A Soft-Modeling Approach to Interpret Thermodynamic and Conformational Transitions of Polynucleotides

Anna de Juan, Anna Izquierdo-Ridorsa, Romà Tauler, Gemma Fonrodona, and Enric Casassas
Departament de Química Analítica, Universitat de Barcelona, Barcelona, Spain

ABSTRACT Multivariate outputs from the experimental monitoring of biochemical processes are usually difficult to interpret applying methods based on a priori chemical models. Curve resolution methods are model-free procedures, generally known as soft-modeling methods, which obtain the concentration profiles and instrumental responses of each individual species involved in a multivariate monitored process without making any kind of external assumption. Of the curve resolution methods available, the alternating least squares (ALS) is proposed here because of its ability to operate on one or on several matrices. Furthermore, ALS allows the introduction of information related to the internal data structure and to the general features of the concentration profiles and instrumental responses through the input of suitable constraints in the iterative resolution procedure. The ALS potential is tested on several data sets coming from the multivariate spectrometric monitoring of polyuridylic (polyU), polycytidylic (polyC), and polyadenylic (polyA) protonation equilibria in dioxane/water 30% (v/v). Information concerning the evolution of the concentration profiles and the spectra of each individual species involved in the acid-base equilibria, the presence and pattern of polyelectrolyte effects, and the presence of conformational transitions associated or not with the proton uptake process is presented.

INTRODUCTION

Currently, the experimental monitoring of biochemical processes is relatively straightforward due to the instrumental techniques and data acquisition systems available to the researcher. The typical output of an instrument used in monitoring the evolution of a process according to the variation of a certain chemical variable consists of arrays of data (e.g., spectra) recorded at certain stages during the reaction (e.g., pH values, temperature, solvent polarity, etc.). These data can be organized in a data matrix, where the rows contain the instrumental responses and the columns reflect the relationship between the variation of the chemical variable and the evolution of the concentration of the species in the process (see Fig. 1).

Despite the availability of such experimental data, univariate monitoring is still widely used in many chemical fields, such as biochemistry. Traditional biochemical studies tend to focus on either obtaining structural information under fixed conditions (e.g., physiological conditions) or on studying dynamic processes by recording univariate measurements (e.g., melting studies using single-wavelength absorbance readings). In the first approach, the information contained in one row of the data matrix described above is used, and ambiguous structural information is obtained as no evidence of coexistence of species can be inferred from a single array of data. Even data treatments such as spectral deconvolution are known to provide highly uncertain results when dealing with this kind of measurement. In the second approach, a column of the data matrix described in Fig. 1 is employed, and the ambiguity related to univariate measurements is again apparent and, although evolutionary information is obtained, no detection of coexisting species is possible and all structural information is lost as only single-wavelength readings are collected. To overcome these shortcomings it is necessary to analyze the whole data matrix. The effective handling of this multivariate output is actually the critical step in the complete interpretation of biochemical processes.

Experiments conducted with small molecules have been successfully interpreted by applying classical iterative least-squares methods based on the refinement of a postulated chemical model to obtain the optimal fit to the experimental data (Leggett, 1977). This approach is often applied to the data coming from the spectrometric monitoring of chemical equilibria and is also known under the name of global analysis. The clear understanding of these simple equilibria allows the following assumptions in the process of model building: 1) the fulfillment of the mass action law (i.e., the validity of a fixed equilibrium constant throughout the reaction) and 2) the one-to-one correspondence between instrumental response and chemical species.

Nevertheless, most of these classical procedures are unable to interpret many biochemical processes due to the macromolecular nature of many biomolecules (e.g., polynucleotides, proteins, etc.), which causes a more complex evolution of the processes. The inapplicability of the classical methods can be explained by 1) the existence of polyelectrolyte effects and 2) the existence of conformational transitions.

Received for publication 17 January 1997 and in final form 8 September 1997.

Address reprint requests to Dr. Anna de Juan, Departament de Química Analítica, Facultat de Químiques, Diagonal, 647, 08028 Barcelona, Spain. Tel.: 34-3-4021286; Fax: 34-3-4021233; E-mail: annaj@zeus.qui.ub.es.

0006-3495/97/12/2937/12 $2.00

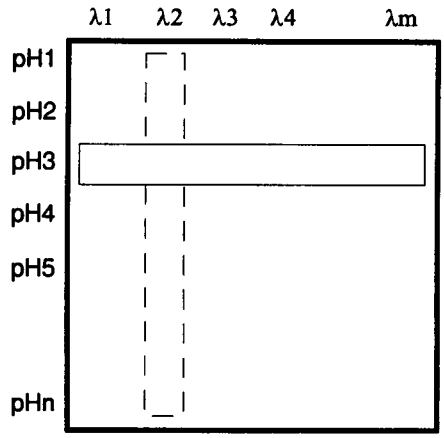

FIGURE 1 Experimental output from the multivariate monitoring of a pH-dependent process (*thick line*), an experiment performed at fixed pH conditions (*thin line*), and a univariate monitoring of a pH-dependent process (*dashed line*).

Existence of polyelectrolyte effects

The mass action law is no longer valid when either important changes in the electric field on the surface of the macromolecule or other effects related to conformational transitions modify the tendency of analogous sites to react. If this happens,

$$\log K = f(\alpha)$$

where α denotes the extent of the reaction. Mathematical expressions that might model this effect cannot be devised as no information concerning its existence and its pattern (i.e., linear or nonlinear) for an unknown macromolecular process is available beforehand.

Existence of conformational transitions

Macromolecular biomolecules can show conformational transitions associated with a chemical reaction or with a spatial rearrangement of the molecule. If the latter phenomenon occurs, the number of structural species will exceed the number of chemical species and the assumption of a one-to-one correspondence between chemical species and spatial configurations will be erroneous.

Of those techniques available for analyzing multivariate data sets, the self-modeling curve resolution methods are the most appropriate for the analysis of evolutionary chemical processes when neither models nor previous information about the number and identity of the species involved are available, as is the case with biomacromolecular equilibria (Kvalheim and Liang, 1992; Malinowski, 1996; Vandeginste et al., 1985; Tauler et al., 1995; Tauler, 1996).

The step common to all resolution methods is the decomposition of the original data matrix **D**, which contains the mixed information of the monitored chemical process, into the product of two smaller matrices **C** and **S**, which contain the concentration profiles and the pure instrumental responses of each individual species present in the reaction, respectively. Once **C** and **S** are known, additional chemical information related to the analyzed data can be obtained. The only requirement of these methods is that the contributions of the pure instrumental responses (be they from a chemical species or a conformer) to the overall measurement behave in accordance with an additive model, in the same way spectrometric measurements do given that they fulfill the Beer-Lambert law. The bilinearity condition (i.e., that the original matrix must be expressed as a sum of additive contributions, formed by the outer product of the concentration profile and the unit spectrum of each individual species in the system), stems from the kind of data decomposition performed by the curve resolution methods. Nevertheless, one must be aware of false nonbilinear situations that can appear as a consequence of a nonappropriate definition of the species. (For example, in this text, if only the chemical species are considered, a species with two conformers would have one concentration profile and two spectra. This apparent nonbilinear situation disappears when the individual contributions are defined as absorbing species and not as chemical species.) The only systems strictly nonbilinear are those for which the instrumental measurements are intrinsically nonadditive.

To illustrate the usefulness of the soft-modeling curve resolution methods in the interpretation of biomacromolecular processes, the spectrometric (UV and CD) study of the acid-base behavior of the homopolynucleotides polyuridylic acid (polyU) (de Juan et al., 1996), polycytidylic acid (polyC) (de Juan et al., 1997a), and polyadenylic acid (polyA) in dioxane/water 30% (v/v) is shown by way of example. Information about the identification and evolution of the chemical species, the existence and pattern of the polyelectrolyte effect, and the conformational transitions associated with the studied reactions is presented. The solvent effect on these equilibria can be determined by comparing the results obtained in the hydroorganic mixture with those previously reported in aqueous solution using the same resolution method.

