Chemometrics and Intelligent Laboratory Systems, 14 (1992) 305–317
Elsevier Science Publishers B.V., Amsterdam

Spectroscopic resolution of macromolecular complexes using factor analysis: Cu(II)–polyethyleneimine system

R. Tauler and E. Casassas *
Department of Analytical Chemistry, University of Barcelona, Diagonal 647, Barcelona 08028 (Spain)

(Received 7 July 1991; accepted 7 October 1991)

Abstract

Tauler, R. and Casassas, E., 1992. Spectroscopic resolution of macromolecular complexes using factor analysis: Cu(II)–polyethyleneimine system. *Chemometrics and Intelligent Laboratory* Systems, 14: 305–317.

The complexes of Cu(II) ion with polyethyleneimine formed during a VIS spectroscopic titration of solutions containing different amounts of Cu(II) and polyethyleneimine at various pH values have been studied using a new self-resolving approach based on different factor analysis techniques, including evolving factor analysis, error in factor analysis, cross-validation, target factor analysis, and rank annihilation. Three different macromolecular complex species between Cu(II) and polyethyleneimine were detected in the system, and their concentration profiles and individual spectra were estimated without any previous knowledge of the underlying chemical model (set of stoichiometric coefficients and set of stability constants). The method developed is proposed for use in the deduction of the metal complexing properties of macromolecular systems.

INTRODUCTION

The study of macromolecular complexes of metal ions is a field of great interest because of its environmental and biological importance [1,2]. However, this study is hindered by the fact that the law of mass action ruling the complexation equilibria is only strictly valid for each one of the reaction sites of the macromolecule, and several additional or secondary effects must be considered. These secondary effects have been classified [3] into three types: (a) polyfunctional effects, referred to as differences in chemical nature of the macromolecule coordination sites, and in the different electrostatic and steric environments of these sites; (b) conformational changes caused by the changes in pH and ionic strength of the medium, or by the amount of complexed ion; and (c) polyelectrolyte effects caused by the ionization of major sites of the macromolecule leading to changes in the local electric field at the macromolecule surface. All these effects contribute to the stability of the species formed, and their relative importance is difficult to define, since it varies with the degree of site occupation (complexation). The interpretation of the experimental data using traditional least-squares curve fitting approaches is consequently rather cumbersome [3–5], and there is a demand for the development of new approaches free from the constraints of the law of mass action and free from the prior assumption of a chemical model.

In the present work a new approach based on factor analysis techniques is proposed for the study of metal–ion complexation by macromolecules. The method is based on our previously developed SPFAC procedure (spectra factor analysis) [6–8] and it is applied here for the first time to the study of the complexation of metal ions by macromolecules. The method is intended to be used for the detection of the number of species present in the system, for the estimation of the concentration profiles and of the individual spectra of the species in equilibrium during a spectroscopic titration and, in this particular case, for the deduction of the complexation properties of the macromolecular ligand. SPFAC is a procedure which does not require the initial assumption of a chemical model, nor the fulfillment of the law of mass action, and although it was initially designed as a complementary tool to least-squares curve fitting approaches in non-polyelectrolyte systems, it can be used as an independent, model-free, and self-resolving tool in the study of the spectroscopic changes produced by the interactions of metal ions with macromolecular systems. In the present work the interaction of Cu(II) with polyethyleneimine is investigated as an example of the application of the method.

The use of spectroscopic techniques in the study of the complexation of metal ions by macromolecules has been criticized [3] because of their ambiguity in determining the concentration of the free ligand sites and because of the difficulties involved in resolving the spectroscopic signal, which is usually a composite average of the different species spectra. However, when used as in the present work, in conjunction with suitable chemometric techniques, and through study of the metal environment instead of the ligand sites environment (i.e. the study of the spectroscopic active *d–d* transitions of the transition metal ions), they become very powerful tools.

EXPERIMENTAL

Reagents and solutions

Polyethyleneimine (PEI, Aldrich) is a synthetic water-soluble polymer with the empirical formula $(C_2H_5N)_n$. Chemical studies show that 25% of the amine nitrogen atoms are primary, 50% secondary, and 25% tertiary. This distribution indicates that PEI is a highly branched, fairly compact polymer [9]. All other reagents and materials used are of analytical grade quality.

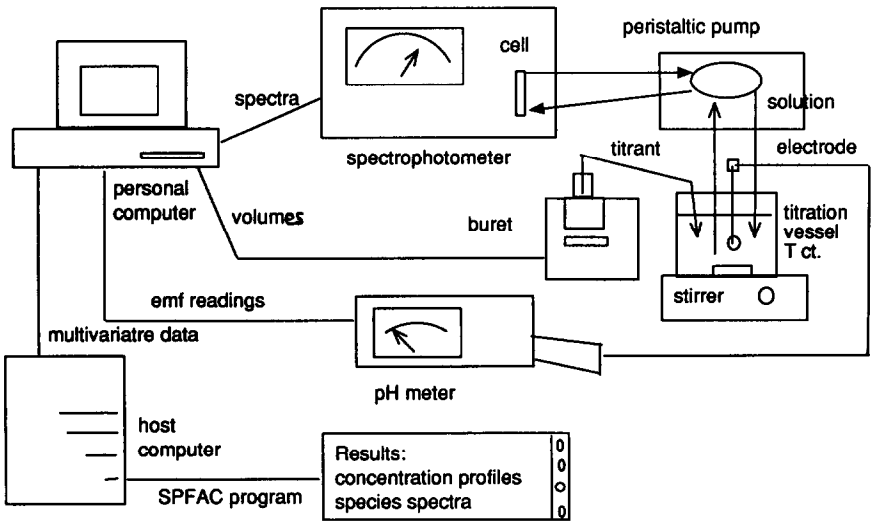

Fig. 1. Experimental setup.

Experimental procedure

The experiments were conducted as shown in Fig. 1. Two different solutions containing different concentration ratios of PEI and Cu(II) ion are titrated by adding small amounts of acid or base to change the pH. At each pH, one spectrum is recorded (400–860 nm). The plots of the whole set of spectra for two titrations are given in Fig. 2 (the respective sets of data colllected in matrices **A** and **B**). In the first case (concentration ratio about 1:1), precipitation of copper hydroxide is observed above pH 5.0, whereas in the second case (ligand site:metal concentration ratio about 4:1) completely transparent solutions are obtained even at basic pH. A more complete data set was used for validation of the obtained results but it is not given here for brevity.

