ELSEVIER

Chemometrics and Intelligent Laboratory Systems 50 (2000) 263–271

Chemometrics and intelligent laboratory systems

www.elsevier.com/locate/chemometrics

pH-Gradient spectrophotometric data files from flow-injection and continuous flow systems for two- and three-way data analysis

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

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

Received 30 August 1999; received in revised form 30 November 1999; accepted 6 December 1999

Abstract

This paper describes a contribution to Chemometrics and Intelligent Laboratory Systems Elsevier's data base of files. Mixtures of nucleosides and a pharmaceutical preparation composed of cytosine and inosine-5′-monophosphate are analyzed using three different pH-gradient spectrophotometric flow-injection and continuous flow systems. Results for one of the flow-injection systems and for the continuous-flow system are reported for the first time in this paper. Resolution and quantitation of the constituents in the mixtures are given using a multivariate curve resolution method based on a constrained alternating least squares optimization.

Keywords: Acid–base spectrophotometric data; Three-way data analysis; Rank deficiency; Multivariate curve resolution; pH-gradients; Flow systems

1. Introduction

Flow-injection analysis and related flow techniques with diode-array spectrophotometric detection provide large amounts of data in a very simple and fast way. This work presents data files from three different flow systems: two flow-injection systems referred to as FIA-A and FIA-B, and one continuous-flow system referred to as CCF-A. These three flow systems were designed to generate an on-line pH-gradient in the time domain.

The analytes selected in this study were four nucleosides (adenosine, cytidine, inosine and uridine) and the components of a pharmaceutical drug (cytosine and inosine-5′-monophosphate). Some of these compounds, such as adenosine, cytosine and cytidine have pK_a values around 4, while uridine, inosine and inosine-5′-monophosphate have pK_a values around 9. Since the spectra of the corresponding acidic and basic species of the same compound are different, pH variations in the flow system lead to changes in the shape of measured spectra over flow time. The set of time-ordered spectra obtained in the analysis of each sample gave a two-way data matrix. If several samples were simultaneously analyzed, a three-way data set was obtained.

Data from FIA-A, FIA-B and CCFA systems were analyzed using a multivariate curve resolution procedure based on alternating least squares (MCR-ALS method) [1–3]. In the present paper, the particular features and modeling strategies to resolve each flow system are outlined. Files giving the results of the

* Corresponding author. Tel.: +34-93-403-4445; fax: +34-93-402-1233.

E-mail address: xavi@zeus.qui.ub.es (J. Saurina).

0169-7439/00/$ - see front matter
PII: S0169-7439(99)00067-2

resolution and quantitation of the individual constituents in the sample mixtures are given for reference and comparison with results obtained using other chemometric methods.

The experimental data in the present study may be of interest to the chemometrics community for the following reasons.

(1) To compare and demonstrate the utility of second-order calibration and three-way data analysis methods in the determination of a particular analyte in an unknown mixture. In particular, the calibration of an analyte in the presence of interferences using a single standard for the analyte is studied in detail (second-order advantage [4]).

(2) To study the effect of rank deficiency in resolution and quantitation of constituents in mixtures. Rank-deficient systems are systems where the number of components detected by rank analysis is lower than the real number of components of the system [5,6]. Two aspects of rank-deficiency systems can be studied in particular:

- The use of matrix augmentation to break rank deficiency and get full rank matrices [6,7].
- The quantitation and resolution of a particular analyte in a rank-deficiency system.

2. Experimental and material

2.1. Reagents

All chemicals were of analytical grade. Ultrapure water was used for all solutions. Titrant solutions were prepared from 85% (w/w) phosphoric acid solution (Fluka) and sodium phosphate (Scharlau). In the flow-injection system A (FIA-A), the acidic titrant solution was 0.025 M phosphoric acid and the basic titrant solution 0.1 M sodium phosphate. In the flow-injection system B (FIA-B), the titrants were a 0.025-M phosphoric acid solution and a 0.025-M sodium phosphate solution. In the continuous-flow system (CCFA), the titrants were a 0.1-M phosphoric acid solution and a 0.1-M sodium phosphate solution.