MATERIALS AND METHODS

Experimental protocol

The experimental protocol followed in the study of the acid-base behavior of the homopolynucleotides includes the combination of spectrometric and potentiometric titrations of the polymers polyA, polyC, and polyU and the comparison of these results with those obtained for their corresponding cyclic nucleotides, adenine-3′,5′-cyclic monophosphate (cAMP), cytosine-3′,5′-cyclic monophosphate (cCMP), and uracil-3′,5′-cyclic monophosphate (cUMP). This section explains in detail only the experimental work related to polyA. Analogous information related to the experiments performed with polyC and polyU has been reported elsewhere (de Juan et al., 1996, 1997a).

Reagents and solutions

Hydrochloric acid and sodium chloride (Merck, Darmstadt, Germany), polyadenylic acid sodium salt (polyA) and adenine-3′,5′-cyclic monophos-

phate sodium salt (cAMP; Sigma Chemical Co., St. Louis, MO), and dioxane (Carlo Erba, Milan, Italy) have been used without further purification.

Stock solutions of the macromolecule and of the cyclic monomer have been prepared by weighing and dissolving a known amount of the solid reagent in the dioxane/water mixture 30% (v/v). The polynucleotide concentration in solution is expressed as the concentration of the monomer unit in the polynucleotide chain. CO_2-free sodium hydroxide (Merck, Darmstadt, Germany) solutions have been prepared following the Kolthoff's procedure and standardized with potassium hydrogenphtalate (Kolthoff et al., 1969).

The ionic strength of all of the solutions has been adjusted at 0.15 M by adding the appropriate amount of sodium chloride. The solutions were stored at 4°C.

Apparatus

UV absorption spectra have been recorded with a Perkin-Elmer λ-19 spectrophotometer. A Jasco J-720 spectropolarimeter has been used to collect the circular dichroism spectra. The original software of both instruments has been used in the instrumental control, data acquisition, and spectra preprocessing.

An ORION 720 pH meter (±0.1 mV precision) has been employed to perform the electromotive force (e.m.f.) measurements. Titrant additions have been carried out with a Metrohm Dosimat 655 autoburette (±0.005 ml precision).

Procedure

Potentiometric titrations have been carried out automatically. The potentiometric setup is connected via an HP 3241 interface with an HP Vectra ES/12 computer having the automation software. The electrode system consists of

GE/WS, I=0.15 M, water-dioxane/RE (KCl_{sat} in water-dioxane)

where GE is the glass electrode, WS is the working solution, and RE is the Ag/AgCl reference electrode, the inner solution of which contains a saturated solution of potassium chloride (KCl_{sat}) in the working hydroorganic mixture to minimize the liquid junction potential.

Neutral or slightly basic solutions with varying concentrations of polyA have been used as the initial titrand solution in the UV and CD spectrometric titrations. After each HCl titrant addition, the pH value of the solution is measured and the corresponding UV or CD spectrum recorded. The spectral wavelength range for both CD and UV experiments is 230–310 nm with $\Delta\lambda = 1$ nm between consecutive absorbance readings.

All titrations have been performed under nitrogen atmosphere, with the working solution placed in a double-walled titration cell, thermostatted at 25°C. Gran's method was used to calibrate the electrode system (Gran, 1952).

Data treatment

Potentiometric titrations

Point-by-point calculation of log K has been applied to the e.m.f. readings from the polynucleotide titrations to detect the presence of polyelectrolyte effect (i.e., changes in the log K value with the degree of protonation). The presence of this effect, as well as the early formation of precipitate in the solution, prevented the application of classical treatments based on the assumption of the mass action law and on the fulfillment of the mass balance to the polyC and polyA macromolecular titrations.

Numerical analysis of the e.m.f. data from the titrations of polyU and all of the cyclic nucleotides has been carried out using the SUPERQUAD program, a least-squares curve fitting method the working procedure of which consists in the refinement of the parameters (e.g., log K) related to a previously proposed chemical model (Gans et al., 1985). Figures of merit provided by the program for the evaluation of the results are the parameter σ, defined as the ratio between the root mean square of the weighted residuals and the estimated error in the working conditions, and the statistical parameter χ^2, based on the distribution of the weighted residuals on e.m.f. readings.

Spectrometric titrations

The measurements from the spectrometric acid-base titration are organized in data matrix **D**, the rows of which contain the spectra recorded at different pH values during the equilibrium process. Both directions of this original matrix are chemically meaningful; thus, the information in successive rows describes the spectral changes in the absorbing species during the titration whereas the information in each column describes the evolution in species concentration with pH.

All resolution methods decompose the matrix **D** into the product of the two smaller matrices **C** and **S**, which contain the concentration profiles and pure spectra related to each absorbing species in the solution, respectively (see Fig. 2). Although the existing procedures differ in their mathematical background, many of them use factor analysis or related techniques in their resolution procedures. Some of them are noniterative and obtain the solutions directly from the abstract analysis of the data (Kvalheim and Liang, 1992; Malinowski, 1996), whereas some others use this abstract information as the starting point of an iterative process that leads to definitive solutions (Vandeginste et al., 1985; Tauler et al., 1995; Tauler, 1996). The alternating least squares (ALS) method belongs to the group of iterative resolution methods, and it works by optimizing initial estimates, often based on the abstract analysis of the data, using suitable constraints related to the internal structure of the data or to the chemical features of the concentration profiles and instrumental responses (Tauler et al., 1995; Tauler, 1996; Gargallo et al., 1996; de Juan et al., 1997b). This method has been chosen due to its great flexibility in the modeling of the **C** and **S** profiles and to the possibility of introducing the chemical knowledge about the process through the input of the appropriate constraints The steps followed by the ALS method are summarized below.

Construction of the **D** data matrix

The spectra from an acid-base titration are sorted in the order in which they have been recorded to form a matrix $\mathbf{D}_i$ ($nph \times nwv$). The number of rows, *nph*, coincides with the number of spectra collected (i.e., one per pH value measured) and the number of columns, *nwv*, equals the number of wavelengths at which the absorbance is read in each spectrum.

ALS allows both a single matrix and several matrices, the so-called three-way data sets, to be analyzed. In the latter case, the initial data matrix is built up by placing the $\mathbf{D}_i$ matrices related to the i titrations one on top of each other. The resulting column-wise augmented **D** matrix has a size of ($npht \times nw$), where *npht* is the total number of spectra from all the appended $\mathbf{D}_i$ matrices.

