Anal. Chem. **2003,** *75,* 641−647

Application of a Combination of Hard and Soft Modeling for Equilibrium Systems to the Quantitative Analysis of pH-Modulated Mixture Samples

Josef Diewok,[†] Anna de Juan,[‡] Marcel Maeder,[§] Romà Tauler,[‡] and Bernhard Lendl*,[†]

Institute of Chemical Technologies and Analytics, Vienna University of Technology, Getreidemarkt 9/164-AC, A-1060 Vienna, Austria, Department of Analytical Chemistry, University of Barcelona, Diagonal 647, E-08028 Barcelona, Spain, and Department of Chemistry, University of Newcastle, Callaghan NSW 2308, Australia

pH modulation of aqueous mixture samples combined with FT-IR detection and a powerful second-order resolution method is proposed for both resolution and quantitation of acid analytes in the presence of similarly behaving interferences. The proposed method allows for the analyte determination in mixtures using a single standard sample per analyte. Due to the very similar pK_a values of the investigated analytes and interferences, the highly correlated concentration profiles of these compounds cannot be successfully resolved with pure soft-modeling second-order approaches. The inclusion of a hard-modeling constraint based on the acid−base equilibrium model in the soft-modeling curve resolution method has allowed for the unambiguous resolution of the analyte profiles and, as a consequence, for the correct quantitation of this compound in the mixture sample. A detailed discussion of the combined hard−soft-modeling approach as well as the analytical problem and the quantitation results is given. Also, strategies to overcome problems associated with variation in pK_a values between different samples are addressed. Due to the flexible implementation of the hard-model equilibrium constraint in the multivariate curve resolution-alternating least squares method, this approach is expected to be useful also for analysis of other complex mixed equilibrium-based chemical systems.

Spectroscopy has proven to be an indispensable tool in the analysis of solution multicomponent samples. In contrast to univariate methods, in which a separation step is necessary for the analysis of mixtures, spectroscopy is by principle a multivariate method, which allows for the analysis and quantification of analytes in complex matrixes without requiring previous separation. Mid-IR spectroscopy in aqueous solution[1] offers richer spectral information than other techniques, such as UV−visible or near-IR, but is limited by the strong absorption of water in its working wavenumber range. However, this problem can be overcome by the use of appropriate measuring techniques such as small-path length transmission cells or attenuated total reflection (ATR) methods.[2]

In most multicomponent analyses, no selective wavenumbers exist for the analytes of interest as the spectral features of the sample constituents overlap. Therefore, multivariate calibration methods[3,4] such as partial least squares (PLS) or principal component regression (PCR) have to be applied to correlate the measured spectra to known concentration values and to predict analyte concentrations in new samples. A drawback of PLS and related methods is that a large number of calibration samples is necessary, as all possible analytes and interferences have to be included in the calibration set in suitable concentration levels to obtain a robust regression model.[5]

When the sample is subjected to certain kinds of modulation and a spectral data matrix is recorded per sample instead of a single spectrum, the situation changes. On one hand, the sample modulation, e.g., acid−base titration in the case of pH-dependent analytes, yields additional and more discriminant spectral information; on the other hand, getting a spectral data matrix per mixture sample allows for the quantification of analytes in the presence of unknown and uncalibrated interferences. The latter feature is known as the second-order advantage[6] and can in principle be achieved by simultaneous analysis of only one pure analyte calibration standard with the mixture sample of interest. Chemometric methods dedicated to the analysis of this kind of data are called "second-order calibration" methods. Examples are generalized rank annihilation (GRAM),[7] parallel factor analysis (PARAFAC),[8] Tucker and second-order curve resolution (CR) methods.[9]

Multivariate curve resolution-alternating least squares (MCR-ALS)[10] has already been applied to very diverse second-order calibration problems, as for example, series of titrations,[11,12]

* Corresponding author. Fax: +43 (0)1 58801 15199. E-mail: blendl@mail.zserv.tuwien.ac.at.

[†] Vienna University of Technology.

[‡] University of Barcelona.

[§] University of Newcastle.

(1) Vonach, R.; Lendl, B.; Kellner, R. *Analyst* **1997**, *122*, 525−530.

(2) *Handbook of Vibrational Spectroscopy*; Chalmer, J. M., Griffith, P. R., Eds.; John Wiley & Sons: Chichester, U.K., 2000; Vol. 2 (Sampling Techniques).

(3) Naes, T.; Martens, H. *TrAC, Trends Anal. Chem.* **1984**, *3*, 266−271.

(4) Malinowski, E. R. *Factor Analysis in Chemistry*, 3rd ed.; John Wiley & Sons: New York, 2002.

(5) Thomas, E. V. *Anal. Chem.* **1994**, *66*, 795A−804A.

(6) Booksh, K. S.; Kowalski, B. R. *Anal. Chem.* **1994**, *66*, 782A−791A.

10.1021/ac026248j CCC: $25.00
Published on Web 12/24/2002

chromatographic runs,[13] or kinetic data.[14] It is a flexible method that takes advantage of known chemical and mathematical information about the data set through the use of constraints and allows for the simultaneous analysis of several data matrices without strong requirements related to the mathematical data structure.

