculated emission profiles are shown in top right panel, calculated excitation intensities are shown in bottom right panel, and calculated concentrations are shown in bottom left panel. Responses of fluorescein monoanion are shown in red and those of the dianion are shown in blue. All calculated spectral responses are normalized to the same area. Monoanion responses are shown in red and dianion responses are shown in blue. The lines in the graphs connect calculated data points and are only to guide the eye.

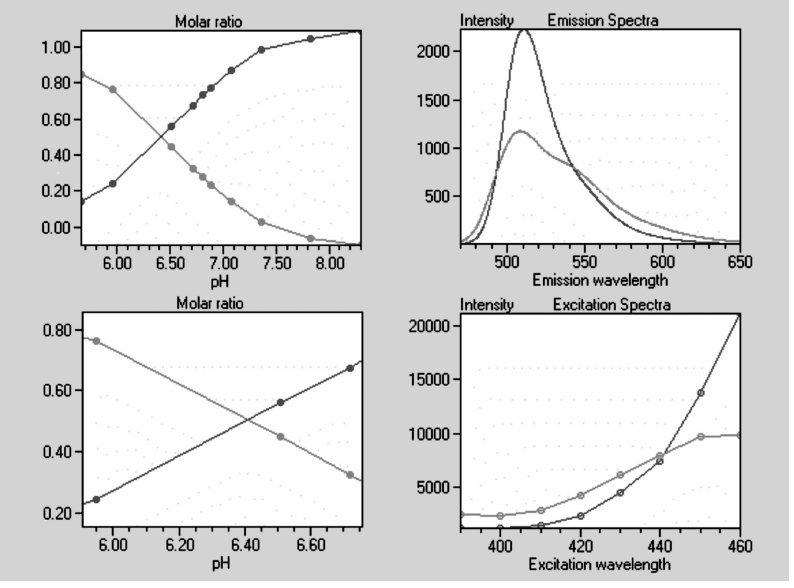

Fig. 6. The results from trilinear decomposition renormalized to molar ratios. Left graphs show calculated molar ratios (bottom panel enlarged scale). Right graphs show calculated emission (top) and excitation (bottom) profiles. Monoanion responses are shown in red and dianion responses are shown in blue. Lines shown connect calculated data points and are only to guide the eye.

assumptions have been made about interactions between the species in the analysis, this is the protolytic constant of the fluorescein monoanion and dianion, and agrees with two decimals precision with previous determinations of fluorescein pK_a at the low ionic strength used [8].

DISCUSSION

With modern spectrophotometers 2-dimensional excitation/emission scans are readily collected. Most instruments can also vary a third parameter such as temperature or switch between samples using a multi-sample holder automatically generating 3-dimensional data. More advanced spectrophotometers can also measure intensity as function of time, which can be either time after excitation (time-resolved measurements) or time after mixing (reaction kinetics). Due to the factorial nature of fluorescence such data are trilinear, and they can be decomposed into components' specific responses without assuming any thermodynamic models or spectral shapes. Here we used trilinear steady-state fluorescence to determine the protolytic constant of fluorescein and the spectral responses of its protolytic species. Comparing with literature data both the spectra and protolytic constant were determined with great precision. This was actually expected due to the very large amount of data analyzed. Five points were collected per nanometer of the emission spectra, which were measured at eight excitation wavelengths in each of ten samples. This gave a total of 80 emission spectra and 72080 data points. Since all spectra are linear combinations of the fluorescein anion and dianion spectral responses trilinear decomposition is mathematically a highly over determined problem, resulting in very high signal to noise ratios in the calculated spectra similar to what is obtained by regular averaging of repeated scans. As a result the calculated spectra have substantially higher signal to noise levels compared to the measured spectra. This very efficient averaging also results in highly accurate determination of concentrations and, hence, of the equilibrium constant.

Trilinear fluorescence has several important advantages compared to traditional fluorescence investigations. Measurements are rather straight forward to perform using modern spectrofluorometers, analysis is model independent, and calculated results are extremely accurate. The drawbacks are that 3-dimensional data are hard to display properly and therefore difficult to inspect visually. Also, most researchers are still not very experienced in designing multidimensional experiments, which, although in essence is not very difficult, takes some practice. But as more examples become available, scientists will get familiar with trilinear methods and we expect a great future in science and technology of all the possible variants of trilinear fluorescence spectroscopy.

REFERENCES

1. Principles of Fluorescence Spectroscopy. Ed. Lakowicz Joseph R. Cloth, Plenum Pub. Co., 1999.
2. W. H. Lawton and E. A. Sylvestre (1971). Elimination of linear parameters in nonlinear regression. *Technometrics* **13**, 617–633.
3. M. Kubista (1990). A new method for the analysis of correlated data using procrustes rotation which is suitable for spectral analysis. *Chemometrics Intell. Lab. Syst.* **7**, 273–279.
4. I. Scarminio and M. Kubista. (1993). Analysis of correlated spectral data. *Anal. Chem.* **65**, 409–418.
5. M. Kubista, J. Nygren, A. Elbergali, and R. Sjöback. (1999). Making reference samples redundon. *Crit. Rev. Anal. Chem.* **29**, 1–28.
6. A. Liwo, P. Skurski, S. Oldziej, L. Lankiewicz, J. Malicka, and W. Wiczk. (1997). A new approach to the resolution of the excitation-emission spectra of multicomponent systems. *Comput. Chem.* **21**, 89–96.
7. R. Bro. (1997). PARAFAC. Tutorial and applications. *Chemometrics Intell. Lab. Syst.* **38**, 149–171.
8. R. Sjöback, J. Nygren, and M. Kubista. (1995). Absorption and fluorescence properties of fluorescein. *Spetrochimica Acta A* **51**, L7–L21.
9. M. Kubista, R. Sjöback, S. Eriksson, and B. Albinsson. (1994). *Anal.* **119**, 417.
10. M. Kasha (1950). Characterization of electronic transitions in complex molecules. *Discuss. Faraday Soc.* **9**, 14.
11. E. R. Malinowski (1991). *Factor Analysis in Chemistry*, 2nd ed., Wiley, New York.
12. A. Elbergali, J. Nygren, and M. Kubista. (1999). An automated procedure to predict the number of components in spectroscopic data. *Anal. Chim. Acta* **379**, 143–158.
13. M. Meloun, J. Capek, P. Miksik, and R. Brereton. (2000). Critical comparison of methods predicting the number of components in spectroscopic data. *Anal. Chim. Acta* **20736**, 1–18.