Anal. Chem. **1999,** *71,* 126−134

Procedure for the Quantitative Determination of Mixtures of Nucleic Acid Components Based on Multivariate Spectrophotometric Acid−Base Titrations

Javier Saurina, Santiago Hernández-Cassou, Romà Tauler, and Anna Izquierdo-Ridorsa*

Department of Analytical Chemistry, University of Barcelona, Diagonal 647, Barcelona 08028, Spain

A new procedure for the quantitative determination of mixtures of nucleic acid components, based on continuous spectrophotometric acid−base titrations and multivariate curve resolution, is proposed. The procedure simultaneously takes into account the spectroscopic and acid−base properties of the compounds, which leads to a higher selectivity. Furthermore, quantitative determination of an analyte in a complex mixture is performed using a synthetic solution as standard containing only the analyte of interest. An intrinsic difficulty in the analysis of spectrometric titration data is the presence of rank deficiency due to closure for the mixtures of two or more compounds. An additional problem can be encountered in some mixtures if species spectra or species concentration profiles are practically identical (rank overlap). However, even in the presence of these rank difficulties, accurate quantitation with prediction errors lower than 5% was obtained. The presence of unknown and uncalibrated interferences in the samples does not affect the quantitative determination of the analyte of interest. The proposed procedure was successfully applied to the analysis of real samples (pharmaceuticals) using synthetic external standards.

The quantitative determination of mixtures of nucleic acid components (nucleic bases, nucleosides, nucleotides) is of fundamental importance in biochemical, biomedical, and pharmaceutical research fields.[1−3] Nucleic acid components are found in physiological fluids, tissues, and cells, as a result of the catabolism of nucleic acids, enzymatic degradation of tissues, or dietary habits among other pathways. Altered concentrations of these compounds indicate the presence of various disease states which cause alterations in the normal purine and pyrimidine metabolic pathways. Furthermore, in certain illnesses nucleic acid components are clinically prescribed, such as in certain hepatic diseases, where the presence of large amounts of nucleic acid components is essential in order to recover the hepatic tissue. Moreover, some nucleoside derivatives are being extensively used as chemotherapeutic drugs. For instance, nucleoside thioketo derivatives and aza nucleosides are powerful anticancer agents, arabinonucleosides have a broad-spectrum antiviral activity against DNA-containing viruses and RNA tumor viruses, and azido nucleosides such as 3′-azido-3′-deoxythimidine (AZT) are some of the most effective drugs against the HIV virus.

These examples demonstrate the interest in the development of new analytical methods for the quantitative determination of such components. The most extensively used techniques are HPLC and capillary electrophoresis, which permit the analysis of both majority and minority bases.[1−2,4−6] However, these analyses are quite time-consuming and other simpler methods are preferred for the determination of majority bases, as in the case of pharmaceutical products. Among these simpler methods, potentiometric acid−base titrations have been proposed for the quantitative analysis of mixtures of compounds that present acid−base properties, such as the nucleic acid components. However, this procedure can only be applied successfully to very simple mixtures. For the analysis of more complex mixtures, first-order multivariate calibration, using the potentiometric values obtained along acid−base titrations as the experimental data, has also been proposed.[7] In another approach, first-order multivariate calibration has been applied to the study of a mixture of nucleic bases with spectrophotometric detection.[8] The experimental data used were the UV spectra obtained for the different mixtures. Methods based on first-order multivariate calibration require the preparation of a calibration matrix using standards that have to be similar (in interferences present, concentration ranges, etc.) to those of unknown mixtures, in order that the relationship between analyte concentration and instrumental response is modeled correctly.

In this paper, a new analytical method based on a second-order multivariate curve resolution data treatment procedure[9] is proposed for the analysis of mixtures of nucleic acid components. This procedure has been previously tested on simulated second-

(1) Simpson, R. C.; Brown, P. R. *J. Chromatogr.* **1986**, *379*, 269−311.
(2) Gehrke, C. W.; Kuo, K. C. *J. Chromatogr.* **1989**, *471*, 3−36.
(3) Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag: New York, 1984.
(4) Di Pierro, D.; Tavazzi, B.; Perno, C. F.; Bartolini, M.; Balestra, E.; Calió, R.; Giardina, B.; Lazzarino, G. *Anal. Biochem.* **1995**, *231*, 407−412.
(5) Meynial, I.; Paquet, V.; Combes, D. *Anal. Chem.* **1995**, *67*, 1627−1631.
(6) Wang, C.; McCann, W. P.; Beale, S. C. *J. Chromatogr., B* **1996**, *676*, 19−28.
(7) Lindberg, W.; Kowalski, B. R. *Anal. Chim. Acta* **1988**, *206*, 125−135.
(8) Tauler, R.; Izquierdo-Ridorsa, A.; Casassas, E. *An. Quím.* **1991**, *87*, 571−579.
(9) Tauler, R.; Izquierdo-Ridorsa, A.; Casassas, E. *Chemom. Intell. Lab. Syst.* **1993**, *18*, 293−300.

10.1021/ac980409v CCC: $18.00
Published on Web 11/21/1998

order data corresponding to acid−base spectrophotometric titrations.[10] In contrast to first-order multivariate calibration, this method takes into account the spectroscopic and the acid−base behavior of the substances simultaneously, giving as a result a higher selectivity. Moreover, second-order multivariate curve resolution avoids the need to use a large number of calibration samples, since a single calibration sample, that containing the analyte of interest, is sufficient.[11,12] The method and the working conditions proposed here can be extended to the analysis of any mixture of compounds whenever they present acid−base behavior and absorption spectra. The experimental data are obtained from spectrophotometric acid−base titrations (at each successive pH value the absorption spectrum is obtained), and they are arranged in an ordered data matrix. The multivariate curve resolution procedure can be applied easily to the simultaneous study of several correlated data matrices, each one obtained in different conditions, and, therefore, providing independent information. If one of those data matrices corresponds to a standard, quantitative determination can be performed by comparison of the results obtained for the standard with those of the sample. Furthermore, quantitation in the presence of unknown and uncalibrated interferences is possible (second-order advantage[13]), in contrast to the need to model completely the interferences in first-order multivariate calibration methods.

