UNIVERSITAT DE BARCELONA

Aplicació de metodologies quimiomètriques a l'estudi de l'efecte del solvent sobre els aspectes termodinàmics i estructurals dels equilibris àcid-base dels polinucleòtids

Anna de Juan Capdevila

DEPARTAMENT DE QUÍMICA ANALÍTICA DE LA UNIVERSITAT DE BARCELONA

PROGRAMA DE DOCTORAT: QUÍMICA ANALÍTICA DEL MEDI AMBIENT I DE LA POL·LUCIÓ. (BIENNI 1988-1990)

APLICACIÓ DE METODOLOGIES QUIMIOMÈTRIQUES A L'ESTUDI DE L'EFECTE DEL SOLVENT SOBRE ELS ASPECTES TERMODINÀMICS I ESTRUCTURALS DELS EQUILIBRIS ÀCID-BASE DELS POLINUCLEÒTIDS

Memòria presentada per Anna de Juan i Capdevila per optar al grau de Doctor en Ciències Químiques.

Directors: Enric Casassas i Simó i Gemma Fonrodona Baldajos.

Barcelona, juliol de 1997.

CHAPTER 6.

THE ACID-BASE EQUILIBRIA OF POLYNUCLEOTIDES IN WATER-DIOXANE MIXTURES

6.1. The use of curve resolution techniques to interpret the multivariate monitoring of biochemical processes: improvement and understanding of the chemometric procedures.

ELSEVIER

Analytica Chimica Acta 18087 (1997) 1–12

ANALYTICA
CHIMICA
ACTA

Assessment of new constraints applied to the alternating least squares method

A. de Juan[a,b,*], Y. Vander Heyden[b], R. Tauler[a], D.L. Massart[b]

[a]*Departament de Química Analítica, Universitat de Barcelona, Diagonal 647, 08028 Barcelona, Spain*
[b]*ChemoAC, Farmaceutisch Instituut, Vrije Universiteit Brussel, Laarbeeklaan 103, B-1090 Brussel, Belgium*

Received 6 December 1996; received in revised form 6 December 1996; accepted 16 December 1996

Abstract

The introduction of constraints in multivariate curve resolution methods, such as the Alternating Least Squares (ALS), is commonly used to limit the span of possible solutions, guiding the iterative process to final result as close as possible to the true situation. In the present work, two modifications of the unimodality constraint and a new constraint for chromatographic concentration profiles related to the prevention of fronting have been checked. Simulated data sets as well as real data have been used to evaluate the effect of these new constraints in the resolution results. The parameters measured to assess the goodness of the constraints are related to the recovery of the concentration profiles and to the quality of the data fit.

Keywords: XXX; XXX; XXX

1. Introduction

The application of constraints is a common operation employed in iterative curve resolution methods, such as the Alternating Least Squares (ALS) [1–3], to narrow the span of feasible solutions to those chemically meaningful [4–6].

In a wide sense, the concept of constraint would include any general feature of the data sets translated into mathematical language. However, a careful attitude has to be adopted in order to avoid false generalizations (e.g. not all the chromatographic peaks have a Gaussian shape) and the researcher has to be aware that the effectiveness of a constraint can be strongly affected by the way it has been implemented.

Within the family of constraints, selectivity is the most important to resolve any kind of data sets [1]. If this feature exists for all the compounds, the ambiguity associated with the factor analysis decomposition of bilinear matrices disappears and unique solutions can be attained [1,8]. However, having partial or null selectivity is the most frequent situation in real data sets. When the latter occurs, the role of some other kind of constraints related to the chemical features of concentration profiles and to the instrumental responses acquires relevance and helps to decrease significantly the domain of possible solutions. The development of new constraints belonging to this last group becomes then specially interesting for the resolution of the most complex data sets.

Numerous works have reported the usefulness of constraints for curve resolution methods like non-

*Corresponding author. E-mail:annaj@zeus.qui.ub.es.

PII S0003-2670(97)00145-1

negativity (applied to both concentration profiles and instrumental responses) [1–7,9–14], unimodality (i.e. presence of an only maximum in each concentration profile) [1–3,5–7,10–14] and closure (i.e. the total amount of the reactive constituents is forced to be the same in all the stages of a process) [1,2,10–13,15,16]. In the present work, two modifications of the unimodality constraint are examined, as well as a new constraint for chromatographic concentration profiles (called hereafter symmetry) related to the prevention of front-tailed peaks.

A two-level full factorial design has been used to generate simulated data sets which could yield a reliable picture of the effect of these new constraints in the resolution procedure. The responses measured to assess the goodness of the constraints are related to the recovery of the qualitative information and the quality of the data fit. All conclusions inferred from this basic study have been afterwards confirmed working with a real example.

2. Theory

2.1. Brief description of the ALS method

The ALS belongs to the family of iterative curve resolution methods. All these techniques share a general working procedure consisting of the refining of some initial estimates (either chromatographic or spectral) by using constraints related to the intrinsic features of the data (i.e. selectivity, zero-concentration regions,...) or to the chemical characteristics of the experimental system (i.e. non-negative concentration profiles or spectra, unimodality,...). In the present work, the initial estimates to be input in the ALS method are always built in the chromatographic direction applying the results obtained in the needle algorithm [17]; then, a constrained alternating least-squares optimization procedure runs till the convergence criterion is reached. The matrices of the concentration profiles and spectra obtained in the last iterative cycle are the definitive solutions of the resolution method when convergence is achieved. A more detailed explanation of the ALS method is out of the scope of this paper and can be found in previous works of the authors [1–3,10–14].

2.2. Presentation of the new constraints

2.2.1. Horizontal unimodality

Equal in concept to the classical unimodal constraint (i.e. no more than one maximum is allowed), this variety differs from the original version in the way it is implemented.

The common steps in both implementations can be summarized as follows:

1. Location of the largest maximum in the concentration profile (m).
2. Suppression of the left local maxima.
3. Suppression of the right local maxima.

The difference lies in the elimination of the secondary maxima: the classical unimodality (a) sets the non-unimodal elements equal to zero and the new implementation (b) equals these elements to the nearest element keeping the unimodal condition. In algorithmic notation, steps 2 and 3 can be expressed as:

2. if $c(m-i) > r \times c(m-i+1)$

(a) $c(m-i) \approx 0$ (for classical unimodality)

(b) $c(m-i) = c(m-i+1)$

(for horizontal unimodality)

3. if $c(m+i+1) > r \times c(m+i)$

(a) $c(m+i+1) \approx 0$ (for classical unimodality)

(b) $c(m+i+1) = c(m+i)$

(for horizontal unimodality)

where $c(m)$ is the maximum value of the concentration profile and the pairs $[c(m-i),\ c(m-i+1)]$ and $[c(m+i+1),\ c(m+i)]$ are the consecutive concentration values to be compared when looking for left and right local maxima, respectively. The parameter r can be optionally larger than 1; if this is the case, small departures of the unimodality are accepted.

From a graphical point of view, the classical constraint cuts the non-unimodal part of the concentration profile *vertically*, whereas the new modification does it *horizontally*. Fig. 1 illustrates the effect of both varieties of unimodality on a concentration profile. The plot stresses the positive behaviour of the new implementation for noisy peaks, normally related to minor compounds. Such peaks contain noisy spikes that are

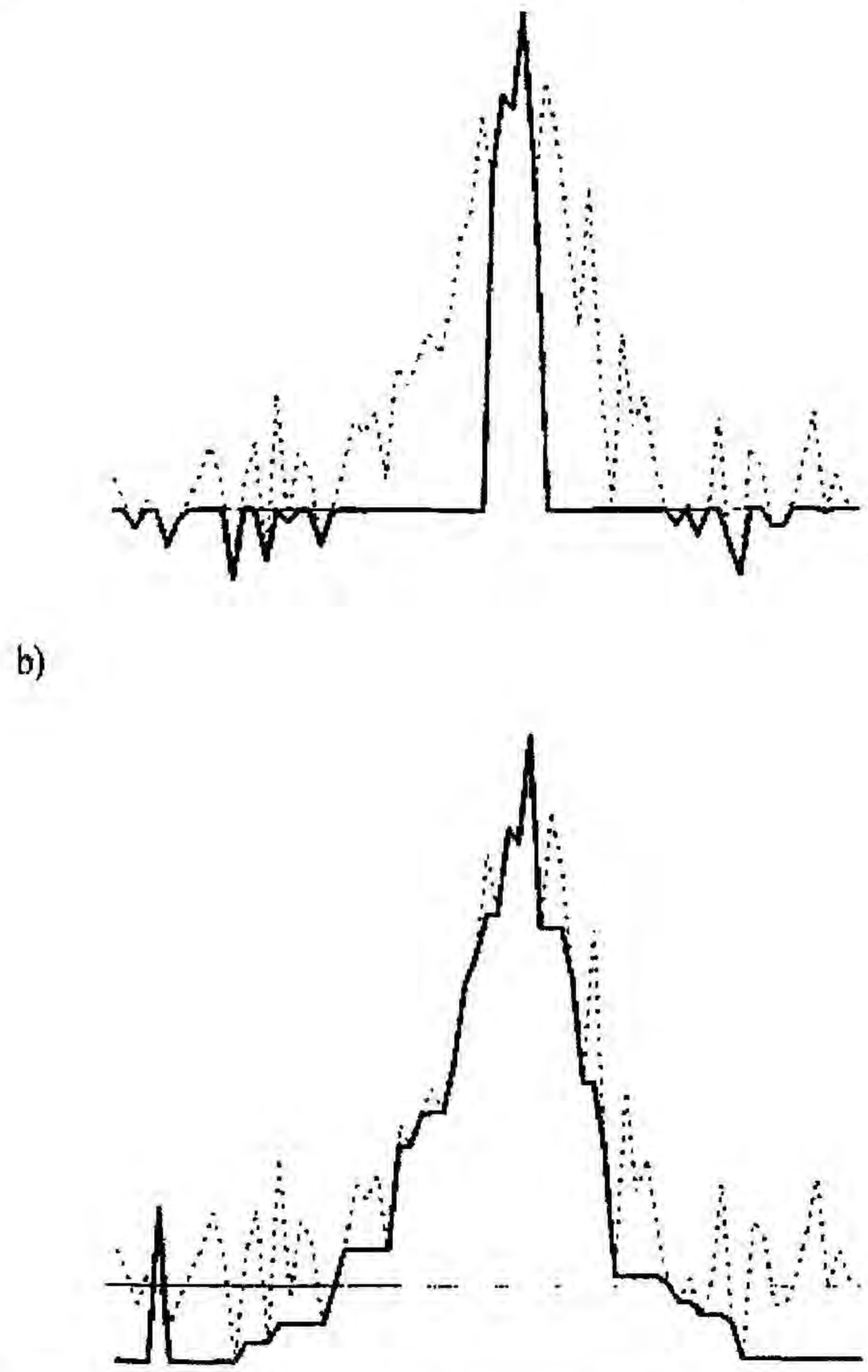


Fig. 1. Effect of the two implementations of the unimodal constraint on a concentration profile: (a) classical unimodality; (b) horizontal unimodality. Dotted line: original peak. Solid line: constrained peak.

detected as secondary maxima, this fact leading to the wrong suppression of a big part of the concentration profile when the classical implementation is applied. In contrast, the horizontal elimination of these maxima allows to keep a constrained concentration profile with a shape much closer to the original peak.

2.2.2. *Localized unimodality*

This constraint has to be considered as a more demanding version of the normal unimodality, whatever implementation applied. Both the existence of an only maximum and its position in the concentration profile are controlled. The positions taken as reference maxima for the different compounds are previously determined by using methods like the needle algorithm [17], OPA [18] or SIMPLISMA [9].

The application sequence of this constraint is detailed below. The text in italics indicates the additional step, absent in the normal unimodality, for which the algorithmic notation is also included.

1. Location of the largest maximum in the concentration profile (m).
2. Comparison between the position of the reference maximum (mt) and the position obtained from the resolution profiles (m). Relocation of the peak maximum if necessary

   if $abs(m - mt) > rpeak$

   $m = mt$.

3. Suppression of the left local maxima.
4. Suppression of the right local maxima.

The maximum shift allowed in the peak position is represented by the parameter *rpeak*, whose value is selected by the user. Analogously to the other varieties of this constraint, small departures of the unimodal condition can be optionally accepted. Steps 3 and 4 have been applied as in the horizontal unimodality.

2.2.3. *Symmetry*

Despite its name, the present constraint does not transform all the concentration profiles into symmetrical signals. Tailed peaks are accepted, provided that this asymmetry in the profile is placed after the peak maxima. Thus, real chromatographic situations where column ageing or other causes produce the distortion of the theoretical Gaussian peaks can be correctly reproduced.

The symmetry constraint is focused on the suppression of front-tailed peaks. When such a peak shape arises from the resolution results, the concentration profile of the affected compound is forced to be symmetrical. Although phenomena of band fronting can occasionally be found in real data, their occurrence is far less common than the appearance of band tailing and it is practically reduced to the domain of ion-pair chromatography [19]. Visual evidence of band fronting in a chromatogram would indicate the need of a modification in the separation parameters

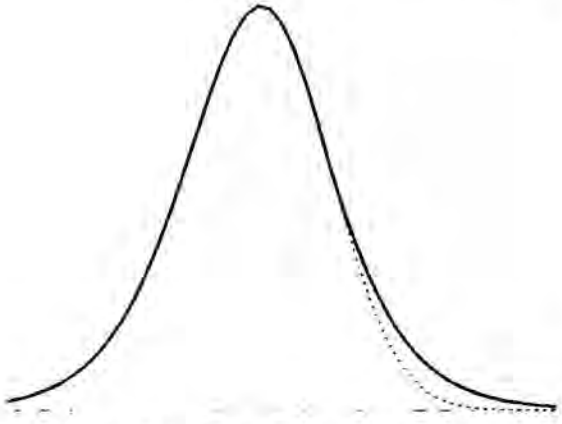


Fig. 2. Effect of the symmetry constraint on a concentration profile. Dotted line: original peak; solid line: constrained peak.

(i.e. increase of temperature, decrease of the amount of injected sample,...). The non-adequacy of the symmetry constraint would be limited to those rare situations where band fronting appears and cannot be chromatographically suppressed. Therefore, the proposal of such a constraint is chemically reasonable in by far the most cases.

The symmetry constraint reshapes the front-tailed peaks by making the back half of the peak symmetrical to the front (see Fig. 2). Gaussian and tailed peaks are not modified. The mathematical formulation of this constraint is very simple and can be explained in two steps:

1. Location of the largest maximum in the concentration profile (m).
2. Detection and suppression of band fronting:

$$\text{if } c(m-i) > r \times c(m+i)$$
$$c(m+i) = c(m-i)$$

where the pair $[c(m-i),\ c(m+i)]$ represents two concentration values equidistant to the peak maxima. Band fronting is present when the values in the left half of the peak are bigger than those in the right half of the peak. The parameter r can be optionally bigger than 1; departures of the symmetry constraint are then accepted.

3. Data sets

3.1. Simulated data sets

Two-compounds simulated data sets have been used to assess the constraints presented above. The choice of this kind of systems is related to the role they play as reference models in peak purity problems and to the fact that many real multicompound samples can be resolved analyzing submatrices of peak clusters which do not contain usually more than two or three overlapped substances. Furthermore, the simple structure of these systems allows a clear interpretation of the effects of the constraints tested, in contrast with the vague conclusions inferred when very complex simulations are employed.

In all the examples the two peaks are slightly tailed and there is a major and a minor compound. Severely overlapped spectra have been chosen on purpose for all the simulated data sets (see Fig. 3). Since the constraints to be assessed affect the concentration profiles, a better evaluation of their real effect can be carried out when no spectral selectivity can influence the resolution procedure.

The general validity of the conclusions related to the quality of the new constraints has been ensured through the non-arbitrary selection of the simulations employed in the testing procedure. A two-level factorial design has been used to determine the features of the generated data sets [20]. Table 1 shows the relevant information concerning the factors and their

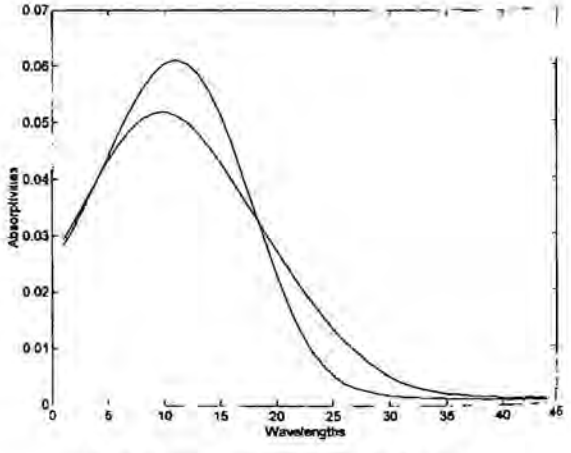


Fig. 3. Spectra used in the simulated data sets.

Table 1
Two-level full factorial design used to generate the simulated data sets. List of the coded properties related to each simulation

Simulation	Factors				
	A	B	C	D	E
1	−	−	−	−	−
2	+	−	−	−	−
3	−	+	−	−	−
4	+	+	−	−	−
5	−	−	+	−	−
6	+	−	+	−	−
7	−	+	+	−	−
8	+	+	+	−	−
9	−	−	−	+	−
10	+	−	−	+	−
11	−	+	−	+	−
12	+	+	−	+	−
13	−	−	+	+	−
14	+	−	+	+	−
15	−	+	+	+	−
16	+	+	+	+	−
17	−	−	−	−	+
18	+	−	−	−	+
19	−	+	−	−	+
20	+	+	−	−	+
21	−	−	+	−	+
22	+	−	+	−	+
23	−	+	+	−	+
24	+	+	+	−	+
25	−	−	−	+	+
26	+	−	−	+	+
27	−	+	−	+	+
28	+	+	−	+	+
29	−	−	+	+	+
30	+	−	+	+	+
31	−	+	+	+	+
32	+	+	+	+	+

Factors	levels (−)	(+)
Constraint (A)	Absennce	Presence
Resolution (B)	0.2	0.8
Ratio minor/major compound (C)	1 : 100	1 : 10
Noise pattern (D)	Homosecdastic	Heteroscedastic
S/N ratio minor compound	20	50

levels, the selected design and the coded properties of all the simulated data sets.

The factors included in the design are parameters whose influence on the results of a resolution method is either proven or at least potential and the two levels set for each of the factors have been chosen trying to cover a wide span of real situations. Thus, the constraint to be checked is introduced as a factor, whose qualitative levels are absence (−) and presence (+). The negative level of this factor has been differently defined according to the constraint to be checked (detailed explanations in Section 4.1.2 and next). The remaining factors are features of the chromatographic system, namely the resolution between peaks, the concentration ratio between major and minor compound, the noise pattern and the signal-to-noise ratio for the minor compound. The heteroscedastic noise has been simulated by adding to a homoscedastic background a scaled contribution of this base noise proportional to the square root of the intensity of the signal, i.e. for each *ij*th point of the data matrix, the noise added can be described as follows:

$$\text{total noise}_{ij} = (\text{homosc. noise})_{ij} + \sqrt{\text{free} - \text{noise signal}_{ij}} \times (\text{homosc. noise})_{ij}. \quad (1)$$

The concept of noise level has been represented through the signal-to-noise ratio for the minor compound instead of using the common measurement % of noise with respect to the largest absorbance. In systems containing major and minor compounds, the *S/N* ratio for the small species has been considered more informative than a general parameter to evaluate the distortion of the minor signals, strongly tied to the possibility of resolving these compounds. For both homoscedastic and heteroscedastic systems, the signal-to-noise ratio for the minor compound follows the expression below:

$$S/N = \frac{\max(\text{free} - \text{noise signal minor compound})}{\sqrt{\frac{\sum_{ij}(\text{homosc. noise})_{ij}}{i \times j}}}, \quad (2)$$

where the free-noise signal for the minor compound comes from the outer product $c_m \times s_m$, being c_m and s_m the simulated concentration profile and spectrum associated with this constituent, respectively. Please note that the apparently large *S/N* values used in the present work are actually the maxima allowed for this parameter in the data sets, since far from the position of the signal maximum the ratio between signal and noise becomes considerably lower. Variations of the minor/

major concentration ratios have been simulated keeping fixed the signal related to the minor compound and modifying the signal of the major compound appropriately. Thus, all the simulations having the same *S/N* ratio for the minor compound have the same amount of added noise as well, since the free-noise signal for the minor compound remains invariant. This strategy simplifies the simulation process and the further interpretation of the results.

3.2. Real data sets

Two-compound data sets, reported as reference systems for peak purity studies [21], are used to confirm the conclusions inferred from the simulated data about the goodness of the checked constraints. These real systems contain hydrocortisone as major compound and prednisone as minor constituent. The spectra of these species are the same used in the simulated data sets, shown in Fig. 3.