Determination of the number of absorbing species in **D**, *ns*

The number of species, *ns*, present in the data matrix can be determined by using principal component analysis (PCA), although certain derived techniques that focus on the analysis of matrices related to evolutionary processes are often used to obtain additional information. Thus, evolving factor analysis (EFA) is able to detect the regions in which the absorbing species are present by performing forward and backward PCA in **D** submatrices, which gradually become larger (Gampp et al., 1986; Keller and Massart, 1991), whereas fixed size moving window evolving factor analysis (FSMW-EFA) works by performing PCA on fixed-sized **D** submatrices and gives a local rank map of the original matrix (i.e., indicates how many species are present in the various regions of the **D** matrix) (Keller and Massart, 1991). An indication that the correct number of absorbing species has been selected is given when the residual standard deviation from the difference between **D** and the matrix reproduced by

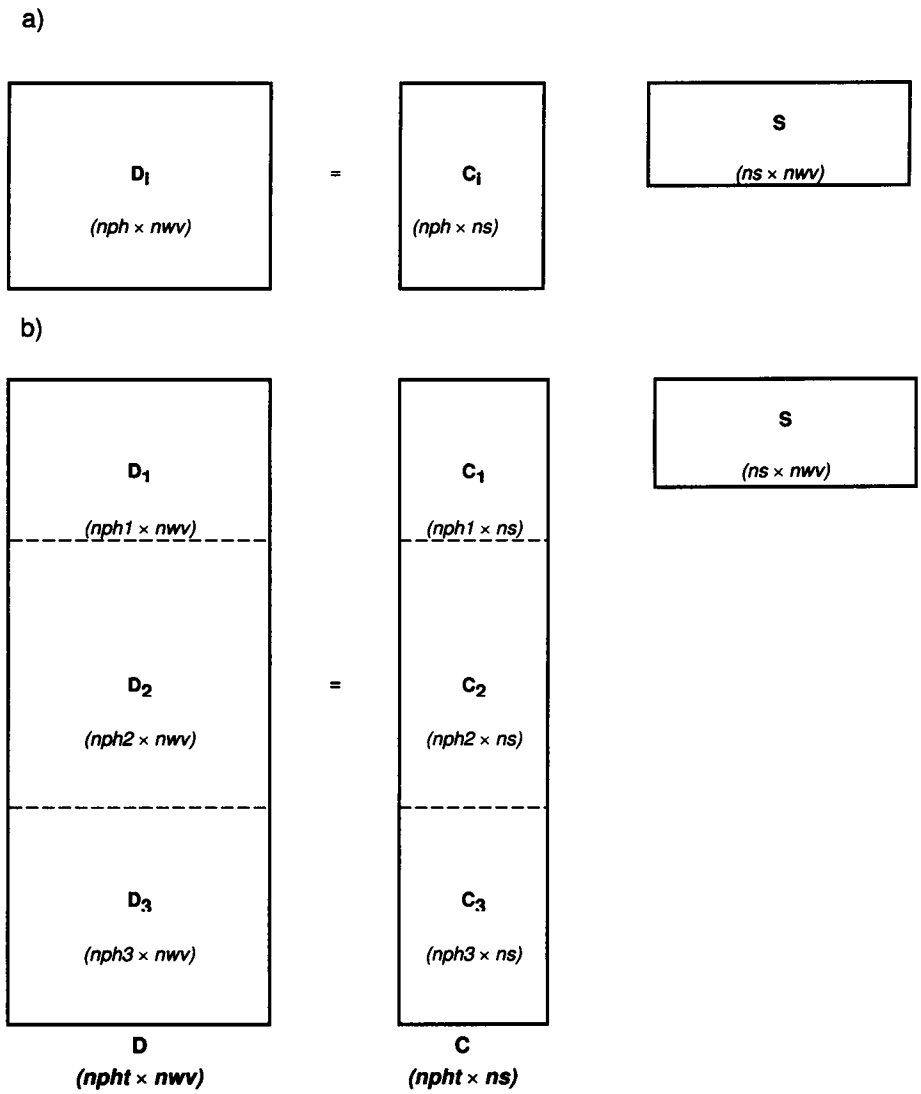

FIGURE 2 ALS working procedure for the resolution of a single pH-dependent experiment (*a*) and several pH-dependent experiments (*b*).

using the first *ns* principal components, $\mathbf{D}^*$, has a magnitude similar to the estimated experimental error.

Building a matrix of initial estimates

To start the least-squares optimization step, a matrix of initial estimates, either containing concentration profiles or spectra, is needed. The use of random estimates has been avoided because techniques such as EFA, FSMW-EFA, and others yield initial estimates based on the structure of the data matrix that are more reliable and closer to the true solutions. The use of these more reliable estimates minimizes the risk that the final solutions might converge to local minima.

For three-way data sets, the matrix of concentration initial estimates $\mathbf{C}$ is built by placing the initial estimates related to each titration $\mathbf{C}_i$ one on top of each other sorted in the same way as the $\mathbf{D}_i$ matrices are to form $\mathbf{D}$.

Application of the ALS optimization to obtain the definitive concentration and spectra matrices

The equation $\mathbf{D} = \mathbf{CS}'$ is solved iteratively by using a constrained least-squares procedure. Working with individual titrations or treating simultaneously several experiments, the resolution process is guided to the right solution by updating the spectra and the concentration profiles obtained by least-squares in each iterative cycle according to the constraints related to the internal structure and to the chemical features of the data set (Tauler et al., 1995; de Juan et al., 1997b). Selectivity is revealed as being the most important constraint in the resolution procedure. Indeed, the presence of selective zones (i.e., zones where only one species is present) for all of the species ensures the recovery of the true solutions. The selective regions for the different species, when present, are determined by using EFA or FSMW-EFA. According to the chemical features of the spectrometric titrations, all of the concentration profiles have been constrained so as to be positive and unimodal (i.e., with only one maximum) whereas each row in the $\mathbf{C}_i$ matrix is constrained so as to fulfill the closure constraint (i.e., the sum of the concentrations of the protonated and deprotonated forms of the polynucleotide at each point of the titration is constant). Nonnegativity is applied to the UV spectra, whereas CD spectra are not subject to this constraint because of the possibility of obtaining negative measurements with this technique. In the simultaneous treatment of several titrations, no constraints that force the shape of the concentration profile of an individual species to be the same in all of the different titrations are applied. In contrast, a single-unit spectrum per species is accepted because all of the titrations analyzed together have been performed in the same conditions of temperature, solvent composition, and ionic strength.

The quality of the ALS results is evaluated by calculating the lack of fit, expressed as follows:

$$\text{lack of fit} = \sqrt{\frac{\sum_{ij} (d_{ij} - d_{ij}^*)^2}{\sum_{ij} d_{ij}^2}}$$

where d_{ij} are the experimental data and d_{ij}* the reproduced data by using the ALS method. Subscripts i and j refer to the rows and columns of the original data matrix, respectively.

The ALS method has been implemented in a set of MATLAB routines written by the authors (MATLAB v. 4.2 (1994), Math Works Inc., Cochituate Place, MA).

The present method shows some advantages when compared with other common methods used for the same purpose in biochemistry, such as global analysis (Beechem, 1992) and singular value decomposition (SVD) (Henry and Hofrichter, 1992; Compton and Johnson, 1986; Henry, 1997). Both SVD and ALS offer clear advantages with respect to global analysis because the complexity of the modeling problem is drastically reduced due to the much smaller number of parameters to be fitted. In imitation to the SVD-based method, ALS uses the algorithm of SVD to remove as much as possible the noise contributions from the original data set and can also use it to quantify the number of chemical contributions in the data matrix, although for this purpose, other methods, such as EFA and SIMPLISMA (Windig and Stephenson, 1992), which are especially meant to define the distinguishable contributions in an evolutionary process (e.g., the spectrometric monitoring of a reaction), are preferably applied. These other methods also provide initial estimates for the later iterative optimizations, which are closer to the real solution than those provided directly by the SVD results. In contrast to SVD, ALS does not require any underlying chemical model (kinetic or thermodynamic) to obtain the concentration profiles and spectra of the species during the optimization step. Actually, the related chemical model is built afterwards from the concentration profiles given by the ALS results. In the case of spectrometric data, ALS does not assume either any mathematical function shape (Henderson-Hasselbach curves for pH profiles (Henry and Hofrichter, 1992) or Lorentzian or Gaussian profiles for spectra (Henry, 1997) in the case of SVD) for the concentration profiles or spectra to be modeled. Only some softer assumptions that are always fulfilled because of the nature of the chemical process under study and of the instrumental technique used, such as nonnegativity for concentration profiles and UV spectra and unimodality and closure for the pH profiles, are introduced in the optimization process. Furthermore, ALS can work simultaneously with several data matrices, thus leading to a significant decrease in the ambiguity associated with the decomposition of bilinear matrices.