The multivariate curve resolution method applied here is based on a constrained alternating least-squares optimization procedure (MCR-ALS), and it has already shown to be a powerful tool for the species resolution and quantitative determination of many types of unresolved chemical mixtures.[10,14−17] MCR-ALS is especially suitable for chemical data structures that do not fulfill the excessively high requirements of trilinearity or strict analyte bilinearity of other three-way and second-order methods, such as the generalized rank annihilation method (GRAM),[18] the direct trilinear decomposition method (DTD),[19] or PARAFAC.[20,21]

Mixtures of substances with acid−base behavior (such as nucleic acid components) give rank-deficient data matrices when spectrophotometric pH titrations are performed.[10,22] A data matrix is rank deficient when the number of significant contributions to the data variance, estimated by using singular value decomposition or other related factor analysis techniques, is lower than the real number of chemical components present in the system. Such a situation occurs quite frequently when, for instance, at the beginning of a chemical reaction or process, more than one component already exists. In this case, instead of detecting all the components present at the beginning of the reaction, only one can mathematically be derived, with its spectrum being a linear combination of the pure spectra of all the components initially present. Therefore, the complete resolution of such a rank-deficient data matrix is hindered.[10,23] Similar situations involving rank deficiencies have also been described for other kinds of experimental data.[23,24]

When the aim of the study is the complete resolution of all the species in the system, the rank deficiency problem has to be eliminated. Recently, two different strategies in order to break the rank deficiencies and obtain a full-rank system[10,23] have been proposed: (1) matrix augmentation by simultaneous analysis of multiple process runs at linearly independent conditions (initial concentrations) and (2) data matrix perturbation by addition of a single component or of mixtures of components during the process. Nevertheless, for quantitative purposes, total resolution of the system is not necessary and only partial resolution involving those species to be quantified is required. This partial resolution is already achieved when the analysis is performed on an augmented matrix that contains the data matrix of the unknown mixture and that of the standard.[10]

In the present paper, mixtures of four nucleosides (adenosine, cytidine, inosine, uridine) have been analyzed. These nucleosides have been selected to evaluate the accuracy of the analytical method we propose, since they have known spectroscopic and acid−base properties. The procedure was subsequently applied to a pharmaceutical drug containing a mixture of inosine 5′-monophosphate (nucleotide) and cytosine (nucleic base). Several synthetic mixtures containing the constituents of the drug were also analyzed.

EXPERIMENTAL SECTION

Reagents and Solutions. Hydrochloric acid and sodium hydroxide (all Merck, ar); cytosine, cytidine, inosine, and inosine 5′-monophosphate (inosinate) (all Sigma, ar); uridine (BDH, ar); and adenosine (Boehringer Mannheim, ar) were used without further purification. Stock solutions of the various compounds were prepared from a known amount of the solid reagent dissolved in 5×10^{-3} M hydrochloric acid.

A drug containing an equimolar proportion of inosine 5′-monophosphate and cytosine (Precurgen from Knoll) was analyzed. A 0.25-g aliquot of the drug was dissolved in 200 mL of water; the resulting solution was filtered through a Nylon membrane of 0.45-μm pore size. This solution contained 10^{-3} mol L^{-1} inosine 5′-monophosphate and 10^{-3} mol L^{-1} cytosine. The working solution to be analyzed by using the proposed titration method was obtained by 50-fold dilution with 5×10^{-3} M hydrochloric acid solution.

Ultrapure water (Millipore) was used in all the experiments.

Table 1 shows the different mixtures studied. In all of them, the concentration of each compound was 2×10^{-5} M. The drug

(10) Izquierdo-Ridorsa, A.; Saurina, J.; Hernández-Cassou, S.; Tauler, R. *Chemom. Intell. Lab. Syst.* **1997**, *38*, 183−196.
(11) Saurina, J.; Hernández-Cassou, S.; Tauler, R. *Anal. Chim. Acta* **1996**, *335*, 41−49.
(12) Saurina, J.; Hernández-Cassou, S.; Tauler, R. *Anal. Chem.* **1997**, *69*, 2329−2336.
(13) Tauler, R.; Smilde, A.; Kowalski, B. R. *J. Chemom.* **1995**, *9*, 31−58.
(14) de Juan, A.; Izquierdo-Ridorsa, A.; Tauler, R.; Fonrodona, G.; Casassas, E. *Biophys. J.* **1997**, *73*, 2937−2948.
(15) Casassas, E.; Marqués, I.; Tauler, R. *Anal. Chim. Acta* **1995**, *310*, 473−484.
(16) Saurina, J.; Hernández-Cassou, S.; Tauler, R. *Anal. Chem.* **1995**, *67*, 3722−3726.
(17) Gargallo, R.; Tauler, R.; Izquierdo-Ridorsa, A. *Anal. Chem.* **1997**, *69*, 1785−1792.
(18) Sanchez, E.; Ramos, L. S.; Kowalski, B. R. *J. Chromatogr.* **1987**, *385*, 151−164.
(19) Sanchez, E.; Kowalski, B. R. *J. Chemom.* **1990**, *4*, 29−45.
(20) Kroonenberg, P.; DeLeeuw, J. *Psychometrica* **1980**, *45*, 69−97.
(21) Gemperline, P. J.; Miller, K. H.; West, T. L.; Weinstein, J. E.; Hamilton, J. G.; Bray, J. T. *Anal. Chem.* **1992**, *64*, 523A−531A.
(22) Norgaard, L.; Ridder, C. *Chemom. Intell. Lab. Syst.* **1994**, *23*, 107−114.
(23) Amrhein, M.; Srinivasan, B.; Bonvin, D.; Schumacher, M. M. *Chemom. Intell. Lab. Syst.* **1996**, *33*, 17−33.
(24) Saurina, J.; Hernández-Cassou, S.; Tauler, R.; Izquierdo-Ridorsa, A. *J. Chemom.* **1998**, *12*, 183−203.