4. Results and discussion

4.1. General remarks

To test each of the constraints, the ALS method has been applied to the series of 32 experiments designed according to the information in Table 1. ALS has always been run forcing non-negative concentration profiles and spectra. Even though the use of selectivity is essential in this resolution method, the information regarding this point has not been taken into account in any of the systems analyzed. In the spectral direction, no selectivity can be found as pointed out above, whereas in the chromatographic direction the detection of this feature is only evident in some systems where the resolution between peaks is large (Rs=0.8) and the signal-to-noise ratio for the minor compound quite favourable. The application of selectivity in these last cases would be unquestionable in a normal analysis; no explicit use of this constraint has been done in the present study because in the cases where it is present its strong effect would mask the influence in the ALS solution coming from the constraints to be checked.

The responses collected to analyze the effects of the proposed constraints in the quality of the resolution results are related to the recovery of the qualitative information (i.e. dissimilarities between actual and recovered concentration profiles) and to the error associated with the definitive solution (i.e. standard deviation of the residuals, σ, and the lack of fit). The mathematical expressions associated with the responses are shown below:

$$\text{dissimilarity} = \sqrt{1 - (\text{correlation coefficient})^2},$$

$$\sigma = \sqrt{\frac{\sum_{ij}(d_{ij} - d_{ij}^*)^2}{i \times j}},$$

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

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

Tables 2–4 list the numerical responses obtained in the study of each of the constraints tested. These data have been used to calculate the main effects of the factors and the 2, 3 and 4-factor interactions [20]. The information associated with the latter calculations is graphically presented in the normal probability plots shown in Figs. 4–6 [20]. These graphs represent the value of each of the calculated effects vs. the expected probability it should have if all the effects were normally distributed. The vertical scale in the plots has been transformed, so that the plotted effects could be fitted with a straight line when changes in the levels of all the factors considered do not cause any noticeable variation in the responses or, if they do, this variation is random and does not follow any definite tendency. Positive or negative points falling far from this line indicate either factors or interactions having an important influence on the responses. The graphical conclusions concerning the significance of the effects have been statistically confirmed through the application of a *t*-test in which it is investigated whether or not an effect is significantly different from zero [20]. To do so, the numerical value of each potentially significant effect is compared with a critical effect value, calculated as the product between the averaged value corresponding to the 3-factor interaction effects (such higher-order interactions are assumed to measure

Table 2
Responses obtained from the ALS runs used to test the horizontal unimodality

Simulations	Responses			
	Dis(c1)[a]	Dis(c2)[b]	σ	Lack of fit (%)
1	1.60E-01	5.26E-01	0.005	6.31
2	1.49E-01	2.09E-01	0.0023	3.06
3	2.51E-03	4.35E-01	0.0004	0.6
4	2.18E-03	1.38E-01	0.00027	0.37
5	2.31E-01	4.07E-01	0.0007	8.6
6	1.10E-01	1.72E-01	0.00034	4.2
7	3.23E-02	4.15E-01	0.00037	4.9
8	3.52E-02	1.55E-01	0.00026	3.45
9	2.21E-01	7.78E-0 1	0.0098	13
10	2.21E-01	3.09E-01	0.0055	7.31
11	3.76E-03	3.25E-01	0.00047	0.62
12	3.31E-03	1.52E-01	0.00035	0.48
13	6.91E-01	6.55E-01	0.0042	52.1
14	3.44E-01	4.61E-01	0. 0016	19.86
15	6.77E-02	7.78E-01	0.00082	10.8
16	7.16E-02	3.14E-01	0.00057	7.57
17	1.12E-01	2.47E-01	0.0012	1.63
18	1.06E-01	7.96E-02	0.0006	0.79
19	1.17E-03	8.02E-02	0.00012	0.16
20	1.19E-03	6.72E-02	0.0001	0.16
21	1.55E-01	7.05E-02	0.00028	3.5
22	1.29E-01	1.65E-01	0.00014	1.71
23	1.04E-02	8.91E-02	0.00013	1.7
24	1.01E-02	6.68E-02	0.00011	1.54
25	1.31E-01	1.97E-01	0.0018	2.4
26	1.49E-01	1.30E-01	0.0012	1.56
27	2.36E-03	8.04E-02	0.00019	0.26
28	1.82E-03	7.82E-02	0.00016	0.21
29	7.35E-01	6.47E-01	0.0027	33
30	2.78E-01	1.82E-01	0.00044	5.41
31	2.28E-02	1.34E -01	0.00025	3.37
32	2.34E-02	9.11E-02	0.00022	2.9

[a] Dissimilarity between the actual and the recovered profile for the minor compound.
[b] Dissimilarity between the actual and the recovered profile for the major compound.

Table 3
Responses obtained from the ALS runs used to test the localized unimodality

Simulations	Responses			
	Dis(c1)[a]	Dis(c2)[b]	σ	Lack of fit (%)
1	2.09E-01	1.49E-01	0.0023	3.06
2	2.09E-01	1.49E -01	0.0023	3.06
3	1.38E-01	2.18E-03	0.00027	0.37
4	1.38E -01	2.18E-03	0.00028	0.37
5	1.72E-01	1.10E-01	0.00034	4.2
6	1.72E-01	1.10E-01	0.00034	4.22
7	1.55E-01	3.52E-02	0.00026	3.45
8	1.55E-01	3.52E-02	0.00026	3.45
9	3.09E-01	2.21E-01	0.0055	7.31
10	4.28E-01	2.13E-01	0.0056	7.51
11	1.52E-01	3.31E-03	0.00035	0.48
12	1.52E-01	3.31E-03	0.00035	0.48
13	4.61E-01	3.44E-01	0.0016	19.86
14	3.22E-01	3. 44E-01	0.0017	21.18
15	3.14E-01	7.16E-02	0.00057	7.57
16	3.14E-01	7.16E-02	0.00057	7.57
17	7.96E-02	1.06E-01	0. 0006	0.79
18	7.96E-02	1.06E-01	0.00059	0.79
19	6.72E-02	1.19E-03	0.0001	0.16
20	6.72E-02	1.19E-03	0.00012	0.16
21	1.65E-01	1.29E-01	0.00014	1.71
22	1.65E-01	1.29E-01	0.00014	1.71
23	6.68E-02	1.01E-02	0.00011	1.54
24	6.68E-02	1.01E-02	0.00011	1.54
25	1.30E-01	1.49E-01	0.0012	1.56
26	1.30E-01	1.49E-01	0.0012	1.56
27	7.82E-02	1.82E-03	0.00016	0.21
28	7.82E-02	1.82E-03	0.00016	0.21
29	1.82E-01	2.78E-01	0.00044	5.41
30	1.85E-01	2.61E-01	0.00044	5.5
31	9.11E-02	2.34E-02	0.00022	2.9
32	1.34E-01	2.28E-02	0.00022	2.9

[a] Dissimilarity between the actual and the recovered profile for the minor compound.
[b] Dissimilarity between the actual and the recovered profile for the major compound.

differences arising from experimental error [20]) and the t-value corresponding to a 95% significance level and a number of degrees of freedom equal to the number of 3-factor interactions in the design. When the value of an effect is larger than the critical reference value, the tested effect is found to be significant.

In all normal probability plots, the main effect caused by the constraint is identified with the letter A. Effects found to be significant through both visual inspection and statistical diagnostic are also labelled with their corresponding capital letters (see Table 1).

The improvement of the ALS results is always connected with a decrease in the numerical value of the responses. Indeed, better recoveries of the qualitative information are reached when the dissimilarities between the actual and the calculated concentration profiles become lower and achievements of more accurate fits occur when the error parameters diminish.

Table 4
Responses obtained from the ALS runs used to test the symmetry constraint

Simulations	Responses			
	Dis(c1)[a]	Dis(c2)[b]	σ	Lack of fit (%)
1	2.09E-01	1.49E-01	0.0023	3.06
2	1.41E-01	3.39E -01	0.0069	9.2
3	1.38E-01	2.18E-03	0.00027	0.37
4	1.38E-01	2.18E-03	0.00027	0.37
5	1.72E-01	1.10E-01	0.00034	4.2
6	1.72E-01	1.10E-01	0.00034	4.2
7	1.55E-01	3.52E-02	0.00026	3.45
8	1.46E-01	4.00E-02	0.0003	4.02
9	3.09E-01	2.21 E-01	0.0055	7.31
10	2.94E-01	2.53E-01	0.0066	8.72
11	1.5 2E-01	3.31E-03	0.00035	0.48
12	2.59E-01	3.34E-03	0.0004	0.53
13	4.61E-01	3.44E-01	0.0016	19.86
14	4.72E-01	3.24E-01	0.0029	13.9
15	3.14E-01	7.16E-02	0.00057	7.57
16	2.50 E-01	8.15E-02	0.00086	8.69
17	7.96E-02	1.06E-01	0.0006	0. 79
18	7.41E-02	2.02E-01	0.0023	3.04
19	6.72E-02	1.19E-03	0.0001	0.16
20	1.07E-01	1.27E-03	0.00014	0.19
21	1.65E-01	1.29E-01	0.00014	1.71
22	1.65E-01	1.29E-01	0.00014	1.7
23	6.68E-02	1.01E-02	0.00011	1.54
24	6.88E-02	1.13E-02	0.00011	1.53
25	1.30E-01	1.49E-01	0.0012	1.56
26	1.14E-0 1	3.58E-01	0.0065	8.59
27	7.82E-02	1.82E-03	0.00016	0.21
28	7.22E-02	1.82E-03	0.00016	0.21
29	6.47E-01	7.35E-01	0.0027	33
30	2.38E-01	3.47E-01	0.00055	6.75
31	1.34E-01	2.28E-02	0.00025	3.37
32	1.34E-01	3.94E-02	0.00032	4.22

[a] Dissimilarity between the actual and the recovered profile for the minor compound.
[b] Dissimilarity between the actual and the recovered profile for the major compound.

Factors or interactions showing significant negative effects indicate an improvement of the ALS results when going from the negative level to the positive level in the simulations. On the contrary, significant positive effects are associated with factors or interactions causing improvements in the ALS results when going from the positive to the negative level. Therefore, the interpretation of the results have to be performed according to the chemical meaning of

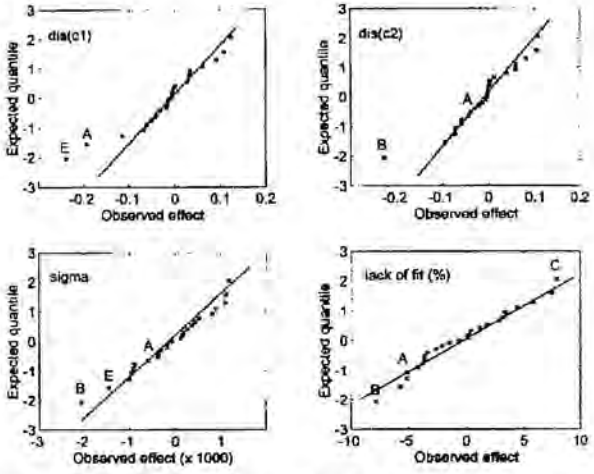


Fig. 4. Normal probability plots related to the assessment of the horizontal unimodality constraint. Significant effects are labelled with capital letters (see Table 1 for identification) and the constraint effect is always identified with the letter A. The responses analyzed are the dissimilarity between the actual and the recovered concentration profile for the minor compound, dis(c1), and the major compound, dis(c2), and the lack of fit.

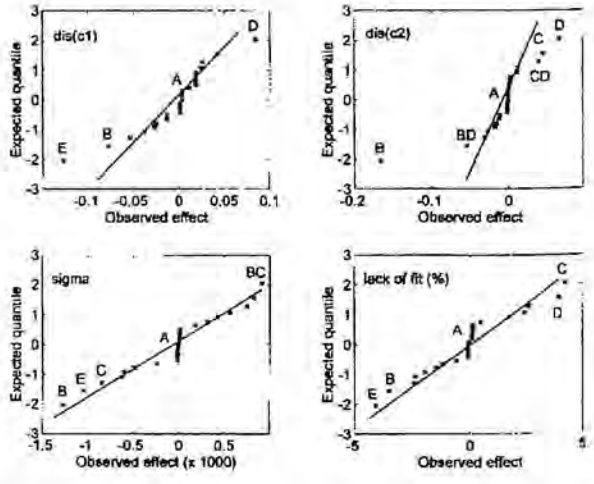


Fig. 5. Normal probability plots related to the assessment of the localized unimodality constraint. Significant effects are labelled with capital letters (see Table 1 for identification) and the constraint effect is always identified with the letter A. The responses analyzed are the dissimilarity between the actual and the recovered concentration profile for the minor compound, dis(c1), and the major compound, dis(c2), and the lack of fit.

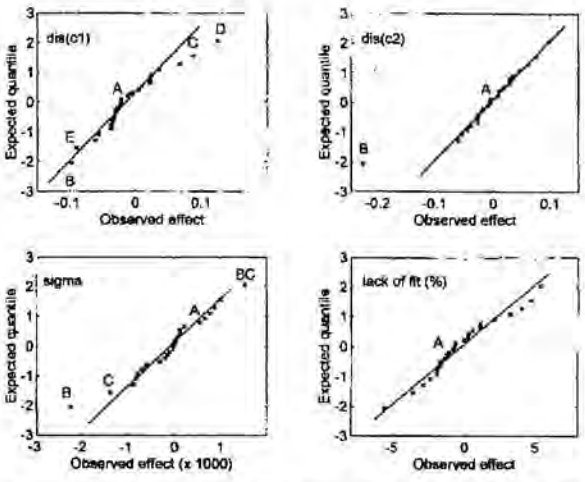


Fig. 6. Normal probability plots related to the assessment of the symmetry constraint. Significant effects are labelled with capital letters (see Table 1 for identification) and the constraint effect is always identified with the letter A. The responses analyzed are the dissimilarity between the actual and the recovered concentration profile for the minor compound, dis(c1), and the major compound, dis(c2), and the lack of fit.

the factors and the definition of their levels in the experimental design.

4.1.1. *Comments about the data features effects on the ALS results*

Figs. 4–6 show some general trends concerning the effects of the factors related to the data features on the quality of the ALS results. As pointed out above, these factors are the resolution between peaks, the minor/major concentration ratio, the noise pattern and the *S/N* ratio for the minor compound and they are identified with the capital letters B, C, D and E (see Table 1), respectively.

As it is shown in the lately mentioned figures, the recovery of the qualitative information is positively influenced by increases in the resolution between peaks and in the signal-to-noise ratio for the minor compound (the first factor being clearly the most important in the shape modelling of the major compound and the second in the modelling of the minor). Bigger dissimilarities between actual and recovered concentration profiles appear when the noise pattern is heteroscedastic and when the minor/major concentration ratio in the binary system increases. The negative action of the latter factor is linked with the decrease of the major compound signal in the simulated data sets. The consequent diminution of the signal-to-noise ratio for this constituent and the comparable contribution of both minor and major signals lead globally to a more merged and noisy binary system, where the distinction and correct modelling of the compounds becomes more difficult.

Most of the normal probability plots related to dissimilarities can be interpreted taking into account only the main effects of the factors. However, a comment related to the interactions resolution-noise pattern (BD) and minor/major ratio-noise pattern (CD) is included, since the effects of these combinations have been found significant in the dissimilarity of the major compound in Fig. 5. Both interactions show that variations in the chromatographic resolution (B) and in the minor/major concentration ratio (C) affect differently the quality of the ALS results when occurring in systems with homoscedastic or with heteroscedastic noise pattern. Thus, the BD interaction implies that a decrease in the resolution between peaks is more critical in the recovery of the concentration profile of the main compound when the noise is heteroscedastic, whereas the CD interaction tells that decreases in the proportion of the major compound in the system affect more strongly the modelling of the concentration profile of this compound in the presence of heteroscedasticity. In both cases the explanation is directly related to the stronger diminution of the *S/N* ratio for the major compound induced by the presence of heteroscedasticity when the resolution decreases (BD) and when the proportion of major compound decreases (CD). An examination of Eq. (1) allows to understand the intense effect of the heteroscedastic pattern in both kind of interactions. This expression consists of two terms, the first one including a homoscedastic background (added to all systems, whatever noise pattern they have) and the second involving properly the heteroscedastic contribution. In data sets with equal peaks and different resolutions, the term referred to the homoscedastic contribution remains invariant, whereas the second term, scaled according to the square root of the signal, increases locally in systems with low resolution because the global signal becomes larger due to the big overlap between compounds. The comparison between systems with different minor/major concentration ratios shows that the variation of the total noise added to the data sets when

going from the low level to the high level of this factor is also larger in systems with heteroscedastic noise pattern because of the contribution of the second term in Eq. (1).

A fast examination of the normal probability plots related to the error parameters (and lack of fit) shows that the significant effects are less pronounced in these responses. The smaller variability of the error parameters in the different ALS runs analyzed is simply explained because of the rotational ambiguity associated with the decomposition of the bilinear matrices when selective information is not available (i.e. many products between matrices of concentration profiles, whose columns are linear combinations of the actual profiles, and spectra matrices, whose rows are linear combinations of the actual spectra, can reproduce the original data matrix with a similar fit) [1]. The latter statement reveals the dissimilarities as a more sensitive indicator of changes in the ALS results and therefore, simulated studies for which actual and recovered profiles are available constitute an advisable starting point to assess the influence of any modification in a resolution technique onto the final solutions. In spite of this, the error parameters are the only ones that can be determined when studying real systems and this is the reason why the knowledge of the effects caused by the different factors on them is necessary. In agreement with the dissimilarity studies, the resolution between peaks is the factor having the clearest negative effect (i.e. an increase in the chromatographic resolution reduces the error parameters in the final results); the *S/N* ratio for the minor compound affects in the same sense as well, whereas the existence of a heteroscedastic pattern yields to a worsening of the results. Variations in the minor/major concentration ratio influence differently the σ and the lack of fit. The more merged and noisy binary system arisen from the decrease in the major signal causes an expected worsening in the lack of fit. The apparently inconsistent improvement of the residual standard deviation (σ) is a consequence of the absolute character of this parameter. The values of the residuals, as such, are taken to be averaged. This causes that small residuals associated with small values of the original data matrix appear to be better than slightly bigger residuals associated with much larger numerical values of the original data matrix. This fact proofs the danger of analyzing the residuals without taking into account the magnitude of the elements of the original data matrix they are associated with. The rest of statistically significant effects are less important than those mentioned above and do not deserve an exhaustive explanation.

4.1.2. *Comments on the assessment of the tested constraints*

4.1.2.1. Horizontal unimodality The ALS method has been run forcing classical unimodality for the concentration profiles in the experiments with negative constraint level. Horizontal unimodality has been applied instead when the constraint level is positive. Small departures of the unimodal constraint have been allowed in both implementations (tolerance parameter, r=1.1).

Fig. 4 shows the normal probability plots related to the assessment of the horizontal unimodality constraint. The letter A marks the point associated with the effect of this constraint in the analyzed responses. Negative values of the constraint effect in all the plots indicate the improvement of the resolution results with the inclusion of the horizontal unimodality constraint compared with the classical implementation. Such a positive influence is specially noticeable in the shape modelling of the minor compound, where the constraint effect is found to be statistically significant. This confirms what was already commented in the theory section about the effect of both implementations of the unimodal constraint in minor peaks. The noisy signals associated with these minor constituents are often reduced to narrow peaks when the classical unimodal constraint is applied, whereas the horizontal unimodality preserves much better the shape of the original signal. The constraint effect is not so essential in the shape recovery of major compounds, since their signals detach clearly from the noise and can be more easily modelled. A remarkable positive influence of the constraint is also detected in the normal probability plot associated with the lack of fit. The value of the constraint effect for this response is not statistically significant, but it is clearly negative (i.e. the use of the horizontal unimodality produces a decrease in the lack of fit). This last conclusion is specially important to confirm the goodness of the analyzed constraint in real data sets, where the lack of fit can be determined in

Table 5
Comparison of the lack of fit for several real data sets after ALS application using different constraints

Data sets		Lack of fit (%)	
Resolution	% Minor comp.	1[a]	2[b]
0.1	5	0.93	2.39
0.1	10	0.58	0.86
0.2	1	0.51	1.72
0.3	1	0.21	0.55
0.4	1	0.79	0.80
0.5	0.5	0.23	0.37
0.8	0.5	0.20	0.33
1.0	0.5	0.30	1.91

[a] Constraints applied in the ALS method: non-negativity in concentration profiles and spectra + horizontal unimodality.
[b] Constraints applied in the ALS method: non-negativity in concentration profiles and spectra + classical unimodality.

contrast with the dissimilarities between actual and recovered concentration profiles.

Table 5 includes the lack of fit related to the resolution of several real data sets, already described in a previous section. The lower values obtained for this parameter when the unimodal condition is applied according to the horizontal modality confirm the usefulness of this new implementation. Some apparent inconsistencies can be observed when the values in the table are examined from top to bottom (i.e. comparing systems with the same amount of minor compound, lower lack of fit can be occasionally observed in systems with less chromatographic resolution). These unexpected reversals often occur when experimental data with low concentration of minor compound are used, since the noise patterns and levels in the different samples are not reproducible. Despite this fact, the validity of the comparison between the ALS results obtained using the classical unimodality and the horizontal unimodality is not questionable because the pairs of values compared (i.e. lack of fit applying ALS using both implementations) are referred to pairs of runs of the ALS method on the same data matrix.