Data treatment

Data treatment has been carried out with the SPFAC program previously developed [6–8]. The aim in the development of the SPFAC procedure is to provide a single computer program which allows the self-resolution of the concentration profiles and species spectra of the components present in spectroscopically ordered data systems using factor analysis techniques. The program is written in FORTRAN 77 and runs in an IBM 3090 mainframe (large sets of data) or in an IBM PC environment (for small sets of data). Essentially the data treatment consisted of the following tasks (see Fig. 3):

1. Selection of the pH and wavelength range of interest and building up of the data matrix **D**(*NSOLN, NWAVE*), being *NSOLN* and *NWAVE*, respectively, the number of solutions and wavelengths analyzed. Optional pretreatment of the experimental data including smoothing and derivation.

2. Factor analysis of the experimental data matrix **D**; calculation of the eigenvalues $\boldsymbol{\lambda}$ and eigenvectors **V** of the matrix $\mathbf{DD}^{\mathrm{T}}$. Calculation of the reproduced data matrix, $\mathbf{D}^{*} = \mathbf{UV}$, where **V** and **U** (calculated from $\mathbf{U} = \mathbf{DV}^{\mathrm{T}}$) are obtained in the eigenvector analysis. The number of components to consider is equal to the number of absorbing species in the system; i.e. it is considered that in the data (UV–VIS absorption spectroscopy of clear solutions) no other source of variation is present, apart from the chemical one.

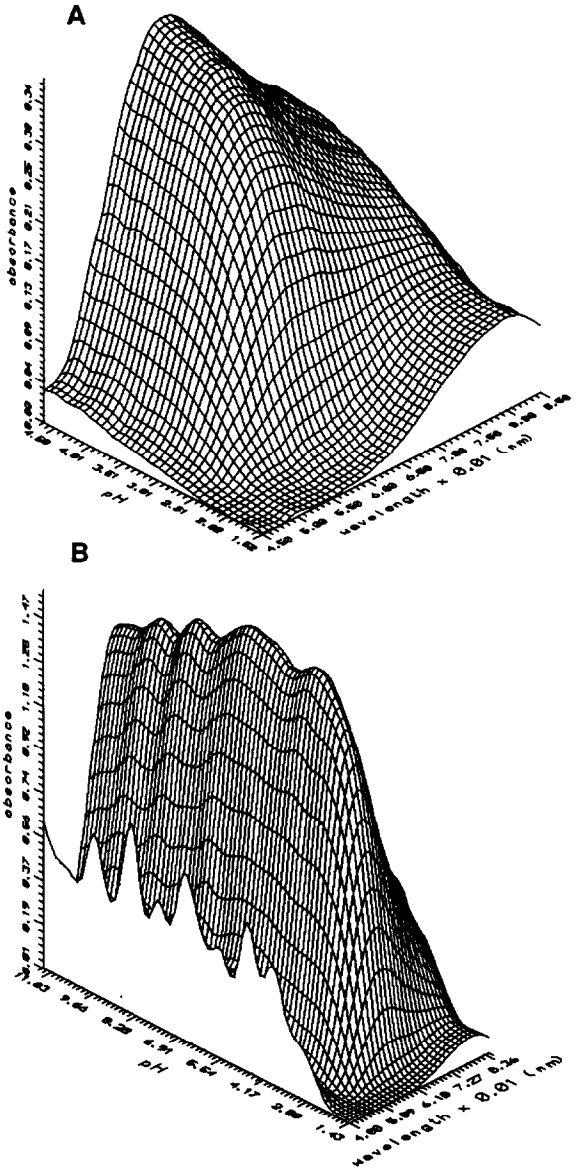

Fig. 2. Examples of experimental three-dimensional data. (A) Plot of the spectroscopic titration data matrix **A**; experimental conditions: Cu(II) concentration 0.00906 *M*; PEI site concentration 0.00995 *M*; pH range 1.52–4.67. (B) Plot of the spectroscopic titration data matrix **B**; experimental conditions: Cu(II) concentration 0.00906 *M*; PEI site concentration 0.04125 *M*; pH range 1.43–11.3.