Table 1. Composition of the Analyzed Mixtures (in mol L^{-1})

data matrix	adenosine (A)	cytosine (S)	cytidine (C)	inosine (I)	inosinate (T)	uridine (U)	pH range studied	no. of spectra
A	2×10^{-5}						2.27−9.94	17
S		2×10^{-5}					2.20−9.26	20
C			2×10^{-5}				2.23−9.19	17
I				2×10^{-5}			2.27−9.85	22
T					2×10^{-5}		2.32−10.05	23
U						2×10^{-5}	2.90−10.19	25
IU				2×10^{-5}		2×10^{-5}	2.29−10.11	18
CU			2×10^{-5}			2×10^{-5}	2.22−10.39	28
AS	2×10^{-5}	2×10^{-5}					2.35−11.18	22
ST		2×10^{-5}			2×10^{-5}		2.30−10.60	37
ACI	2×10^{-5}		2×10^{-5}	2×10^{-5}			2.23−10.25	32
ISU		2×10^{-5}		2×10^{-5}		2×10^{-5}	2.28−10.68	44
ACIU	2×10^{-5}		2×10^{-5}	2×10^{-5}		2×10^{-5}	2.25−10.56	34
drug		$\sim 2 \times 10^{-5}$			$\sim 2 \times 10^{-5}$		2.56−10.59	37

solution was also 2×10^{-5} M in each active constituent. All solutions were prepared in an acidic medium (5×10^{-3} M HCl).

Apparatus and Procedure. UV absorption spectra were recorded on a Perkin-Elmer Lambda-19 spectrophotometer. pH was measured with an Orion model 701 A pHmeter (with a precision of ±0.1 mV) using a combined pH electrode Orion 9103SC with an inner Ag/AgCl reference electrode. The titrant solution was added with a Metrohm 655 Dosimat autoburet (with a precision of ±0.005 mL). The calibration of the electrode was performed with two buffer solutions (a pH = 4.01 solution of KH_2PO_4/Na_2HPO_4 and a pH = 6.86 solution of potassium hydrogen phthalate).

Spectrophotometric acid−base titrations of the solutions listed in Table 1 were performed. Since all the solutions contained 5×10^{-3} M HCl, the acidic species of each compound was the one initially present in the system. The solution in the titration vessel was continuously pumped through the flow cell in the spectrophotometer. After each titrant addition, the pH of the solution was measured and the corresponding UV spectrum recorded. The spectral wavelength range was 220−350 nm, with a $\Delta\lambda = 1$ nm between consecutive absorbance readings. For each spectrum, the absorbance value obtained at 350 nm was subtracted from the whole spectrum in order to correct the baseline drift, since at this wavelength none of the compounds of the mixtures absorbed.

Systems under Study. The compounds under study can be classified in accordance with the nature of their functional groups in the nucleic base moiety that participates in an acid−base equilibrium. In this way, uridine, inosine, and inosinate have amide-like nitrogens with pK_a values around 9 (see Table 2). In contrast, cytidine, cytosine, and adenosine contain ring nitrogens, with pK_a values around 4 (see Table 2). Consequently, for the compounds with different types of nitrogen groups, the distribution of species with pH will be different, whereas those compounds with the same type of nitrogen group will show similar acid−base properties. One of the compounds analyzed, inosinate, has another functional group that participates in an acid−base equilibrium: the second acidic group of the phosphate residue.

Since the procedure proposed in the present paper takes advantage of both differences in the concentration and spectral profiles of the compounds to be analyzed, a comparison between the spectral behavior of these compounds was also carried out. Figure 1 shows the absorption spectra for the protonated and deprotonated forms of the four nucleosides. There is a high degree of overlap throughout the whole wavelength range (no spectral selective regions). Table 3 contains the similarity (correlation) values between the spectra of the different compounds. Each compound has been differentiated into acidic and basic forms, since each has a different absorption spectrum. From these results, certain features should be stressed. It is noticeable that the spectra of the acidic and basic forms of adenosine are very similar (with a similarity value of 0.9977). In fact, the differences in both spectra are mostly due to intensity differences and not to shifts in the band maximums. A high similarity between the spectra of inosine and inosinate is also observed. This result can be attributed to the fact that the main absorbing chromophore in these compounds is the nitrogen base, which is not directly affected by the acid−base equilibrium in the phosphate group. It can be concluded then that inosine and inosinate are largely indistinguishable in working conditions. For all other compounds, the similarities between spectra are less. The fewest similarities were found between the spectra of cytosine derivatives and inosine derivatives, groups of compounds that also present major differences in their acid−base behavior.

Table 2. Acid−Base Properties of the Compounds Analyzed

		working conditions		
compound	pK_a	temp (°C)	I (M)	ref
adenosine	3.6 (N(1))	25	0.1	28
cytosine	4.56 (N(3))	37	0.15	29
	11.7 (N(1))	25	0.1	30
cytidine	4.13 (N(3))	37	0.15	29
inosine	8.50 (N(1))	37	0.15	31
inosine 5′-monophosphate	6.00 (phosphate)	34	0.5	32
	9.27 (N(1))	34	0.5	32
uridine	8.855 (N(3))	37	0.15	31

The mixtures to be analyzed (see Table 1) were chosen according to the similarities and differences in both the acid−base and spectral behavior of the compounds, trying to cover all degrees of complexity of resolution from the study of a limited number of examples. Binary mixtures of inosine + uridine and cytosine + adenosine were studied in order to evaluate the predictive ability of the proposed procedure when the compounds have very similar acid−base behavior. The cytidine + uridine mixture was taken as an example of the situation in which there are differences in the acid−base behavior of the compounds. And

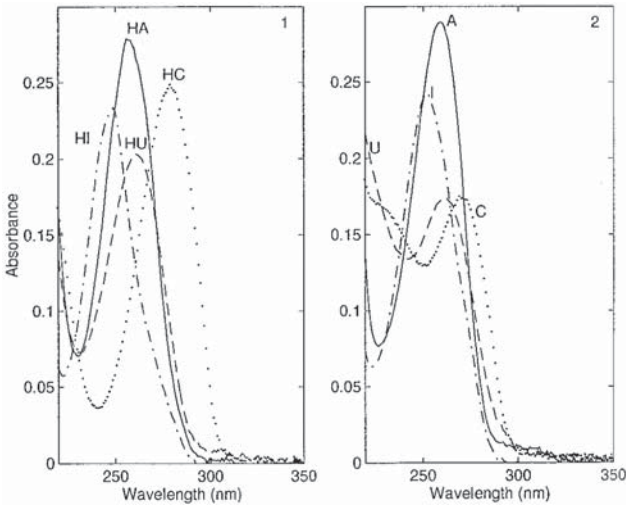

Figure 1. (1) Absorption spectra for the protonated forms of adenosine (HA), cytidine (HC), inosine (HI), and uridine (HU). (2) Absorption spectra for the deprotonated forms of adenosine (A), cytidine (C), inosine (I), and uridine (U).

finally, the cytosine + inosinate mixture was selected as it is present in the drug that was analyzed. Two ternary mixtures (adenosine + cytidine + inosine and cytidine + inosine + uridine) and one quaternary mixture (adenosine + cytidine + inosine + uridine) were also prepared and analyzed in order to study more complex situations.