4.1.2.2. Localized unimodality The ALS method has been run forcing horizontal unimodality for the concentration profiles in the experiments with negative constraint level. Localized unimodality has been applied instead when the constraint level is positive. Small departures of the unimodal constraint have been allowed in both varieties (tolerance parameter, r=1.1).

Fig. 5 shows the normal probability plots related to the assessment of the localized unimodality constraint. None of the plots shows a significant effect of this constraint in the resolution results, not even a noticeable positive or negative effect. Hence, in its current implementation and for the conditions spanned by the designed experiments, the localized unimodality does not seem to affect the quality of the resolution results. Additional studies revealed that no improvements were obtained when the constraint was only applied to the minor compound.

4.1.2.3. Symmetry The experiments with a negative constraint level have been resolved applying the horizontal unimodality with a tolerance parameter, r=1.1. When the constraint level is positive, ALS is run by using the horizontal unimodality and the symmetry constraint. Small departures from both constraints are allowed (tolerance parameter, r=1.1).

Fig. 6 shows the normal probability plots connected with the symmetry constraint. None of the responses is affected significantly by the introduction of the symmetry constraint; so that, no clear modifications can be noticed in the resolution results. The effect of this constraint changes its sign according to the response observed (positive for sigma and negative for the rest of responses); however, this opposed behaviour is not relevant because of the small magnitude of the constraint effect in all the plots examined.

5. Conclusions

The proposal of new constraints related to the experimental features of the data sets has been used as a strategy to improve the solutions coming from the resolution methods. All the constraints presented are related to the modelling of the concentration profiles. Among the constraints, the horizontal unimodality has a more general application and the localized unimodality and symmetry constraint are more focused on the resolution of hyphenated chromatographic systems.

An exhaustive study on a wide span of simulated data sets showed the goodness of the horizontal unim-

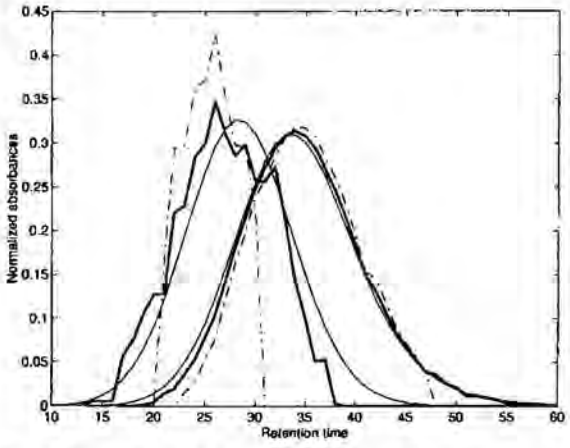


Fig. 7. Comparison of the recovered normalized concentration profiles related to a binary chromatographic system (Rs=0.2, ratio minor : major compound 1 : 100 and *S/N* for the minor compound equal to 20) by using different implementations of the unimodal constraint. True concentration profiles (single lines), profiles recovered using horizontal unimodality (thick lines) and profiles recovered using vertical unimodality (dashed lines). The minor compound is the first eluting.

odality in both the recovery of the shape of the concentration profiles and the error associated with the final solutions. The better quality of the profiles recovered using this kind of unimodality is shown in Fig. 7, where profiles obtained using this new implementation are compared with those obtained using the classical vertical unimodality. The clear decrease of the lack of fit detected in the study performed with simulated data sets allowed the confirmation of the usefulness of this constraint with real data.

Since the horizontal unimodality is exactly equal in concept to the classical unimodality, this better implementation has been applied in the test of the more demanding localized unimodality and symmetry constraints.

Neither the localized unimodality nor the symmetry constraint offered any kind of visible improvement on the resolution results according to the study performed in the present work. Since the real information included in these constraints can be potentially helpful in the modelling of concentration profiles, future research could be oriented to find more effective implementations for these latter constraints. The postulation of new constraints based on different properties of the experimental data could also be explored.

Despite the theoretical validity of the constraints proposed for most of the real chromatographic data sets, the introduction of these constraints in the resolution procedure is always optional and must be supported on the chemical knowledge of the researcher about his data. Evidence of weird behaviours (i.e. fronting phenomena) justify completely the non-application of any of the constraints proposed.

References

[1] R. Tauler, A.K. Smilde and B.R. Kowalski, J. Chemometrics, 9 (1995) 31.
[2] R. Tauler, Chemom. Intell. Lab. Sys., 30 (1995) 133.
[3] R. Tauler and D. Barceló, TrAC, 12 (1993) 319.
[4] W.H. Lawton and E.A. Sylvestre, Technometrics, 13 (1971) 617.
[5] B.G.M. Vandeginste, W. Derks and G. Kateman, Anal. Chim. Acta, 173 (1985) 253.
[6] P.J. Gemperline, Anal. Chem., 58 (1986) 2656.
[7] J. Craig Hamilton and P.J. Gemperline, J. Chemometrics, 4 (1990) 1.
[8] O.M. Kvalheim and Y.Z. Liang, Anal. Chem., 64 (1992) 936.
[9] W. Windig and J. Guilment, Anal. Chem., 63 (1991) 1425.
[10] R. Tauler, A. Izquierdo-Ridorsa, R. Gargallo and E. Casassas, Chemom. Intell. Lab. Sys., 27 (1995) 163.
[11] A. de Juan, G. Fonrodona, R. Gargallo, A. Izquierdo-Ridorsa, R. Tauler and E. Casassas, J. Inorg. Biochem., 63 (1996) 155.
[12] J. Saurina, S. Hernández-Cassou and R. Tauler, Anal. Chem., 67 (1995) 3722.
[13] E. Casassas, R. Tauler and I. Marqués, Macromolecules, 27 (1994) 1729.
[14] S. Lacorte, D. Barceló and R. Tauler, J. Chrom A, 697 (1995) 345.
[15] H. Gampp, M. Maeder, C. Meyer and A.D. Zubergühler, Talanta, 32 (1985) 1133.
[16] H. Gampp, M. Maeder, C. Meyer and A.D. Zubergühler, Anal. Chim. Acta, 193 (1987) 287.
[17] A. de Juan, B. van den Bogaert, F. Cuesta Sánchez and D.L. Massart, Chemom. Intell. Lab. Sys., 33 (1996) 133.
[18] F. Cuesta Sánchez, J. Toft, B. van den Bogaert and D.L. Massart, Anal. Chem., 68 (1996) 79.
[19] J.W. Dolan and L.R. Snyder, Troubleshooting LC Systems: A comprehensive approach to troubleshooting LC equipment and separation, Humana Press, US, 1989.
[20] G.E.P. Box, W.G. Hunter and J.S. Hunter, Statistics for Experimenters: An introduction to Design, Data Analysis and Model Building, John Wiley and Sons, US, 1978.
[21] H.R. Keller and D.L. Massart, Anal. Chem., 58 (1993) 471.

Comparison between the Trilinear Decomposition (TLD) and the Alternating Least Squares (ALS) methods for the resolution of three-way data sets.

Anna de Juan[1,*], Sarah C. Rutan[2], Romà Tauler[1] and D. Luc Massart[3].

1. Departament de Química Analítica. Universitat de Barcelona, Barcelona, Spain.
2. Department of Chemistry. Virginia Commonwealth University, Richmond, US.
3. ChemoAC. Farmaceutisch Instituut. Vrije Universiteit Brussel, Brussel, Belgium.

Abstract

Trilinear Decomposition (TLD) and Alternating Least Squares (ALS) methods are two of the most representative three-way resolution procedures. The former, non-iterative, is based on the resolution of the generalized eigenvector/eigenvalue problem and the latter, iterative, is focused on the optimization of initial estimates by using data structure and chemical constraints.

TLD and ALS have been tested on a variety of three-way simulated data sets having common sources of variation in real response profiles, such as signal shift, broadening or shape distortions caused by noise. The effect of these factors on the resolution results has been evaluated through the analysis of several parameters related to the recovery of both qualitative and quantitative information and to the quality of the overall data description.

Conclusions inferred from the simulated examples help to clarify the performance of both methods on a real example and to provide some general guidelines to understand better the potential of each method.

---

* Corresponding author.

Introduction

Curve resolution methods are usually applied in the framework of Mixture Analysis. Extended work has been reported concerning the resolution of single matrices and the necessary conditions to recover the real response profiles have recently been stated by Manne in two theorems [1]. When these requirements are not fulfilled, the decomposition of a bilinear matrix is subject to ambiguities [2,3] and the true solutions are not obtained.

Many of the limitations associated with the resolution of a single matrix are partially or completely overcome when several matrices (the so-called three-way data sets) are treated together [4]. Three-way resolution methods always introduce a significant improvement in the recovery of the true response profiles and have the additional benefit of providing quantitative information.

Two tendencies prevail within the family of three-way resolution methods: the use of non-iterative procedures, the solutions of which are based on the resolution of a generalized eigenvalue/eigenvector problem [5,6] and the application of iterative methods, focused on the optimization of initial estimates by using suitable data structure and chemical constraints [3,7].

Trilinear Decomposition [6] (TLD) and Alternating Least Squares [3] (ALS) methods are good representatives of the former and the latter tendencies, respectively. General advantages and drawbacks are recognised for both methods. TLD is fast, user-friendly and furnishes unique solutions. However, trilinear data structure is always assumed (i.e. common concentration profiles and spectra for the same species in the different appended matrices) and it is not possible to input data structure and chemical information to prevent meaningless solutions. ALS optionally assumes trilinearity and therefore can work with non-trilinear matrices sharing only one order in common (either rows or columns). Input of external information in the resolution procedure is also allowed. Despite these advantages, unique solutions are not always guaranteed, the iterative optimization slows down the resolution process and more demanding user intervention is required.

Successful chemical applications of TLD [8,9] and ALS [10-13] have been reported, but there is a lack of comparative literature about the performance of these two methods. This work presents the results of both methods on a variety of simulated three-way data sets including common sources of variation in real chemical data, such as signal shifts, broadening or distortions caused by the presence of noise. Knowledge of these synthetic data sets allows a sound assessment of the performance of both methods in terms of recovery of the response profiles, quantitation and quality of the overall data description. Conclusions inferred from this theoretical study are used afterwards to interpret the TLD and ALS results when applied to a real example [8] and some general guidelines are suggested to take advantage of the capabilities of both procedures.

Theory

Both TLD and ALS work with the original tensor or stack of matrices (**D**) to obtain three smaller matrices (**X, Y** and **Z** for TLD and **C, S** and **Z** for ALS) containing information associated with the evolution of the pure compounds profiles in each of the three directions of the initial three-way data set. Thus, the results coming from a typical data set formed by several high-performance liquid chromatography with UV diode array detection (HPLC-DAD) samples would be a first matrix having pure chromatographic profiles, a second including pure spectra and a third one with the information about the relative concentration of each compound in the different samples.

The three-way data set decomposition is carried out differently by TLD and ALS, as shown in Figure 1. The picture shows how trilinearity is inherently assumed in TLD (**X, Y** and **Z** provide only one profile per compound for all the stacked matrices), whereas ALS allows independent modelling for the concentration profiles in $\mathbf{D}_1$, $\mathbf{D}_2$,... unless the trilinear constraint is introduced in the resolution process. Detailed explanations about TLD [6,14], ALS [3,4] and their related algorithms are reported in the literature and will not be given in this section; only the main steps of each method will be mentioned for a better understanding of the comparative study.

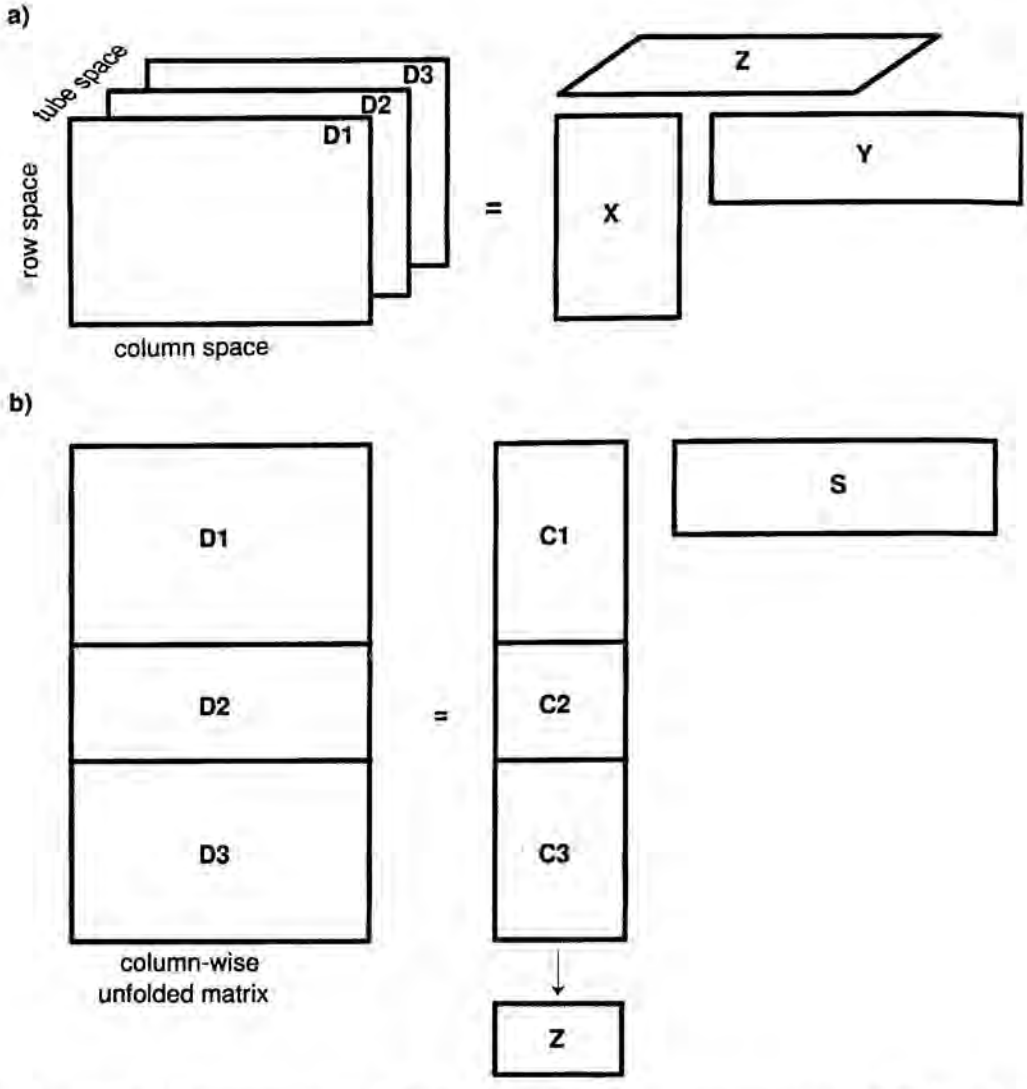


Figure 1. Three-way data decomposition according to a) Trilinear Decomposition (TLD) and b) Alternating Least Squares (ALS) resolution methods.

The working procedure of Trilinear Decomposition can be summarized as follows:

(i) Application of Singular Value Decomposition to the row-wise, column-wise and tube-wise unfolded matrices coming from the original tensor.

(ii) Construction of the basis sets formed by the row space scores (**U**), the column space scores (**V**) and the first two vectors of the tube space scores (**W**).

(iii) Determination of two representative pseudosamples $\mathbf{G}_1$ and $\mathbf{G}_2$ by projection of the original tensor on the (**U, V, W**) basis sets.

(iv) Determination of **X** and **Y** matrices from the resolution of the generalized eigenvalue/eigenvector problem for matrices $\mathbf{G}_1$ and $\mathbf{G}_2$.

(v) Least-squares estimation of **Z**, given **X** and **Y**.

Alternating Least Squares operates following the sequence below:

(i) Determination of the number of compounds of the column-wise unfolded matrix.

(ii) Building of the initial estimates matrix (either concentration profiles (**C**) or spectra (**S**)).

(iii) Selection of the constraints to be input in the iterative resolution process (e.g. non-negativity, unimodality, selectivity, trilinearity,...).

(iv) Optimization of the initial estimates by using a constrained alternating least squares procedure till the convergence criterion is fulfilled. Each iterative cycle includes the calculation and constraint of the **C** and **S** matrices (see Figure 1).

(v) Determination of the relative concentration of each particular compound in the different data matrices (**Z** matrix) as the ratio between the area of its resolved concentration profile in a matrix $\mathbf{C}_i$ and the area of the same compound in another data matrix $\mathbf{C}_j$ included in the same simultaneous analysis and taken as reference.

Both TLD and ALS algorithms have been implemented in a set of MATLAB routines [15].

Data sets

Simulated data sets

All the three-way data sets are formed by two appended matrices ($\mathbf{D}_A$ and $\mathbf{D}_B$) sized 110 × 44 representing a typical binary chromatographic system. In all the examples, the areas of the concentration profiles for compounds 1 and 2 are 25:75 in $\mathbf{D}_A$ matrix and 70:40 in $\mathbf{D}_B$ matrix. The chromatographic profiles have been simulated as slightly tailed gaussian peaks (σ = 5 channel units) and a 0.01% A_{max} homoscedastic noise level has been added to all data matrices, unless stated otherwise.

Four basic trilinear systems are proposed to represent the variety of real situations, namely:

System 1: resolution between peaks, Rs = 0.2, similar spectra.

System 2: Rs = 0.2, dissimilar spectra.

System 3: Rs = 0.8, similar spectra.

System 4: Rs = 0.8, dissimilar spectra.

Figure 2 shows the spectra used in the different simulations. Variations in real data like shift effects, signal broadening or noise addition are generated by modifying $\mathbf{D_B}$ or both $\mathbf{D_B}$ and $\mathbf{D_A}$ in the chromatographic direction according to the effect to be studied as explained below. No more than one cause of variation is simulated in each three-way data set for the sake of simplicity in the interpretation of the results. None of the simulated data sets includes changes in the spectral direction.

a)

b)

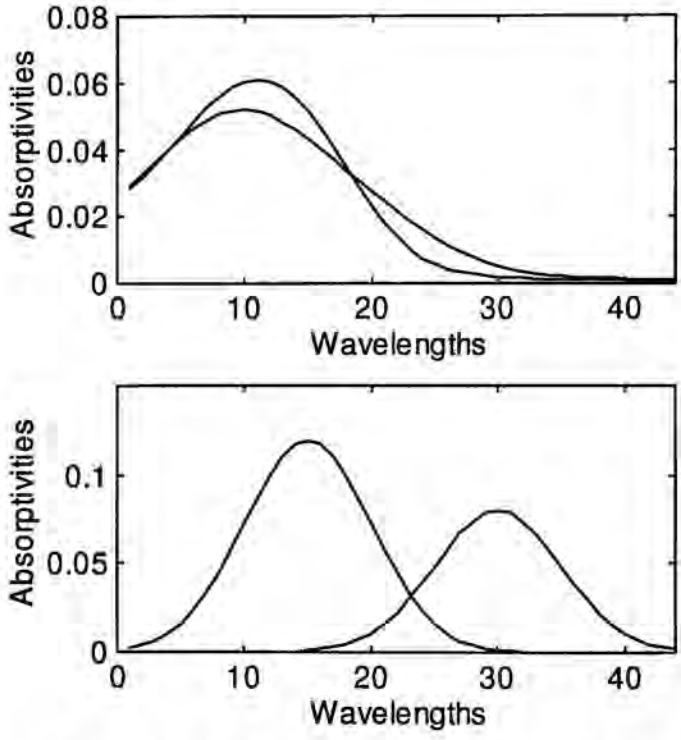


Figure 2. Spectra used in the simulations of a) systems 1 and 3 and b) systems 2 and 4.

Simulations related to signal shift

For each basic system, four three-way examples representing different signal shift levels are generated. In all of them, $\mathbf{D_A}$ matrix is not modified, whereas the peaks related to compounds 1 and 2 in $\mathbf{D_B}$ matrix are shifted apart from their initial position by 1, 2, 4 or 6 channels, depending on the shift intensity. The peak shapes are not modified in any of the two appended matrices.

Simulations related to changes in width signal

For each basic system, four three-way examples representing different width signal changes are generated. $\mathbf{D_A}$ matrix remains invariant, whereas width signals of compounds 1 and 2 are modified in matrix $\mathbf{D_B}$ by changing the σ value used in the generation of the peaks. Thus, $\Delta\sigma = -1$ and -2 reproduce situations of signal narrowing and $\Delta\sigma = +1$ and +2 account for signal broadening (please note that $\Delta\sigma$ is expressed in channel units). No changes in the peak positions are introduced in any of the two appended data matrices.

Simulations related to signal distortions caused by noise.