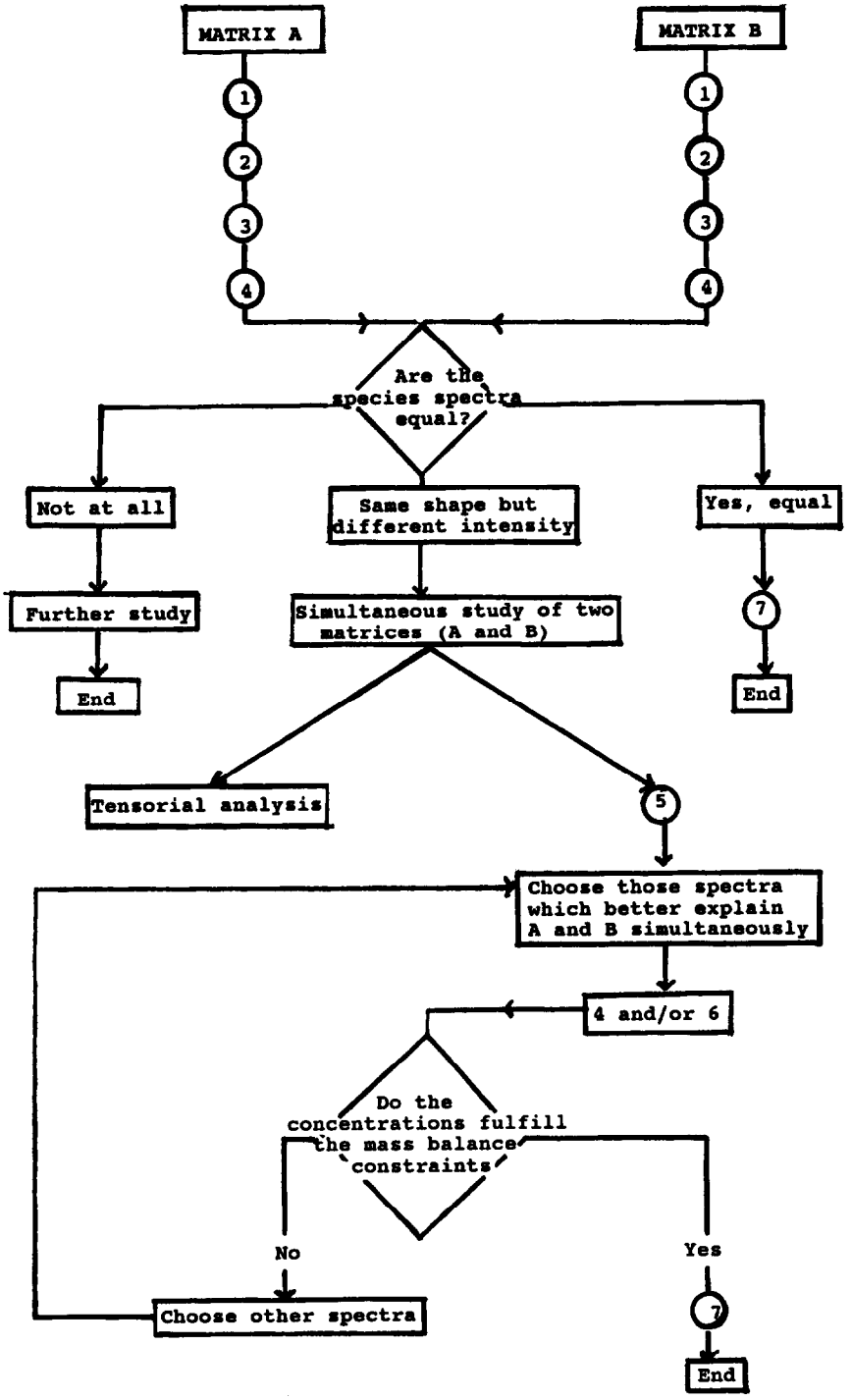

Fig. 3. Flow diagram of data treatment using SPFAC (spectra factor analysis). See ref. 8 for more detail about the structure of the computer program. 1. Data acquisition and pretreatment. 2. Determination of the number of factors: (a) principal component analysis (PCA); (b) error in factor analysis; and (c) cross-validation. 3. First estimation of species distribution by evolving factor analysis (EFA). 4. Determination of the best set of species spectra and of species distribution for the unknown components from alternating least squares (ALS) using EFA results. 5. Identification of the components by target factor analysis (TFA) of candidate spectra. 6. Determination of the concentration of the components with known spectra through rank annihilation (RAEFA, using results of TFA and EFA). Estimation of the concentration of a known component in the mixture and subtraction of its spectral contribution to the whole data set. 7. Deduction of the complexation properties of the system.

As this number is a critical point in the whole analysis, it is estimated using different independent methods: (1) from the changes in magnitude of the eigenvalues, from the plot of the indicator IND function and from Fischer variance ratio tests using the principles of the theory of error in factor analysis as proposed by Malinowski [10–14]; (2) from the plot of the PRESS (prediction error sum of squares) function obtained from cross-validation of the complete spectral matrix data as proposed by Wold [15]; the $\mathbf{DD}^T$ matrix is partitioned into two equal groups; one group is used to obtain the factors and the other is used for the prediction; (3) from abstract distribution plots obtained from submatrix analysis [16] and in evolving factor analysis [17] (see below).

The results obtained are compared and examined critically before starting the self-resolution iterative method. In cases where no conclusive results concerning the exact number of components are obtained, the analysis is carried on considering one choice and repeating the analysis for another among the most probable number. Finally, that number of species is adopted which best fit the data.

3. Evolving factor analysis [17,18]; the experimental spectra data matrix $\mathbf{D}(NSOLN, NWAVE)$ is partitioned on a set of submatrices obtained successively increasing the pH and decreasing the pH as follows:

For N = 2 to NSOLN
  $\mathbf{D} = [\mathbf{A}, \mathbf{B}]$ where $\mathbf{A}(N, NWAVES)$ and $\mathbf{B}(NSOLN - N + 2, NWAVES)$
  Compute the eigenvalues of $\mathbf{S} = \mathbf{AA}^T$ and $\mathbf{R} = \mathbf{BB}^T$
  Plot all the eigenvalues at varying pH and draw lines through them for each factor at each pH (forward and backward)
Next N

The initial estimation of the actual concentration profiles for the species distribution of the system (abstract distribution plot) is based on the evolution with pH of the magnitudes of the eigenvalues during the spectroscopic titration, and it is obtained by junction of the appropriate factor lines on the plot obtained in the factor analysis of the submatrices. The factor lines describe the changes in the spectral contributions of the system. It has been shown [7,8] that the method has two critical points: the selection of the correct number of components and the rule for crossing the lines related to the upsurging (forward analysis) and decreasing (backward analysis) factor lines.

4. From the initial estimation of the concentration profiles obtained using evolving factor analysis, the generalized Beer law equation in matrix form is solved iteratively by least-squares to obtain the matrices of species spectra **A**, and of concentration profiles **C** which best fit the data.

$$\mathbf{A} = (\mathbf{C}^T\mathbf{C})^{-1}\mathbf{C}^T\mathbf{D}^*$$

$$\mathbf{C} = \mathbf{D}^*\mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}$$

The concentration profiles in matrix **C** should be unimodal and positive; only a certain amount (given in the input data) of departure from the unimodal condition is allowed at any instant. Negative values of absorbances obtained in the iterative least-squares estimation of **A** are set equal to zero. The pH ranges or windows of existence of each species determined using evolving factor analysis can be kept constant at all times, or they can be allowed to change.

At each iteration the residuals matrix is obtained from the difference between the data matrix reproduced by factor analysis and the product of absorbance and concentration matrices currently evaluated. The process is repeated iteratively until convergence is achieved or until a certain number of cycles has been run.

5. Species which are suspected to be present in the titration and have a known spectrum can be tested independently using target factor analysis [10,14]. The disagreement between the target known spectrum and the estimated spectrum using target factor analysis (very low correlation between both spectra) excludes the possibility of the presence of such species in the system. By the contrast, a high correlation between the target and the estimated spectrum indicates that such species could be present in the system. In the present work, target testing has been proved to be a very useful tool in solving the dichotomy

between the intensity of the species spectrum and its concentration (see Results and Discussion).