DATA TREATMENT

In acid−base titration systems, each species is defined by a single spectrum and concentration profile in all the titrations. As a result, acid−base titration data from multiple runs give a trilinear structure. To get such a structure experimentally, the working methodology has to allow the synchronization of the wavelength order and the pH order. The synchronization of wavelengths means that all spectra have to be registered at the same wavelengths. The synchronization in the pH order requires that spectra have to be registered at the same pH values in any titration run.

The experimental methodology to perform the spectrophotometric titrations allows synchronization in the wavelength order (here, all spectra are recorded from 220 to 350 nm in steps of 1 nm) but not in the pH order (e.g., the working range of pH, the number of spectra, and the individual pH values are different in each run). Thus, the data obtained are not trilinear since there is not synchronization in the pH order. The following data pretreatment was performed in order to synchronize the pH order in the matrices analyzed simultaneously and thus, to transform data in trilinear: the data matrices of the standards were linearly interpolated taking the pH values of the matrix of the sample mixture as reference.

Both original and interpolated data were analyzed with the MCR-ALS method. This method is implemented in a small set of MATLAB functions. It has been extensively explained elsewhere,[25−27] and the different steps will only be summarized here, emphasizing the main features for resolving the systems analyzed in this paper:

Data Arrangement. All the individual matrices, obtained using the experimental procedure described above, share their wavelength space, since the absorbance data were obtained at the same wavelength values for all the acid−base titrations. However, since the experimental procedure of the acid−base titrations hinders a synchronization in the pH direction between the different data matrices, when dealing with the original data matrices, matrix augmentation can only be performed in the wavelength direction (wavelengthwise-augmented matrices) (see Figure 2). These wavelengthwise-augmented matrices have a number of rows equal to the total number of acquired spectra in the different acid−base titrations and a number of columns equal to the number of wavelengths present in each single matrix.

The pH-wise augmentation is only possible with the interpolated data, since, as mentioned above, interpolation allows the pH order to be synchronized. Even in this case, wavelengthwise augmentation is preferred because rank deficiencies produced by closure (as occurs in the cases under study, see below) can be solved.[23,24]

Wavelengthwise-augmented matrices were built from the data matrix of a mixture (considered as a problem sample) and one standard data matrix for each different component to be quantified.

Initial Estimates of Species Profiles. Both absorption spectra and concentration profiles can be used as initial estimates in the ALS optimization. In this paper, initial concentration estimates were preferred. The main advantage lies in the possibility of applying two decisive constraints on the ALS optimization steps: selectivity and zero concentration windows.[10]

Alternating Least-Squares Optimization. The linear model (generalized Beer law) equation in matrix form is solved iteratively by ALS to obtain the matrix of pure spectra $\mathbf{S}^T$ and the matrix of concentration profiles $\mathbf{C}$ which best fit the data. For the augmented data matrices, $\mathbf{D}_{aug}$, the linear model can be written as

$$\mathbf{D}_{aug} = \mathbf{C}_{aug}\mathbf{S}^T + \mathbf{E} \qquad (1)$$

where $\mathbf{E}$ is the matrix of the residuals not explained by the chemical species or conformations in $\mathbf{C}$ and $\mathbf{S}^T$ and should be close to the experimental error. In the present work, data are assumed to follow a trilinear model. Under this condition, theoretically, the resolution of $\mathbf{D}^*_{aug}$ is unique.

(25) Tauler, R.; Smilde, A. K.; Henshaw, J. M.; Burgess, L. W.; Kowalski, B. R. *Anal. Chem.* **1994**, *66*, 3337−3344.
(26) Tauler, R.; Izquierdo-Ridorsa, A.; Gargallo, R.; Casassas, E. *Chemom. Intell. Lab. Syst.* **1995**, *27*, 163−174.
(27) Casassas, E., Gargallo, R.; Izquierdo-Ridorsa, A.; Tauler, R. *React. Polym.* **1995**, *27*, 1−14.
(28) Tribolet, R.; Sigel, H. *Eur. J. Biochem.* **1987**, *163*, 353.
(29) Gargallo R. Ph.D. Thesis, University of Barcelona, 1996.
(30) Reinert, H.; Weiss, R. *Z. Physiol. Chem.* **1969**, *350*, 1310.
(31) Casassas, E.; Gargallo, R.; Giménez, I.; Izquierdo-Ridorsa, A.; Tauler R. *J. Inorg. Biochem.* **1994**, *56*, 187−199.
(32) Scheller, K. H.; Scheller-Krattiger, V.; Martin, R. B. *J. Am. Chem. Soc.* **1981**, *103*, 6833.

Table 3. Similarity (Correlation) Values between the Spectra of the Different Compounds[a]

		adenosine		cytosine		cytidine		inosine		IMP		uridine	
		acid	base	acid	base	acid	base	acid	base	acid	base	acid	base
adenosine	acid		0.9977	0.7464	0.8713	0.6518	0.8757	0.9024	0.9705	0.9072	0.9678	0.9857	0.9186
	base			0.7322	0.8625	0.6397	0.8687	0.9061	0.9734	0.9109	0.9705	0.9826	0.9112
cytosine	acid				0.7795	0.9721	0.8543	0.5461	0.6384	0.5514	0.6335	0.8337	0.7764
	base					0.7411	0.9832	0.8264	0.8458	0.8310	0.8473	0.9162	0.9925
cytidine	acid						0.8239	0.4857	0.5550	0.4909	0.5512	0.7512	0.7323
	base							0.8320	0.8528	0.8358	0.8535	0.9308	0.9781
inosine	acid								0.9763	0.9998	0.9794	0.8712	0.8692
	base									0.9784	0.9998	0.9414	0.8959
IMP	acid										0.9815	0.8763	0.8740
	base											0.9387	0.8967
uridine	acid												0.9479
	base												

[a] Acid and base refer, respectively, to the protonated and deprotonated forms of the compound.