RESULTS AND DISCUSSION

All of the experiments have been carried out in a dioxane/water 30% (v/v) mixture. This hydroorganic mixture is often employed in biocoordination studies to emulate low polar biological microenvironments (Sigel et al., 1985; Liogang et al., 1988; Sigel, 1993; de Juan et al., 1996). Table 1 shows the experimental conditions used in studying the systems polyA-H, polyC-H, and polyU-H and their respective cyclic nucleotides. In the pH working ranges, a single protonation site has to be considered in each nitrogen base of the homopolynucleotide. Thus, the N(1) ring nitrogen of adenine in polyA protonates giving a positively charged polymer; polyC also gives an analogous charged structure when the cytosine N(3) nitrogen ring takes a proton. In contrast, the deprotonation of the uracil N(3) amide-like nitrogen is responsible for the negatively charged structure of polyU at basic pH values.

ALS has been applied to the UV and CD acid-base titrations of polyA, polyC, and polyU. For each homopolynucleotide, all of the titrations from the same spectrometric technique have been processed together. The ALS results and the chemical information inferred from them are shown and commented in detail below.

ALS results

Concentration profiles and pure unit spectra of the individual species involved in the acid-base equilibrium

The number of species involved in each acid-base equilibrium and the initial estimates to be input in the ALS optimization have been determined by using EFA. The acid-base equilibria of polyC and polyU were explained with two species, related to the protonated and the deprotonated species, whereas the polyA protonation process needed three species to be described. Figs. 3 and 4 show the concentration profiles and spectra obtained after the ALS optimization for the three polynucleotides from the UV and CD titrations, respectively. The agreement between the distribution plots obtained with both spectrometric techniques for the same polynucleotide protonation process reflects the reproducibility of the experimental work and the good performance of the ALS method on this kind of biochemical data.

TABLE 1 Description of the experiments performed in dioxane/water 30% (v/v)

System	Number of experiments	Technique	[L] range (M × 10^3)	pH range
cUMP-H	3	Potentiometry	0.5–0.7	5–10.5
cCMP-H	3	Potentiometry	1.4–1.8	2.8–7.2
cAMP-H	3	Potentiometry	0.8–1.3	3.0–5.2
PolyU-H	3	Potentiometry	1.5–3.0	5–10.5
	3	UV	0.05–0.09	4–11
	3	CD	0.08–0.12	5.5–11
PolyC-H*	3	UV	0.07–0.13	4.0–7.7
	3	CD	0.06–0.12	4.0–7.5
PolyA-H*	3	UV	0.07–0.10	9.5–4.5
	3	CD	0.07–0.11	9.5–4.1

[L], ligand concentration. For polynucleotides, [L] means concentration in moles of monomer unit per liter.
*Potentiometric experiments, although performed, have not been listed as no quantitative results were obtained.

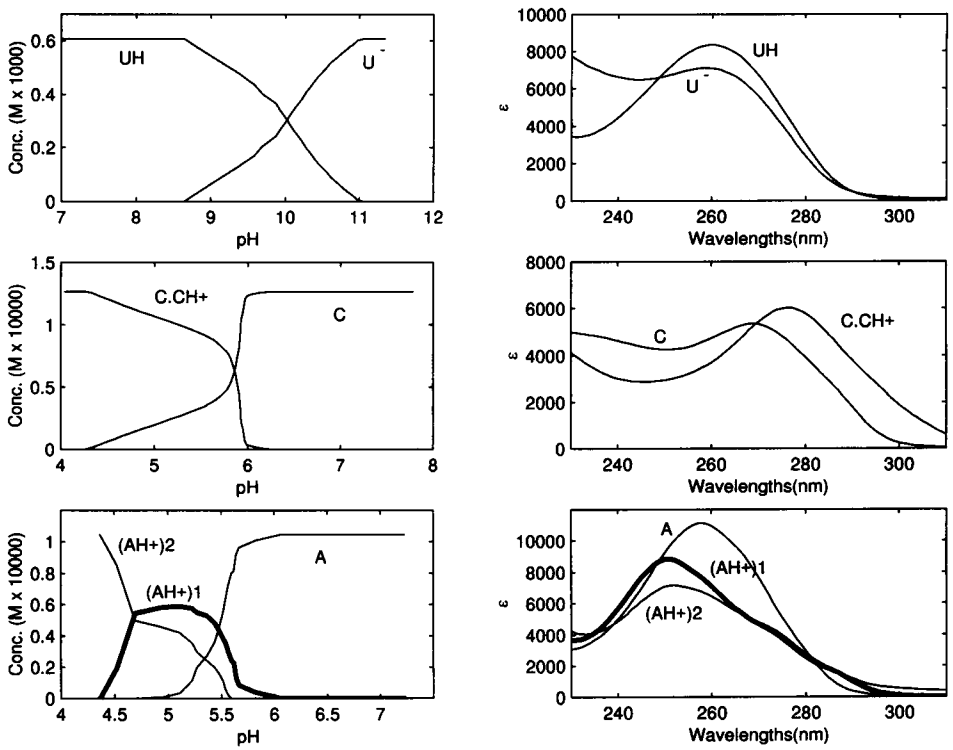

FIGURE 3 ALS results from the UV titrations in dioxane/water 30% (v/v): concentration profiles and UV spectra related to the polyU-H, polyC-H, and polyA-H systems. ϵ values are molar absorptivities.

Detection and description of conformational transitions

The concentration profiles diagram obtained with ALS provides complete information about the evolution of all of the species present in the acid-base equilibria. This includes transitions between chemical species that do not involve changes in the spatial structure of the molecule, conformational changes associated with the proton uptake process, and changes in the spatial configuration of a single chemical species that do not alter the protonation state of the molecule. Identifying which kind of transition takes place and which conformations are involved depends on the chemical knowledge about the process being analyzed. PolyU, polyC, and polyA are good examples of the three transitions mentioned above.

Chemical literature on polyU generally accepts the existence of random coil structures related to both protonated and deprotonated species (Saenger, 1988; Casassas et al., 1993). The similar shape of the UV and CD spectra related to the poly(UH) and poly(U^-), where the shift between their absorption maxima is the main difference, is in agreement with the hypothesis of a protonation process without conformational changes.

Previous studies of the polyC structure in aqueous solution suggest three different chemical species and conformations depending on the protonation form of the molecule, namely, a single-helical deprotonated polyC, a double-stranded helical half-protonated [poly(C) · poly(CH^+)] and a fully protonated random coiled poly(CH^+) (Saenger, 1988; Hartman and Rich, 1965; Adler et al., 1967; Arnott et al., 1976; Broido and Kearns, 1982; Garriga et al., 1992; Antao and Gray, 1993; Tanigawa and Yamaoka, 1995). Only two species are detected when this equilibrium takes place in the hydroorganic mixture because of the narrower pH working range, limited by the precipitation of the charged polynucleotide at acid pH values. The most plausible identification of the two existing species includes the presence of the deprotonated and half-protonated polyC species. Apart from the similar shape of the spectra and the similar pH region of existence of the species presented in Figs. 3 and 4 and those found in aqueous solution, the absence of the characteristic decrease in the CD ellipticities associated with a helix → random coil transition, and the more favorable situation of a charged species partially stabilized with the formation of an interstrand hydrogen bond N-H^+····N in comparison with a species with a net charge NH^+ in a low polar solution, support the identification proposed. Thus, the polyC-H system is a good example of acid-base equilibrium with a conformational change associated with the proton uptake process.