The effect of noise pattern and noise level is taken into account in the different simulations. Peak shapes and positions in matrices $\mathbf{D_A}$ and $\mathbf{D_B}$ are not modified. Two noise levels (0.01% A_{max} and 0.5% A_{max}) and three noise patterns (homoscedastic, heteroscedastic $\propto$ signal and heteroscedastic $\propto \sqrt{\text{signal}}$) are combined in the six simulations generated for each basic system, as shown below:

hol $\rightarrow$ noise level: 0.01% $A_{max.}$ Noise pattern: homoscedastic.

hsl $\rightarrow$ noise level: 0.01% $A_{max.}$ Noise pattern: heteroscedastic $\propto$ signal.

hql $\rightarrow$ noise level: 0.01% $A_{max.}$ Noise pattern: heteroscedastic $\propto \sqrt{\text{signal}}$.

hoh $\rightarrow$ noise level: 0.5% $A_{max.}$ Noise pattern: homoscedastic.

hsh $\rightarrow$ noise level: 0.5% $A_{max.}$ Noise pattern: heteroscedastic $\propto$ signal.

hqh $\rightarrow$ noise level: 0.5% $A_{max.}$ Noise pattern: heteroscedastic $\propto \sqrt{\text{signal}}$.

Results and discussion

Comments on the simulated data

Both TLD and ALS have been run assuming the existence of two compounds in all the three-way data sets. Trilinear Decomposition has been applied in the different unfolding directions and the results presented in this work are those related to the direction giving the best data fit. The Alternating Least Squares method has always been used introducing the non-negativity constraint in both chromatographic and spectral directions and the unimodal constraint in the chromatographic profiles. All the examples have adopted chromatographic profiles as initial estimates. The initial concentration profiles for systems 1,3 and 4 have been built applying the needle search methodology [16]. The spectral selectivity detected by using

local rank detection methods, such as FSMW-EFA [17], has been taken into account to build the initial estimates of system 2. In this case, column vectors (chromatograms) placed in spectral selective zones of the original data matrices have been selected to be input as initial chromatographic profiles. The Alternating Least Squares method has been run twice for each example: first, forcing trilinear structure in the data set and second, without doing so. To make a difference between the results coming from these two different options, the acronyms ALSf and ALS will be used hereafter to refer respectively to the presence or absence of the trilinear constraint in the iterative optimization procedure.

The performance of Trilinear Decomposition and the Alternating Least Squares methods have been compared through the observation of several parameters related to the quality of the final results. The recovery of the qualitative information, i.e. chromatographic and spectral profiles, is assessed by looking at the dissimilarities between the true response profiles and the profiles obtained by using the resolution methods.

$$dis(x_{calc}, x_{true}) = \sqrt{1 - r^2(x_{calc}, x_{true})}$$

where x_{calc} and x_{true} are the response profile calculated with ALS or TLD and the true response profile, respectively, and r is the correlation coefficient between these two factors.

The quality of the overall data description is related to the percent of lack of fit, calculated as follows:

$$\%\ \text{lack of fit} = 100 \times \sqrt{\frac{\sum_{i,j,k} r_{ijk}^2}{\sum_{i,j,k} d_{ijk}^2}}$$

where r_{ijk} is the residual of the ijkth element of the three-way data set, found as the difference between the element in the original tensor and the same element reproduced by using the results of the resolution method and d_{ijk} is the ijkth element of the original three-way data set. The quantitation ability has been evaluated with a relative parameter of comparison between the true concentration ratios of a compound in the $\mathbf{D_A}$ and $\mathbf{D_B}$ matrices and the same

concentration ratios obtained from the resolution results. The mathematical expression of this last parameter is shown below (the ith subscript indicates the different compounds in the data set).

$$\% \text{ rsdz} = 100 \times \sqrt{\frac{\sum_i \left(\left(\frac{c_{i,A}}{c_{i,B}} \right)_{true} - \left(\frac{c_{i,A}}{c_{i,B}} \right)_{calc} \right)^2}{\sum_i \left(\frac{c_{i,A}}{c_{i,B}} \right)^2_{true}}}$$

The conclusions inferred from the comparison of the Trilinear Decomposition and Alternating Least Squares methods change completely if the variation introduced in the three-way data sets is associated with a loss of the trilinear data structure (i.e. changes in width and signal position) or if it is not (i.e. signal distortions produced by noise).

As a general rule, ALS works much better than TLD and ALSf in all aspects when trilinearity is broken, owing to its greater flexibility in the modelling of the response profiles. The best performance of ALS in these situations is specially clear when the analyzed three-way data sets have selectivity in the chromatographic direction (data sets coming from system 3), in the spectral direction (data sets coming from system 2) or in both of them (data sets coming from system 4), as shown in the example of Figure 3. Indeed, when selectivity is present, there are no ambiguities in the decomposition of the augmented matrix by ALS [3] and the minimal dissimilarities between the true and the calculated profiles are due to the noise added to all data sets (please note that for binary systems, the conditions required by Manne's theorems for the correct resolution of data matrices coincide with the presence of selectivity for both compounds). When neither spectral nor chromatographic selectivity exist, as in system 1, ALS still performs better in most of the cases, as shown in Figures 4 and 5. Only some reversals in this tendency can be found when the departures of trilinearity are extremely small or when these departures are associated with an increase in the overlapping between compounds (i.e. moderate positive signal broadening). In the latter case, there is a critical zone ($+2 > \Delta\sigma > +1$) where the negative effects caused by the increase of correlation in the chromatographic direction on ALS and by the loss of trilinearity on TLD and ALSf are comparable. Such a behaviour is specially noticed in the recovered response profiles, which are actually linear combinations of the pure profiles. This distortion in the profiles shape also influences the

quantitation ability in the Trilinear Decomposition and the Alternating Least Squares procedures. On the one hand, variations in the form of the profiles (i.e. in matrices **X** and **Y**) modify the least-squares calculation of the **Z** matrix in TLD and on the other hand, these variations are also associated with changes in the peak area of the different compounds, this parameter being used to obtain the quantitative information in the ALS and ALSf. In contrast to the parameters above, the lack of fit is only severely affected by losses of trilinearity. As TLD and ALSf only use the trilinear information present in the first two compounds to reproduce the data matrix, the non-trilinear contribution excluded from the model is responsible for the increase in the lack of fit. Owing to the greater flexibility in the modelling of the profiles, ALS always inputs more information in the data reproduction, though using linear combinations of the real solutions when there is no selective information in the system, and therefore the data fit is always better than for methods with trilinear structure imposed. This behaviour and the possibility to calculate this parameter for real data sets provides an additional method to check the existence of trilinear structure in real three-way data sets since non-trilinear systems will always have a significantly greater lack of fit for TLD than for ALS when the same number of compounds is used in the resolution process.

Figures 4 and 5 show clearly the factors that affect negatively the different resolution methods. Whereas TLD and ALSf are mainly influenced by trilinearity losses (i.e. by increases in the signal shift or in the changes of width signal, whatever their sign), ALS is more sensible to variations in the correlation between the response profiles of the data set. Thus, improvements in the ALS results are associated with a decrease in the overlapping between compounds (i.e. big signal shifts or signal narrowing) and the opposite effect is promoted by increases in the correlation between profiles (i.e. signal broadening).

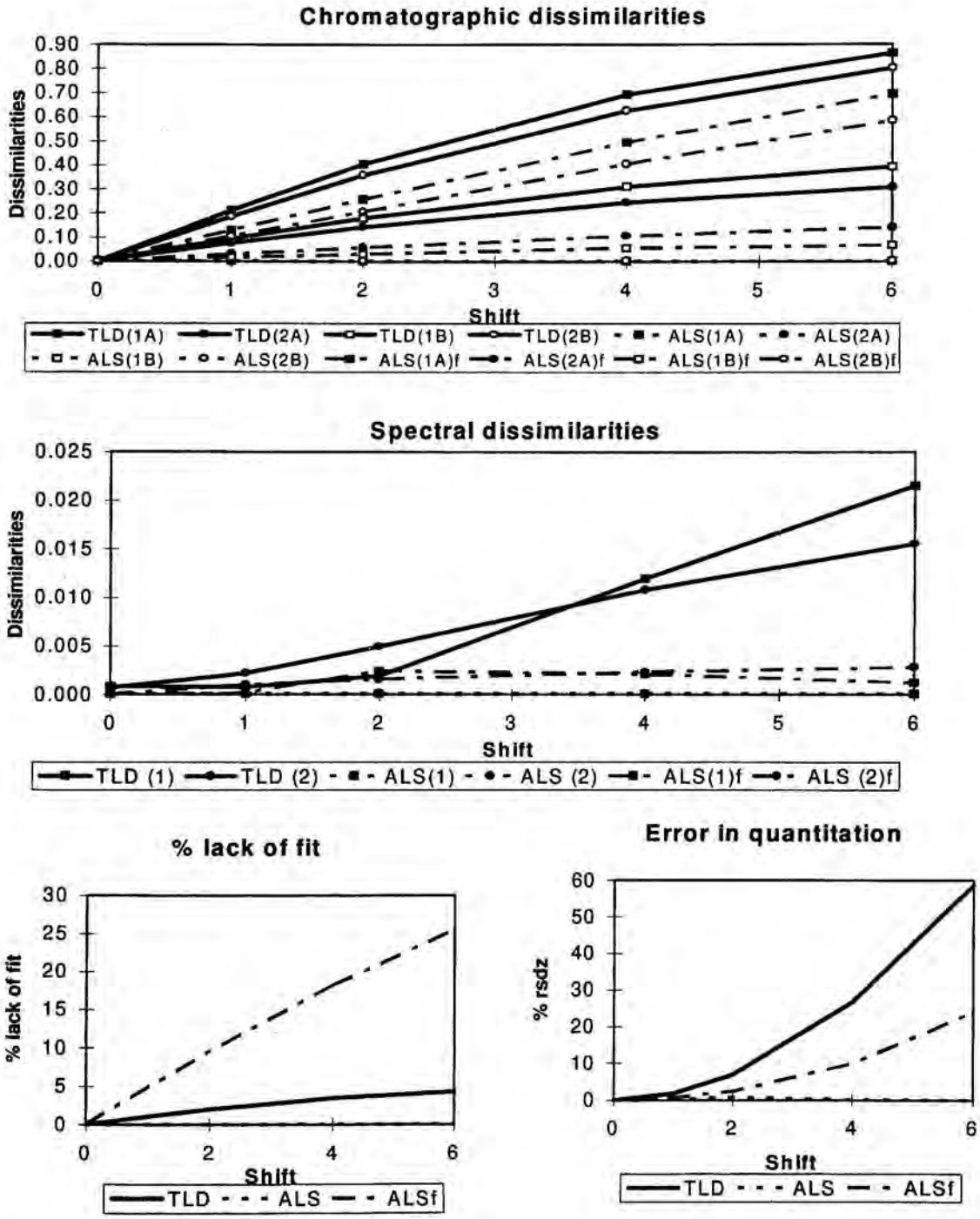
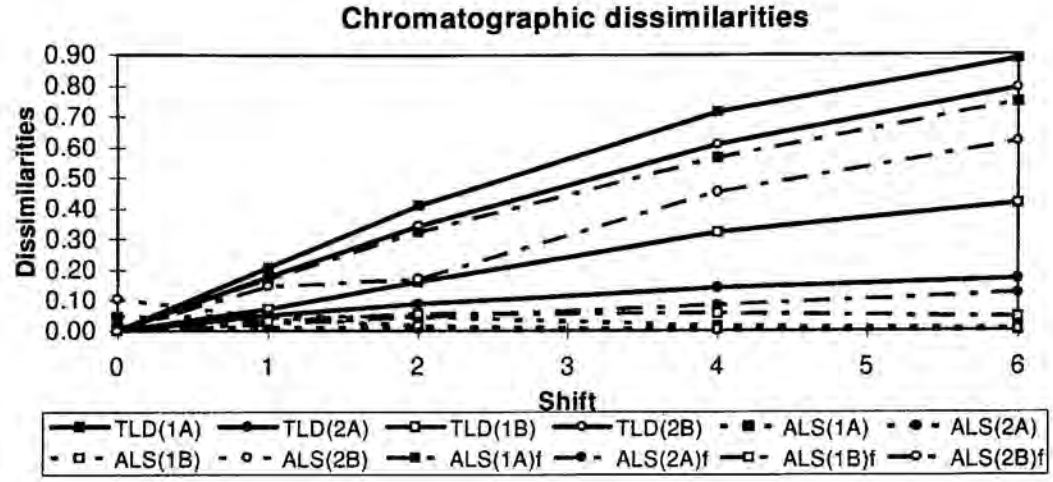
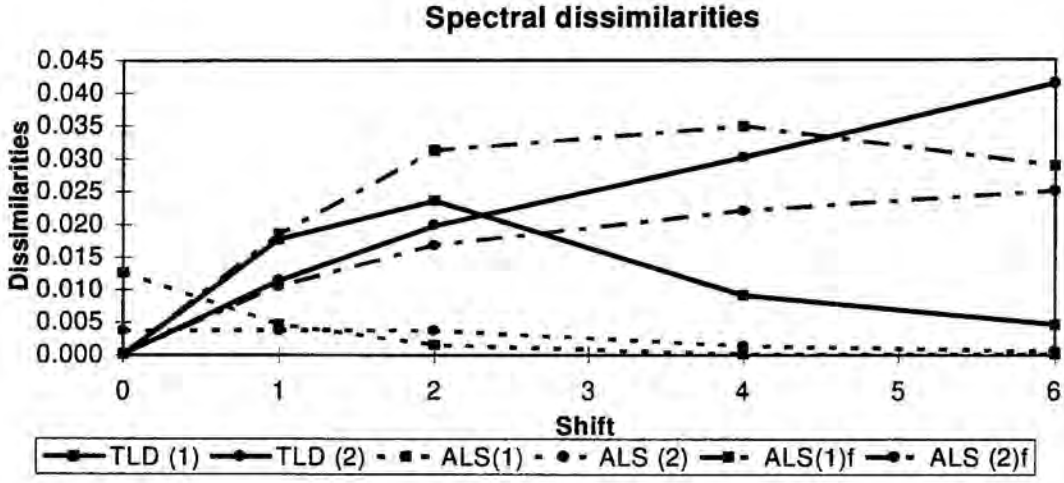
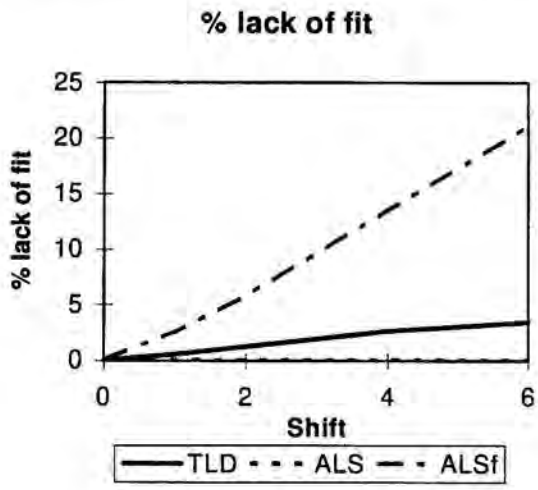


Figure 3. Effect of the signal shift on the resolution of the three-way data sets generated by modifying the basic system 3. Solid lines: TLD. Pointed lines: ALS. Dashed lines: ALSf. Compound one: squares. Compound two: circles. Compounds in $\mathbf{D}_A$ matrix: filled symbols. Compounds in $\mathbf{D}_B$ matrix: empty symbols. Legends: e.g., TLD(1A), the number and the letter between parentheses when existing indicate the compound (1 or 2) and the data matrix ($\mathbf{D}_A$ or $\mathbf{D}_B$) for which a certain parameter has been determined.







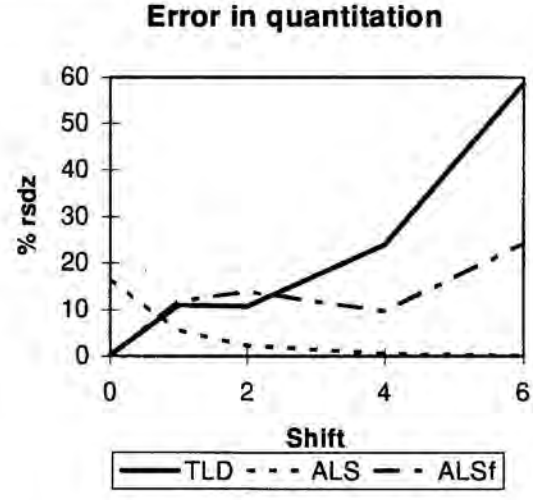
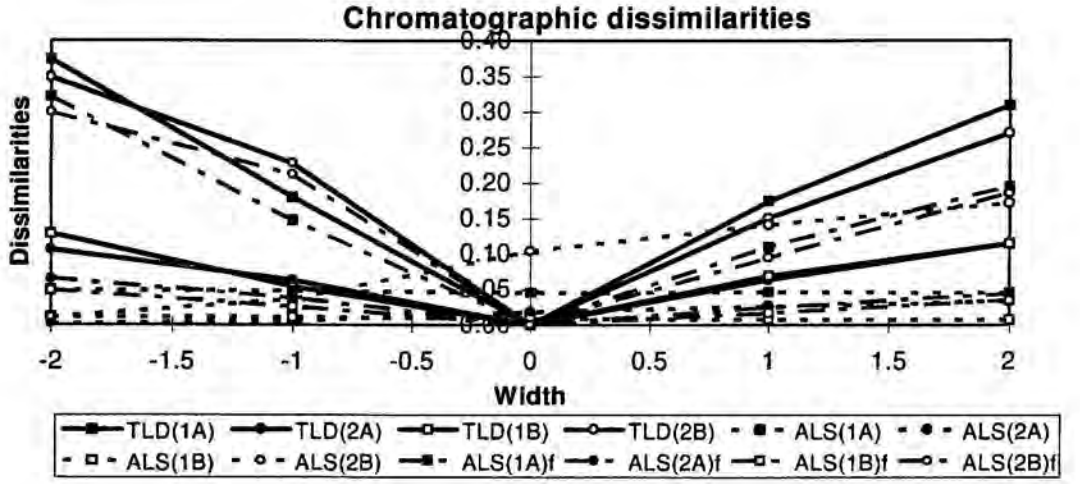
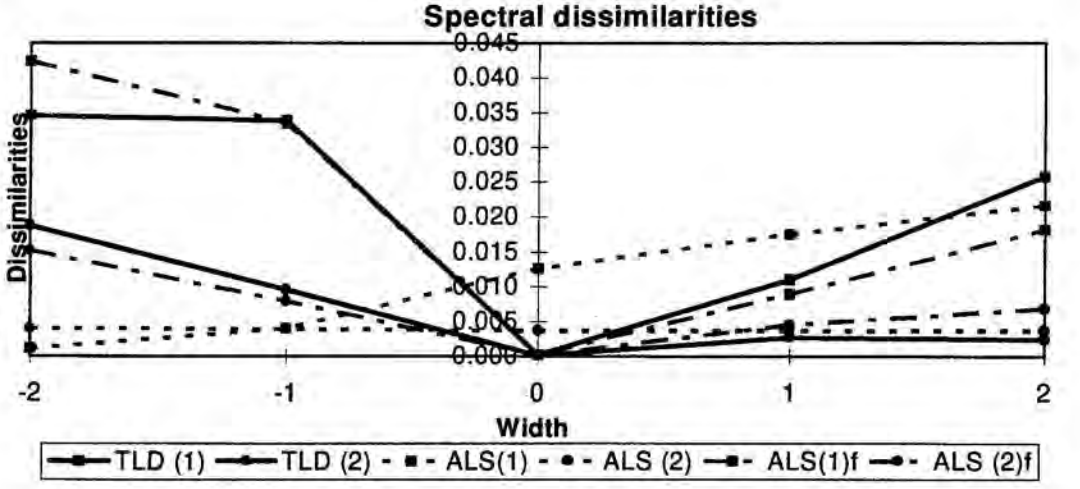


Figure 4. Effect of the signal shift on the resolution of the three-way data sets generated by modifying the basic system 1. Legends, line and symbol patterns are the same as in Figure 3.