Pure standards for adenosine, cytidine, inosine, uridine, cytosine and inosine-5′-monophosphate were prepared, which were noted as A, C, I, U, S and T, respectively (see Fig. 1 for their structures). Binary, ternary and quaternary mixture solutions of nucleosides used as unknown solutions where: cytidine + uridine, noted as CU sample; inosine + uridine, as IU; cytidine + inosine + uridine, as CIU; cytidine + inosine + adenosine, as CIA; cytidine + inosine + adenosine + uridine, noted as CIAU. In all cases, the concentration of each individual compound was 5×10^{-5} M.

A drug consisting of an equimolar proportion of inosine-5′-monophosphate and cytosine (Precurgen® from Knoll) was analyzed. An unknown solution containing about 5×10^{-5} M of each component was prepared. This sample solution was referred to as Drug.

2.2. Apparatus

A Hewlett-Packard HP8452A diode array spectrophotometer using a Hellma flow cell of 10 mm pathlength and 18 μl volume was used. The acquisition and storage of spectrophotometric data was carried out with a Hewlett-Packard software by using a Hewlett-Packard Vectra N2 4/50 computer. A Scharlau HP4 Peristaltic pump and a Sirtex VM-8 injection valve were also used.

2.3. Flow-injection and continuous-flow systems

The flow system used to generate the data has been described in detail elsewhere [3]. The two-channel FIA-A injected 300 μl of alkalinized sample (sample solution + 0.1 M Na_3PO_4 solution) into a 0.025-M H_3PO_4 acidic carrier. The three channel FIA-B injected a bolus of sample (175 μl) between a 0.025-M H_3PO_4 acidic solution (in the front of the FI peak) and a Na_3PO_4 basic solution (in the tail of FI peak). CCFA aspirated continuously the sample solution, which is mixed with either a H_3PO_4 or a Na_3PO_4 solution (see Fig. 2).

Flow rates for all channels were 0.5 ml/min. In all these cases, the controlled partial dispersion of acidic and basic zones, which occurred in a 200 cm × 0.5 mm i.d. reaction coil, was responsible for the generation of on-line pH-gradients in a reproducible manner (see Fig. 2 for an approximation to the shape of each pH-gradient). Spectrophotometric flow data

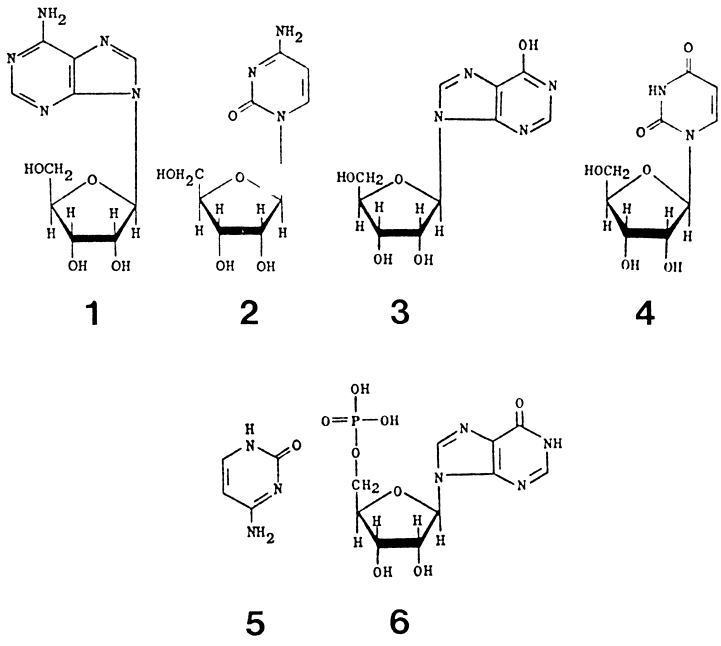

Fig. 1. Scheme of compound under study. Assignment: 1 = Adenosine; 2 = Cytidine; 3 = Inosine; 4 = Uridine; 5 = Cytosine; 6 = Inosine-5′-monophosphate.

were monitored in the range 220 to 350 nm at regular steps of 2 s for 4 min (FIA-A and FIA-B systems) or for 2 min (CCFA system).