Polyadenylic acid behaves in a way that is somewhat more complex than that of the two previous polynucleotides. Although the single helical conformation of deprotonated polyA and its first transition to double-stranded protonated polyA is widely accepted (Rich et al., 1961; Leng and Felsenfeld, 1966; Adler et al., 1969; Saenger et al., 1975; Saenger, 1988; Antao and Gray, 1993; Maggini et al., 1994; Casassas et al., 1994), different conformational transitions between double-helical configurations of the protonated polynucleotide have also been proposed (Antao and Gray, 1993; Maggini et al., 1994; Casassas et al., 1994).

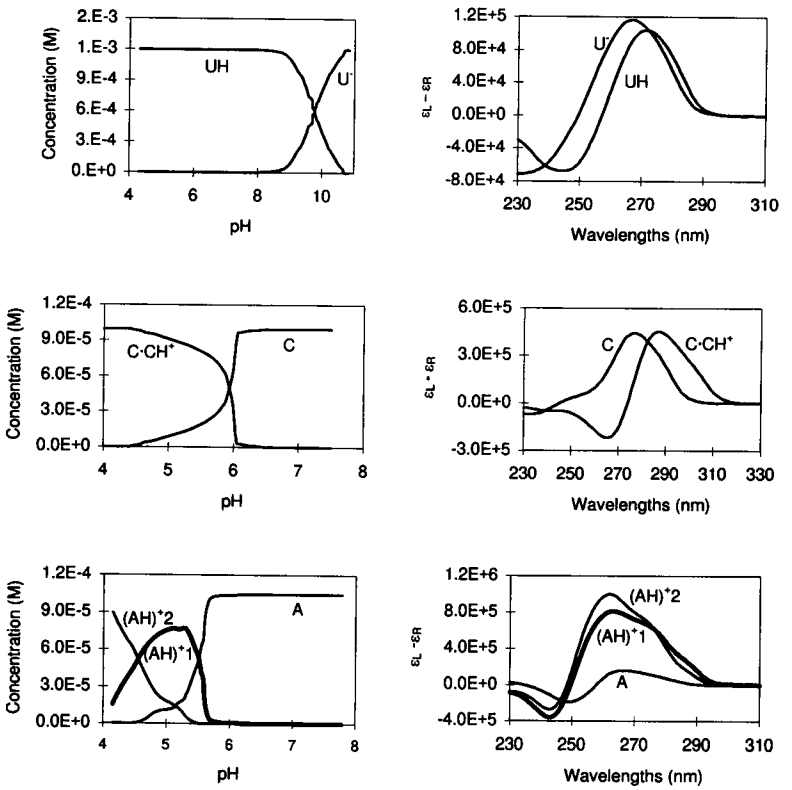

FIGURE 4 ALS results from the CD titrations in dioxane/water 30% (v/v): concentration profiles and CD spectra related to the polyU-H, polyC-H, and polyA-H systems. ϵ_L-ϵ_R values are molar ellipticities.

Figs. 3 and 4 show three different species associated with the polyA-H system in the dioxane/water mixture. The similar shape of the two species occurring at more acidic pH values seems to identify them as different double-helical protonated configurations, whereas the species present at higher pH values can probably be attributed to the single helical deprotonated polyA. The transition between the deprotonated and the first protonated species, poly(AH^+)1, is associated with the proton uptake process and, therefore, very reproducible. The transformation between both protonated forms, poly(AH^+)1 $\rightarrow$ poly(AH^+)2, presents more irregular concentration profiles owing to the lack of chemical reaction and to the probable time dependency of the process (Maggini et al., 1994). Nevertheless, CD and UV spectra show coherent hyperchromism and hypochromism, respectively, when going from poly(AH^+)1 to poly(AH^+)2. Such a phenomenon could be explained by the gradual minimization of the electrostatic repulsion of neighboring phosphate groups due to the increase in the stabilizing electrostatic interactions between the negative charges of these groups and the protonated sites of the adenine bases as the protonation proceeds. Such a stabilizing effect would probably allow the formation of a more compact and ordered structure with the phosphate groups occurring closer together and with a consequent stronger base stacking.

The presence of highly ordered polymeric structures could be confirmed by comparing the polynucleotide spectra with the spectra of their respective cyclic nucleotides, as shown in Fig. 5. The marked hyperchromism in the CD spectra and hypochromism in the UV spectra of polyA and polyC with respect to cAMP and cCMP indicate a significant stacking in both polynucleotides and, therefore, the existence of ordered structures. The slight differences in terms of intensity between the polyU and cUMP support the hypothesis of random structure for this polymer.

Existence and pattern of polyelectrolyte effect

Once the species in the polynucleotide protonation process are identified, the equilibrium constant of this reaction can be properly evaluated. The distribution plot obtained by applying the ALS method provides the concentration values of all the protonated and the deprotonated species for each pH measured. A log K value is then calculated for each titration point bearing in mind that a fixed log K value for each functional site of the polynucleotide during the whole protonation process cannot be obtained unless no polyelectrolyte effect exists. Plotting the log K values versus their corresponding protonation degrees (α) is a graphical way of studying both the existence and the pattern of a polyelectrolyte effect. If this effect is revealed, the so-called apparent constant (K_{app}) of analogous sites of the polynucleotide changes as the protonation process advances and the log K value usually given is not a thermodynamic constant, but an intrinsic constant (K_{int}) defined as the extrapolated K_{app} value for a protonation degree equal to zero, i.e., for the theoretical point where no effects of neighboring protonated sites are present.

Bearing in mind that the equilibrium process followed is the protonation of the site in the monomer unit of the

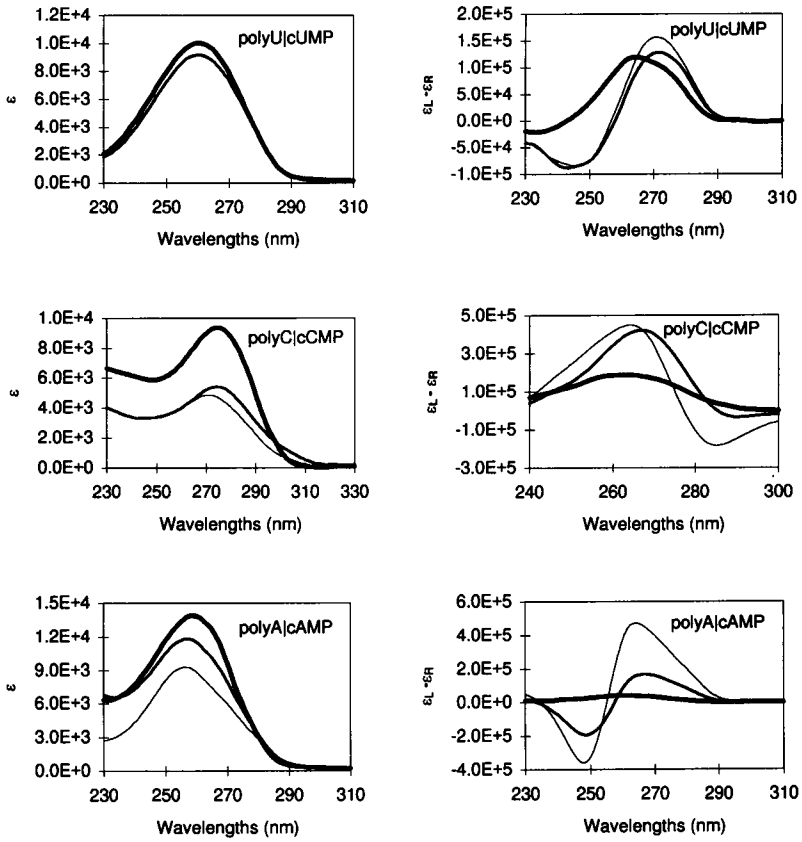

FIGURE 5 Solvent effect on the UV and CD spectra of protonated polyU, deprotonated polyC, and deprotonated polyA. The thin lines are spectra of the polynucleotides in water. The medium lines are spectra of the polynucleotides in dioxane/water 30% (v/v). The thick lines are the spectra of the respective cyclic nucleotides cUMP, cCMP, and cAMP in dioxane/water 30% (v/v).

macromolecule, the concentration of the polynucleotide is expressed as moles of monomer per volume unit, and therefore K is defined as

$$K = \frac{[\text{protonated monomer}]}{[\text{deprotonated monomer}][H^+]}$$

and consequently, the protonation degree α as

$$\alpha = \frac{[\text{protonated monomer}]}{[\text{total monomer}]}$$

In the expressions below referring to the protonation constants of each polynucleotide, the names between brackets indicate the form of the macromolecule where the protonated or deprotonated monomers are placed.