6. Once the presence of a certain species with known spectrum is confirmed, its concentration profile can be estimated using rank annihilation procedures provided its spectral contribution to the experimental data matrix is significant. The implemented rank annihilation algorithm is based on the rank annihilation by evolving factor analysis (RAEFA) as proposed by Gampp et al. [19]. In this method the submatrix containing the known component is removed, giving a new data matrix of reduced rank. The eigenvectors obtained for this new data matrix are augmented with the known spectrum of the pure known component to form a set of basis vectors in the spectral domain that spans the full rank of the entire original data matrix. From the pseudo-inverse of the basis vectors it is then possible to estimate the concentration profile of the known component directly. Once the estimation of the concentration of the known component has been made using this procedure, the spectral contribution of this species is subtracted from the original

TABLE 1

Complexation properties of polyethyleneimine deduced from the SPFAC treatment of the experimental data matrix **B**

NSOLN = Solution number; pH of this solution; X0 = fraction of free (non-complexed) metal; X1 = fraction of complexed metal; Y0 = fraction of free (non-complexed, non-protonated) macromolecule sites; Y1 = fraction of non-complexed macromolecule sites; Y2 = fraction of complexed and protonated macromolecule sites; Y3 = fraction of complexed (but non-protonated) macromolecule sites; LK1, LK2 = the log of the pseudostability constants calculated from the log of the quotients of the concentration of complexed metal ion divided by the product of the concentration of free metal and the concentration of non-complexed macromolecule sites (LK1), or of non-complexed, non-protonated macromolecule sites (LK2), to the power of four; LK3 is the log of the quotient of the concentrations referred to the substitution equilibrium: $4LH + Cu = CuL_4 + 4H^+$, where LH is the protonated macromolecule site and CuL_4 is the metal bonded to four macromolecule sites.

NSOLN	pH	X0	X1	Y0	Y1	Y2	Y3	LK1	LK2	LK3
1	1.430	1.000	0.000	0.230	1.000	0.770	0.000	–	–	–
2	1.500	1.000	0.000	0.240	1.000	0.760	0.000	–	–	–
3	1.600	1.000	0.000	0.240	1.000	0.760	0.000	–	–	–
4	1.700	1.000	0.000	0.250	1.000	0.750	0.000	–	–	–
5	1.790	0.989	0.011	0.248	0.990	0.752	0.010	3.596	6.004	−3.064
6	1.940	0.983	0.017	0.256	0.985	0.744	0.015	3.796	6.136	−3.441
7	2.050	0.972	0.028	0.254	0.975	0.746	0.025	4.044	6.384	−3.633
8	2.130	0.952	0.048	0.249	0.958	0.751	0.042	4.318	6.658	−3.679
9	2.260	0.897	0.103	0.246	0.910	0.754	0.090	4.762	7.037	−3.731
10	2.350	0.839	0.161	0.232	0.859	0.768	0.141	5.086	7.361	−3.767
11	2.470	0.738	0.262	0.216	0.770	0.784	0.230	5.544	7.755	−3.766
12	2.550	0.655	0.345	0.195	0.697	0.805	0.303	5.885	8.097	−3.744
13	2.680	0.530	0.470	0.164	0.587	0.836	0.413	6.410	8.621	−3.740
14	2.760	0.457	0.543	0.152	0.523	0.848	0.477	6.737	8.887	−3.708
15	2.880	0.376	0.624	0.131	0.452	0.869	0.548	7.138	9.289	−3.787
16	2.980	0.347	0.653	0.124	0.426	0.876	0.574	7.296	9.446	−4.029
17	3.160	0.288	0.712	0.112	0.374	0.888	0.626	7.640	9.731	−4.380
18	3.360	0.228	0.772	0.097	0.322	0.903	0.678	8.035	10.127	−4.785
19	3.610	0.203	0.797	0.093	0.299	0.907	0.701	8.228	10.263	−5.567
20	3.850	0.150	0.850	0.081	0.253	0.919	0.747	8.681	10.661	−6.049
21	4.630	0.000	1.000	0.043	0.121	0.957	0.879	–	–	–
22	6.630	0.000	1.000	0.064	0.121	0.936	0.879	–	–	–
23	8.500	0.000	1.000	0.109	0.121	0.891	0.879	–	–	–
24	9.730	0.000	1.000	0.121	0.121	0.879	0.879	–	–	–
25	10.660	0.000	1.000	0.121	0.121	0.879	0.879	–	–	–
26	11.030	0.000	1.000	0.121	0.121	0.879	0.879	–	–	–
27	11.300	0.000	1.000	0.121	0.121	0.879	0.879	–	–	–

data matrix calculated from both its concentration profile and its species spectrum. Once the subtraction is performed, a new run of the SPFAC procedure is initiated with the new reduced data matrix in order to solve the unknown contributions. The method, however, can only be used successfully for those species which have a significant and separate concentration domain in the distribution plots. This happens with the first species, which is often the free metal ion; or with species formed at the highest pH, which is often the saturated metal complex.

7. The study of the complexation ability of the macromolecules is obtained directly from the concentration profiles of the different absorbing species deduced in the treatment of the experimental data. Apparent stability constants can be evaluated from the evolution of the degree of metal complexation with pH. This degree is evaluated from the quotient of bonded to total metal ion concentrations, and/or from the quotient of the bonded to total concentration sites of the macromolecule. The study of the complexation ability of this kind of ligand requires to follow the evolution with pH of the stability quotient between complexed metal ion concentration and the product of concentrations of free metal ion and non-complexed macromolecule sites (see Table 1 for more detail).

RESULTS AND DISCUSSION

The complexation between Cu(II) and polyethyleneimine is directly seen from the strong absorption changes produced in the spectroscopic titrations of solutions containing both components as pH is varied (Fig. 2). The application of the previously described procedure to the two data set matrices **A** and **B** give the following results. (See Fig. 3, which will help to follow this discussion more clearly; the numbering in the present section is the same as the numbering in the section *Data treatment*, above).

1. Optional pretreatment of data in the case under study is reduced to the selection of the ranges and steps of the wavelength and pH measurements, where changes due to the phenomena under study take place (see the Experimental section). No other data pretreatment was needed in this case because the data do not contain other significant sources of variance, apart from the

TABLE 2

Error in factor analysis applied to data matrix **A**

NUM = Solution number; RSD, RMS, IE, XE and IND are the residual standard deviation, root of mean-square, imbedded error, extracted error and indicator function, respectively, of the error in factor analysis (see refs. 10 and 11); EIGENV = the eigenvalues, TRACE = sum of the residual eigenvalues; PROB = the significance level associated with the calculated value of F for the variance ratio test F at GL1 and GL2 degrees of freedom (see refs. 12 and 13). Lines marked in bold show the transition of real factors to error factors.