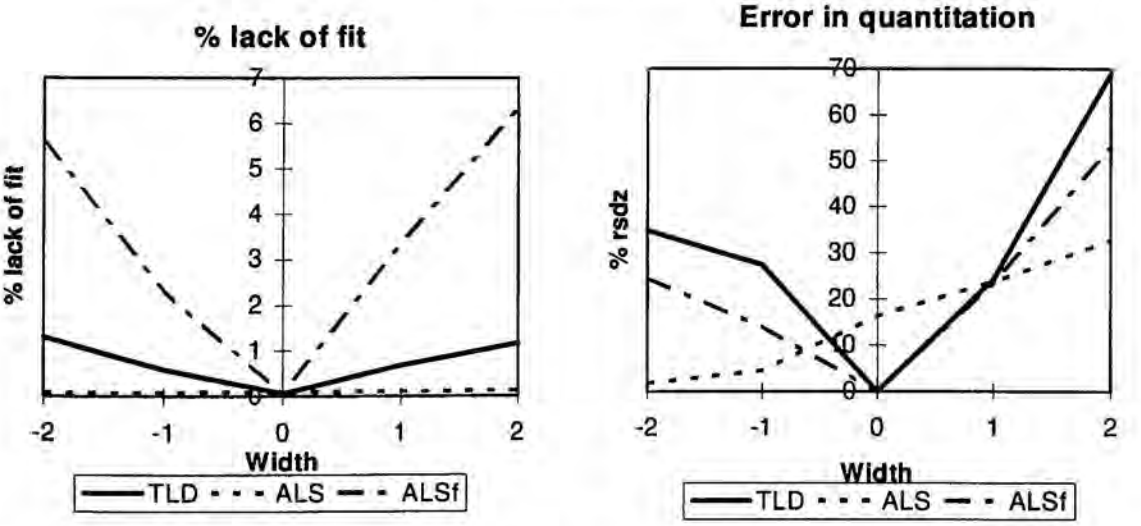


Figure 5. Effect of width signal changes on the resolution of the three-way data sets generated by modifying the basic system 1. Legends, line and symbol patterns are the same as in Figure 3.

In all the comments referred to non-trilinear systems, TLD and ALSf have been put in the same group when compared with ALS. However, noticeable differences between them derive from the distinct way to include the trilinear condition in their resolution processes. TLD works on the whole data set to obtain the trilinear combination of response profiles that best fit the original data. This strategy gives priority to the data fit over the recovered shape for the response profiles. ALSf operates one by one on each column of the augmented concentration matrix to get the best common and constrained profile for each compound [18]. In contrast to TLD, ALSf is more focused on the quality of the individual response profiles than on the global fit. The explanations above are in agreement with the general tendencies observed in figures 4 and 5, where ALSf provides smaller dissimilarities and errors in quantitation, whereas TLD shows lower lack of fit. The bigger lack of fit in the ALSf models is related to the removal of both the non-trilinear contributions and the trilinear information not in agreement with the constraints imposed in the optimization procedure (e.g. non-unimodal and negative parts of the concentration profile,...). The larger distortions in the response profiles found when TLD is applied are due to the acceptance of any kind of profile shape, provided that it keeps the trilinear condition.

An unavoidable element in the real data sets is the presence of noise. This factor by itself does not induce trilinearity losses unless it appears simultaneously with some of the data variation sources mentioned above. The simulated examples are used to analyze the stability of the resolution methods against the only introduction of different noise levels and patterns in the data sets.

All the tested three-way resolution methods are more affected by increases in the noise level than by variations in the noise pattern. Generally speaking, both TLD and ALSf provide more satisfactory results than ALS for the original noisy data sets.

The best performance of TLD and ALSf is more evident when there is no selectivity in the 3-way data set, as shown in Figure 6. In this case, owing to the intrinsic trilinear structure of the data sets and the forced trilinear character of the resolution results, both methods are able to reach the real unique solutions. In contrast, ALS suffers from the ambiguity associated with the decomposition of the bilinear matrices lacking selective information and the profiles

obtained, though correctly reproducing the original data, are linear combinations of the true response profiles. TLD and ALSf being compared, the former method gives slightly better results. For the sake of speed and easiness of application, TLD is more recommendable to treat trilinear three-way data sets with no selective information.

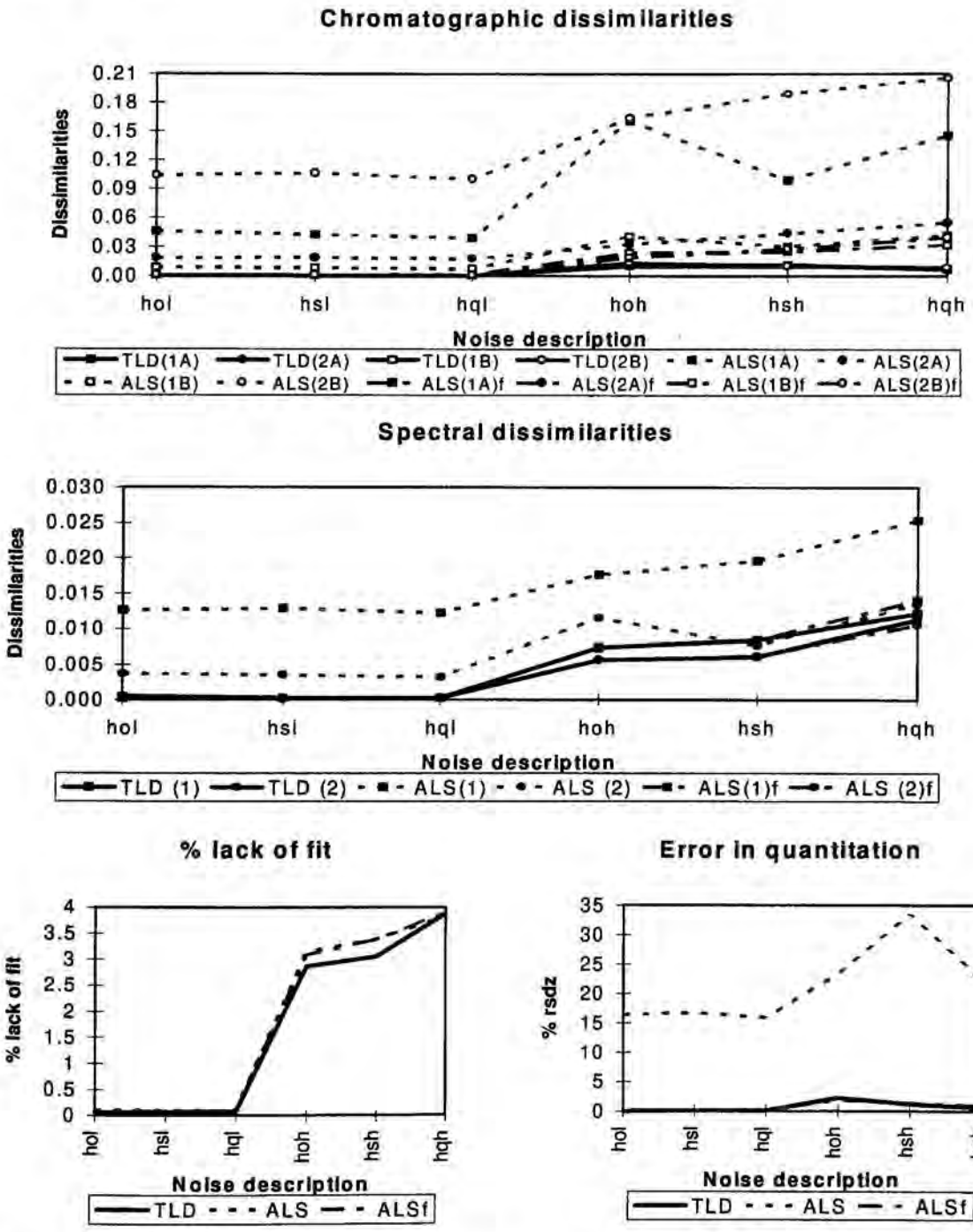


Figure 6. Effect of noise on the resolution of the three-way data sets generated by modifying the basic system 1. Legends, line and symbol patterns are the same as in Figure 3. Noise description of the data sets: **hol**→ noise level: 0.01% $A_{max.}$ Noise pattern: homoscedastic. **hoh**→ 0.5% $A_{max.}$, homoscedastic. **hsl**→ 0.01% $A_{max.}$, heteroscedastic $\propto$ signal. **hsh**→ 0.5% $A_{max.}$, heteroscedastic $\propto$ signal. **hql**→ 0.01% $A_{max.}$, heteroscedastic $\propto \sqrt{\text{signal}}$. **hqh**→ 0.5% $A_{max.}$, heteroscedastic $\propto \sqrt{\text{signal}}$.

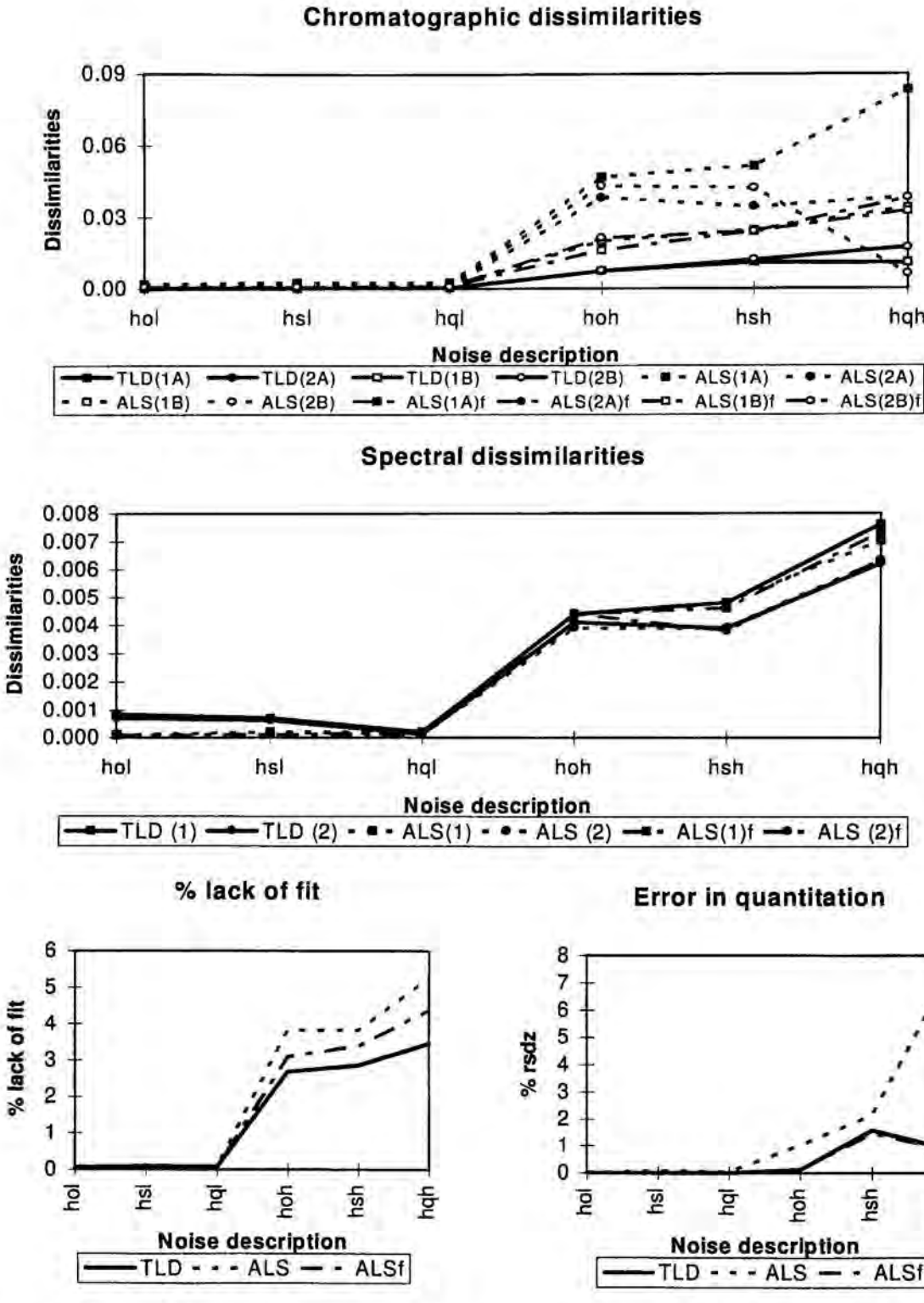


Figure 7. Effect of noise on the resolution of the three-way data sets generated by modifying the basic system 3. Line and symbol patterns are the same as in Figure 3. Noise description are the same as in Figure 6.

The application of the tested resolution methods to data sets with selective information, i.e. to the examples coming from systems 2,3 and 4 offers a rather different situation (see Figure 7). TLD, ALS and ALSf all get unique solutions in a very good agreement with the true profiles. Uniqueness in ALS is due to the presence of selectivity in the data sets. Nevertheless, TLD and ALSf usually show a subtle superiority with respect to ALS, although some occasional reversals in this behaviour can be found in some concrete simulations. Such a difference can likely come from the lesser introduction of noise in the profiles recovered by TLD and ALSf. Therefore, TLD, ALSf and all other resolution methods modelling the response profiles by only taking the trilinear information of the data set, i.e. the common information shared by all the matrices treated together, perform a smoothing effect on the profiles obtained owing to the non-inclusion of the diverse noise contributions in the resolution process. In consequence, the profiles so recovered are more similar to the true noise-free signals than those obtained by using more flexible methods, such as ALS, where the presence of noise in the final profiles can only be suppressed if it distorts the profiles in such a way that the constraints imposed in the optimization procedure are not respected.

Comments on a real example

The selected example consists of thin layer chromatographic data with fluorescence detection [8]. The matrices contain glycine and/or glutamine as analytes and two contributions owing to background effects, which are also modelled as additional compounds. The three-way data sets are formed by three matrices of glycine standards, three of glutamine standards and three having mixtures of both analytes. The two background contributions are present in all matrices. To compare the performance of the resolution methods, the data sets called experiment 1a and experiment 2 have been chosen. Experimental details about these data can be found in the original work.

A prior exploratory analysis shows the presence of a high noise level in the data sets. The application of EFA and FSMW-EFA reveals the lack of selectivity in any of the directions of the 3-way data sets. Such a feature is due to the existence of background contributions during the whole elution process and to the big chromatographic and spectral overlap between the two analytes, glycine and glutamine.

The internal structure of the data sets can be known through the comparison of the lack of fit related to the ALS and TLD results. ALS has been run using the non-negativity constraint in both chromatographic and spectral directions. Unimodality has been applied to the concentration profiles of glycine and glutamine, whereas the background profiles have not been forced to respect this constraint. TLD results are those obtained in the unfolding direction giving the best data fit. ALS gives a lack of fit significantly lower than TLD for both exp. 1a (8.29% vs. 17.3%) and exp. 2 (9.58% vs. 15.46%) and therefore, the data sets are clearly non-trilinear.

Exp. 1a and exp.2 are classified within the group of non-trilinear systems without selective information. According to the conclusions inferred from the simulated examples, the choice of ALS would be clear if there was complete or partial selectivity in the system. As it is not so, this real example can be placed in the critical zone where both TLD and ALS solutions, the first because of the non-trilinear structure of the data and the second because of the big overlap between compounds and the high noise level, are comparably damaged. In this situation, all the resolution methods can fail to find out the true solutions, but the application of both TLD and ALS can provide complementary information in order to obtain solutions as close as possible to the real ones.

A rather general observation is the best performance of ALS to recover the qualitative information. This is due to the inclusion of the non-trilinear contributions to build up the concentration profiles and the spectra and to the more reproducible procedure followed in the modelling of the response profiles by using the suitable chemical constraints. Figure 8 shows the robustness of the ALS modelling procedure through the comparison of the glycine and glutamine spectra recovered for the two three-way data sets studied independently, exp. 1a and exp. 2. As TLD works to have the best trilinear combination of response profiles and no constraints are used to control the profiles shape, the variability of the spectra between different analogous data sets is much larger and depends on how far from the trilinear situation are the different data sets.

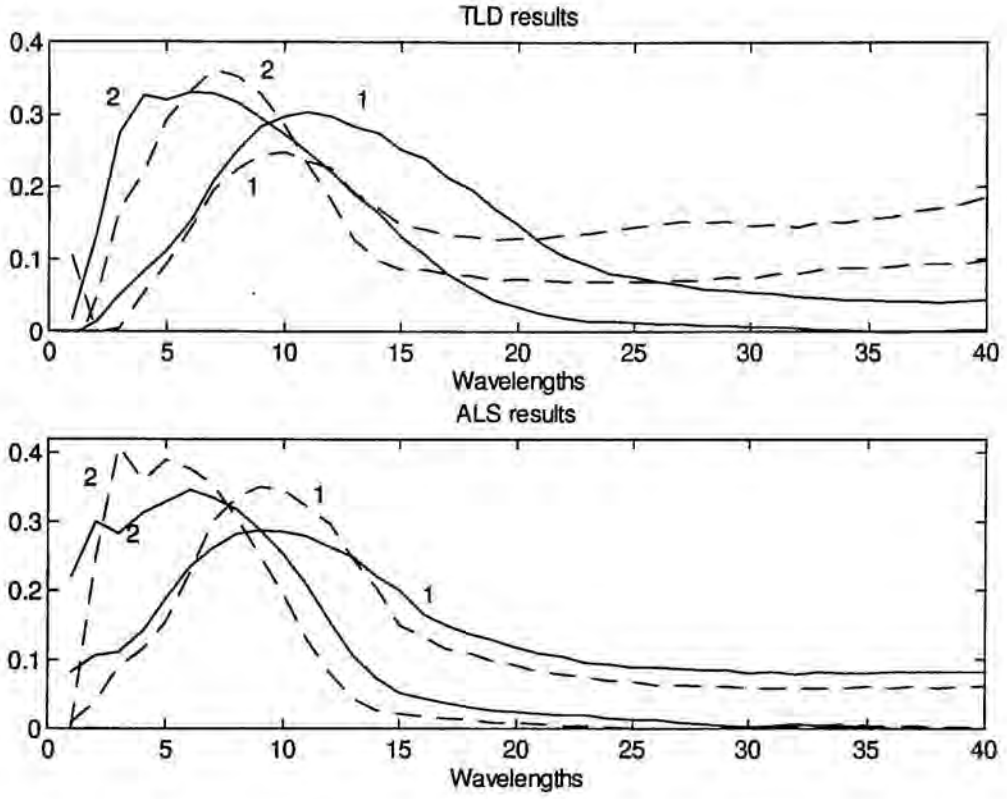


Figure 8. Spectra recovered for glycine (1) and glutamine (2) in exp. 1a (solid lines) and exp. 2 (dashed lines) by using TLD and ALS.

In contrast to the qualitative aspects of the resolution methods, there is no fix rule to know which method will provide a best quantitation. Depending on how far is the system from the trilinear condition and on some other data features, such as the noise level, the quantitation ability can turn out to be better either applying ALS or using TLD. There is no objective way to know which method must be chosen to quantify a real data set unless, as in the present example, some standards are available. In this case, the method having a better predictive ability for the known standards will consequently quantify more accurately the unknown samples. Table 1 shows the quantitative information found by TLD and ALS for the exp. 1a and exp. 2 data sets. TLD is the best method to quantify the first data set, that is likely close to be trilinear, and ALS behaves better to quantify the second. The poor quantitative performance of TLD in this last case must be attributed to the larger departure from trilinearity of data set exp. 2.

Table 1. Correlation coefficients from the calibration lines obtained comparing actual and predicted concentrations.

Experiment	TLD		ALS	
	Glycine	Glutamine	Glycine	Glutamine
1a	0.93	0.993	0.836	0.860
2	0.362	0.693	0.871	0.920

Conclusions

The existence of selectivity and the internal structure of the three-way data set (trilinear or non-trilinear) are the key features in the selection of the most suitable resolution procedure. Therefore, the first step to do is classifying the real data set according to these two properties. The comparison of the lack of fit obtained with TLD and ALS is proposed as a simple method to confirm the presence or absence of trilinearity in a three-way data set, i.e., if the lack of fit for ALS is clearly lower than for TLD using models with the same number of components, the system is non-trilinear, and vice versa if the lack of fit for TLD is equal or lower than for ALS. The presence of selectivity can be easily found applying local rank analysis, such as FSMW-EFA.

TLD and ALSf work much better than ALS with trilinear systems lacking selectivity. The presence of selectivity in the data sets makes the ALS results comparable to those coming from TLD and ALSf, though the more smoothed solutions of the latter are usually slightly better. Within TLD and ALSf, the first one is normally chosen because of the speed and the easiness of application.

ALS is the best method to deal with non-trilinear data sets. Systems lacking selectivity and having a slight departure of trilinearity, as the real example presented, are the only ones showing solutions of similar quality with all the tested resolution methods. In these situations, ALS appears to be more robust in the modelling of the response profiles whereas either TLD or ALS can provide a best quantitation.

Additional information

Graphical results completing those shown in the text about shift effect, change of width signal and noise addition on the resolution results are available on request.

Acknowledgements

A.J. and R.T. acknowledge the financial support of the Spanish government through the DGICYT PB93-0744 and DGICYT PB93-1055 projects.