Fig. 6 shows in thick lines the log K versus α plots for polyA, polyC, and polyU in dioxane/water 30% (v/v). The plots include the theoretical lines obtained after the polynomial fitting of the experimental (α, log K) values. The related equations log $K = f(\alpha)$ can be found in Table 2.

According to the identification of species in the ALS distribution plots, the polyU protonation constant has been determined as follows:

$$K = \frac{[\text{poly(UH)}]}{[\text{poly}(U^-)][H^+]}$$

No polyelectrolyte effect has been detected, and therefore, a thermodynamic protonation constant can be given for all of the analogous protonation sites in the macromolecule. This is probably due to the random coiled structures associated with the protonated and the deprotonated species. These disordered structures are more flexible and allow spatial rearrangements of the macromolecule to minimize the between-sites effect during the protonation process.

The polyC protonation constant has been calculated by using the following equation:

$$K = \frac{[\text{poly(C)} \cdot \text{poly}(CH^+)]/2}{([\text{poly(C)}] + [\text{poly(C)} \cdot \text{poly}(CH^+)]/2)[H^+]}$$

Please note that the [poly(C) · poly(CH^+)] concentration is always divided by 2 when included as the protonated and deprotonated form due to the existence of one protonated base and one deprotonated base per base pair. There is a nonlinear pattern in the polyelectrolyte effect owing to the cooperative action between the protonation process and the formation of the double-stranded helical structure. Thus, the formation of this helix stabilizes the protonated base because of the interstrand hydrogen bond N-H^+ · · · N, and at the same time, this base pair arrangement is responsible for the growth and stabilization of the helical structure. As the protonation process advances, the intensity of the positive polyelectrolyte effect decreases, becoming negative for α_p values higher than 0.3 (de Juan et al., 1997a) because the increase of charge density in the macromolecular structure means that the repulsive effect between the protonated sites

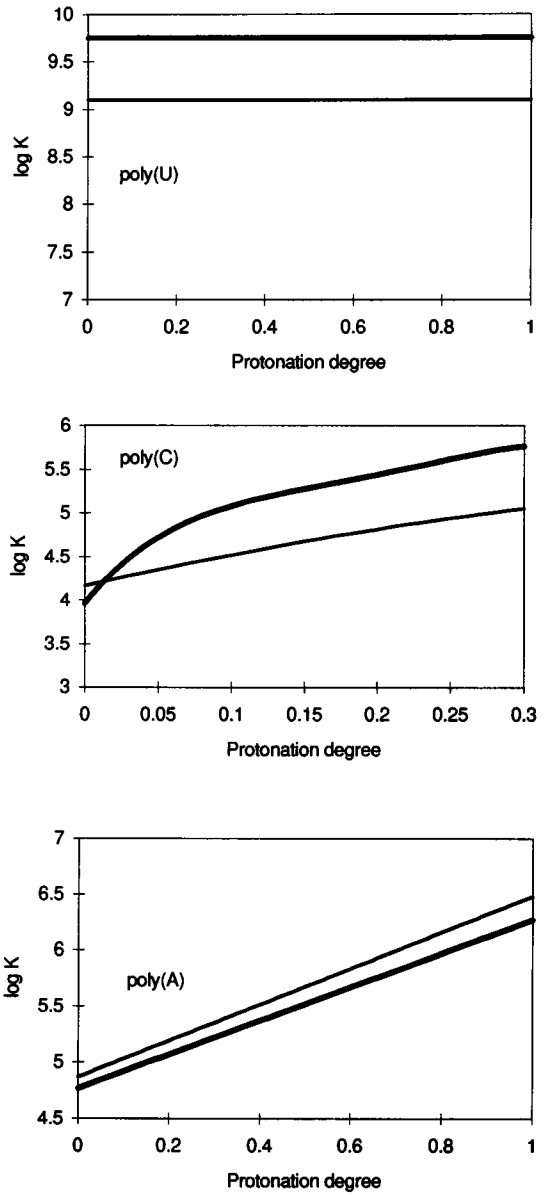

FIGURE 6 Theoretical models, log $K = f(\alpha)$, of the polyelectrolyte effect related to the polyU, polyC, and polyA protonation processes in dioxane/water 30% (v/v) (*thick line*) and in aqueous solution (*thin line*).

is more important than the stabilization caused by the formation of the double-helical structure.

The equation related to the polyA protonation process is shown below:

$$K = \frac{([\text{poly}(\text{AH}^+)1] + [\text{poly}(\text{AH}^+)2])}{[\text{poly}(\text{A})][\text{H}^+]}$$

A linear positive polyelectrolyte effect caused by the stabilizing electrostatic interactions between the negatively charged phosphate groups and the protonated adenine sites is shown. The difference between the polyelectrolyte effect patterns of polyC and polyA depends on the role of the protonation sites in the formation of the double-stranded helix of the polynucleotide. PolyC protonated sites are responsible for the formation of one of the interstrand hydrogen bonds, whereas polyA protonated sites, although having a positive effect on the stabilization of the double-helical structure because of the minimization of the electrostatic repulsion of the phosphates, do not participate directly in the interstrand hydrogen bonding (Antao and Gray, 1993). Thus, a nonlinear pattern of the polyelectrolyte effect can in all likelihood be attributed to the presence of a cooperative mechanism associated with the protonation process, as occurs in polyC, whereas the more common linear pattern would appear to be caused mainly by purely electrostatic interactions, as shown in polyA and in other simpler polyelectrolytes, such as polyacrylic acid.

A good agreement between the numerical values of polyU and of cUMP protonation constants and between the intrinsic protonation constant of polyC and the protonation constant of cCMP is shown, the slight differences between the values of K_{polyU} and K_{cUMP} being due to hindrance effects (see Table 2). No such agreement is noticed when the cAMP protonation constant and the polyA intrinsic protonation constant are compared. The significantly lower value of the cAMP protonation constant can probably be explained by the absence of the positive influence of the surrounding phosphate groups, present on the protonation of all the adenine sites of polyA, on the protonation process of the cyclic nucleotide cAMP.

Solvent effect on the acid-base equilibria of polynucleotides

The acid-base equilibria of polyA, polyC, and polyU were also studied in aqueous solution (Casassas et al., 1993, 1994, 1995). The comparison of the results obtained in water with those obtained in dioxane/water 30% (v/v) allows the inference of the solvent effects on the protonation processes of the polynucleotides. The system dioxane/water was used because, in comparison with some other hydroorganic mixtures, it provides a larger decrease of polarity with the inclusion of smaller amounts of organic co-solvent. No experiments were performed at dioxane proportions higher than 30% (v/v) because some of the polynucleotides studied presented precipitation phenomena.