NUM	RSD	RMS	IE	XE	IND	EIG	TRACE	LOG(EIG)	PROB	F	GL1	GL2
1	2.031	2.009	0.299	2.009	0.001002973	26.157379200	0.445505440	1.418	0.00	552.89	1.0	23.00
2	0.474	0.463	0.099	0.463	0.000244717	0.421802998	0.023703035	−0.375	0.00	160.95	1.0	22.00
3	**0.082**	**0.079**	**0.021**	**0.079**	**0.000044222**	**0.023013115**	**0.000689958**	**−1.638**	**0.00**	**289.20**	**1.0**	**21.00**
4	**0.074**	**0.071**	**0.022**	**0.071**	**0.000042084**	**0.000134445**	**0.000555514**	**−3.871**	**33.76**	**2.01**	**1.0**	**20.00**
5	0.068	0.064	0.022	0.064	0.000040385	0.000102030	0.000453484	−3.991	39.04	1.78	1.0	19.00
6	0.062	0.058	0.022	0.058	0.000038856	0.000082441	0.000371043	−4.084	41.97	1.67	1.0	18.00
7	0.057	0.052	0.022	0.052	0.000037285	0.000070014	0.000301029	−4.155	42.45	1.66	1.0	17.00
8	0.052	0.047	0.022	0.047	0.000035902	0.000055912	0.000245118	−4.252	46.07	1.54	1.0	16.00
9	0.048	0.043	0.021	0.043	0.000034876	0.000042684	0.000202434	−4.370	52.73	1.35	1.0	15.00
10	0.043	0.038	0.020	0.038	0.000033489	0.000039686	0.000162748	−4.401	49.15	1.46	1.0	14.00
11	0.040	0.035	0.020	0.035	0.000032747	0.000027574	0.000135174	−4.560	61.12	1.14	1.0	13.00
12	0.037	0.032	0.019	0.032	0.000031814	0.000024806	0.000110368	−4.605	60.33	1.17	1.0	12.00

changes in composition and in absorbance caused by the chemical equilibria under study. The data matrices used for the steps 4–8 of the procedure are the reduced data matrices for the number of factors (species) considered, and they therefore contain less noise than the actual experimental matrices.

2. In Table 2 the results of applying error in factor analysis and the F statistical ratio test as proposed by Malinowski [12,13] for the deduction of the number of principal components are given for the experimental data collected in matrix **A**. Values of the F ratio test show that three independent contributions are present in the data matrix. In Fig. 4A the plot of the PRESS function (prediction error sum of squares in cross-validation) as proposed by Wold [15], and in Fig. 4B that of the IND function proposed by Malinowski [10] are given for the analysis of the data matrix **A**. Fig. 4C and D presents the plots of the same functions in the study of data matrix **B**. These plots indicate the presence of three components for the data matrix **A** and of four components for the data matrix **B**. The difference observed in the number of components in the two sets of data is caused by the broader pH range studied in the second case. As it is assumed that the number of independent components of the data variance is caused by the presence of an equal number of independently absorbing species, this is a first relevant result obtained from the factor analysis techniques as applied to the complexation by macromolecular ligands.

3. Fig. 5 displays the abstract distribution plots obtained by submatrix analysis [16] and evolving factor analysis [17,18] when applied to the two sets of data. These plots are very useful because they provide: (a) the number of significant species, deduced directly from the evolution of the factors related to the larger systematic variations in the data matrix, which clearly differentiate from those factors related to the smaller or random variations; (b) a first estimation of the pH range where factors contribute significantly to data variation, related to the pH range where the concentration of absorbing species varies; and (c) a first estimation of the concentration profiles of the species which contribute significantly to the data variation, which is obtained by drawing the lines between the log of the appropriate eigenvalues obtained at each pH (see thick lines in plots).

4. The next step in the proposed procedure is to obtain the combination of concentration profiles and species spectra which best explain the data matrix. This is carried out by starting with the concentration profiles initially estimated by evolving factor analysis, then improving them and the related species spectra by an alternating least-squares curve fitting method. In the case where the concentration profiles and the species spectra of the different species are different (not linearly dependent), convergence is achieved in a few iterations. However, it is found that when several titrations at different concentrations of the metal ion and of the macromolecule are analyzed independently, the species spectra obtained have the same shape and spectral features, but a different intensity, which depends on the ratio of the ligand site to Cu(II) ion concentration in the titrated solution. This is a result which can be interpreted, on one hand, as if the number of ligand sites coordinated per central metal ion is higher than one, and/or, on the other hand, as if the intensity of the species spectra and the concentration of the related species are exchangable to a certain extent for a similar fit of the experimental data. To remove this ambiguity, the different experimental data matrices of spectra obtained at different ligand site to metal concentration ratios should be analyzed together, and the analysis of this higher dimensional data (four dimensions: pH, wavelength, absorbance and ligand site to metal concentration ratio) should be carried out directly using tensorial resolution [20,21], or indirectly using target factor analysis. In this work the latter, simpler approach is adopted, and the former left for further investigation.

5. The three first species spectra obtained in the analysis of the data obtained in excess of ligand sites concentration (matrix **B**) are used as targets for the data set obtained at lower concentration of ligand sites (matrix **A**). As the fourth species found in the analysis of the data of matrix **B** is not detected in the titration data of matrix **A**, since the pH range of study is narrower in this