2.4. Description of files

Two types of MATLAB files (in MATLAB version 5.0) are provided.

· Data files: FIAA-DAT.MAT, FIAB-DAT.MAT and CCFA-DAT.MAT are the original files of the pH-gradient spectrophotometric flow data for FIA-A, FIA-B and CCFA systems, respectively. The concentration of each constituent is in all samples equal to 5×10^{-5} M.

In each data file, the following matrices were included: (a) Data matrices of pure single compounds: A = adenosine; C = cytidine; I = inosine; U = uridine; S = cytosine; T = inosinate. (b) Data matrices of mixtures: CU = cytidine + uridine; IU = inosine + uridine; CIU = cytidine + inosine + uridine; CIA = cytidine + inosine + adenosine; CIAU = cytidine + inosine + adenosine + uridine; DRUG = pharmaceutical drug containing inosinate and cytosine.

· Results files: FIAA-RES.MAT, FIAB-RES.MAT and CCFA-RES.MAT have the concentration profiles and spectra recovered for each case using MCR-ALS for FIA-A, FIA-B and CCFA systems, respectively. The nomenclatures in these results files were as follows: First letter: a = spectral profiles; c = concentration profiles. Uppercase letters refer to the matrix (see Table 1). Last letter refers to the standard used in the matrix augmentation: a = adenosine; c = cytosine; i = inosine; u = uridine; s = inosinate; t = cytosine.

2.5. Data matrices for study

Matrices of FIAA-DAT.MAT, FIAB-DAT.MAT and CCFA-DAT.MAT files can be studied for rank analysis, resolution and quantitation purposes. Apart from the individual matrices listed above, the following column-wise augmented data matrices were analyzed: [CU;C], [CU;U], [IU;I], [IU;U], [DRUG;S], [DRUG;T], [CIU;C], [CIU;I], [CIU;U], [CIA;C], [CIA;I], [CIA;A], [CIAU;C], [CIAU;I], [CIAU;A] and [CIAU;U]. For instance, the column-wise data matrix [CU;C] is a composite of the two individual data ma-

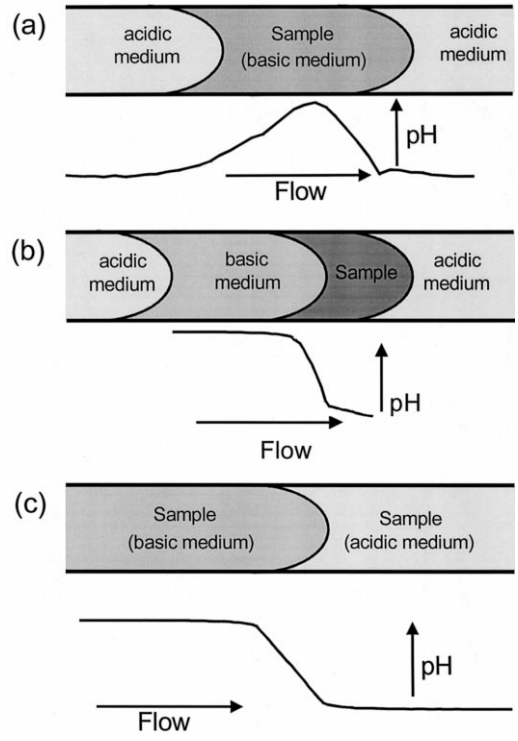

Fig. 2. Scheme of the flow segments of sample, acidic and basic flow zones and shapes of the pH gradient profiles. (a) FIA-A system; (b) FIA-B system; (c) CCFA system.

trices CU and C, setting matrix CU on top of matrix C, sharing the same column (wavelengths) indexes and the same column vector space (spanned by a common vector basis). In this particular case, both individual matrices have in common cytidine, C, which is considered the analyte. Uridine, U, is only present in matrix CU, and acts as interference. Therefore, in the augmented matrix [CIAU;C], the analyte is cytidine, C, and there are three interferences inosine, adenosine, and uridine. The analytical problem to be solved is the determination of cytidine in the mixture of this compound with inosine, adenosine and uridine giving matrix CIAU. The only information given is about cytidine in the data matrix C which only contains it. Note, however, that other different combinations of column-wise or row-wise augmented data matrices can be built by the user for further studies.