References

[1] R. Manne, Chemom. and Intell. Lab. Sys., 27 (1995) 89.
[2] W.H. Lawton and E.A. Sylvestre, Technometrics,13 (1971) 617.
[3] R. Tauler, A.K. Smilde and B.R. Kowalski, J. Chemom., 9 (1995) 31.
[4] R. Tauler, Chemom. and Intell. Lab. Sys., 30 (1995) 133.
[5] E. Sánchez and B.R. Kowalski, Anal. Chem., 58 (1986) 496.
[6] B.E. Wilson, E. Sánchez and B.R. Kowalski, J. Chemom., 3 (1989) 493.
[7] A.K. Smilde, R. Tauler, J.M. Henshaw, L.W. Burgess and B.R. Kowalski, Anal. Chem., 66 (1994) 3345.
[8] M. Gui, S.C. Rutan and A. Agbodjan, Anal. Chem., 67 (1995) 3293.
[9] K.S. Booksh, Z. Lin, Z. Wang and B.R. Kowalski, Anal. Chem., 66 (1994) 2561.
[10] R. Tauler, A.K. Smilde, J.M. Henshaw and B.R. Kowalski, Anal. Chem., 66 (1994) 3337.
[11] R. Tauler, A. Izquierdo-Ridorsa, R. Gargallo and E. Casassas, Chemom. and Intell. Lab. Sys., 27 (1995) 163.
[12] A. de Juan, G. Fonrodona, R. Gargallo, A. Izquierdo-Ridorsa, R. Tauler and E. Casassas. J. of Inorg. Biochem. , 63 (1996) 155.
[13] X. Saurina, S. Hernández-Cassou and R. Tauler, Anal. Chim. Acta, 67 (1995) 3722.
[14] E. Sánchez and B.R. Kowalski, J. Chemom., 4 (1990) 29.
[15] MATLAB version 4.2. The Mathworks Inc., 1994.
[16] A. de Juan, B. van den Bogaert, F. Cuesta Sánchez and D.L. Massart, Chemom. and Intell. Lab. Sys., 33 (1996) 133.
[17] H.R. Keller and D.L. Massart, Anal. Chim. Acta, 246 (1991) 379.
[18] R. Tauler and D. Barceló, Trends Anal. Chem., 12 (1993) 319.

6.2. Application of curve resolution techniques to the study of pH-dependent transitions of some homopolynucleotides in water-dioxane mixtures.

ELSEVIER

Application of a Self-Modeling Curve Resolution Approach to the Study of Solvent Effects on the Acid-Base and Copper(II)-Complexing Behavior of Polyuridylic Acid

A. de Juan, G. Fonrodona, R. Gargallo, A. Izquierdo-Ridorsa, R. Tauler, and E. Casassas

Departamento de Química Analítica, Universitat de Barcelona, Barcelona, Spain

ABSTRACT

The solvent effect on the acid-base and complexation behavior of the homopolynucleotide polyuridylic acid (poly(U)) has been studied by means of potentiometric and spectrometric titrations (circular dichroism and UV-VIS) in water and in 30 and 50% (v/v) dioxane–water media.

The potentiometric studies revealed the absence of polyelectrolytic effects in the acid-base equilibrium, and the spectrometric experiments detected only a random coil conformation associated with both the protonated and deprotonated species. The common behavior observed in the three media seems to indicate the weakness of both intramolecular interactions, i.e., base stacking, and solute/solvent interactions, i.e., hydrogen-bonding, and consequently their small effect during the protonation process. Differences regarding the solubility of the deprotonated species in the solvents used are due to the difficult stabilization of such a charged species in the low polar environment of the dioxane–water mixtures.

Complexation has been exhaustively studied in aqueous media, and no conformational changes have been noticed in the only copper(II)-poly(U) complex detected. The inclusion of the copper(II) ion in the macromolecular skeleton of the polynucleotide does not contribute to an ordination of the structure, which remains as a random coil. No comparison between this equilibrium in aqueous solution and in the hydroorganic mixtures could be carried out since the limited pH range of the soluble complex in those solvent mixtures prevented a rigorous quantitative monitoring of such a chemical process.

Address reprint requests to: Dr. A. Izquierdo-Ridorsa, Departament de Química Analítica, Universitat de Barcelona, Diagonal 647, 08028 Barcelona, Spain.

Journal of Inorganic Biochemistry, **63**, 155–173 (1996)

655 Avenue of the Americas, NY, NY 10010

0162-0134/96/$15.00
SSDI 0162-0134(95)00211-1

INTRODUCTION

This work is in the frame of a wider study concerning the interpretation of metal ion and proton interactions with polynucleotides and their constituents under physiological conditions [1–6].

Traditionally, water has been considered the solvent which best represents the biological conditions. Although this is a general assumption, a lower polarity has been detected in some biochemical microenvironments, such as active sites of enzymes and side chains in proteins, sometimes located in lower dielectric cavities [7–9]. In these cases, the selection of other solvents seems more recommendable to properly emulate the real medium features. Among the wide range of eligible solvents, hydroorganic mixtures have been found especially suitable because they simultaneously show a low polar character and a partially aqueous content, always present in all the biological systems. In this work, dioxane–water mixtures have been selected because of the extremely different nature of their constituents, which provides solvent mixtures offering a wide diversity of features and behavior.

Equilibria involving macromolecular ligands can be affected by the presence of some important side effects. Conformational changes, polyfunctional, and polyelectrolytic effects [10] can cause significant variations in the nature and stability of the species present in the equilibria, depending on the degree of deprotonation or site occupation of the macromolecular structure. Consequently, the numerical treatment of such complex processes cannot be carried out via the traditional least-squares curve fitting methods, since all of them are supported on the postulation of a chemical model and on the fulfillment of the mass action law. A good alternative for these systems is the application of soft-modeling techniques, such as multivariate curve resolution methods based on factor analysis, that work without any assumption of a fixed model. Within this group, a self-modeling curve resolution approach, the so-called Alternating Least-Squares (ALS) method [4–6, 11–16], has been applied successfully to the spectrometric titrations. This procedure can be advantageously used for the study of macromolecular multiequilibria systems [4–6, 16] to define the number of absorbing species present in the system, and to evaluate their concentration profiles and their individual spectra. It has also proved to be a good tool to monitor the conformational changes associated either with one or more chemical species that often occur in such macromolecular systems [6]. These conformational changes are strongly influenced by experimental conditions, i.e., polymer concentration, pH, temperature, solvent, ionic strength and nature of the background electrolyte, and concentration and nature of other metal ions present in the system. Since conformational changes may have slow kinetics, time is also a critical parameter to be taken into account. Many published results concerning homopolynucleotide conformations have been obtained at fixed working conditions, but there is a lack of quantitative information related to the equilibria between different conformations. To study the evolution of such chemical processes when a variable is modified, i.e., pH in the present work, all of the other experimental conditions have to be fixed during the whole experiment.

In this work, the solvent effect on the acid-base and copper(II) complexing behavior of a pyrimidinic homopolynucleotide, the polyuridylic acid (poly(U)),

has been evaluated. This synthetic polynucleotide contains identical monomeric units, whose nitrogen base is uracil (see Fig. 1). For the sake of comparison, the monomer uridine-3′,5′-cyclic monophosphate has also been studied. This monomer can be considered as a model compound as it contains the same coordinating centers as poly(U). The media selected for such study have been water and two different dioxane–water media (30 and 50% v/v). Sigel et al. [17, 18] also used this hydroorganic system in many of their works in biocoordination chemistry, and they pointed out that the active site properties of certain kind of metalloenzymes [7] could be well represented by the mixture dioxane–water 50% (v/v). The experimental monitoring of the protonation and complexation processes has been carried out by using potentiometric, UV-visible, and circular dichroism (CD) techniques. The numerical treatment of the spectrometric titrations has been performed with the above-mentioned Alternating Least-Squares method.

EXPERIMENTAL

Reagents and Solutions

Hydrochloric acid, copper(II) oxide, and sodium chloride (Merck, a.r.), polyuridylic acid sodium salt (poly(U)) and uridine-3′,5′-cyclic monophosphate (cUMP) sodium salt (Sigma, a.r.), and dioxane (Carlo Erba, a.r.) were used without further purification.

Stock solutions of copper(II) ion were standardized by iodometric titration. Stock solutions of poly(U) and cUMP were prepared from a known amount of the solid reagent and dissolution in the corresponding solvent. The concentration of the polynucleotide solutions is referred to the concentration of the nucleotide cyclic monophosphate, which is the monomeric unit in the polynucleotide chain. Solutions of poly(U) were stored at 4°C. CO_2-free sodium hydroxide (Merck, a.r.) solutions were prepared following Kolthoff's procedure [19] and standardized with potassium hydrogenphtalate.

The ionic strength of the measured solutions was kept at 0.15 mol·L-1 by adding, when necessary, the appropriate amount of sodium chloride.

Apparatus

UV and visible absorption spectra were recorded on a Perkin-Elmer lambda-19 spectrophotometer. Circular dichroism spectra were collected with a Jasco J-720 spectropolarimeter. Instrumental control, data acquisition, and spectra preprocessing were carried out via the personal computers containing the software related to the different instruments.

For all kind of titrations, the EMF measurements were performed with an ORION 720 pH meter (with a precision of ± 0.1 mV), the titrant was added with a Metrohm Dosimat 655 autoburette (precision of volume addition, ± 0.005 mL), and the working solution was placed in a double-walled thermostatted titration cell, stirred magnetically under a continuous nitrogen flow. The temperature was set to 37°C for the experiments in aqueous solutions and to 25°C for those carried out in hydroorganic mixtures.

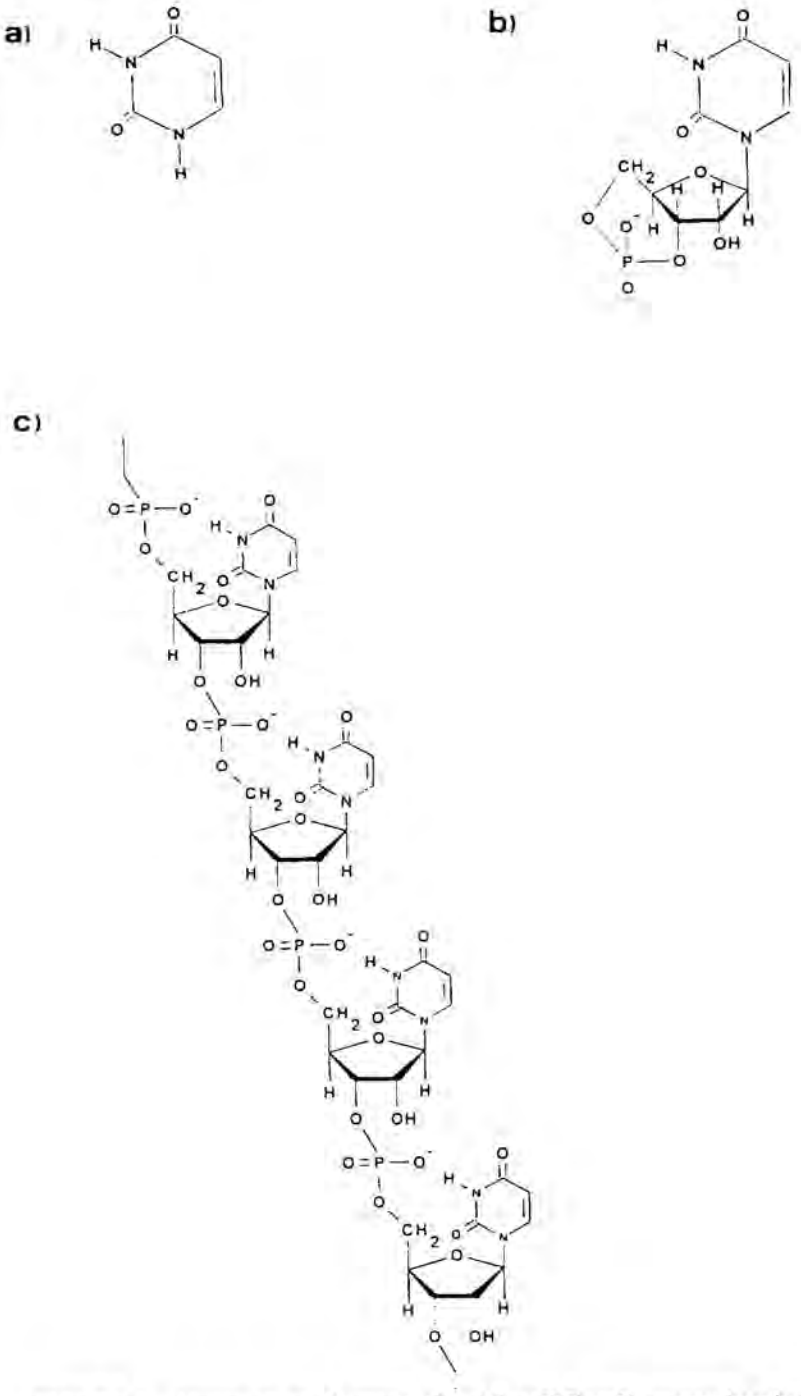


FIGURE 1. Structure of: a) uracil, b) uridine-3′,5′-cyclic monophosphate (cUMP), and c) polyuridylic acid.

Procedure

Potentiometric titrations were carried out automatically. The potentiometric setup was connected with an HP 8916 computer containing the automation software through an HP 3241 interface. For the water–dioxane mixtures, the electrodic system consisted of: G.E./W.S., I = 0.15 M, water–dioxane/R.E.(KClsat in water–dioxane), where G.E. is the glass electrode, W.S. is the working solution, and R.E. is the Ag/AgCl reference electrode, whose inner solution contains the working hydroorganic mixture in order to minimize the liquid junction potential. For the titrations in aqueous medium, a combined Ross pH electrode (ORION 81-02) was used.

The solutions employed in the UV and CD spectrometric titrations were slightly acidic solutions containing different ligand concentrations for the study of the poly(U)-H system, and different ligand-to-metal ion concentration ratios for the copper(II)-poly(U) system. For all of the UV titrations, the solution in the titration vessel was continuously pumped through the flow cell in the UV spectrophotometer. The CD study was not performed continuously because of the spectropolarimeter working procedure. After each titrant addition, the pH of the solution was measured and the corresponding UV or CD spectrum recorded. The spectra wavelength range for both CD and UV experiments was 230–310 nm, with 1 nm of difference between consecutive absorbance readings.

For all kinds of titrations, a previous calibration of the potentiometric system was carried out by Gran's method [20].

DATA TREATMENT

UV and CD Spectrometric Titrations

The interpretation of the experimental UV and CD spectrometric data has been carried out with a multivariate curve resolution approach, the Alternating Least-Squares (ALS) method [5, 11, 21]. This method has been implemented in a small set of MATLAB [22] routines. Essentially, the data treatment consists of the following steps:

1) Building up the data matrix D ($NSOLN \times NWAVE$): The experimental results of a single titration can be collected in a data matrix $\mathbf{D}_i$, whose $NSOLN_i$ rows are the spectra recorded at the successive pH values of the titration process. Each of these spectra contains $NWAVE$ absorbance readings. When several titrations are analyzed simultaneously, the augmented data matrix **D** contains the row-appended data matrices corresponding to each individual titration, $\mathbf{D}_i$. This global matrix **D** will keep the same number of columns $NWAVE$ since the spectra in all the titrations are measured at the same wavelength range, and a number of rows equal to the total number of spectra recorded in all the titrations $NSOLN = \Sigma.NSOLN_i$.

2) Determination of the number of chemical species, NS, present in each spectroscopic titration: Such a number is usually estimated by using Principal Components Analysis or some related techniques based on Factor Analysis, such as Evolving Factor Analysis (EFA) or Fixed Size Moving Window Evolving Factor Analysis (FSMW-EFA) [23–29]. The number of significant principal

components is considered equal to the number of absorbing species in the system. If this number is correctly selected, the difference between the original data matrix (**D**) and that reproduced with Principal Component Analysis (**D***), i.e., the residual variance not explained by the selected number of species, must be similar to the estimated experimental error.

3) Application of the Alternating Least-Squares (ALS) procedure to obtain the concentration profiles and spectra of each titration: The generalized Beer law equation in matrix form, **D** = **C***S, where **C** is the matrix of the concentration profiles of the different chemical species in the different titrations and **S** the matrix of their associated unit spectra, is solved iteratively by using a constrained least-squares procedure. An initial estimation of the augmented **C** data matrix is obtained by setting the concentration profiles obtained using Evolving Factor Analysis (EFA) [27–28] for each individual titration one on top of the other in the same order as they are in **D**. Either to work with individual titrations or to simultaneously treat several experiments, the iterative process is guided to the right solution by updating the spectra and the concentration profiles obtained by least-squares in each iteration according to some known constraints related to the nature of the systems studied. Thus, the concentration profile of each species is forced to be positive and unimodal, and the whole system has to fulfill the closure constraint (i.e., the sum of the concentrations of all the species in each point of the titration is constant). According to the features of the instrumental signals, the UV spectra are forced to be positive, while the CD spectra do not have to respect this requirement. In the simultaneous treatment of several titrations, the intensity and shape of the concentration profile of a chemical species present in more than one titration are allowed to change in the different titrations. This is the normal situation in the study of multiequilibria systems, where the shape of the concentration profiles in the distribution plot varies depending on the species present and on the concentration ratio among them. On the contrary, the unit spectra of the common species in the different titrations are considered to be equal and described in a unique matrix **S** (see Fig. 2). This can be assumed since the physical external conditions such as temperature, solvent composition, and ionic medium are kept constant in all of the titrations studied together.

The accuracy of the results obtained with the ALS procedure is indicated by the value of the percentage of fitting error (%error) and by the standard deviation (σ), calculated as

$$\%\text{error} = \left(\left(\Sigma(d_{ij} - d_{ij}^{*})^2\right)^{1/2} \Big/ \left(\Sigma d_{ij}^{2}\right)^{1/2}\right) * 100$$

$$\sigma = \left(\Sigma(d_{ij} - d_{ij}^{*})^2/(r*c)\right)^{1/2}$$

where d_{ij} and d_{ij}^{*} are the experimental and the ALS calculated absorbances, respectively, at pH i and wavelength j, r is the number of pH values, and c is the number of absorbance readings.

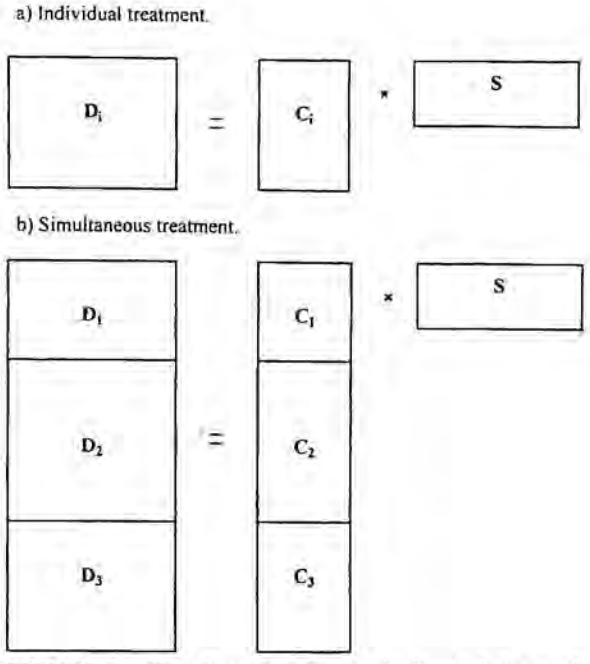


FIGURE 2. The Beer law in matrix form a) for an individual spectrometric titration, and b) for the simultaneous treatment of several titrations.

Potentiometric Titrations

The numerical analysis of the experimental EMF data from the potentiometric titrations has been carried out with the computer program SUPERQUAD [30]. The accuracy of the results obtained is indicated by the value of the parameter SIGMA, which is the ratio between the root mean square of the weighted residuals and the estimated error in our working conditions (0.005 mL for the autoburette volume readings and 0.1 mV for the EMF readings), and by the value of the statistical parameter χ^2, which is based on the distribution of the weighted residuals on EMF readings.

RESULTS AND DISCUSSION

Solvent Features of the Water–Dioxane Mixtures

Some previous studies were carried out by the authors to find the solvent effect on the acid-base and complexation behavior of some monomeric substances in dioxane–water mixtures [31, 32]. Such chemical processes are mainly affected by the solvent present in the solvation sphere of the solutes, which presents properties rather different from those of the bulk solvent. This microscopic environment can be well described through the solvatochromic parameters of Kamlet and Taft, α, β, and π^*, related to the hydrogen-bond acidity, the hydrogen-bond basicity, and the polarity of the solvent, respectively [33–35].

The solvatochromic parameters of water, as well as those related to the dioxane–water mixtures used in the present work, are listed in Table 1. When

TABLE 1. Solvatochromic Parameters of the Solvents used [32]

Solvent	α	β	π^*
Water	1.17	0.47	1.09
Dioxane–water 30% (v/v)	0.74	0.38	1.08
Dioxane–water 50% (v/v)	0.61	0.474	0.98
Dioxane	0.0	0.36	0.54

the values of such parameters for the different solvents are compared, a decrease in the global solute/solvent interaction ability is noticed in the solvent mixtures, especially in terms of hydrogen-bond acidity and polar contributions. The inclusion of dioxane does not involve the introduction of specific interactions, but the weakening of those developed by water. This fact, together with the confirmed existence of a dioxane specific solvation around the solutes in such mixtures [36], helps to emulate the real biological situations, where the difficult accessibility of water to the low polar environments severely reduces the strength of the interactions that this solvent can develop.

Poly(U)-H System

The relevant information related to the experiments performed to study this system is collected in Table 2.