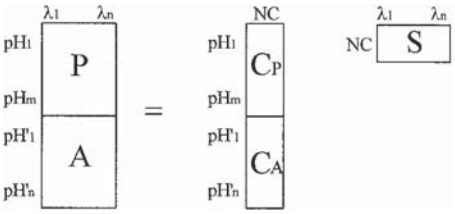

Figure 2. Arrangement of the experimental data in the augmented wavelengthwise matrices. **P** refers to the matrix of an unknown mixture, and **A** refers to the matrix of the standard. **NC** is the number of larger factors. Augmented matrices are indicated using Matlab notation, [**P**;**A**].

The iterative procedure starts with the evaluation of the unknown species spectra using the initial estimation of the $\mathbf{C}_{aug}$ data matrix:

$$\mathbf{S}^T = (\mathbf{C}_{aug})^+ \mathbf{D}^*_{aug} \quad (2)$$

where $\mathbf{D}^*_{aug}$ is the data matrix reproduced for the number of factors considered and $(\mathbf{C}_{aug})^+$ is the pseudoinverse of $\mathbf{C}_{aug}$.

In the second stage, a new estimate of the species concentrations is obtained by least squares using the equation

$$\mathbf{C}_{aug} = \mathbf{D}^*_{aug}(\mathbf{S}^T)^+ \quad (3)$$

where $(\mathbf{S}^T)^+$ is the pseudoinverse of the $\mathbf{S}^T$ matrix.

The steps are repeated until the data matrix $\mathbf{D}^*_{aug}$ is well explained within experimental error.

Depending on the nature and structure of the data, different constraints can be applied during the ALS optimization. The following constraints were applied in this paper to the ALS optimization from concentration initial estimates: (i) Concentration profiles must be unimodal; (ii) zero concentration windows, i.e., the concentration in those regions where the species is known not to be present is kept at zero; (iii) there is correspondence between common species in the different data matrices; (iv) pure spectra of common species present in different titrations are forced to be equal; (v) concentration profiles of common species present in different titrations have equal shapes; (vi) closure constraint, i.e., the sum of the concentrations of the protonated and deprotonated forms for a certain compound at each point of the titration is constant.

The closure constraint can only be applied to the matrices of the standards, for which the total concentration is known.

Constraint v is fulfilled a priori in the case of acid−base equilibria, where the shape of the species distribution profiles does not depend on concentration. However, it cannot be applied to the original experimental data under study, due to the lack of synchronization between pH values. It can only be applied to the interpolated data. The fulfillment of constraints iv and v implies second-order data with a trilinear structure,[19] whenever there is both pH and wavelength synchronization.

Depending on the sample mixture, the optimal combination of constraints to be applied in the ALS optimization may differ. Consequently, and since the composition of the standards was known perfectly, the accuracy of the results obtained for the standards was taken to indicate whether the best quantitation conditions had been achieved. It was assumed that, given the correspondence between species in the different individual matrices, an improved resolution for the matrices of the standards directly implied a better resolution for that of the sample.

Quantitation Step. When the data matrix of the unknown sample is analyzed simultaneously with that of a standard, quantitative determination of a certain compound in the unknown mixture can be performed from comparison of the concentration profiles obtained for this compound in each matrix.

The concentration profiles for the acidic (HA) and basic (A^-) form of a compound are the representation of the following expressions versus pH:

$$[HA] = c[H^+]/(K_a + [H^+]) \quad (4)$$

$$[A^-] = cK_a/(K_a + [H^+]) \quad (5)$$

where c is the total concentration of the compound ($c = [HA] + [A^-]$) and K_a is its acid dissociation constant.

These expressions are general for any acid−base equilibrium of a monoprotic system, and they can be deduced by applying the mass balance and the mass action law.

Table 4. Singular Values for the Individual Data Matrices[a]

matrices	singular values 1	2	3	4	5	6	real no. of species
[**A**]	**5.7915**	**0.1484**	0.0440	0.0328			2
[**S**]	**4.4437**	**1.1244**	0.0267	0.0109			2
[**C**]	**5.0849**	**1.1084**	0.0208	0.0133			2
[**I**]	**5.3635**	**0.4115**	0.0561	0.0323			2
[**T**]	**5.4927**	**0.3462**	0.0363	0.0181			2
[**U**]	**5.7018**	**0.5858**	0.0400	0.0338			2
[**IU**]	**8.8103**	**0.3977**	**0.1021**	0.0326	0.0156	0.0101	4
[**CU**]	**11.4261**	**1.8086**	**0.5047**	0.0412	0.0131	0.0116	4
[**AS**]	**10.5833**	**1.1981**	**0.1099**	0.0141	0.0049	0.0044	4
[**ST**]	**11.8984**	**1.5602**	**0.4313**	0.0239	0.0141	0.0073	4
[**ACI**]	**19.5362**	**1.5320**	**0.4493**	0.0582	0.0447	0.0124	6
[**ISU**]	**19.3177**	**1.8787**	**0.4642**	**0.1256**	0.0240	0.0075	6
[**ACIU**]	**25.3734**	**1.7107**	**0.3676**	**0.1123**	0.0587	0.0273	8
[**drug**]	**12.0887**	**1.4534**	**0.3949**	0.0464	0.0247	0.0136	4

[a] See Table 1 for the notation used to describe the different data matrices. Bold numbers correspond to singular values associated to chemical species.

Consequently, at a certain pH value, the ratio between the value obtained from the concentration profile for a particular species (the protonated or the deprotonated form) in the sample mixture and that obtained from the concentration profile of that same species in the standard will give the ratio between concentrations in the two titrations. Since one concentration is known (that in the standard), the absolute concentration in the unknown sample can then easily be evaluated. Independently of the chosen pH value, this concentration corresponds to that of the compound (see eq 4 or 5). Quantitative determination can then be performed from the concentration profile of either the protonated or the deprotonated form at any pH value. In this paper, both concentration profiles were used in the quantitation process for the sake of comparison.