2.6. Characteristics of data

- A single run of any of the three flow spectrophotometric systems gives a two-way data matrix.
- N(3) deprotonation of pyrimidinic nucleosides (cytidine and uridine) produced the spectral changes monitored in this study. Deprotonation of hydroxylic groups of ribose residues at pH > 12 did not detect any changes since this process occurred far away from the chromophore moiety. Purinic nucleosides have an additional deprotonation in N(7) at very acidic medium (pK_a is < 1); thus, this process is out of the pH range of the study, and only one equilibrium is monitored spectrophotometrically. With respect to the components of the drug, cytosine pK_as are 4.56 (N(3)) and 11.7 (N(1)), and both processes can be followed spectrophotometrically since they affect the chromophore group. However, N(1) deprotonation is strongly interfered by the presence of the

Table 1
Singular values for the individual and augmented data matrices for the FIA-B system
Bold numbers correspond to singular values associated to significant contributions.

Matrices	Singular values								Real number of contributions
	1	2	3	4	5	6	7	8	
I	**1.4176**	**0.2251**	**0.0874**	0.0261	0.0187	0.0092	0.0028	0.0016	3
C	**1.5196**	**0.3525**	**0.1459**	0.0405	0.0178	0.0080	0.0030	0.0010	3
CI	**2.5021**	**0.3047**	**0.1420**	**0.0557**	0.0317	0.0145	0.0129	0.0024	6
CIU	**3.6965**	**0.4435**	**0.1299**	**0.0597**	0.0255	0.0214	0.0124	0.0023	9
CIAU	**4.5791**	**0.3978**	**0.1792**	**0.0669**	0.0219	0.0178	0.0049	0.0025	12
[CI;C]	**2.9258**	**0.4713**	**0.3703**	**0.1255**	**0.0730**	**0.0565**	0.0265	0.0100	6
[CIU;C]	**3.9703**	**0.5427**	**0.5075**	**0.1202**	**0.0669**	**0.0507**	0.0288	0.0119	9
[CIAU;C]	**4.7958**	**0.5775**	**0.5102**	**0.1404**	**0.0663**	**0.0473**	0.0240	0.0070	12

phosphate titrant absorption, so that the spectrophotometric study of the N(1) deprotonation is masked. Despite having a phosphate group with a $pK_a = 6.00$, inosine-5′-monophosphate behaves similarly to inosine as the corresponding deprotonation does not produce changes in the spectral region under study.

· Two-way data acquired in one experimental run are bilinear since Beer's linearity law holds for each of the species at equilibrium.

· The three pH-gradient flow systems gave data with a run-to-run pH gradient synchronization; in addition, every species in acid–base equilibrium is defined by the same concentration profile and pure spectrum. As a result, a trilinear data structure is expected when several two-way data runs are simultaneously analyzed (three-way data). All chemometric methods based on the compliance of a trilinear model may be applied to the data (i.e., GRAM [4,8], PARAFAC [9], MCR-ALS [1,2,10]).

· Mixtures of components with acid–base behavior give rank-deficient data matrices when a pH gradient is applied [3,6,7]. This feature is encountered in data matrices from binary, ternary and quaternary mixtures of nucleosides and drug components. Data matrices with a single component are full rank data matrices.

2.7. Methods

2.7.1. Rank analysis

The chemical rank of individual and/or augmented data matrices was estimated by singular value decomposition (SVD). It was assumed that larger singular values corresponded to the more significant chemical components (see Table 1 for rank-analysis examples).

2.7.2. Resolution and quantitation

The concentration profiles and spectra of the constituents in the mixtures were recovered using the MCR-ALS method [1,2]. Initial estimates of the concentration profiles of all components are given. For each analyte, they were obtained from the analysis of the matrix of its standard with techniques based on the detection of purest variables. For the interferences, they were obtained in a similar way from the analysis of the sample matrix. During the MCR-ALS optimization, constraints were applied to get chemically meaningful solutions (e.g., non-negativity for spectra and concentration profiles, correspondence between common species in different runs and trilinearity [10]). For a more detailed description of the MCR-ALS method for three-way data analysis and for rank-deficient systems, see Refs. [1–3,6,7,10].