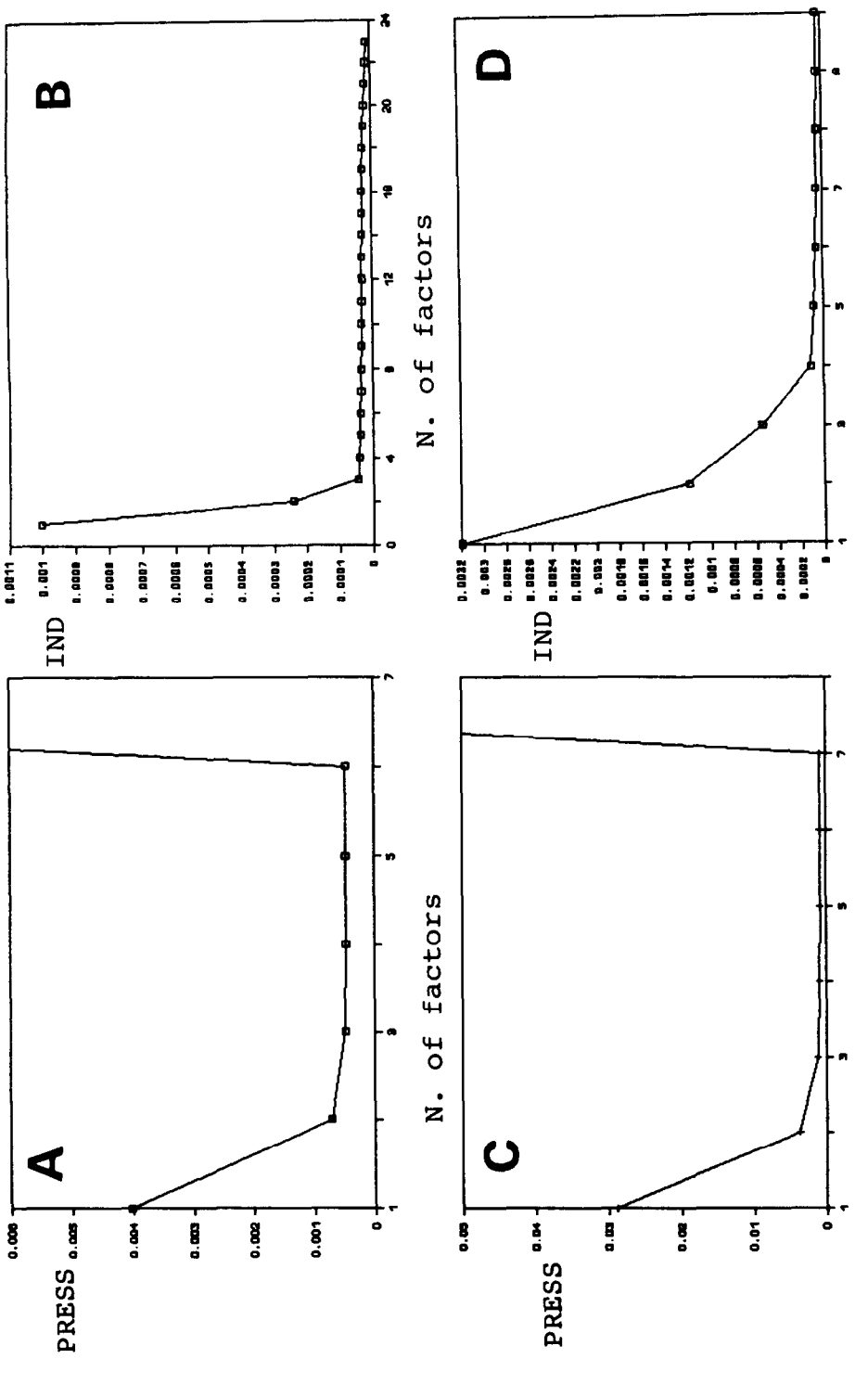

Fig. 4. Determination of the number of components: (A) Plot of the PRESS function versus the number of factors for data matrix A; (B) plot of the IND function for matrix A; (C) plot of the PRESS function versus the number of factors for matrix B; (D) plot of the IND function for matrix B.

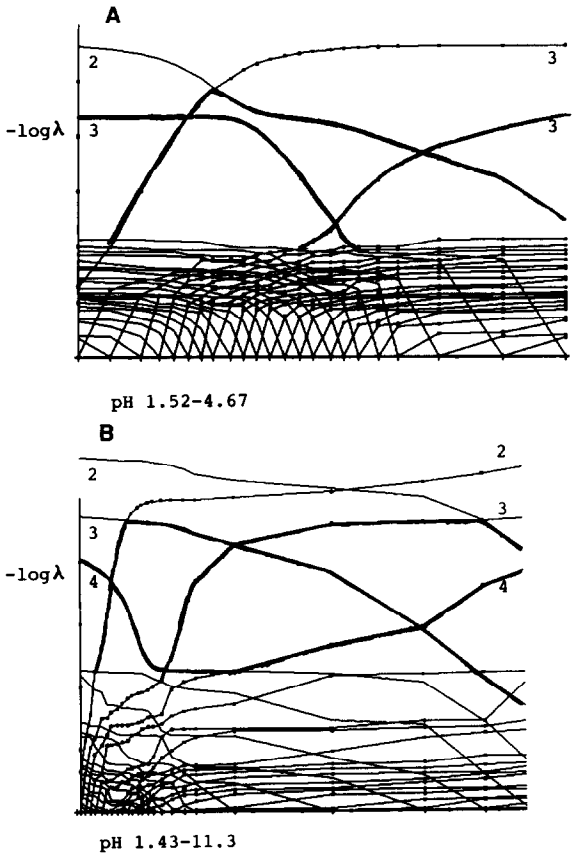

Fig. 5. Abstract distribution plots obtained by submatrix and evolving factor analysis for (A) data matrix **A** and (B) data matrix **B**. Thick lines are the initial estimated concentration profiles of the different species. Numbers on the left and on the right give the factor line obtained joining the upsurging factors (right) and the decreasing factors (left). See refs. 6–8 and 16–19 for details concerning the way it was obtained.

case (see the Experimental section), this species was not included as a target. The projections of the three tested targets onto matrix **A** are given in Fig. 6. As can be seen in this figure, there is a good agreement between these targets and the estimated spectra, which shows that the latter can be used to explain simultaneously both sets of data (matrix **A** and **B**). Indeed, this is what is obtained from the analysis of the experimental data, whereas the reverse is not true: the species spectra obtained from the analysis of matrix **A** cannot reasonably explain the data of matrix **B** because the deduced species concentrations would be too high and greatly exceed the constraints of the total known amounts of metal concentration at each titration point. Consequently, in order to describe both sets of data with a unique set of species, the species spectra finally chosen are the species spectra obtained in the analysis of matrix **B** once they have been projected onto matrix **A**. The corresponding distribution plots are then obtained using least-squares. It is found that the standard deviation of the residuals between the factor analysis reduced data matrix and the calculated data matrix using the best set of species spectra and of species concentrations is below 0.01 absorbance units.