When the original experimental data were analyzed, there was no synchronization between pH values. The pH chosen was either that at the beginning of the titration (the most acidic pH value) or the final value (the most basic pH value), where quantitative determination was performed from the concentration profile of HA or A^-. In the pH region at which pH $\ll$ pK_a, [HA] was independent of the pH and equal to c, while the same was reported for [A^-] in the pH region at which pH $\gg$ pK_a. So, in these regions, a comparison can be made between titrations, even though synchronization between pH values is incomplete. From the results obtained with interpolated data, instead of using a single value of the concentration profile, quantitative determination can be performed from the areas of the concentration profiles, since there is synchronization in the pH order between matrices. Quantitation from areas implies, in a certain way, averaging individual results, which yields a more robust estimator of the concentration than in the previous case.

RESULTS AND DISCUSSION

Rank Analysis of the Individual and Augmented Matrices. Table 4 shows the singular values obtained for all the individual matrices (those of a single compound and those of mixtures of compounds). The inspection of the magnitude of singular values provides an estimation of the number of chemical species. However, our experience and chemical knowledge about the system is also important. As a result, the number of species finally chosen is that which provides a more reliable resolution of the system from a chemical point of view.

For all individual matrices of a single compound, two larger singular values were found. This number agrees with the number of different species in the systems analyzed (an acidic and a basic one) with the exception of the inosinate system. As mentioned above, inosinate has two protonation sites to take into account in the working pH range. However, the deprotonation process of the phosphate residue cannot be detected spectrophotometrically, since it does not affect the main absorbing center in the molecule (which is mainly related to the nitrogen base).

For the binary mixtures, with four different absorbing species, three larger singular values were obtained. This rank deficiency is due to the chemical nature of these systems: they are closed reacting systems, in which the different concentration profiles are linearly dependent due to the governing equilibrium reactions (rank deficiency caused by closure). It has already been demonstrated[10,23] that in these systems the rank of the data matrix **D** is rank (**D**) $\leq$ min (R + 1, S), where R is the number of independent reactions and S is the number of absorbing species. In the binary cases under study, R is 2 and S is 4, and thus, the rank of the data matrix is 3.

In the ternary mixtures, R is 3 and S is 6. In these cases, the rank of the data matrices should be equal to 4, due to the presence of the rank deficiency due to closure. However, for one of the analyzed ternary mixtures (that of ACI), the rank was 3 (see Table 4). This mixture is especially difficult because the spectra of the acid and basic adenosine forms are practically identical (as pointed out above) and also because of the great similarities between the spectra of adenosine and inosine species (with similarities higher than 0.9 for all the species). In this particular case, the experimental rank is due not only to the presence of closure but also to rank overlap between spectral profiles.[24]

The same explanation can be extended to the quaternary mixture, ACIU, where there are eight species and four independent reactions. Consequently, the rank expected, taking into account the presence of rank deficiency by closure, should have been 5 (R + 1). However, the experimental rank was 4 since

Table 5. Singular Values for the Augmented Data Matrices[a]

matrices	singular values[b] 1	2	3	4	5	6	7	8	real no. of species
[**ACI;A;C;I**]	**21.1649**	**3.3108**	**1.5416**	**0.9327**	**0.1961**	**0.1417**	0.0959	0.0588	6
[**ACI;A;C**]	**20.5374**	**3.0409**	**1.3186**	**0.6854**	**0.1722**	**0.1272**	0.0544	0.0228	6
[**ACI;A;I**]	**20.6997**	**1.9907**	**1.4585**	**0.7311**	**0.1473**	**0.1266**	0.0671	0.0206	6
[**ACI;C;I**]	**20.4028**	**3.2568**	**0.9582**	**0.9186**	0.0990	0.0794	0.0740	0.0212	6
[**ACI;A**]	**20.0495**	**1.6640**	**1.1829**	**0.1735**	**0.1377**	0.0606	0.0184	0.0140	6
[**ACI;C**]	**19.7519**	**2.9290**	**0.8944**	**0.4423**	0.0773	0.0610	0.0199	0.0137	6
[**ACI;I**]	**19.9147**	**1.9839**	**0.9239**	**0.4322**	0.0740	0.0403	0.0156	0.0139	6
[**ISU;I;S;U**]	**21.1487**	**2.9758**	**1.6909**	**0.8239**	**0.2386**	**0.1433**	0.0579	0.0266	6
[**ISU;I;U**]	**20.7669**	**2.1911**	**1.5199**	**0.5349**	**0.2239**	**0.1326**	0.0387	0.0205	6
[**ISU;S;U**]	**20.5307**	**2.6343**	**1.4071**	**0.5793**	**0.1617**	0.0999	0.0319	0.0132	6
[**ISU;I;S**]	**20.3777**	**2.9273**	**1.6078**	**0.6554**	**0.2147**	0.0666	0.0329	0.0142	6
[**ISU;I**]	**19.9840**	**2.0937**	**1.4088**	**0.3332**	**0.1870**	0.0520	0.0118	0.0112	6
[**ISU;S**]	**19.7321**	**2.6043**	**1.3086**	**0.2993**	0.0681	0.0324	0.0128	0.0109	6
[**ISU;U**]	**20.1305**	**1.9385**	**0.7610**	**0.4476**	**0.1167**	0.0388	0.0198	0.0117	6
[**ACIU;A;C;I;U**]	**27.1093**	**3.3750**	**1.5925**	**1.2603**	**0.4321**	**0.1984**	**0.1339**	0.0991	8
[**ACIU;A**]	**25.6417**	**1.7583**	**1.0279**	**0.3691**	**0.1244**	0.0698	0.0421	0.0172	8
[**ACIU;C**]	**25.3965**	**2.9969**	**1.1111**	**0.3312**	**0.1041**	0.0692	0.0393	0.0183	8
[**ACIU;I**]	**25.5180**	**2.1909**	**1.0261**	**0.3132**	**0.2316**	0.0724	0.0304	0.0163	8
[**ACIU;U**]	**25.6343**	**1.7267**	**0.8635**	**0.4719**	**0.1519**	0.0696	0.0300	0.0248	8

[a] Augmented data matrices are indicated using notation [$\mathbf{R_1}$;$\mathbf{R_2}$;$\mathbf{R_3}$] to refer to columnwise (wavelengthwise) augmentation (see Figure 2). [b] Bold numbers correspond to singular values associated to chemical species.

rank overlap in the spectral profiles was also present in this mixture in the spectral profiles.