3. Results and discussion

For FIA-A data, a detailed description of the resolution strategy and of the results achieved in the quantitative analyses of the experimental data were given in Ref. [3]. Results for FIA-B and CCFA data systems are reported for the first time in this paper.

3.1. FIA-A data analysis

In this case, single constituent data matrices were modeled with two components for the analyte (one component for the acidic species and one for the basic species). In contrast to the other two flow systems described below, the pH and phosphate concentration in the dispersed bolus of sample was sufficiently low to avoid the formation of further background contributions. Sample mixtures were modeled with four components, one component for the analyte acidic species, one component for the analyte basic species and two additional components for the interfering contributions. As a general rule, quantitation using the acidic species was more accurate than using the basic species. Using the acidic species, the overall prediction error in the determination of each particular analyte in the binary, ternary and quaternary mixtures was 2.7%, 5.7%, and 6.6%, respectively.

3.2. FIA-B data analysis

Fig. 3 shows the experimental spectra obtained in the titration of a cytidine solution using the FIA-B system. According to the composition of phosphoric acid and phosphate solutions, the pH range under study was from 2 to 12, approximately. Three different successive pH flow zones with the predominance of the acidic species (0 to 80 s), the basic species (80 to 120 s) and the background contributions (from 120 s) were detected. Rank analysis of data matrices from

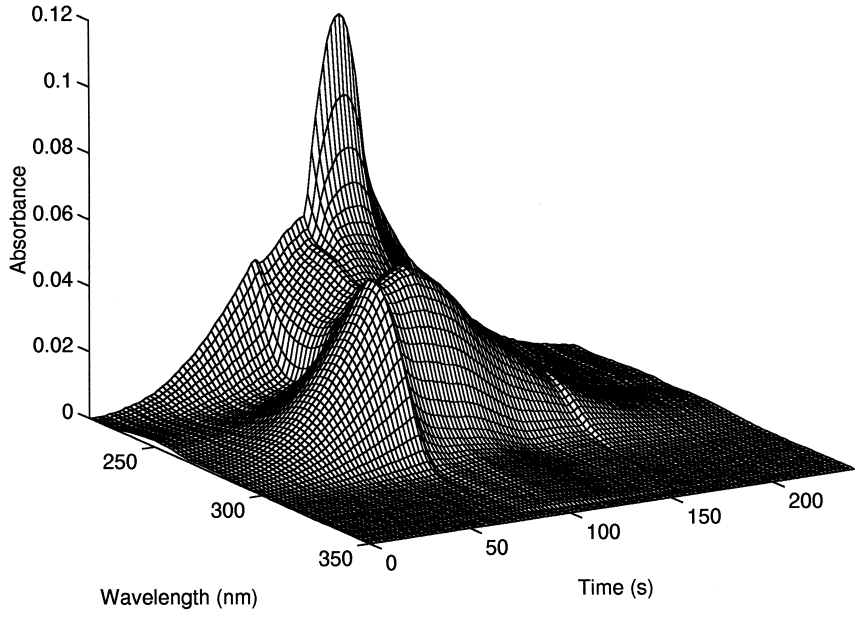

Fig. 3. 3-D plot of FIA-B data from a 5×10^{-5} M cytidine solution.

a single nucleic acid component confirmed this, since the number of significant components estimated by SVD was equal to three. Data matrices corresponding to the mixtures of nucleosides or drug components were rank-deficient [5–7]. As is shown in Table 1, when column-wise augmented matrices were analyzed, rank deficiency was solved only for binary mixtures. For ternary and quaternary mixtures rank deficiency was still present.