6. Rank annihilation by evolving factor analysis [8,19] has also been used to determine directly the concentration profiles associated with certain species spectra, giving similar results. In principle the advantage of using this method here is that it allows the independent evaluation of the concentration of each species once its spectrum is known, without the full model requirements of least-squares approaches. The proposed method should be used for those species whose concentration spans a part of the full data matrix and not to those species which span the complete data matrix, nor to those which do not contribute appreciably to data variance. In the examples studied the first and third species fulfill this requirement. As an example, when the third species spectrum

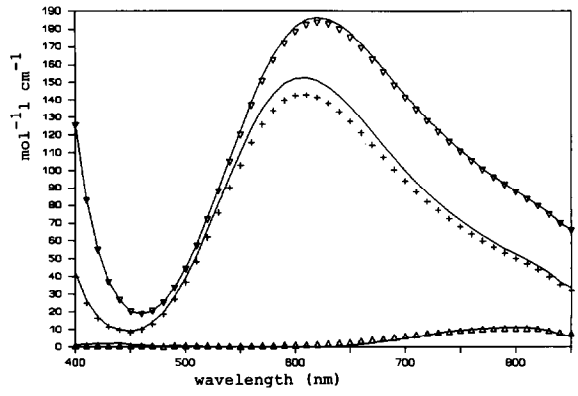

Fig. 6. Target factor analysis applied to species spectra obtained in the analysis of the two titrations. Continuous lines refer to the estimated species spectra for data matrix **A** using target factor analysis. Symbols refer to spectra obtained in the SPFAC least-squares analysis of matrix **B** (targets).

is considered as known, and its concentration is estimated using rank annihilation (the rank of the data matrix diminished by one), the distribution plot obtained for this species is similar to that obtained using the full model least-squares approach (see Fig. 7).

7. The distribution plots finally obtained for the data matrices **A** and **B** are given in Fig. 8. In the first case, the concentration of free Cu(II) is still high at pH 5, in agreement with the fact that precipitation of copper hydroxide commences up at this point. Conversely, in the case of matrix **B**, copper ion is completely bonded to the macromolecule even at acidic pH, and virtually no free Cu(II) remains at pH 5.0, in agreement with the fact that no precipitation is observed. This is also proof that the complexation involves several ligand sites from the commencement of the titration at acid pH. In the case where the ratio of ligand site to metal ion is approximately one, not enough ligand sites are available to chelate all the metal ion present in solution (more than one binding site coordinating to the same metal ion) and a considerable amount of copper remains uncomplexed. Conversely, in the second titration, where the ratio of concentrations of ligand sites to metal ion is about four, all the metal ion is already completely bonded to the macromolecule and no precipitation is observed.

Another source of valuable chemical information comes from the species spectra resolved in

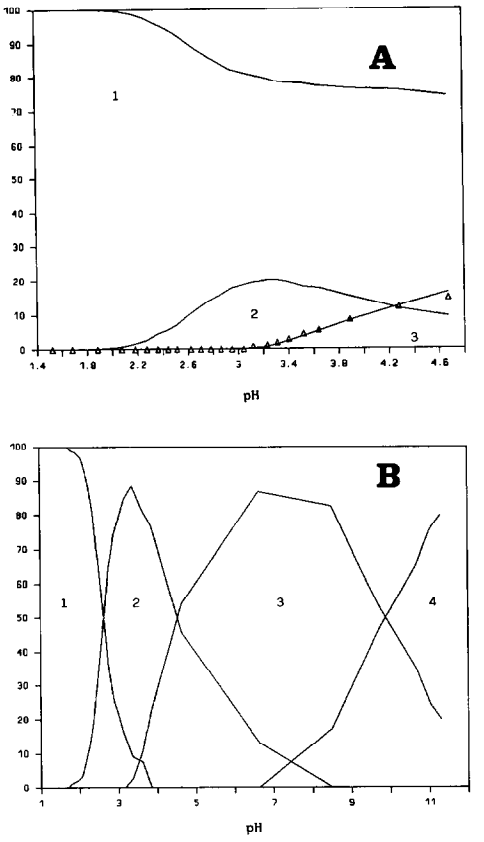

Fig. 8. Species distribution obtained in the treatment of (A) data matrix **A** and of (B) data matrix **B**. (1) Free Cu(II) species; (2–4) successive complex species between Cu(II) and PEI. The distribution of the species labelled 3 estimated using rank annihilation for matrix **A** is given by symbols (△).

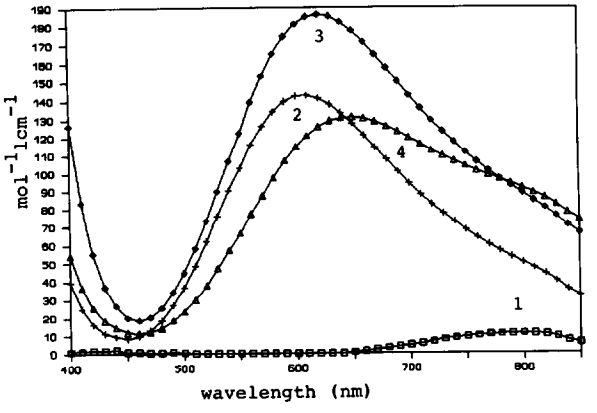

Fig. 7. Species spectra finally obtained in the analysis of the two sets of data. (1) Cu(II) estimated species spectra; (2–4) successive complex species between Cu(II) and PEI.

the treatment of the data (Fig. 7). The wavelength of the absorbance band maximum of the first complex is aproximately 600 nm. When this value is compared with the values observed for other known copper–ammine complexes [22] and for the free aquo–metal ion, it is deduced that in this first complex species three or four N ligand sites of the macromolecule should be bonded to the metal ion. The shifts observed for the higher complexes to longer wavelengths reflect the fact that coordination numbers of five or even higher are attained [23], including the possibility of hydroxocomplex formation.

As was mentioned above, the evaluation of the concentrations of all the absorbing species leads

to a knowledge of the chemical composition of the system, but, furthermore, it allows the complexing properties of polyethyleneimine to be deduced. As an example, in Table 1 a brief description of these properties is given for the results obtained from the analysis of the data matrix **B** up to pH 3.85, where the metal ion is already nearly completely complexed (96%). From the results of a simultaneous investigation of the protonation of polyethyleneimine [24], the number of protonated sites at each pH can be roughly estimated. If it is assumed that each metal ion is coordinated to four sites of the macromolecule (i.e. if the amount of ligand bonded to the metal ion is calculated to be equal to four times the amount of metal bonded) the apparent stability constants or concentration quotients at equilibrium I–III can be calculated. The values obtained are collected in Table 1. Equilibrium III describes the competitive reaction of copper and hydrogen ions for the sites of the macromolecule.

Equilibrium I $\quad Cu + 4L^* = CuL_4^*$

$$K1 = [CuL_4^*]/[Cu][L^*]^4$$

Equilibrium II $\quad Cu + 4L' = CuL_4'$

$$K2 = [CuL_4']/[Cu][L']^4$$

Equilibrium III $\quad Cu + 4LH = CuL_4 + 4H$

$$K3 = [CuL_4][H]^4/[Cu][LH]^4$$

where L, L* and L′ are the free ligand sites, the non-complexed ligand sites and the non-complexed nor non-protonated ligand sites, respectively; LH are the protonated ligand sites.