In a previous work,[12] we reported the successful quantitation of mixtures of diprotic organic acids by Fourier transform infrared (FT-IR) flow titration and MCR-ALS. Quantitation of mixtures of one or two pH-dependent analytes in the presence of an inert unknown interfering agent was successful using only one standard per analyte in the MCR-ALS calibration. However, when the interference was a pH-evolving compound with a behavior very similar to the analyte, i.e., very close pK_a values, correct resolution and quantitation of the analytes were not possible any more. The reasons for the failure of the pure soft-modeling MCR-ALS approach are the very similar pK_a values of the acids and the resulting highly correlated concentration profiles among the different acid species. In such a situation, more restrictive constraints are required in order to obtain correct MCR-ALS resolution and quantitation.

In the present work, we describe a novel combination of hard and soft modeling for successful quantitative analysis of the complex mixture titration data of pH-dependent analytes and interferences monitored by FT-IR. A hard equilibrium model is implemented as an additional and optional constraint in the soft MCR-ALS algorithm. The analyte concentration profiles are nonlinearly fitted[15] according to an acid−base equilibrium model, where the pK_as and total concentration of the system are the parameters to be optionally modified. Comparable examples of combined hard and soft modeling have been reported for the analysis of kinetic data.[16−18]

Another novelty of this work is that the hard-modeling constraint is not focused on the recovery of the physicochemical model, i.e., obtaining the equilibrium constants, but on the improvement of the analyte quantitation. Thus, on one hand, the known physicochemical model of the analyte guarantees the correct shape of the concentration profiles of the different analyte acid species; on the other hand, the hard model yields the analytical information of interest, i.e., the total concentration of the analyte in the mixture sample, as one of the fitted parameters. This work is not only a new example of the combination of a mixed hard- and soft-modeling approach but, mainly, a way to prove that this strategy clearly surpasses the pure physicochemical interest and can have a much wider analytical application. Thus, a hard-modeling constraint should not be seen as a manner to exclusively unravel a reaction pathway from a complex system, but as a way to significantly improve the resolution and quantitation of identified compounds (analytes) in complex samples through the inclusion of their known behavior model.

(7) Sanchez, E.; Ramos, L. S.; Kowalski, B. R. *J. Chromatogr.* **1987**, *385*, 151−164.
(8) Harshman, R. A. *UCLA Work. Pap. Phonetics* **1970**, *16*, 1−84.
(9) de Juan, A.; Casassas, E.; Tauler, R. Soft-Modeling of Analytical Data. In *Encyclopedia of Analytical Chemistry*, 1st ed.; Meyers, R. A., Ed.; John Wiley & Sons: New York, 2000; Vol. 11, pp 9800−9837.
(10) Tauler, R. *Chemom. Intell. Lab. Syst.* **1995**, *30*, 133−146.
(11) Saurina, J.; Hernandez-Cassou, S.; Tauler, R.; Izquierdo-Ridorsa, A. *Anal. Chem.* **1999**, *71*, 126−134.
(12) Diewok, J.; de Juan, A.; Tauler, R.; Lendl, B. *Appl. Spectrosc.* **2002**, *56*, 40−50.
(13) Gargallo, R.; Tauler, R.; Cuesta-Sanchez, F.; Massart, D. L. *TrAC, Trends Anal. Chem.* **1996**, *15*, 279−286.
(14) Saurina, J.; Hernandez-Cassou, S.; Tauler, R. *Anal. Chem.* **1997**, *69*, 2329−2336.
(15) Maeder, M.; Zuberbuehler, A. D. *Anal. Chem.* **1990**, *62*, 2220−2224.
(16) de Juan, A.; Maeder, M.; Martinez, M.; Tauler, R. *Chemom. Intell. Lab. Syst.* **2000**, *54*, 123−141.
(17) Bijlsma, S.; Smilde, A. K. *J. Chemom.* **2000**, *14*, 541−560.
(18) Bezemer, E.; Rutan, S. C. *Chem. Intell. Lab. Syst.* **2001**, *59*, 19−31.

EXPERIMENTAL SECTION

Reagents. All chemicals were of analytical reagent grade. L-(+)-Tartaric acid and a 2 N NaOH solution were purchased from Riedel-de Haën and DL-malic acid was from Aldrich. Sucrose and lactic acid were obtained from Merck and HCl (37%) from Fluka. Deionized water was used in all solutions. Samples were prepared by dissolving the appropriate amount of organic acids or sucrose in ∼200 mL of water and adjusting to a pH of 11 with NaOH solution for complete deprotonation of the acids. Then, the samples were filled with water to a volume of 250 mL in a graduated flask and transferred into the titration beaker.

Instrumentation. A SenTix 61 pH electrode (working range pH 0−14) and a pH-meter 320 (both WTW GmbH) were used to measure the pH changes during the FT-IR titrations. A peristaltic pump Minipuls 3 (Gilson) with a flow rate of 2.05 mL/min was used for the on-line sample transfer from the titration beaker to the transmission cell in the spectrometer and back to the titration beaker. The IR transmission cell was built of 2-mm-thick CaF_2 windows with a 25-μm PTFE spacer. The cell was connected to the peristaltic pump tubing and the outlet tubing with standard FIA tubing and fittings, respectively. The tubing length was kept as short as possible.