In the MCR-ALS analysis of the augmented data matrices, the component of interest (that which is in both individual matrices present in the column-wise augmented matrix) was modeled with three contributions. As interferences were expected to behave similarly to the analyte, three interfering contributions were also taken into account in any case. Then, a six-species model was used for the resolution of any augmented matrix, even though the number of real chemical species from the interference was higher than three for ternary and quaternary mixtures (six and nine interfering species, respectively). This did not hinder the correct resolution of the two concentration and spectra profiles of the analyte (one for the acidic and one for the basic species). However, the concentration and spectra profiles recovered for the generic interference were, in fact, linear combinations of the real concentration and spectra profiles of the interferences. Models with five and seven species were also tested. However, the best resolution from a chemical point of view was obtained with six-species models. In addition, quantitations using six species were also more accurate.

As in Sections 3.1 and 3.3, the most accurate quantitations were obtained using the acidic species. Table 2 gives the results obtained in the determination of nucleosides and drug components from concentration profiles of the acidic species. The percentage of error in the prediction of binary mixtures was about 5%, and increased up to 13.6% when more complex samples were analyzed. As an example, Fig. 4 shows the concentration profiles and spectra recovered in the resolution of the augmented matrix [CIAU;C] (three interferences) using MCR-ALS and flow injection system FIA-B.

3.3. CCFA data analysis

Fig. 5 shows the spectra obtained in the analysis of a cytidine solution by using the continuous flow

Table 2
Determination of nucleic acid compound in unknown mixtures in the presence of interference using pure standards of the analyte by the MCR-ALS method
Results from the analysis of the spectrophotometric acid–base titration obtained with the FIA-B system.

Matrices	Analyte	Concentration found	Quantification error (%)	% Fitness error
[CU;C]	cytidine	4.66	6.8	16.48
[CU;U]	uridine	4.65	6.9	6.54
[IU;I]	inosine	4.88	2.3	8.40
[IU;U]	uridine	5.18	3.6	7.90
[Drug;S]	cytosine	6.18	23.7	11.8
[Drug;T]	inosinate	4.53	9.4	7.05
[CIU;C]	cytidine	8.47	69.7	7.69
[CIU;I]	inosine	4.59	8.3	10.7
[CIU;U]	uridine	5.28	5.6	11.4
[CIA;I]	inosine	4.80	4.0	13.2
[CIA;A]	adenosine	4.92	1.6	10.7
[CIAU;C]	cytidine	4.90	2.1	4.32
[CIAU;I]	inosine	3.63	26.5	6.93
[CIAU;A]	adenosine	5.21	4.2	6.39
[CIAU;U]	uridine	5.01	0.1	6.85

method. pK_a values reported in the literature for cytidine were 4.13 (N(3)) and 12.5 (sugar). However, the ribose deprotonation practically does not affect the chromophore absorption since its equilibrium

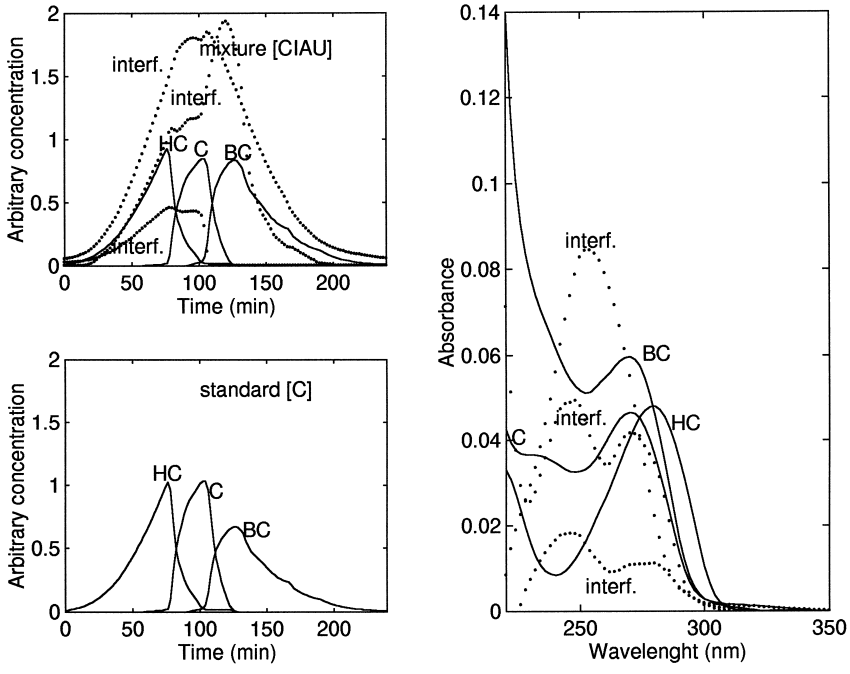

Fig. 4. Concentration profiles and spectra recovered in the resolution of the augmented matrix [CIAU;C] obtained with the FIA-B using the MCR-ALS method. Species identification: HC, acidic cytidine species; C, basic cytidine species; BC, background species; interf., interfering species.