For the drug, the experimental rank obtained was 3, in accordance with its theoretical rank considering that there are four absorbing species (the acid and basic forms of inosinate and cytosine) linked by two independent reactions (the two acid−base equilibria); i.e., a rank deficiency due to closure is also present in this case.

Table 5 includes the singular values obtained for some of the augmented matrices analyzed in the paper. When the rank deficiency is caused by closure, the inclusion of a new individual matrix within the augmented one can increase the rank of the previous matrix by one.[10,23,24] In the rank-deficient systems of Table 5, the inclusion of new information from the standard increases by one the number of independent concentration profiles (i.e., the number of independent reactions) and thus, the rank of the augmented matrices. It can be deduced that the rank of these augmented wavelengthwise matrices will then be equal to $r + k$, for $k < n$, where r is the rank of the initial matrix (that of the sample mixture), k is the number of different matrices of standards (each standard corresponds to the titration of a single compound) added to the system, and n is the total number of compounds present in the sample mixture. When the information concerning n − 1 titrations of different standards is included in the augmented data matrix, this matrix becomes full rank. However, since in the mixtures studied here a certain rank overlap both in the spectral profiles (see the case of adenosine before mentioned) and in the concentration profiles (for those compounds with similar pK_a) was also present, the rank results obtained for the different augmented matrices do not agree completely with those expected. Thus, for the system containing the ASI mixture matrix, full rank is only achieved for the augmented matrices [**ACI;A;C;I**], [**ACI;A;C**], and [**ACI;A;I**], i.e., for those augmented matrices with the information of adenosine as a pure standard. On the other hand, for the ISU system full rank is only achieved for the [**ISU;I;S;U**] and [**ISU;I;U**] augmented matrices. In this case, both the information relative to the uridine and to the inosine standards seem to be necessary in order to completely break the rank deficiencies. These two compounds are very similar with respect to acid−base properties and have similar acid and basic spectra (see Table 2). The rank analysis values of several augmented matrices containing the quaternary mixture are also included in Table 5.

Quantitation in Binary Mixtures. In the analysis of the binary mixtures, the following strategy was applied: one of the compounds is the analyte to be quantified while the other is considered as an unknown interference. In this case, the matrix of the binary mixture is simultaneously analyzed with the matrix of one of the analytes (standard); i.e., no information relative to the interference is introduced, for example, matrix [IU;I] or matrix [IU;U], depending on which is the analyte to be quantified. All these augmented matrices are full rank. So, the full resolution of the binary mixture system is possible, a priori, i.e., the recovery of the concentration and spectral profiles for all the species present, even for those not included as standards. Figure 3 shows, as an example, the concentration profiles obtained using MCR-ALS procedure in the analysis of the full-rank augmented matrix [**drug;S**]. The recovery of the concentration profiles of both the analyte (adenosine) and the interference (cytosine) was achieved correctly.

The quantitation results obtained for different analytes in binary mixtures are given in Table 6. This table also includes the percentage of ALS fitting error, which gives information related to the fit between the experimental absorbance data and the absorbance data calculated from the concentration and spectral profiles recovered mathematically by ALS. Quantitation is usually better from the acidic concentration profiles than from the basic ones. This is attributed to the fact that in the acidic region there is selectivity in all the matrices of the standards (both for compounds with pK_a values around 4 and for compounds with

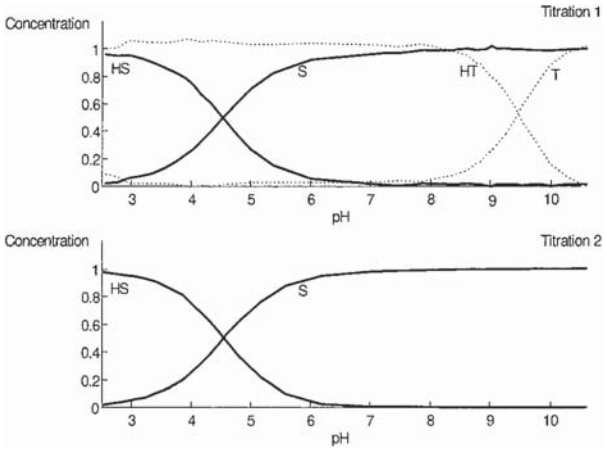

Figure 3. Concentration profiles obtained from the analysis of the augmented matrix [**drug;S**]. Titration 1: concentration profiles for matrix **drug**. Titration 2: concentration profiles for matrix **S**. HT, T, HS, and S refer to the protonated and deprotonated forms of inosinate and cytosine, respectively.

Table 6. Percentage of Prediction Error Obtained in the Quantitative Determination of the Analytes in Binary Mixtures, Using the Experimental Data Matrices[a,b]

augmented data matrices	species[c]		% fitting error[d]
	acidic	basic	
[CU;C]	4.77 (C)	7.40 (C)	0.69
[CU;U]	2.29 (U)	7.52 (U)	1.34
[AS;A]	0.40 (A)	2.37 (A)	0.85
[AS;S]	6.67 (S)	1.46 (S)	0.43
[ST;S]	0.52 (S)	0.09 (S)	0.44
[ST;T]	2.29 (T)	21.29 (T)	0.51
[IU;I]	24.0 (I)	23.1 (I)	2.05
[IU;U]	22.0 (U)	37.8 (U)	1.06
[drug;S]	2.81 (S)	0 (S)	0.62
[drug;T]	0.10 (T)	8.97 (T)	0.76

[a] The percentage of prediction error is calculated as follows: % PE $= 100(c_i^* - c_i)/c_i$, where c_i^* and c_i refer respectively to the calculated and real concentrations of the nucleic base that is quantified. [b] In parentheses, the analyte which is quantified in each case. See Table 1 for the notation of the analytes and matrices. [c] Quantitation of each analyte is performed using the concentration profile of the acidic species and of the basic species. [d] The percentage of fitting error is calculated as follows: % fitting error $= 100(\Sigma\Sigma(d_{if}^* - d_{if})^2/\Sigma\Sigma d_{if}^2)^{0.5}$, where d_{if}^* and d_{if} refer respectively to the calculated and real absorbance data.