All experiments were monitored with a Bruker IFS 66 FT-IR spectrometer (Bruker Optik GmbH) with a mercury−cadmium−telluride (MCT) detector. To increase the signal-to-noise (S/N) ratio in the spectral region of interest (1600−1000 cm^{-1}) a low-pass filter with a 5% cut on at 1900 cm^{-1} and an aperture of 6 mm were used. Each spectrum was the result of 128 coadded scans, recorded with 8-cm^{-1} resolution, Blackman−Harris three-term apodization, and a scanner velocity of 100-kHz HeNe modulation frequency.

Titrations and Spectra Acquisition. A spectrum of a flow of neutral water was taken as a reference spectrum immediately before each titration experiment. Then, the sample was magnetically stirred and titrated in the pH range from 11 to 1 with 3 and 6 N HCl. HCl was used as concentrated as possible to keep sample dilution effects negligible. Sample spectrum recording was triggered manually approximately every 0.2 pH unit when the wanted pH value was reached. This led to ∼25−35 spectra/sample at defined pH values and a typical titration time of 30 min. All spectra were recorded with the OPUS IR software. The obtained data were converted into ASCII files by use of an OPUS macro for further data analysis in Matlab 5.3.

DATA TREATMENT

Data Pretreatment. Major baseline changes were observed in the original FT-IR spectra during the titrations that stem from changes in water association with varying H_3O^+ and salt concentration.[12] These baseline contributions included shifts and complex curvatures and their pH dependency was highly correlated with the concentration profiles of the acid species. Therefore, they

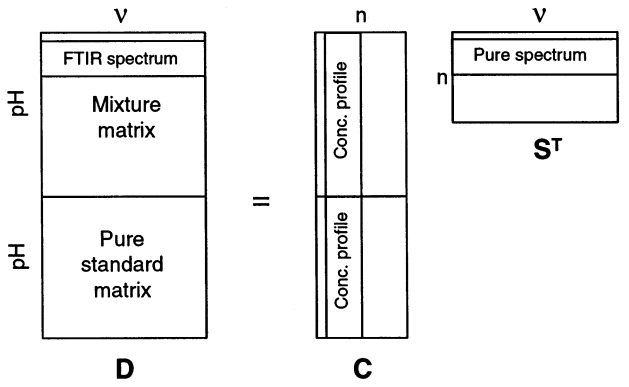

Figure 1. Column-wise matrix augmentation for simultaneous MCR-ALS analysis of mixture and standard sample in second-order calibration.

could not be corrected for with conventional baseline correction methods. The pretreatment that proved to completely eliminate the baseline contributions in the 1540−1000-cm^{-1} range and allow quantitative analysis of the titration data sets was calculation of second-derivative spectra.

For all data analysis, MATLAB 5.3 was used.[19] Second-derivative spectra were calculated according to the method of Savitzky and Golay.[20] For MCR-ALS, an equilibrium hard-model adaptation of the algorithm implemented by the authors as MATLAB code was applied.[21]

MCR-ALS. MCR-ALS[10] is a flexible, iterative curve resolution method for decomposing a mixture data matrix **D** into the pure contributions of all (significant) components in the system along the two measurement directions:

$$\mathbf{D} = \mathbf{C}\mathbf{S}^{\mathrm{T}} + \mathbf{E} \tag{1}$$

In this work, **D** is the spectral data matrix containing rows of FT-IR spectra recorded during a titration experiment. The columns of the **C** matrix are the pure pH-dependent concentration profiles of the modeled components and the rows in the $\mathbf{S}^{\mathrm{T}}$ matrix are their related pure FT-IR spectra. **E** is the matrix of residuals not explained by the MCR-ALS model.

For the purpose of second-order calibration and prediction, two or more matrices from different experiments, e.g., mixture and standard samples, should be analyzed simultaneously (Figure 1). The application of MCR-ALS to quantitative analysis of FT-IR titration data has been explained in detail in the previous work[12] and will only be described briefly.

Rank Analysis. The number of components to be modeled with MCR-ALS is determined by singular value decomposition (SVD).[22] Chemical components give rise to bigger singular values than noise or minor instrument contributions. Therefore, the chemical rank can be estimated by the number of singular values larger than singular values associated with noise.

In data sets coming from one titration with more than one closed equilibrium system (e.g., two different acids) or with an inert interfering agent (sugar) and closed equilibrium systems simultaneously present, the chemical rank is unavoidably underestimated; i.e., the number of significant components estimated by SVD is lower than the true number of spectroscopically absorbing species. This phenomenon is known as rank deficiency[23,24] and completely hinders the correct resolution of the respective data matrix. However, rank deficiency can be removed by matrix augmentation, i.e., by simultaneous MCR-ALS analysis of the respective matrix together with one or more additional full-rank standard matrices. The issues of rank deficiency and rank augmentation by matrix augmentation have been addressed in detail in our previous work[12] for FT-IR titration data of mixtures of organic acids and sugar.