Potentiometric Study

For the covered pH range, the poly(U) acid has only one protonation site to be considered: the N(3) in the nitrogen base moiety (see Fig. 1). Since this site is repeated along a macromolecular structure, the existence of polyelectrolytic effects has to be checked before applying any numerical method based on the fulfillment of the mass action law to calculate the protonation constant.

The previous potentiometric study of the poly(U)-H system in aqueous solution [4] showed the absence of appreciable polyelectrolytic effects. Similar conclusions have been inferred for the same system when studied in 30 and 50% (v/v) dioxane–water media. Figure 3 includes as an example the variation of pKa values with the degree of protonation of poly(U) in a 30% dioxane medium. Such pKa values have been calculated from the experimental EMF data for each titration point obtained in the different titrations. The horizontal lines in the plot show that the value of the protonation constant does not change depending on the number of deprotonated positions in the macromolecule; hence, the possible occurrence of polyelectrolytic effects can be discarded.

To reach a more precise evaluation of the stability constants, the application of the SUPERQUAD program, based on the traditional least-squares curve fitting approach to a postulated chemical model, has been carried out. The results obtained are included in Table 3, together with their associated statistical parameters (SIGMA and χ^2). For the sake of comparison, the values calculated for the protonation equilibrium of the monomer uridine-3',5'-cyclic monophosphate (cUMP) have also been included. Figure 4 shows the experimental formation curves for the poly(U)-H system in 50% dioxane–water mixture,

TABLE 2. Working Conditions of the Experiments Carried Out

System	Solvent	Technique	No. of Titrations	pH Range	[L] Range	$[Cu^{2+}]$ Range	$L:Cu^{2+}$ Range
cUMP-H	Water	Potentiometry	4	2.4–9.5	1.6–6.3	—	—
	D/W 30% (v/v)	Potentiometry	3	5–10.3	$4.5\cdot10^{-4}$–$6.5\cdot10^{-4}$	—	—
	D/W 50% (v/v)	Potentiometry	3	5–10.5	$2\cdot10^{-3}$–$5\cdot10^{-3}$	—	—
Poly(U)-H	Water	Potentiometry	3	3–10.0	$2.2\cdot10^{-3}$–$3.9\cdot10^{-3}$	—	—
		CD	2	3–10.6	$5.6\cdot10^{-5}$–$6.8\cdot10^{-5}$	—	—
	D/W 30% (v/v)	Potentiometry	3	5–10.5	$1.5\cdot10^{-3}$–$3\cdot10^{-3}$	—	—
		UV-VIS	3	4–11	$5\cdot10^{-5}$–$9\cdot10^{-5}$	—	—
		CD	3	5.5–11	$8\cdot10^{-5}$–$1.2\cdot10^{-4}$	—	—
	D/W 50% (v/v)	Potentiometry	3	3–10.3	$2\cdot10^{-3}$–$5\cdot10^{-3}$	—	—
		UV-VIS	3	4–10.7	$6.5\cdot10^{-5}$–$1.1\cdot10^{-4}$	—	—
		CD	3	4.5–10.3	$9\cdot10^{-5}$–$1.5\cdot10^{-4}$	—	—
Poly(U)-Cu	Water	Potentiometry	4	3.0–7.0	$2.1\cdot10^{-3}$–$3.2\cdot10^{-3}$	$2.1\cdot10^{-3}$–$1.2\cdot10^{-3}$	1:1–2.2:1
		UV-VIS	3	3.0–10.0	$8.2\cdot10^{-3}$–$1.3\cdot10^{-2}$	$4.8\cdot10^{-3}$–$1.1\,10^{-2}$	1:1—2.6:1
		CD	2	3.2–8.2	$5.6\cdot10^{-5}$–$6.1\cdot10^{-5}$	$2.8\cdot10^{-5}$–$6.1\cdot10^{-5}$	1:1–2:1

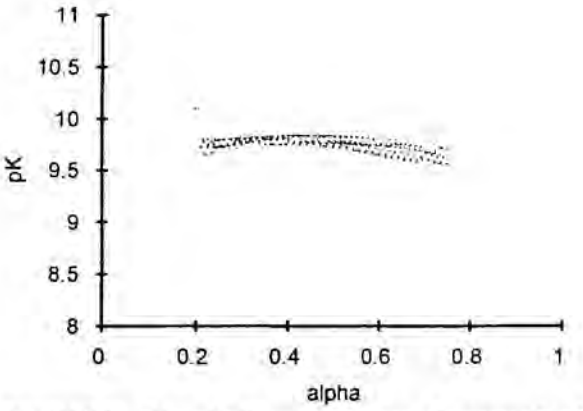


FIGURE 3. Plot of pKa vs. the degree of protonation for poly(U) in 30% dioxane–water mixture.

together with that calculated from the protonation constant value. There is good agreement between experimental and calculated curves for pH values lower than 10.5. At higher pH values, alkaline hydrolysis of the phosphate ribose ester bond takes place [3]. Furthermore, in 50% dioxane–water medium, a precipitate appears in the solutions at pH values around 10.3. This precipitate formation, which has not been observed in water or in 30% dioxane–water, has been attributed to the low polarity of the medium with the highest dioxane content, which does not stabilize the negative charges that appear on the macromolecule surface as deprotonation takes place. As expected, the higher the dioxane content, the higher the value obtained for the logarithm of the protonation constant, again a direct consequence of the decreasing polarity of the solutions. The slightly higher protonation constant of poly(U) with respect to cUMP (see Table 3) in each of the solvents studied has been attributed to steric hindrance effects in the macromolecule.

UV and CD Spectrometric Study

The macromolecular equilibria can present some structural peculiarities; to emphasize among them, there is the possibility to have several spectroscopic forms associated with only a chemical equilibrium. The presence of such different structural species with a common chemical nature, the so-called conformers, cannot be detected with a potentiometric procedure since no proton exchange, i.e., no EMF variation, may be involved when transitions between them occur. To follow both the intramolecular and intermolecular conformational variations, and to see which effect can cause a solvent change to them, the spectroscopic monitoring of the equilibria processes is absolutely necessary. Therefore, the poly(U) protonation equilibrium has also been followed by means of UV and CD spectrometric titrations.

Principal Component Analysis of both the experimental UV and CD data for the H-poly(U) system in the three studied media revealed the formation of only two different absorbing species in the working pH range, corresponding to the N(3) protonated and deprotonated ligands. This indicates that every chemical species only has a conformation since no additional absorbing species are

TABLE 3. Stability Constants Evaluated from the Experimental Data

System	Solvent	Technique	Data Treatment	Stability Ct.[a]	σ_{cv}, χ^{2c}	σ_{ALS}, % Error[d]	References
cUMP-H	Water	Potentiometry	CV[b]	8.524			[3]
	D/W 30% (v/v)	Potentiometry	CV	9.250(4)	2.14, 60.09	—	This work
	D/W 50% (v/v)	Potentiometry	CV	9.761(3)	2.62, 16.57	—	This work
Poly(U)-H	Water	Potentiometry	CV	9.364			[4]
		CD	ALS	9.1(2)	—	0.2, 5.9	This work
	D/W 30% (v/v)	Potentiometry	CV	9.756(4)	1.40, 80	—	This work
		UV-VIS	ALS	9.9(2)	—	0.02, 6.2	This work
		CD	ALS	9.75(6)	—	0.2, 5.8	This work
	D/W 50% (v/v)	Potentiometry	CV	9.914(2)	2.03, 10.66	—	This work
		UV-VIS	ALS	9.8(2)	—	0.012, 2.7	This work
		CD	ALS	—	—	—	This work
Poly(U)-Cu	Water	Potentiometry	CV	3.1			[4]
		UV-VIS	ALS	3.0			[4]
		CD	ALS	3(1)	—	0.2, 6.5	This work

[a] pKa for the acid-base equilibria and log K for the complexation processes.
[b] CV: classical curve fitting involving a postulated chemical model [30].
[c] Error parameters associated with the CV procedure [30].
[d] Error parameters associated with the ALS method [21].

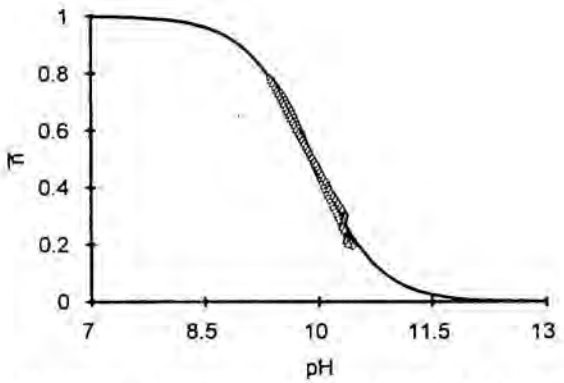


FIGURE 4. Formation curves for the poly(U)-H system in 50% dioxane–water mixture. $\bar{n}$ is the degree of protonation. Symbols: experimental data. Continuous line: theoretical curve calculated with the found protonation constant.

observed in the system while deprotonation occurs. Such a conclusion is also confirmed when Evolving Factor Analysis is applied to the different data sets. Figure 5 shows the EFA plots and the derived concentration profiles to be input in the ALS iterative procedure for UV and CD acid-base spectrometric titrations in 30% dioxane–water. Similar plots are obtained for the other titrations and solvents studied. Figure 6 includes the concentration profiles and UV spectra obtained for the protonated and deprotonated macromolecule in 50% dioxane–water after the simultaneous treatments of all of the UV titrations obtained in this medium. Few subtle changes in shape and intensity can be detected between the spectra of protonated and deprotonated species in the different media studied, although in the 30% dioxane–water mixture, a solvatochromic shift between these species is noticeable. These small differences suggest that the protonation process does not include any conformational change between the two species involved in the equilibria. Note that the spectra shown for these chemical species in the hydroorganic mixtures can be slightly distorted due to some remaining absorbance of the dioxane in the lowest wavelength region after the background correction.

The use of CD to monitor the poly(U)-H system has the advantage that the dioxane, due to its symmetrical structure, does not absorb. On the other hand, CD spectra are more noisy than UV spectra, and the necessary step of filtering also can remove small differences in shape among spectra, preventing the posterior resolution of species with very similar spectra. This has been the case for the protonation equilibria of poly(U) in 50% dioxane–water mixture, where the CD-spectroscopic titrations could not be resolved. The concentration profiles and CD spectra obtained for the protonated and deprotonated macromolecule in 30% dioxane–water after the ALS application are included in Figure 7. The spectral similarities between species, extremely high in the 50% dioxane–water mixture, confirm the conclusions obtained from the UV study, which attribute the same conformation to both protonated and deprotonated species. The quantitative results associated with the spectroscopic monitoring of the protonation equilibrium of poly(U) are collected in Table 3.

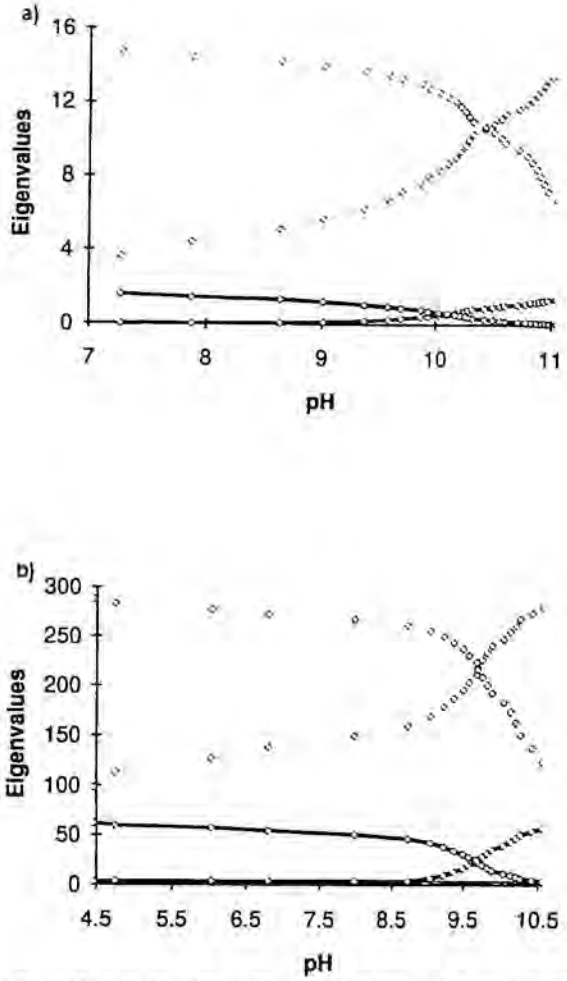


FIGURE 5. Evolving Factor Analysis plot and initial concentration profiles for the poly(U)-H in 30% dioxane–water. a) UV-VIS titration. b) CD titration. Symbols: EFA plot. Thick lines: initial estimates of the species distribution.

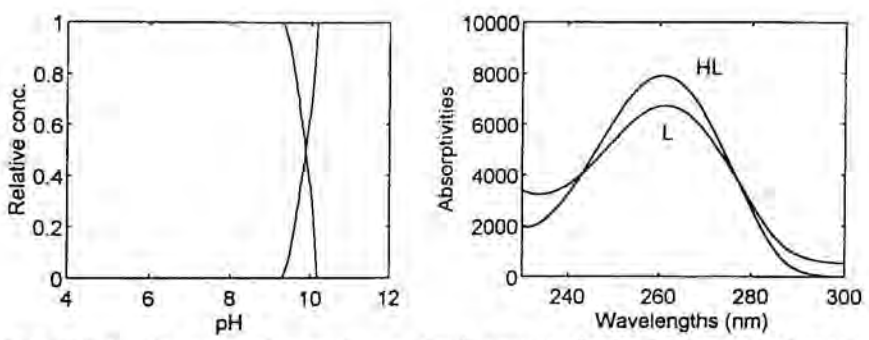


FIGURE 6. Concentration profiles and unit spectra obtained from a UV-VIS titration of the poly(U)-H system, after simultaneous treatment by means of the ALS method, in a 50% (v/v) dioxane–water solvent.

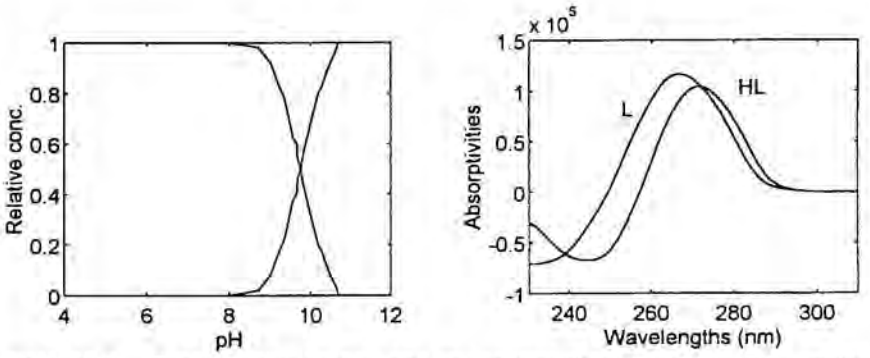


FIGURE 7. Concentration profiles and unit spectra obtained from a CD titration of the poly(U)-H system in 30% (v/v) dioxane–water after simultaneous treatment by means of the ALS method.

It has been reported that poly(U) can form highly ordered structures under certain experimental conditions, particularly at low temperatures and high ionic strength. Under these conditions, poly(U) is supposed to adopt a double-helical structure attributed either to the folding of the molecule on itself or to the association of two poly(U) single strands [37]. However, at room temperature, aqueous medium and moderate ionic strength poly(U) is considered to be present in a random coil conformation. To find which is the conformation that poly(U) shows in our experimental conditions, the effect of some special interactions has to be taken into account. One of the forces that may stabilize ordered polynucleotide structures is base stacking. It is well known that while base stacking is negligible in nonpolar solvents, it dominates in water, where other base–base interactions, such as hydrogen bonding, are greatly suppressed due to competition of the binding sites by water molecules. An increase in the intensity of base stacking, and consequently in the ordination of the structure, can be detected through both a hypochromicity in the UV spectrum and a hyperchromicity in the CD spectrum of the polynucleotide analyzed. Figure 8 shows the UV and CD spectra of the protonated poly(U) in the three media studied. The results obtained indicate the weakness of the base stacking interactions in aqueous solution, since no relevant variations in the absorption intensity are observed when less polar solvents are used. (Note that a change in conformation in CD from a random coil to an ordered structure would cause a variation in CD absorptivities of more than one order of magnitude.) An additional way to confirm the lack of structure of a polynucleotide is the comparison between its spectrum and the one obtained for its monomeric unit, which cannot develop stacking interactions, or if it does, they will be comparable in strength to those developed by a disordered macromolecular structure. Figure 9 shows simultaneously the UV and CD spectra of both the protonated poly(U) and the protonated cUMP in aqueous solution. The slight difference in intensity among the compared spectra allows us to conclude that, under the working conditions, the polyuridylic acid, poly(U), is present in a random coil conformation, in both protonated and deprotonated form, in all of the solvents studied.

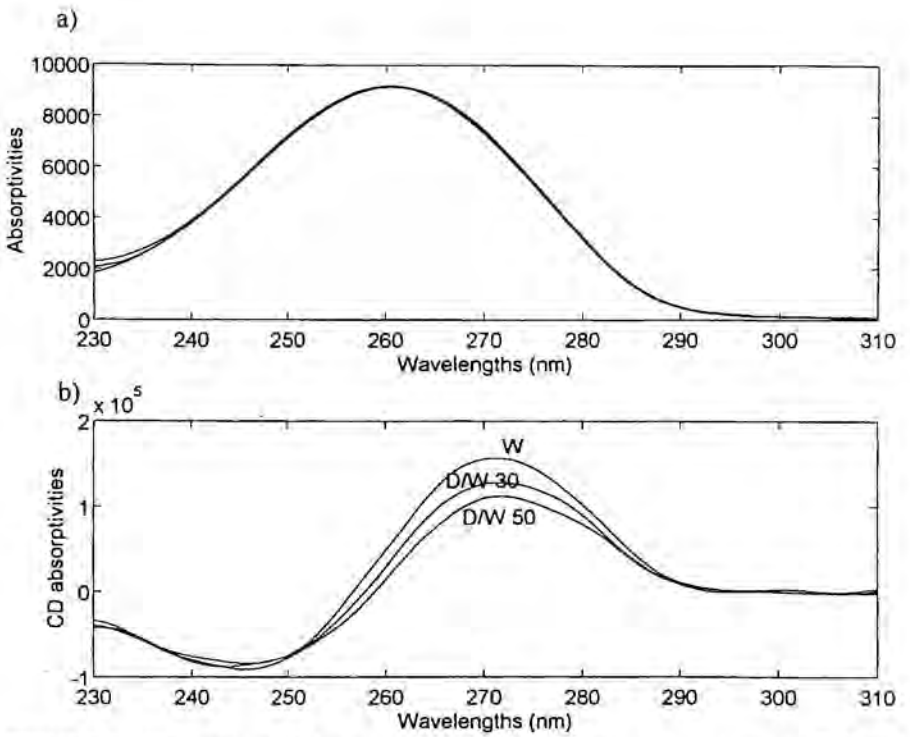


FIGURE 8. Unit spectra of the protonated poly(U) in water, dioxane–water 30% (v/v), and dioxane–water 50% (v/v). a) UV-VIS spectra. b) CD spectra.

Cu(II)-poly(U) System

The relevant information related to the experiments performed to study this system is collected in Table 2.

Previous studies focused on the effect of metal ion complexation on the polynucleotide conformations concluded that the conformations observed strongly depend on the nature of the metal ion studied [38, 39].