pK_a values around 9). On the other hand, at basic pH values, the deprotonation of other acidic groups, such as hydroxyl groups in the ribose moieties, can occur.[3] Consequently, in some cases, the titrations were stopped at pH values low enough to neglect this second deprotonation step. Under these experimental conditions, the formation of the basic species was not complete and an important proportion of the acidic form of those compounds with pK_a values around 9 was still present (lack of selectivity). If titrations were carried on up to higher pH values, new species evolved, making the species resolution more difficult. All these difficulties are present in the quantitative determination of those analytes with pK_a values around 9 when using their basic concentration profiles. Quantitation from the acidic concentration profiles was the method chosen to apply to all the mixtures being analyzed (binary, ternary, and quaternary mixtures). In these conditions, prediction errors were in most of the cases lower than 5% (Table 6). The worst results were obtained in the analysis of matrix **IU**. This mixture was the most difficult since both analytes have pK_a values around 9. Thus, in the analysis of this mixture, the absorbance measured was constant until pH values around 8, where changes in the composition of the mixture began to appear due to the acid−base reactions. At this basic pH, resolution of the species profiles was more difficult. To improve the results obtained for this mixture, the analysis of the pH-synchronized augmented matrices was attempted (see data treatment). The prediction errors obtained are included in Table 7. A significant improvement was obtained for the system IU. For all other systems, the results do not differ from those obtained with the original data.

Table 7. Percentage of Prediction Error Obtained in the Quantitative Determination of the Analytes in Binary Mixtures, Using Interpolated Data Matrices[a−c]

augmented data matrices	% prediction error	% fitting error[d]
[**CU;C**]	5.78 (C)	0.60
[**CU;U**]	3.07 (U)	1.23
[**AS;A**]	6.46 (A)	0.91
[**AS;S**]	4.61 (S)	0.80
[**ST;S**]	2.77 (S)	0.43
[**ST;T**]	0.02 (T)	0.71
[**IU;I**]	0.22 (I)	2.32
[**IU;U**]	1.62 (U)	0.88
[**drug;S**]	0.62 (S)	0.58
[**drug;T**]	0.02 (T)	0.80

[a] As in Table 6. [b] As in Table 6. [c] Quantitation is performed from the acidic concentration profiles. [d] As in Table 6.

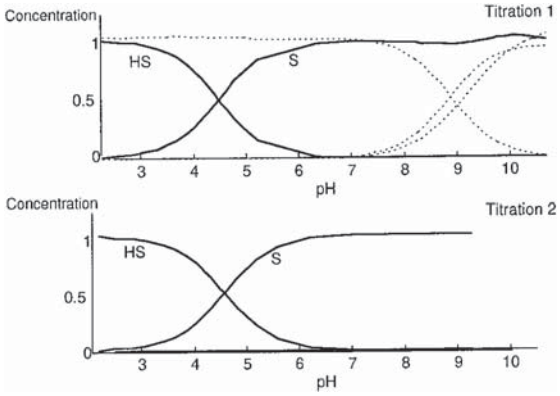

Figure 4. Concentration profiles obtained from the analysis of the augmented matrix [**ISU;S**]. Titration 1: concentration profiles for matrix **ISU**. Titration 2: concentration profiles for matrix **S**. HS and S refer to the protonated and deprotonated forms of cytosine, respectively.

Quantitation in Ternary and Quaternary Mixtures. In the analysis of the ternary and quaternary mixtures, the same strategy as in the analysis of the binary mixtures was considered: here one of the compounds to be quantified is the analyte while the others are considered as unknown interferences. In these cases, the analysis was performed on rank-deficient augmented matrices (i.e., matrices [**ACI;A**], [**ACI;C**], or [**ACI;I**], for the quantitation of adenine, cytidine, or inosine, respectively). Full resolution of

Table 8. Percentage of Prediction Error Obtained in the Quantitative Determination of the Analytes in Ternary and Quaternary Mixtures, Using the Experimental Data Matrices[a−c]

augmented data matrices	% prediction error	% fitting error[d]
[**ACI;A**]	5.57 (A)	0.60
[**ACI;C**]	1.01 (C)	0.43
[**ACI;I**]	4.09 (I)	1.13
[**ISU;I**]	5.27 (I)	1.13
[**ISU;S**]	1.12 (S)	0.63
[**ISU;U**]	2.44 (U)	0.60
[**ACIU;A**]	2.23 (A)	1.00
[**ACIU;C**]	4.43 (C)	1.81
[ACIU;I]	0.35 (I)	1.04
[ACIU;U]	1.44 (U)	0.55

[a−d] As in Table 7.

the rank-deficient systems was not possible, but quantitation of the analyte can be performed since the recovery of its concentration and spectral profiles (present in both matrices, that of the mixture and that of the standard) was still possible. Figure 4 shows, as an example, the concentration profiles obtained by MCR-ALS for the rank-deficient augmented matrix [**ISU;S**]. Table 8 includes the percentage of prediction error obtained for the different analytes in the ternary and quaternary mixtures. The results (prediction error not higher than 5% in most cases) show that accurate quantitation can be obtained in rank-deficient systems too. These results confirm the use of the method in determining the analyte in complex systems and in the presence of unknown interferences.

CONCLUSIONS

Second-order multivariate curve resolution was successfully applied to the quantitative analysis of mixtures of nucleic acid components using spectrophotometric acid−base titrations.

Quantitative determination of analytes in the presence of unknown and uncalibrated interferences was correctly achieved, with prediction errors around 5% or even lower. The proposed procedure has many advantages over first-order multivariate calibration procedures. A single synthetic standard is enough to perform quantitation when spectroscopic and acid−base properties are both simultaneously taken into account (second-order data with higher chemical selectivity). Furthermore, due to the simplicity of the experimental methodology and the instrumentation, the procedure proposed constitutes a feasible inexpensive alternative to HPLC or capillary electrophoresis for quantitative analysis of nucleic acid component mixtures, such as those present in pharmaceuticals.

As the acid−base titration is the most time-consuming step in the procedure proposed here, further work is being carried out in order to increase the speed of the analyses and to obtain experimental synchronization both in the wavelength and in the pH orders. Flow injection analysis and continuous-flow analysis seem to be the best strategies to optimize the experimental procedure.

ACKNOWLEDGMENT

This research has been funded by DGICYT (Spain) project (PB96-0377).

Received for review April 13, 1998. Accepted September 28, 1998.

AC980409V