Initial Estimates. MCR-ALS requires initial estimates of either spectral (S-type) or concentration profile type (C-type) for all modeled components. S-Type estimates can, for example, be taken from mixture spectra. C-Type estimates can be obtained from evolving factor analysis (EFA)[25] or can be calculated for analyte profiles according to the suitable equilibrium model.

Constrained Alternating Least-Squares Optimization. Equation 1 is solved in an alternating least-squares (ALS) optimization by minimizing the residual matrix **E**. The two alternating steps in the iterative optimization are

$$\mathbf{S}^{\mathrm{T}} = (\mathbf{C})^{+}\mathbf{D}^{*} \tag{2}$$

$$\mathbf{C} = \mathbf{D}^{*}(\mathbf{S}^{\mathrm{T}})^{+} \tag{3}$$

where $(\mathbf{C})^{+}$ and $(\mathbf{S}^{\mathrm{T}})^{+}$ are pseudoinverses of **C** and $\mathbf{S}^{\mathrm{T}}$. $\mathbf{D}^{*}$ is the PCA-reproduced data matrix for the number of modeled components. To minimize the ambiguity of the solution,[26] the iterative calculation of **C** and $\mathbf{S}^{\mathrm{T}}$ is always subject to constraints. These constraints are based on chemical and mathematical properties of the data and can be individually applied to profiles in **C**, $\mathbf{S}^{\mathrm{T}}$, or both and, in data sets formed by several matrices, to submatrices of **C**, $\mathbf{S}^{\mathrm{T}}$, or both. Constraints applicable to FT-IR titration data include nonnegativity, selectivity, local rank, unimodality, and closure for concentration profiles. Nonnegativity of spectra cannot be used because of the second-derivative pretreatment.

The ALS optimization is stopped when the relative difference in lack of fit (LOF) between consecutive iterations is below a threshold value. The lack of fit (LOF) is defined as

$$\mathrm{LOF}\ (\%) = 100 \times \sqrt{\frac{\sum_{\mathrm{i,j}} e_{\mathrm{ij}}^{2}}{\sum_{\mathrm{i,j}} d_{\mathrm{ij}}^{2}}} \tag{4}$$

where d_{ij} is the original element in the data set and e_{ij} is its related residual.

(19) MATLAB 5.3, The Maths Work Inc., Natick, MA, 1999.
(20) Savitzky, A.; Golay, M. J. E. *Anal. Chem.* **1964**, *36*, 1627−1639.
(21) Tauler, R.; de Juan, A. *Multivariate Curve Resolution*−Alternating Least Squares (MCR-ALS), MATLAB code; University of Barcelona, Barcelona, Spain, 1999; http://www.ub.es/gesq/mcr/mcr.htm.
(22) Golub, G. H.; Reinsch, C. *Numer. Math.* **1970**, *14*, 403−420.
(23) Amrhein, M.; Srinivasan, B.; Bonvin, D.; Schumacher, M. M. *Chemom. Intell. Lab. Syst.* **1996**, *33*, 17−33.
(24) Izquierdo-Ridorsa, A.; Saurina, J.; Hernandez-Cassou, S.; Tauler, R. *Chemom. Intell. Lab. Syst.* **1997**, *38*, 183−196.
(25) Gampp, H.; Maeder, M.; Meyer, C. J.; Zuberbuehler, A. D. *Talanta* **1985**, *32*, 1133−1139.
(26) Tauler, R.; Smilde, A.; Kowalski, B. *J. Chemom.* **1995**, *9*, 31−58.

Equilibrium Constraint. As mentioned in previous works related to kinetic examples,[16,17] a hard-model constraint selects the concentration profiles in the **C** matrix involved in the known reaction process as input for a nonlinear hard-modeling fit. The selected profiles in the matrix **C** (or in the suitable submatrix of **C** in the case of second-order data) are fitted to a kinetic model in every iteration of the MCR-ALS. The fitted profiles update the input profiles in every iteration, and the kinetic parameters optimized in the nonlinear fit are obtained as additional information.

In this work, we introduce for the first time a hard-modeled *equilibrium* constraint into a CR algorithm. The underlying physicochemical model describes the acid−base equilibria of simple molecules. In a general manner, the concentration profiles of the species involved in the acid−base equilibrium of an *n*-protic acid can be described as follows:

$$[H_nA] = c_A \frac{[H^+]^n}{[H^+]^n + K_1[H^+]^{n-1} + K_1K_2[H^+]^{n-2} + \cdots + K_1K_2 \cdots K_n}$$

$$[H_{n-1}A^-] = c_A \frac{K_1[H^+]^{n-1}}{[H^+]^n + K_1[H^+]^{n-1} + K_1K_2[H^+]^{n-2} + \cdots + K_1K_2 \cdots K_n}$$

$$\vdots$$

$$[A^{n-}] = c_A \frac{K_1K_2 \cdots K_n}{[H^+]^n + K_1[H^+]^{n-1} + K_1K_2[H^+]^{n-2} + \cdots + K_1K_2 \cdots K_n} \quad (5)$$

As can be seen, the concentration profiles of all involved species are expressed as a function of two kinds of parameters, the total concentration of the acid−base system, c_A, related to the analytical information sought, and the suitable acidity constants, K_i (also denoted by the logarithmic pK_i), related to the physicochemical behavior of the compound. Those are the parameters that may be modified in the nonlinear fitting. The algorithm used is an adapted Newton−Gauss−Marquardt algorithm, as described by Maeder and Zuberbuehler,[15] and it allows for the refinement of some or all the parameters in the model.