When comparing the experiments mentioned above, a general observation is the greater instability of charged macromolecular structures in the hydroorganic medium. Indeed, some species detected in aqueous solutions are apparently not formed in the lower polar medium because the macromolecule is precipitated at suitable pH values. This is the case of the fully protonated poly(CH^+), the formation of which involves the appearance of two net charges in solution, one from the new protonated site and the other from the breaking of the interstrand hydrogen-bonding N-H^+····N. The precipitation of the macromolecule from dioxane/water solution also takes place for the protonated polyA at pH values lower than 4, and this prevents the formation of other postulated double-helical configurations

TABLE 2 Results of the experiments performed

System	Solvent	Technique	Data treatment	log K	$\sigma^{\dagger}$	$\chi^{2\dagger}$	Lack of fit (%)#	Fit log $K = f(\alpha)$* Model	r^2
cUMP-H	D/W 30%	Potentiometry	CV§	9.250 (4)¶	2.14	60.09	–	–	–
cCMP-H	D/W 30%	Potentiometry	CV	3.766 (3)	2.41	34.03	–	–	–
cAMP-H	D/W 30%	Potentiometry	CV	3.195 (8)	1.29	30.55	–	–	
	Water‖	Potentiometry	CV	9.364	–	–	–	log K = 9.364	–
	Water	CD	ALS	9.1 (2)	–	–	5.9	log K = 9.1	–
PolyUH	D/W 30%	Potentiometry	CV	9.756 (4)	1.40	80	–	log K = 9.756	–
	D/W 30%	UV	ALS	9.9 (2)	–	–	6.2	log K = 9.9	–
	D/W 30%	CD	ALS	9.75 (6)	–	–	5.8	log K = 9.75	–
PolyC-H**	Water‖	UV	ALS	4.21 (5)	–	–	1.3	log $K = 4.21 + 4.6\alpha - 4.5\alpha^2$	0.980
	Water	CD	ALS	4.16 (4)	–	–	3.6	log $K = 4.16 + 3.9\alpha - 3.1\alpha^2$	0.985
	D/W 30%	UV	ALS	4.04 (2)	–	–	1.6	log $K = 4.04 + 17\alpha - 110\alpha^2 + 350\alpha^3 - 400\alpha^4$	0.996
	D/W 30%	CD	ALS	3.96 (2)	–	–	3.5	log $K = 3.96 + 21\alpha - 140\alpha^2 + 480\alpha^3 - 600\alpha^4$	0.996
PolyA-H**	Water‖	CD, UV	ALS	4.87 (4)	–	–	2.0	log $K = 1.61\alpha + 4.87$	0.94
D/W 30%	UV	ALS	4.74 (6)	–	–	3.3		log $K = 1.57\alpha + 4.74$	0.97
D/W 30%	CD	ALS	4.78 (4)	–	–	4.9		log $K = 1.46\alpha + 4.78$	0.99

*α means protonation degree.

†Figures of merit related to the curve fitting program SUPERQUAD (see text for definitions).

#Figure of merit related to ALS (see text for definition).

§CV: classical least-squares curve fitting procedure (SUPERQUAD program).

¶Numbers in parentheses are the errors associated with the last figure.

‖Results in aqueous solution are taken from Casassas et al., 1993, for polyU; Casassas et al., 1995 for polyC; and Casassas et al., 1994 for polyA.

**Log K values for polyC and polyA are always intrinsic constants.

at more acidic pH values. Disordered charged structures, such as deprotonated polyU, are more stable than ordered forms in dioxane/water solution due to the greater ability of the macromolecule to reach spatial arrangements suitable for accommodating the net charges.

All of the species detected in dioxane/water 30% (v/v) are present in water with the same spatial structure. The only difference is the formation of more relaxed configurations in the dioxane/water mixtures because of the weakening of the base stacking interactions caused by the low polar solvents, as can be seen in Fig. 5. This is clearly seen in the single-helical structures of polyC and polyA, the latter showing a greater unstacking because of the stronger purinic base stacking in aqueous solution.

The polyelectrolyte effect, if absent or linear, maintains the same behavior in both solvents studied, as shown in Table 2 and in Fig. 6. Thus, the polyU acid-base behavior does not present a polyelectrolyte effect either in water or in the working dioxane/water mixture, and the only difference between the polyU equilibrium in both media is the higher value of the protonation constant in the hydroorganic mixture owing to the lower stabilizing effect of this low polar solvent on the negatively charged deprotonated polynucleotide. Linear models with fairly similar slopes describe the polyelectrolyte effect associated with the polyA protonation process in hydroorganic and in aqueous solutions. Nevertheless, there is not enough information to determine whether the small difference between slopes is due to the intrinsic similarity of the polyelectrolyte effects in both media or whether the stabilizing effect of the inert salt counterions around the negatively charged phosphates conceals the real solvent effect on the polyA protonation process. To clarify this point, it would be necessary to work at lower ionic strengths.

There is a clear solvent influence on the pattern of the polyelectrolyte effect when a cooperative mechanism is involved in the polynucleotide protonation process, as can be seen in the polyC protonation process. Although not visible in Fig. 6, where only the mathematically fitted positive polyelectrolyte effect is shown, there is a change in the sign of the polyelectrolyte effect associated with the polyC protonation process in both water and water/dioxane solutions (de Juan, 1997a). Solvent effect is first noticed in the ranges of existence of the positive and the negative polyelectrolyte effects. Whereas the change of sign appears in α values around 0.3 in the hydroorganic solution, this negative behavior is not seen until an α value of higher than 0.5 in water. The greater stability of charged structures in water explains that the negative polyelectrolyte effect does not appear in this solvent until the breakdown of the stable double-stranded helical structure takes place, whereas an increase in the charge density of the double-stranded helix is enough to change the tendency of the polyelectrolyte effect in less polar media. The second difference concerns the pattern of the positive effect in both media: in the hydroorganic mixture, a fourth-order polynomial is needed to explain the variation of log K with α, whereas a second-order polynomial is enough to explain the same data in water, as shown in Table 2. A much steeper effect for low α values is detected in water/dioxane because of the easier formation of the interstrand hydrogen bond N-H$^+$····N due to the weaker

competition of the solvent molecules in the development of these interactions in a hydroorganic mixture less polar than water. As the protonation degree increases, this tendency is inverted and the polyelectrolyte effect in water/dioxane becomes less and less pronounced, being close to a plateau for α values near to 0.3. The smoothing in the evolution of the polyelectrolyte effect comes from the gradual balance between the favorable tendency to form the hydrogen bond $N\text{-}H^+\cdots N$ and the destabilizing influence associated with the increase of charge density in the macromolecular structure.

CONCLUSIONS

The ALS method has been found to be a powerful tool in dealing with multivariate data sets related to complex biochemical processes. A complete interpretation of all of the simultaneously occurring thermodynamic and conformational transitions in the polynucleotide equilibria can be obtained from the information in the ALS concentration profiles and spectra matrices C and S.

ALS has also been successfully applied in the interpretation of other biochemical processes, such as multivariate spectrometric melting monitoring (Gargallo et al., 1997) and metal-peptide interactions (Mendieta et al., 1996). The same procedure is also now being used to tackle more complex problems, such as protein folding, and is proposed as a suitable method to be applied to any other process followed by techniques the instrumental response of which follows an additive model.

This work has received financial support from the project INTAS 93-2771 Ext.

REFERENCES

Adler, A., Grossman, L., and Fasman, G. D. 1967. Single-stranded oligomers and polymers of cytidylic and 2′-deoxycytidylic acids: comparative optical rotatory studies. *Proc. Natl. Acad. Sci. U.S.A.* 57:423–430.

Adler, A., Grossman, L., and Fasman, G. D. 1969. Polyriboadenylic and polydeoxyriboadenylic acids: optical rotatory studies of pH-dependent conformations and their relative stability. *Biochemistry.* 8:3846–3858.

Antao, V. P., and Gray, D. M. 1993. CD spectral comparisons of the acid-induced structures of poly[d(A)], poly[r(A)], poly[d(C)], and poly[(rC)]. *J. Biomol. Struct. Dyn.* 10:819–838.