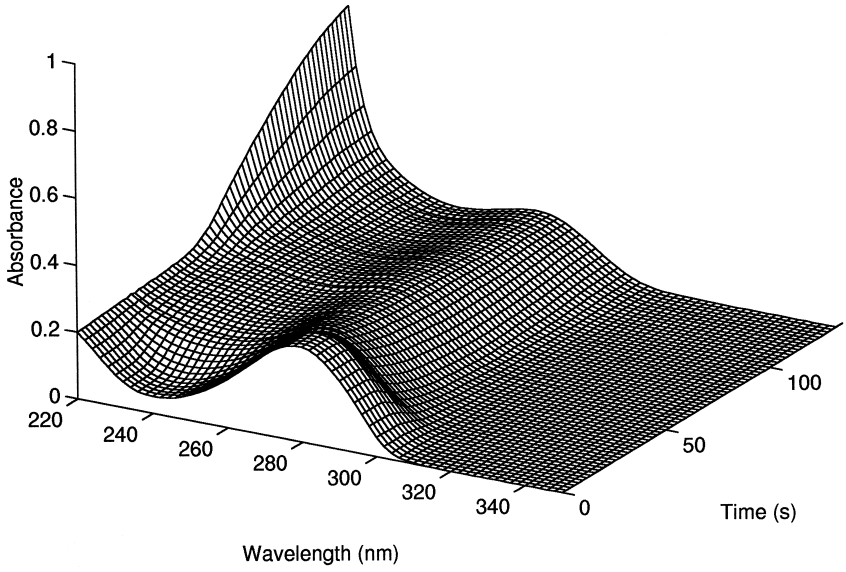

Fig. 5. 3-D plot of CCF-A data from a 5×10^{-5} M cytidine solution.

cannot be detected spectrophotometrically. As a result, spectral changes observed in Fig. 5 for cytidine are explained as follows: initially, the acidic species of the nucleoside was the only species present in the system, whereas the basic species and background contributions appeared later. This corroborated the

Table 3
Determination of nucleic acid compound in unknown mixtures in the presence of interference using pure standards of the analyte by the MCR-ALS method
Results from the analysis of the spectrophotometric acid–base titration obtained with the CCF-A system.

Matrices	Analyte	Concentration found	Quantification error (%)	% Fitness error
[CU;C]	cytidine	4.91	1.9	1.07
[CU;U]	uridine	4.70	6.0	1.18
[IU;I]	inosine	4.45	11.0	2.02
[IU;U]	uridine	5.06	1.3	1.81
[Drug;S]	cytosine	5.12	2.4	1.74
[Drug;T]	inosinate	4.89	2.2	1.61
[CIU;C]	cytidine	6.95	39.4	1.05
[CIU;I]	inosine	4.70	6.1	1.52
[CIU;U]	uridine	4.84	3.3	1.50
[CIA;C]	cytidine	5.18	3.6	1.31
[CIA;I]	inosine	6.90	38.2	2.68
[CIA;A]	adenosine	5.01	0.2	1.27
[CIAU;C]	cytidine	5.46	9.2	1.45
[CIAU;I]	inosine	4.59	8.3	2.52
[CIAU;A]	adenosine	5.13	2.6	2.44
[CIAU;U]	uridine	5.24	4.8	2.28