Given the flexibility of both the MCR-ALS algorithm and the nonlinear fitting algorithm integrated in it, the possibilities to apply the hard-modeling equilibrium constraint to a chemical system are extremely diverse. Thus, (1) all or some of the concentration profiles in **C** can be included in the nonlinear fitting. (2) The different submatrices in **C** (related to different titrations) can be treated differently: (a) The number of acid−base systems to be fitted can vary (e.g., submatrix of mixture of analytes and submatrices of pure standards). (b) The nature of the acid−base systems to be fitted can vary (e.g., different combinations of analytes in the different titrations). (c) The selection of the parameters to be fixed or loose in the nonlinear fit may vary (e.g., in a pure standard matrix, the total concentration of analyte is known and, therefore, fixed, whereas it is left loose in a sample matrix).

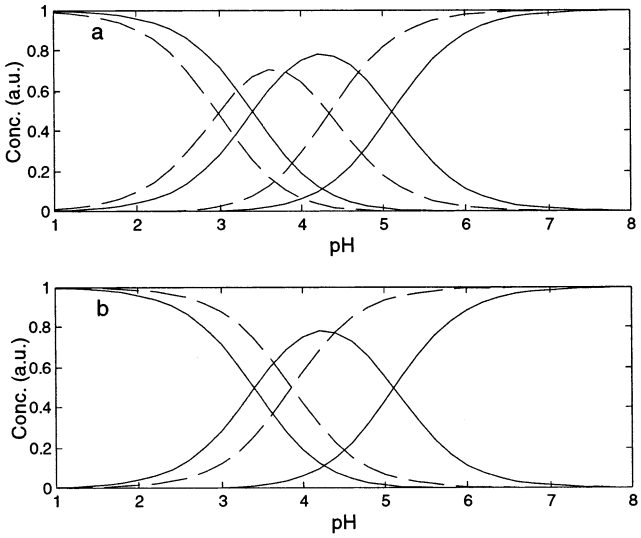

Figure 2. Concentration profiles of acid species with pH modulation, calculated from literature pK_a values. Both plots depict mixture systems that correspond to experimental samples of this work: (a) malic acid (−), tartaric acid (- -); (b) malic acid (−), lactic acid (- -).

In the example of quantitation of analytes in the presence of interferences, the hard-modeling constraint applies only to the concentration profiles of the analytes, the only compounds of known identity and formally known physicochemical model. The concrete application of this constraint to the real examples in this work will be explained in Results and Discussion.

CHEMICAL SYSTEMS AND DATA SETS

Two diprotic organic acids, L(+)-tartaric and DL-malic acid, were studied as analytes in this work. These two acids, the monoprotic lactic acid, and the pH-inert sucrose were also used as interfering substances in the mixture samples. Mixture and standard samples were titrated in the pH range 11−1, and FT-IR spectra were recorded at defined pH values in steps of ∼0.2 pH unit. In the course of the titration, the diprotic acids are characterized by three acid species, namely, the fully deprotonated A^{2-}, the intermediate HA^-, and the fully protonated H_2A. The pH dependency of the H_2A, HA^-, and A^{2-} species is determined by the equilibrium constants pK_{a1} and pK_{a2} for the corresponding protonation steps. H_2M, H_2T, and HL will be used throughout the article for malic, tartaric, and lactic acid species, respectively. The literature pK_a values[27] are 3.40 and 5.11 for DL-malic acid, 2.98 and 4.34 for L(+)-tartaric acid, and 3.85 for lactic acid. Concentration profiles for the species of the different acids are calculated with these pK_a values and plotted in Figure 2. It can be seen that some profiles, e.g., H_2M and H_2T, HM^- and HT^-, or H_2M and HL, are very similar and separated by no more than 0.4 pH unit. This corresponds to the pH interval in which three FT-IR spectra were recorded in the titration experiments. The concentration profiles of analogous species from different acids are extremely correlated due to the very similar pK_a values of the acids, which is also confirmed by the correlation coefficients between the different species of malic, tartaric, and lactic acid that are given in Table 1, with values often exceeding 0.80. In contrast to the

(27) *CRC Handbook of Chemistry and Physics*, 81st ed.; Lide, D. R., Ed.; CRC Press: New York, 2000/2001; Section 8.