In Table 1 the evolution of these apparent stability constants with pH is given. These quotients are significant only up to pH 3.0. Above this pH the contribution of other species with higher coordination numbers was found to be important (see Fig. 8), and the related equilibrium constants should consider them. Although they can also be studied from the results obtained in the data treatment with SPFAC, since the concentrations of such species are also obtained, they are not studied here and work is currently being pursued in this direction. As can be seen from the values of the apparent stability constants given in Table 1, the dependence of these values on the degree of ligand site occupation and on pH shows the extent of the secondary effects, mentioned in the introduction, which are present in the complexation by macromolecules. A knowledge of these dependences opens the door to a deeper study of such effects in the complexation of metal ions by macromolecules, and by polyethyleneimine in particular.

CONCLUSIONS

The application of several techniques based on factor analysis, and especially on evolving factor analysis, allows the investigation of complexation between Cu(II) and polyethyleneimine. The proposed method allows the deduction of the number of independent macromolecular metal complex species present during an acid–base titration of solutions containing Cu(II) and polyethyleneimine, and also allows the estimation of the concentration profiles and spectra of the species formed during the titration. Three or four different species were detected, depending on the pH range and on the excess of ligand site concentration. The ambiguity found in the self-resolving treatment of the data between the intensity of the species spectra and the concentration of the related species was solved by simultaneous target factor analysis of different titrations obtained over a broad range of concentration ratios of ligand site to metal ions. The values of the absorbance maxima, and also the values of the concentration quotients describing the equilibria of formation of the different species, show that the complexation process involves chelate formation through several N atoms from the commencement of the process; four ammine groups of polyethyleneimine interact from the start with the central metal ion (first detected species) and higher coordination is present in higher complexes. An estimate of the values of the concentration quotients of the different equilibria present in the system is obtained directly from the concentration profiles of the different species deduced at each pH value.

REFERENCES

1 J.A. Marinsky and J. Ephraim, Unified physicochemical description of the protonation and metal ion complexation equilibria of natural organic acids (humic and fulvic acids). 1. Analysis of the influence of polyelectrolyte properties on protonation equilibria in ionic media: fundamental concepts, *Environmental Science and Technology*, 20 (1986) 349–354.

2 H. Sigel (Editor), *Metal Ions in Biological Systems Series*, Marcel Dekker, New York, 1973 et seq.

3 J. Buffle, *Complexation in Aquatic Systems: An Analytical Approach*, Ellis Horwood, Chichester, 1988.

4 R. Barbucci, M. Casolaro, M. Nocentini, S. Corezzi, P. Ferruti and V. Barone, Acid–base and metal ion complex formation properties of polymers containing amino acid residues, *Macromolecules*, 19 (1986) 37–42.

5 J.A. Marinsky, Ion binding in charged polymers, *Coordination Chemistry Reviews*, 19 (1976) 125–171.

6 R. Tauler and E. Casassas, Principal component analysis applied to the study of successive complex formation data in Cu(II)–ethanolamine systems, *Journal of Chemometrics*, 3 (1988) 151–161.

7 R. Tauler and E. Casassas, Application of principal component analysis to the study of multiple equilibria systems. Study of copper(II)/salicylate/mono-, di- and triethanolamine systems, *Analytica Chimica Acta*, 223 (1989) 257–268.

8 R. Tauler, E. Casassas and A. Izquierdo-Ridorsa, Self-modelling curve resolution in studies of spectrometric titrations of multi-equilibria systems by factor analysis, *Analytica Chimica Acta*, 248 (1991) 447–458.

9 T.W. Johnson and I.M. Klotz, Fluorine nuclear magnetic resonance study of the conformation of a polyethylenimine derivative, *Journal of Physical Chemistry*, 75 (1971) 4061–4064.

10 E.R. Malinowski and D.E. Howery, *Factor Analysis in Chemistry*, Wiley, New York, 1980.

11 E.R. Malinowski, Theory of error in factor analysis, *Analytical Chemistry*, 49 (1977) 606–612; Determination of the number of factors and the experimental error in data matrix, *Analytical Chemistry*, 49 (1977) 612–617.

12 E.R. Malinowski, Theory of the distribution of error eigenvalues resulting from principal component analysis with applications to spectroscopic data, *Journal of Chemometrics*, 1 (1987) 33–40.

13 E.R. Malinowski, Statistical *F*-test for abstract factor analysis and target testing, *Journal of Chemometrics*, 3 (1988) 39–60.

14 E.R. Malinowski, Theory of error for target factor analysis with applications to mass spectrometry and nuclear magnetic resonance spectrometry, *Analytica Chimica Acta*, 103 (1978) 339–354.

15 S. Wold, Cross validatory estimation of the number of components in factor and principal component models, *Technometrics*, 20 (1978) 397–405.

16 H. Cartwright, Determination of the dimensionality of spectroscopic data by submatrix analysis, *Journal of Chemometrics*, 1 (1987) 111–120.

17 H. Gampp, M. Maeder, C.J. Meyer and A.D. Zuberbühler, Calculation of equilibrium constants from multiwavelength spectroscopic data. III. Model free analysis of spectrophotometric and ESR titrations, *Talanta*, 32 (1985) 1133–1139.

18 H. Gampp, M. Maeder, C.J. Meyer and A.D. Zuberbühler, Calculation of equilibrium constants from multiwavelength spectroscopic data. IV. Model free least-squares refinement by use of evolving factor analysis, *Talanta*, 33 (1986) 943–951.

19 H. Gampp, M. Maeder, C.J. Meyer and A.D. Zuberbühler, Quantification of a known component in an unknown mixture, *Analytica Chimica Acta*, 193 (1987) 287–293.

20 E. Sanchez and B.R. Kowalski, Tensorial resolution: A direct trilinear decomposition, *Journal of Chemometrics*, 4 (1990) 29–45.

21 B.R. Kowalski and M.B. Seasholtz, Recent developments in multivariate calibration, *Journal of Chemometrics*, 5 (1991) 129–145.

22 E.J. Billo, Copper(II) chromophores and the rule of average environment, *Inorganic Nuclear Chemistry Letters*, 10 (1974) 613–617.

23 J. Bjerrum, C.J. Ballhausen and C. K. Jorgensen, Studies on absorption spectra, *Acta Chemica Scandinavica*, 8 (1954) 1275–1289.

24 R. Tauler and J. Barceló, unpublished results.