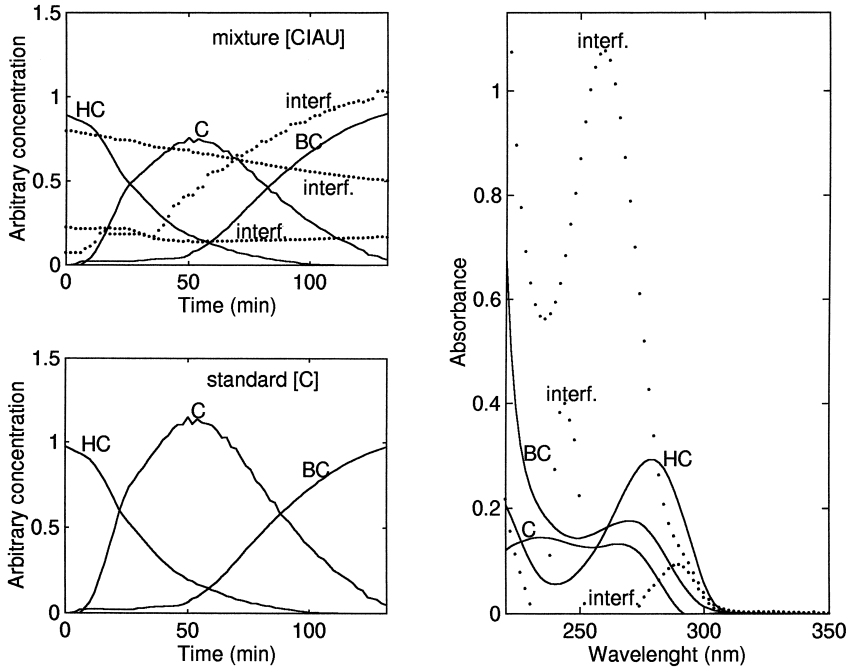

Fig. 6. Concentration profiles and spectra recovered in the resolution of the augmented matrix [CIAU;C] obtained with the CCF-A system using MCR-ALS method. Species identification: HC, acidic cytidine species; C, basic cytidine species; BC, background species; interf., interfering species.

SVD results of the corresponding data matrix, which suggested also the presence of three significant components. Similar conclusions were obtained from the study of other pure standards.

Similarly to FIA-B data, the resolution of the augmented matrices considered three species for the analyte and three more for the interference. This was the model assumed in all cases, even though in the analysis of the ternary and quaternary mixtures, more interfering contributions were really present.

The overall prediction error with the acidic species was 8.8%, and with the basic species 19.9%. Table 3 shows the results of the quantitative determination of various nucleoside mixtures and the components of the pharmaceutical drug using the concentration profiles recovered for the acidic species.

Concentration profiles and spectra obtained in the resolution of the augmented matrix [CIAU;C] using the MCR-ALS method are given in Fig. 6. Three contributions for acidic, basic and background species were obtained for the standard sample of cytidine, whereas three additional profiles associated with a linear combination of the interference of inosine, adenosine and uridine were obtained in the CIAU mixture.

References

[1] R. Tauler, A. Izquierdo-Ridorsa, E. Casassas, Chemom. Intell. Lab. Syst. 18 (1993) 293–300.
[2] R. Tauler, A. Smilde, B.R. Kowalski, J. Chemom. 9 (1995) 31–58.
[3] J. Saurina, S. Hernández-Cassou, R. Tauler, A. Izquierdo-Ridorsa, Anal. Chem. 71 (1999) 2215–2220.
[4] E. Sanchez, B.R. Kowalski, J. Chemom. 4 (1990) 29–45.
[5] M. Amrhein, B. Srinivasan, D. Bonvin, M.M. Schumacher, Chemom. Intell. Lab. Syst. 33 (1996) 17–33.
[6] A. Izquierdo-Ridorsa, J. Saurina, S. Hernández-Cassou, R. Tauler, Chemom. Intell. Lab. Syst. 18 (1993) 293–300.
[7] J. Saurina, S. Hernández-Cassou, R. Tauler, A. Izquierdo-Ridorsa, J. Chemom. 12 (1998) 183–203.
[8] E. Sanchez, B.R. Kowalski, Anal. Chem. 58 (1986) 496–499.
[9] R. Bro, Chemom. Intell. Lab. Syst. 38 (1997) 149–171.
[10] R. Tauler, I. Marqués, E. Casassas, J. Chemom. 12 (1998) 55–75.