Table 1. Correlation Coefficients between Theoretical Concentration Profiles of the Different Acid Species[a] with pH

	H_2M	HM^-	M^{2-}	H_2T	HT^-	T^{2-}	HL	L^-
H_2M	1	−0.26	−0.79	0.98	0.11	−0.92	0.98	−0.98
HM^-		1	−0.39	−0.37	0.80	−0.12	−0.09	0.09
M^{2-}			1	−0.70	−0.62	0.95	−0.88	0.88
H_2T				1	−0.06	−0.84	0.93	−0.93
HT^-					1	−0.49	0.31	−0.31
T^{2-}						1	−0.98	0.98
HL							1	−1.00
L^-								1

[a] Abbreviations: H_2M, HM^-, M^{2-} (malic acid); H_2T, HT^-, T^{2-} (tartaric acid); HL, L^- (lactic acid).

Table 2. Sample Composition (in g/L) for the Different Acid−Base Titrations

sample	malic acid	tartaric acid	lactic acid	sucrose
[M]	1.500			
[T]		1.500		
[MS1]	0.735			24
[MS2]	1.626			24
[ML1]	1.509		1.715	
[ML2]	1.080		2.010	
[MT1]	2.908	2.801		
[MT2]	1.458	1.462		
[MT3]	2.920	1.368		
[MT4]	1.272	2.503		
[MT5]	5.522	0.804		
[MT6]	0.813	5.591		
[MTS]	2.044	0.958		24

organic acid concentration profiles, the shapes of the related second-derivative FT-IR spectra were significantly less correlated but there was a strong overlap in all the working wavenumber range and no clear selective regions could be detected. Samples of varying complexity were titrated in order to systematically evaluate the performance of the new hard−soft-modeling algorithm. The compositions and the data matrix names of the different titration samples are given in Table 2. Samples included pure standards of malic and tartaric acid, samples of malic acid with sucrose or the monoprotic lactic acid as interferences and, finally, mixtures of the two diprotic acids, malic and tartaric acid, with and without sucrose, where one of the diprotic acids also acts as interfering agent. All data matrices contained 25−35 FT-IR spectra in rows that were evaluated in the 1540−1000-cm^{-1} range, thus yielding matrices of dimension 25 × 210 to 35 × 210.

RESULTS AND DISCUSSION

Pure Soft-Modeling MCR-ALS Approach. In a previous work,[12] we reported on the quantitative analysis of FT-IR titrations of mixtures of diprotic organic acids and sucrose. The pure soft-modeling approach was successful when malic acid and tartaric acid were quantified simultaneously by augmenting the mixture sample matrix with standard matrices of both acids. Clear advantage was taken of the inert behavior of sucrose during titration and relative prediction errors clearly below 5% could be achieved for the acids even in the presence of 30 times higher sucrose concentrations. No pH measurements were required for MCR-ALS analysis of the titrations.

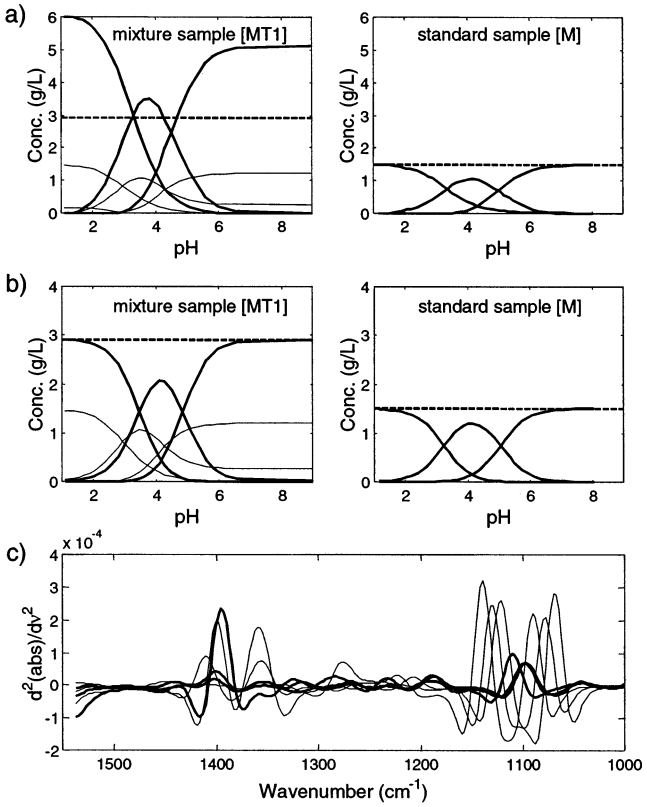

Figure 3. Malic acid quantitation results for sample [MT1;M]. Malic acid profiles (thick lines), interference profiles (thin lines). The horizontal dashed lines in (a) and (b) indicate the true concentration level of malic acid in the samples. (a) Concentration profiles from pure soft-modeling MCR-ALS. Left plot, [MT1] mixture; right plot, [M] standard. (b) and (c) Concentration profiles (left plot, [MT1] mixture; right plot, [M] standard) and second-derivative spectra from MCR-ALS with equilibrium constraint.

However, if only one diprotic acid is regarded as an analyte, and hence, only one standard sample titration is included in the analysis and the other acid is considered as an unknown interference, pure soft-modeling MCR-ALS does not yield acceptable resolution and quantitation. The main reasons for this failure are attributed to the extremely high correlation between the malic and tartaric acid species concentration profiles, the big overlap in the spectral direction, and the impossibility of applying the nonnegativity constraint in the second-derivative spectra.