Arnott, S., Chandrasekaran, R., and Leslie, A. G. W. 1976. Structure of the single-stranded polyribonucleotide polycytidylic acid. *J. Mol. Biol.* 106: 735–748.

Beechem, J. 1992. Global analysis of biochemical and biophysical data. *Methods Enzymol.* 210:37–54.

Broido, M. S., and Kearns, D. R. 1982. ^{1}H NMR evidence for a left-handed helical structure of poly(ribocytidylic) acid in neutral solution. *J. Am. Chem. Soc.* 104:5207–5216.

Casassas, E., Gargallo, R., Giménez, I., Izquierdo-Ridorsa, A., and Tauler, R. 1993. Application of an Evolving Factor Analysis-based procedure to speciation analysis in the copper(II)-polyuridylic acid system. *Anal. Chim. Acta.* 283:538–547.

Casassas, E., Gargallo, R., Izquierdo-Ridorsa, A., and Tauler, R. 1995. Application of a multivariate curve resolution procedure for the study of the acid-base and copper(II) complexation equilibria of polycytidylic acid. *Reactive Polymers.* 27:1–14.

Casassas, E., Tauler, R., and Marqués, I. 1994. Interactions of H^+ and Cu(II) ions with poly(adenylic) acid: study by factor analysis. *Macromolecules.* 27:1729–1737.

Compton, L. A., and Johnson, W. C. 1986. Analysis of protein circular dichroism spectra for secondary structure using a simple matrix multiplication. *Anal. Biochem.* 155:155–167.

de Juan, A., Fonrodona, G., Gargallo, R., Izquierdo-Ridorsa, A., Tauler, R., and Casassas, E. 1996. Application of a self-modeling curve resolution approach to the study of solvent effects on the acid-base and copper(II)-complexing behaviour of polyuridylic acid. *J. Inorg. Biochem.* 63: 155–173.

de Juan, A., Izquierdo-Ridorsa, A., Gargallo, R., Tauler, R., Fonrodona, G., and Casassas, E. 1997a. Three-way curve resolution applied to the study of solvent effect on the thermodynamic and conformational transitions related to the protonation of polycytidylic acid. *Anal. Biochem.* 249: 174–183.

de Juan, A., Vander Heyden, Y., Tauler, R., and Massart, D. L. 1997b. Assessment of new constraints applied to the alternating least squares (ALS) method. *Anal. Chim. Acta.* 346:307–318.

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

Gans, P., Sabatini, A., and Vacca, A. 1985. SUPERQUAD: an improved general program for computation of formation constants from potentiometric data. *J. Chem. Soc. Dalton Trans.* 1195–1200.

Gargallo, R., Cuesta-Sánchez, F., Izquierdo-Ridorsa, A., and Massart, D. L. 1996. Application of eigenstructure tracking analysis and SIMPLISMA to the study of the protonation of cCMP and several polynucleotides. *Anal. Chem.* 68:2241–2247.

Gargallo, R., Tauler, R., and Izquierdo-Ridorsa, A. 1997. Application of a multivariate curve resolution method to the study of the melting behaviour and conformational changes of synthetic and natural polynucleotides. *Anal. Chem.* 69:1785–1792.

Garriga, P., García-Quintana, D., and Manyosa, J. 1992. Study of polynucleotide conformation by resolution-enhanced ultraviolet spectroscopy of poly(rC) and poly(dC). *Eur. J. Biochem.* 210:205–210.

Gran, G. 1952. Determination of the equivalence point in potentiometric titrations. *Analyst.* 77:661–671.

Hartman, K. A., and Rich, A. 1965. The tautomeric form of helical polyribocytidylic acid. *J. Am. Chem. Soc.* 87:2033–2039.

Henry, E. R. 1997. The use of matrix methods in the modeling of spectroscopic data sets. *Biophys. J.* 72:652–673.

Henry, E. R., and Hofrichter, J. 1992. Singular value decomposition: application to analysis of experimental data. *Methods Enzymol.* 210: 129–192.

Keller, H. R., and Massart, D. L. 1991. Peak purity control in liquid chromatography with photodiode array detection by a fixed size moving window evolving factor analysis. *Anal. Chim. Acta.* 246:379–390.

Kolthoff, I. M., Sandell, E. B., Meehan, E. J., and Bruckenstein, S. 1969. Quantitative Chemical Analysis. Collier-MacMilland Canada, Toronto.

Kvalheim, O. M., and Liang, Y. Z. 1992. Heuristic evolving latent projections: resolving 2-way multicomponent data. I. Selectivity, latent-projective graph, datascope, local rank and unique resolution. *Anal. Chem.* 64:936–946.

Leggett, D. J. 1977. Numerical analysis of multicomponent spectra. *Anal. Chem.* 49:276–281.

Leng, M., and Felsenfeld, G. 1966. A study of polyadenylic acid at neutral pH. *J. Mol. Biol.* 15:455–466.

Liogang, G., Tribolet, R., and Sigel, H. 1988. Ternary complexes in solution. 50. Dependence of intramolecular hydrophobic ligand-ligand interactions on ligand structure, geometry of the coordination sphere of the metal ion, and solvent composition: opposing solvent effects. *Inorg. Chem.* 27:2877–2887.

Maggini, R., Secco, F., Venturini, M., and Diebler, H. 1994. Kinetic study of double-helix formation and double-helix dissociation of polyadenylic acid. *J. Chem. Soc. Faraday Trans.* 90:2359–2363.

Malinowski, E. R. 1996. Automatic window factor-analysis: a more efficient method for determining concentration profiles from evolutionary spectra. *J. Chemometrics.* 10:273–279.

Mendieta, J., Díaz-Cruz, M. S., Tauler, R., and Esteban, M. 1996. Application of multivariate curve resolution to voltammetric data. II. Study of metal-binding properties of the peptides. *Anal. Biochem.* 240:134–141.

Rich, A., Davies, D. R., Crick, F. H. C., and Watson, J. D. 1961. The molecular structure of polyadenylic acid. *J. Mol. Biol.* 3:71–86.

Saenger, W. 1988. Principles of Nucleic Acid Structure. Springer-Verlag, New York.

Saenger, W., Riecke, J. and Suck, D. 1975. A structural model for the polyadenylic acid single helix. *J. Mol. Biol.* 93:529–534.

Sigel, H. 1993. Interactions of metal ions with nucleotides and nucleic acids and their constituents. *Chem. Soc. Rev.* 22:255–267.

Sigel, H., Martin R. B., Tribolet, R., Häring, U. K., and Malini-Balakrishnan, R. 1985. An estimation of the equivalent solution dielectric constant in the active-site cavity of metalloenzymes: dependence of carboxylate-metal ion complex stabilities on the polarity of mixed aqueous/organic solvents. *Eur. J. Biochem.* 152:187–193.

Tanigawa, M., and Yamaoka, K. 1995. Electro-optical and hydrodynamic properties of synthetic polyribonucleotides in solutions as studied by electric birefringence. *Bull. Chem. Soc. Jpn.* 68:481–492.

Tauler, R. 1996. Multivariate curve resolution applied to second order data. *Chemom. Intell. Lab. Sys.* 30:133–146.

Tauler, R., Smilde, A., and Kowalski, B. 1995. Selectivity, local rank, three-way data analysis and ambiguity in multivariate curve resolution. *J. Chemometrics.* 9:31–58.

Vandeginste, B. G. M., Derks, W., and Kateman, G. 1985. Multicomponent self-modeling curve resolution in high-performance liquid chromatography by iterative target transformation analysis. *Anal. Chim. Acta.* 173:253–264.

Windig, W., and Stephenson, D. A. 1992. Self-modeling mixture analysis of 2nd order derivative of near-infrared spectral data using the SIMPLISMA approach. *Anal. Chem.* 64:2735–2742.