Pure soft modeling of the diprotic acid mixture titration data does not allow for quantitation of the analyte acid as rotational ambiguous MCR-ALS solutions[26] are obtained. Using S-type initial estimates, the recovered analyte species profiles do not normally fulfill diprotic acid behavior; hence, no quantitative answer can be extracted from the MCR-ALS model. Using C-type initial estimates, reasonably shaped concentration profiles may be obtained but the estimated concentrations, i.e., the heights of the profiles in relation to the ones in the standard sample, can be far from the true ones, depending on the scaling of the initial estimates (Figure 3a). This is because the data structure leads to severe rotational ambiguity and the algorithm cannot ensure convergence to the correct resolution with the exclusive application of soft constraints (nonnegativity, unimodality, selectivity, closure).

MCR-ALS Approach with Equilibrium Constraint. When the novel equilibrium constraint was applied in MCR-ALS, correct resolution for the analyte diprotic acid was achieved, hence allowing reliable quantification. As shown in the Data Treatment section, this constraint has general applicability to any acid−base system but discussion is limited to diprotic acids, the analytes investigated in this work. Only three parameters were necessary to describe the concentration profiles of the H_2A, HA^-, and A^{2-} species as a function of pH. These parameters were c_A, the total concentration of the acid−base system, and the two equilibrium constants, pK_{a1} and pK_{a2}. The equilibrium constraint was only applied to the analyte in both the mixture and the related standard matrix and not to the interferences, since they were regarded as unknowns. Within each titration, this constraint guaranteed that the concentration profiles of the three analyte acid species summed up to a constant total concentration at each pH value and that the profile shapes obeyed a pH equilibrium model. No additional constraints other than nonnegativity for all concentration profiles had to be used in MCR-ALS.

In all analyses, the mixture sample matrix was augmented with one standard sample matrix of the analyte diprotic acid, e.g., [MT1;M] (see also Figure 1). Initial concentration profile estimates for the interfering acid could be taken from pure soft modeling of this augmented data matrix. Concentration estimates for the analyte diprotic acid in the standard and the mixture sample were calculated according to the diprotic equilibrium model with pK_a values obtained from analysis of a pure standard sample. The crucial point in building the analyte estimates, however, was the scaling of the profiles in the mixture relative to the scaling in the standard. This ratio could not be known a priori, as it was the analytical information sought for. A successful algorithm should give correct results, i.e., the true ratio between analyte in standard and mixture sample, independently of this initial scaling ratio (R_{init}), and should also converge from extreme initial ratios or ratios far from the true ones.

For successful quantitation of the diprotic acid mixture samples, a two-step resolution process was developed. In the first step ("pK_a fixed"), very high and low R_{init} were used to build initial analyte estimates in order to cover the possible analyte concentrations in the mixture sample. The pK_as of the analyte acid in the mixture were fixed to the same known values as in the pure standard sample. Only the analyte acid concentration, c_A, in the mixture sample was left loose for fitting in the equilibrium constraint. The MCR-ALS converged fast to solutions and c_A values that were not accurate, but much closer to the true ones than the initial estimates. The obtained rough estimates of c_A were averaged and used to build a new initial estimate for the second step ("pK_a optimizing") of the analysis. Now, the pK_a values for the analyte in the mixture sample were left loose together with c_A; i.e., they were optimized in the nonnegative least-squares fit of the equilibrium constraint. The so-obtained diprotic acid concentration predicted accurately the analyte content of the mixture samples with prediction errors mostly lower than 5% (see Table 3).

This two-step resolution process combines the advantages of the individual fixed pK_a and optimizing approaches. In the fixed pK_a approach, the algorithm converges fast, even from extreme R_{init}. The obtained c_A values are only a first estimate of the true analyte concentration as the assumption of identical pK_as in the mixture and the standard samples is an approximation. Due to various factors (total concentration of acid in the samples, differences in salt concentration during titration and in analyte/interference ratios, shifts in pH measurement), the pK_a values of the diprotic acids change to a small extent between different titrations. Therefore, in the second, pK_a optimizing, step the pK_a values of the analyte in the mixture sample are left loose and can adopt their true values. As a result, correct prediction of analyte concentration in the mixture samples is obtained.

Table 3. Prediction Results and pK_a Values for Malic and Tartaric Acid from MCR-ALS Analysis with Equilibrium Constraint

sample[a]	LOF[b] %	pred concn, g/L	pred error, %	pK_{a1}	pK_{a2}
analyte: malic acid					
[MS1;M]	0.31	0.713	−3.1	3.29	5.14
[MS2;M]	0.40	1.665	2.4	3.30	4.96
[ML1;M]	2.61	1.572	4.2	3.12	4.78
[ML2;M]	2.54	1.071	−0.8	3.28	5.10
[MT1;M]	1.48	2.928	0.7	3.45	4.83
[MT2;M]	2.50	1.374	−5.8	3.40	4.93
[MT3;M]	1.85	2.870	−1.7	3.49	4.89
[MT4;M]	1.74	1.205	−5.3	3.42	4.89
[MT5;M]	1.19	5.505	−0.3	3.43	4.83
[MT6;M]	0.94	0.812	−0.2	3.50	4.83
[MTS;M]	0.68	1.989	−2.7	3.48	4.89
				3.40[c]	5.11[c]
analyte: tartaric acid					
[MT1;T]	1.41	2.772	−1.0	2.86	3.96
[MT2;T]	2.49	1.458	−0.3	2.89	3.85
[MT3;T]	1.59	1.352	−1.2	2.88	4.03
[MT4;T]	1.66	2.453	−2.0	2.85	4.05
[MT5;T]	1.06	0.848	5.4	2.90	3.99
[MT6;T]	0.99	5.483	−1.9	2.85	3.97
[MTS;T]	0.67	1.000	4.4	2.91	4.08
				2.98[c]	4.34[c]

[a] Mixture samples are given together with the standard sample used in MCR-ALS analysis with equilibrium constraint. The resulting augmented matrices are given in Matlab notation, i.e., semicolon holds for columnwise augmentation of the data set. [b] LOF as defined in Data Treatment Section. [c] Literature pK_a values for comparison.

In Table 3, the quantitation results for the two-step MCR-ALS are given for samples of increasing complexity, for both malic and tartaric acid as analytes. They ranged from samples with sucrose as interfering agent ([MS1], [MS2]) that can also be resolved by pure soft-modeling MCR-ALS and samples with the monoprotic lactic acid interference ([ML1], [ML2]) to samples with a diprotic acid interference ([MT1] − [MT6]) and also diprotic and sucrose interference ([MTS1]). In all cases, the resolution with MCR-ALS was successful and the predicted analyte concentrations obtained from the equilibrium model were very close to the true ones. A look at the obtained pK_a values for the malic and tartaric analytes in the different samples shows that the pK_a values were not identical but changed slightly from sample to sample as was mentioned before. However, the range of pK_as was close enough to the literature values and confirmed that the hard equilibrium constraint was appropriate and did not introduce artifacts into the data. In Figure 3b and c, the resolved concentration profiles and second-derivative spectra for the [MT1;M] analysis are depicted. The malic acid concentration profiles in Figure 3b obeyed the diprotic equilibrium model and the heights of the profiles corresponded to the true ones, in contrast to the

pure soft-modeling results of Figure 3a that gave a malic acid concentration much higher than the true one. The interference profiles (tartaric acid) were not correct in the higher pH range as they were not constrained at all, except nonnegativity for the concentration profiles. Still, this rotational ambiguity[26] of the interference that could be much higher in other samples did not hinder the correct resolution and quantitation of the analyte. From Figure 3b it can also be seen how overlapped the analyte and interference contributions to the measured matrix [MT1] were and that in a wide pH range all acid species of the analyte and interference were present simultaneously. Even the rank-deficient sample [MTS1] in which a chemical rank of only six could be detected by singular value decomposition instead of the seven spectroscopically contributing species[12] was correctly resolved for the analytes. In this sample, the three interference components accounted simultaneously for the interference diprotic acid and the sucrose contribution to the spectral data matrix. Nevertheless the hard-modeling constraint ensured the correct recovery of the analyte profiles and limited the rotational ambiguities to the components of the sample that were not of analytical interest.

An additional asset of the novel hard-model constraint is that it implicitly includes several other conventional soft constraints. Unimodality, closure, and selectivity/local rank information are covered by the equilibrium model and do not have to be applied separately anymore. When initial estimates for the analyte are built based on its approximate acid−base equilibrium model, the application of this single constraint and the inclusion of the measured pH values are sufficient for correct resolution of the analyte in the FT-IR titration data.

CONCLUSIONS

In this work, we demonstrated how second-order resolution and calibration can be performed successfully, even when the concentration profiles are extremely correlated for analytes and interferences and classical pure soft-modeling approaches fail because of severe rotational ambiguities in the resolved concentration profiles and spectra. The knowledge about the chemical behavior of the analytes is used to implement a hard-modeling equilibrium constraint in the MCR-ALS algorithm that ensures the correct resolution and quantitation for the analyte and concentrates the remaining rotational ambiguities of the system in the profiles of the uninteresting interference components. This holds even for samples where the interference contribution is rank deficient.

In contrast to previous examples of combination of hard- and soft-modeling methods, this work shows that the inclusion of a hard-modeling constraint into soft-modeling methods does not necessarily require a physicochemical focus on the model and its parameters but can have a direct analytical application. Indeed, the analyte concentration is obtained as one of the fitted parameters in the equilibrium model and can be directly used for analyte prediction in the presence of unknown interferences. Due to the flexible implementation of the hard-model equilibrium constraint in MCR-ALS, this approach is expected to be useful also for analysis of other mixed equilibrium-based chemical systems.

ACKNOWLEDGMENT

The authors acknowledge gratefully the financial support of this work by the Austrian Science Fund within the project P13350 ÖCH.

Received for review October 21, 2002. Accepted November 21, 2002.

AC026248J