The potentiometric, spectrophotometric, and ESR study of the Cu(II)-poly(U) system in aqueous solution [4] showed a strong complexation in the pH range 5–7 for different rM/L (metal-to-ligand ratio). A global complexation constant was calculated from the potentiometric data with the SUPERQUAD program, and from the spectrophotometric data with the ALS procedure: log $\beta_{22-2} = 3.1$ and 3.0, respectively, where 2:2: −2 refers to polynucleotide:metal ion:hydroxide ion stoichiometry, which was the one included in the postulated model that gave the best fitting. The spectrophtometric study was carried out at the visible wavelength range, between 520 and 820 nm, at which copper(II) species absorb. However, in order to evaluate the effect that copper(II) complexation has upon the conformation of the polynucleotide, it is more convenient to carry out the study at the wavelength range between 230 and 310 nm, at which the poly(U) containing species absorb. Thus, in addition to all the known results related to the copper(II)-poly(U) system in water, in this paper, this system has been

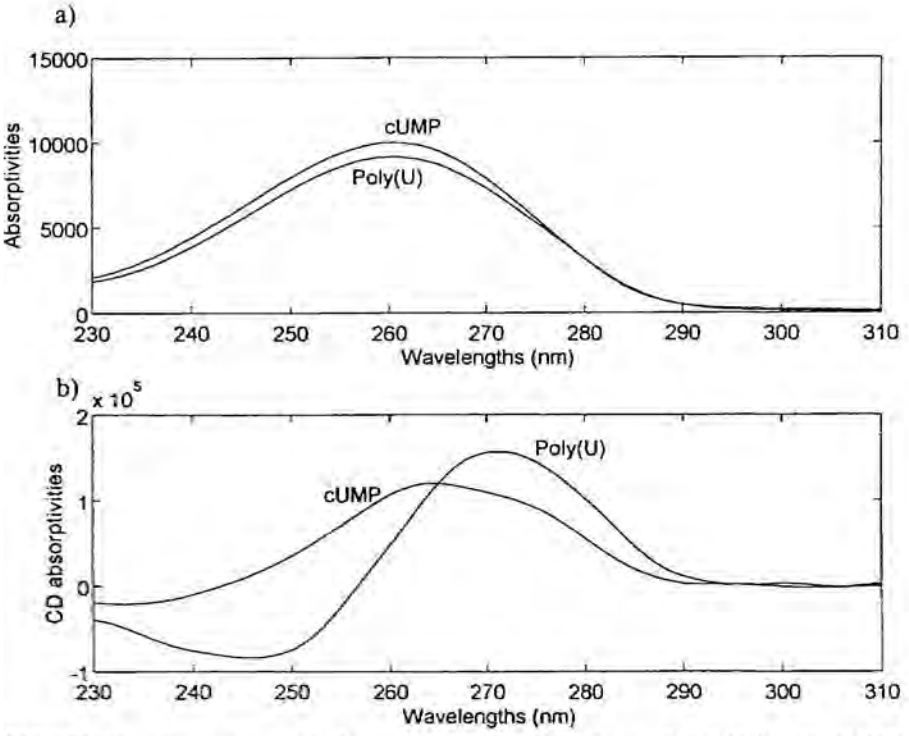


FIGURE 9. Comparison of the unit spectra of protonated poly(U) and cUMP in aqueous solution. a) UV-VIS spectra. b) CD spectra.

studied by means of CD spectrometric titrations (see Table 2) at the wavelength range between 230 and 310 nm.

As in the H-poly(U) system, the Cu(II)-poly(U) study started with the determination of the number of components. The application of Principal Component Analysis to the experimental data matrices showed two different species in the Cu(II)-poly(U) system in aqueous solution. Figure 10 shows the real concentration profiles and their associated individual spectra for the poly(U)-Cu system. The first species, present from the beginning of the titration, can be related to the protonated ligand. The second species appears at pH value around 6.5, and it corresponds to the complex species between copper(II) ions and the deprotonated ligand. The CD spectra obtained for both the protonated ligand and the complex are rather similar, and this fact indicates the absence of significant conformational changes during the complexation process. The global complexation constant obtained from the experimental concentration profiles and considering a $22-2$ stoichiometry is in good agreement with the one calculated from the potentiometric and visible data [4] (see Table 3).

No quantitative complexation monitoring could be carried out in the dioxane–water mixtures since the complex formed is quite insoluble in these media. The restricted pH range for which this species is in solution and the low concentration that it presents in it prevents the use of these experimental data to evaluate reliable stability constants related to the complexation equilibria.

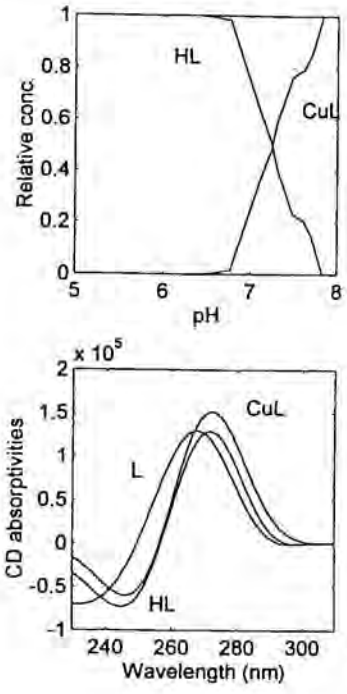


FIGURE 10. Spectra and concentration profiles obtained from the CD titrations of the poly(U)-Cu system in aqueous solution after simultaneous treatment by means of the ALS method.

However, the existence of a complex species can be confirmed qualitatively through the appearance of a blue intense precipitate, rather different in color from the insoluble copper hydroxocomplexes, which exhibit a fade blue–green color in dioxane–water mixtures. The instability of such a complex species in these hydroorganic media can be attributed to its charged character (a positive charge is present in each complexation site for the simplest 1:1 poly(U)-Cu complex). Since the polarity of the dioxane–water media is rather low, the only possibility to stabilize such a positive charged structure ought to come from a significant increase in the basic hydrogen-bonding interactions performed by the solvent on the complexation sites. The slight variation of the basicity solvatochromic parameter β in all the solvents studied, as shown in Table 1, is certainly not enough to balance the decrease in both α and π^* parameters, responsible for the acidic hydrogen-bonding and the polar solute/solvent interactions respectively, when the amount of dioxane grows in the hydroorganic mixture. Therefore, the poly(U)-Cu complex cannot remain in solution in the dioxane–water mixtures, and consequently, the complexation process cannot be followed.

CONCLUSIONS

Some dioxane–water mixtures were used to emulate physiological environments of low polarity. The solvent effect caused on the protonation and copper complexation equilibria of the polyuridylic acid was studied by means of potentiometric and spectroscopic titrations.

No variations in terms of polyelectrolytic effect or conformational changes were detected when dioxane–water mixtures were used. For the protonation equilibria, the polyelectrolytic effect was absent, and both the protonated and deprotonated species showed a random coil conformation. This null structural variation of the different species among solvents indirectly revealed the weak base stacking interactions of the uracil base in aqueous solution that could have led, if stronger, to more ordered structures. The introduction of the copper(II) ion in the macromolecular structure did not induce the formation of different combinations either.

Some charged species, such as the deprotonated poly(U) ligand and the complex Cu-poly(U), were hardly soluble in some of the hydroorganic mixtures. The low polarity of the medium and the absence of sufficiently strong hydrogen-bond interactions prevented the stabilization of such structures in solution, and consequently, the complete monitoring of their related equilibria.

This research was supported by the C.I.C.Y.T., Spain (Grant Nos. PB93-0744 and PB93-1055). One of the authors (RG) acknowledges an FI grant from Generalitat de Catalunya.

REFERENCES

1. R. Tauler, J. F. Cid, and E. Casassas, *J. Inorg. Biochem.* **39**, 277 (1990).
2. E. Casassas, A. Izquierdo-Ridorsa, and R. Tauler, *J. Inorg. Biochem.* **39**, 327 (1990).
3. E. Casassas, R. Gargallo, I. Giménez, A. Izquierdo-Ridorsa, and R. Tauler, *J. Inorg. Biochem.* **56**, 187 (1994).
4. E. Casassas, R. Gargallo, I. Giménez, A. Izquierdo-Ridorsa, and R. Tauler, *Anal. Chim. Acta* **286**, 538 (1993).
5. R. Tauler, A. Izquierdo-Ridorsa, R. Gargallo, and E. Casassas, *Chemometr. Intell. Lab. Syst.* **27**, 163 (1995).
6. E. Casassas, R. Tauler, and I. Marqués, *Macromol.* **27**, 1729 (1994).
7. H. Sigel, R. B. Martin, R. Tribolet, U. K. Häring, and R. Malini-Balakrishnan, *Eur. J. Biochem.* **152**, 187 (1985).
8. R. Kanski and C. J. Murray, *Tetrahedron Lett.* **34**, 2263 (1993).
9. Y. K. Li, A. Kuliopulos, A. S. Mildvan, and P. Talalay, *Biochemistry* **32**, 1816 (1993).
10. J. Buffle, *Complexation Reactions in Aquatic Systems. An Analytical Approach*, Ellis Horwood, Chichester, 1988.
11. R. Tauler, A. Izquierdo-Ridorsa, and E. Casassas, *Chemometr. Intell. Lab. Syst.* **18**, 293 (1993).
12. R. Tauler and E. Casassas, *J. Chemometrics* **3**, 151 (1988).
13. R. Tauler and E. Casassas, *Anal. Chim. Acta* **223**, 257 (1989).
14. R. Tauler, E. Casassas, and A. Izquierdo-Ridorsa, *Anal. Chim. Acta* **248**, 447 (1991).
15. R. Tauler and E. Casassas, *Analusis* **20**, 255 (1992).
16. R. Tauler and E. Casassas, *Chemometr. Intell. Lab. Syst.* **14**, 305 (1992).
17. G. Liogang, R. Tribolet, and H. Sigel, *Inorg. Chem.* **27**, 2877 (1988).
18. H. Sigel, *Chemical Society Reviews*, 255 (1993).
19. I. M. Kolthoff, E. B. Sandell, E. J. Meehan, and S. Bruckenstein, *Quantitative Chemical Analysis*, Collier-Macmillan Canada, Toronto, 1969.
20. G. Gran, *Analyst* **77**, 661 (1952).
21. R. Tauler, A. K. Smilde, and B. Kowalski, *J. Chemometrics* **9**, 31 (1995).
22. Matlab, The Math Works, Inc., Natick, MA.

23. E. R. Malinowski and D. E. Howery, *Factor Analysis in Chemistry*, 2nd ed., Wiley, New York, 1991.
24. E. R. Malinowski, *J. Chemometrics* **1**, 33 (1987).
25. E. R. Malinowski, *J. Chemometrics* **3**, 49 (1988).
26. S. Wold, *Technometrics* **20**, 397 (1978).
27. H. Gampp, M. Maeder, C. J. Meyer, and A. D. Zuberbühler, *Talanta* **32**, 1133 (1985).
28. H. Gampp, M. Maeder, C. J. Meyer, and A. D. Zuberbühler, *Talanta* **33**, 943 (1986).
29. H. R. Keller and D. L. Massart, *Anal. Chim. Acta* **246**, 379 (1991).
30. P. Gans, A. Sabatini, and A. Vacca, *J. Chem. Soc., Dalton Trans.*, 1195 (1985).
31. E. Casassas, G. Fonrodona, and A. de Juan, *J. Sol. Chem.* **21**, 147 (1992).
32. E. Casassas, N. Domínguez, G. Fonrodona, and A. de Juan, *Anal. Chim. Acta* **283**, 548 (1993).
33. M. J. Kamlet, J. L. Abboud, and R. W. Taft, *J. Amer. Chem. Soc.* **99**, 6027 (1977).
34. R. W. Taft and M. J. Kamlet, *J. Amer. Chem. Soc.* **98**, 2886 (1976).
35. M. J. Kamlet and R. W. Taft, *J. Amer. Chem. Soc.* **98**, 377 (1976).
36. E. Casassas, G. Fonrodona, and A. de Juan, *Inorg. Chim. Acta* **187**, 187 (1991).
37. W. Saenger, *Principles of Nucleic Acid Structure*, Springer-Verlag, New York, 1988.
38. R. Ménard and M. Zador, *Can. J. Chem.* **66**, 178 (1988).
39. V. A. Sorokin, I. V. Levchenko, V. A. Valeev, G. O. Gladchenko, Y. P. Blagoi, and R. G. Li, *Makromol. Chem.* **190**, 1595 (1989).

Received August 7, 1995; accepted September 22, 1995

ANALYTICAL BIOCHEMISTRY **249,** 000–000 (1997)
ARTICLE NO. AB972175

Three-Way Curve Resolution Applied to the Study of Solvent Effect on the Thermodynamic and Conformational Transitions Related to the Protonation of Polycytidylic Acid

Anna de Juan, Anna Izquierdo-Ridorsa, Raimundo Gargallo, Romà Tauler, Gemma Fonrodona, and Enric Casassas
Departament de Química Analítica, Facultat de Químiques, Universitat de Barcelona, Diagonal, 647, 08028 Barcelona, Spain

Received December 12, 1996

Solvent effect on the acid–base behavior of polycytidylic acid [poly(C)] is studied by means of potentiometric and spectrometric (uv and CD) procedures. Low-polarity biological environments are mimicked by using a 30% (v/v) dioxane–water mixture. Experiments performed in this medium are compared with previously reported results in aqueous solution in order to determine changes in both thermodynamic and structural aspects associated with modifications of the biomolecular surroundings. Potentiometric and spectrometric studies reveal the presence of a nonlinear polyelectrolytic effect associated with the protonation of poly(C) in the hydroorganic mixture, different from the analogous effect in aqueous solution. The curve-resolution method alternating least squares is applied to the poly(C) spectrometric data. Concentration profiles and spectra of both deprotonated poly(C) and half-protonated [poly(C) · poly(CH^+)] are thus obtained. The fully protonated species poly(CH^+) precipitates in the hydroorganic medium at pH values lower than 4. Evidence for the ordered structure of both poly(C) and [poly(C) · poly(CH^+)] species is seen through the comparison of the macromolecule spectra with those of the cytidine-3′,5′-cyclic monophosphate monomer. Polarity decreases around the macromolecule produce significant hypochromicity in the CD spectra and hyperchromicity in the uv spectra, both signs of a disordering effect in the macromolecular structure due to the weakening of base-stacking interactions.

Homopolynucleotides are macromolecules used to mimic the behavior of nucleic acids in living organisms. Their capacity for emulation explains their increasing use as therapeutic agents owing to their antiviral, antitumor, and immunostimulatory properties (1, 2). The activity of these biomolecules is strongly conformation dependent. Therefore, monitoring the structural variation of these compounds when subject to modification of chemical variables, such as pH and solvent polarity, provides very valuable information.

Many studies on polynucleotides report instrumental measurements performed under fixed experimental conditions. This research strategy only provides a qualitative and rough idea of the macromolecular structure in concrete situations. Indeed, the information included in one measurement (e.g., a single spectrum) does not tell anything about the possible coexistence of several species in the sample nor about the possible structural transitions between them. Several recent works (3–6) have shown that continuous spectrometric monitoring of metal ion and pH-dependent polynucleotide transitions in aqueous solution can solve the limitations associated with traditional approaches.

Though water is by far the most suitable solvent for reproducing biological conditions, low-polarity biochemical microenvironments, such as active sites of enzymes or side chains in proteins located in low dielectric cavities, need other kinds of solvents to be properly represented (7–9). Water accessibility to these special environments is reduced and therefore the solute/solvent interactions are much weaker than those developed in an aqueous solution. Use of hydroorganic mixtures is a good artificial way to limit the activity of water in the macromolecule surroundings. Within the large group of hydroorganic mixtures, water–dioxane mixtures have been chosen due to the diversity of their features and behavior and to the rather passive role of

0003-2697/97 $25.00

dioxane, almost merely limited to produce an exclusionary effect on water.

Any acid–base polynucleotide study must involve the consideration of side effects associated with the macromolecular nature of these compounds. Thus, changes in the degree of protonation are accompanied by variations in the stability and structure of the species present when polyelectrolytic effects and/or conformational transitions occur. In consequence, traditional least-squares curve-fitting methods based on the fulfilment of the mass action law (i.e., one value of the equilibrium constant during the whole reaction) and on the assumption of a one-to-one correspondence between structural and chemical species are not adequate to deal with treatment of data related to macromolecular processes. An effective alternative to tackle this problem is the use of multivariate curve-resolution methods, such as alternating least squares (ALS)[1] (3–6, 10–12). This model-free procedure is able to produce the concentration profiles and instrumental responses of each individual species without any previous chemical assumption. ALS allows the detection of all conformational transitions, i.e., those associated with proton uptake and those which are not, and further treatment of the data coming from the concentration profiles furnishes information about both the existence and pattern of the polyelectrolytic effect.

The present work is devoted to studying the solvent effect on the acid–base behavior of polycytidylic acid, poly(C). This homopolynucleotide is formed by identical monomeric units, whose nitrogen base is cytosine (see Fig. 1). Dioxane–water [30% (v/v)] has been chosen as the lowest polarity solvent mixture where a quantitative monitoring of poly(C) protonation can be performed; higher concentrations of organic cosolvent cause the precipitation of the macromolecule. The results obtained in the potentiometric and spectrometric (uv and CD) experiments have been compared with those reported in aqueous solution in order to see the effect of a polarity decrease on the protonation of the homopolynucleotide. Series of spectra measured in solvent mixtures with different percentages of dioxane have been performed to see in more detail the solvent effect in the evolution of the macromolecular structure. A study of the cytosine-3′,5′-cyclic monophosphate (cCMP) protonation has also been carried out. This small molecule has the same protonation sites as the monomeric units in poly(C). Therefore, differences between cCMP and poly(C) behavior can be attributed to the presence of macromolecular effects on the poly(C) protonation process.

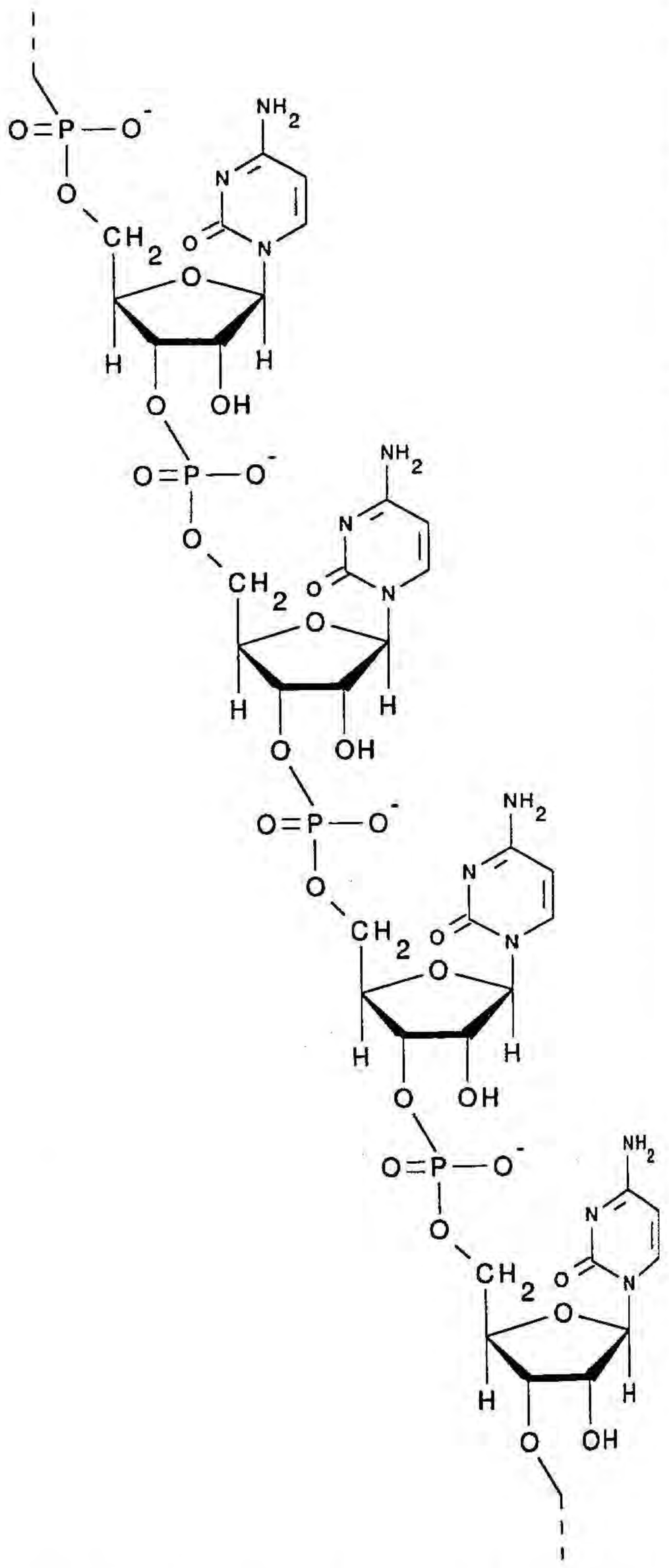


FIG. 1. Structure of the polycytidylic acid [poly(C)].

[1] Abbreviations used: poly(C), polycytidylic acid; cCMP, cytidine-3′,5′-cyclic monophosphate sodium salt; α_p, protonation degree; α, Kamlet and Taft parameter related to solvent hydrogen-bonding acidity; β, Kamlet and Taft parameter related to solvent hydrogen-bonding basicity; π^*, Kamlet and Taft parameter related to solvent polarity–polarizability; CD, circular dichroism, CV, classical least-squares curve-fitting method; σ, ratio between the root mean square of the EMF-weighted residuals and the estimated error in the working conditions; χ^2, measure related to the distribution of the weighted residuals on EMF readings; PCA, principal component analysis; EFA, evolving factor analysis; FSMW-EFA, fixed-size moving window evolving factor analysis; ALS, alternating least-squares method; **D**, original data matrix; d_{ij}, ijth element of the original data matrix **D**; **C**, concentration profiles matrix; **S**, spectra matrix; **D***, matrix reproduced as the product of the **C** and **S** matrices obtained with the ALS method; d_{ij}*, ijth element of the reproduced data matrix **D***.

EXPERIMENTAL

Reagents and Solutions

Hydrochloric acid and sodium chloride (Merck, a.r.), polycytidylic acid sodium salt [poly(C)] and cytidine-3′,5′-cyclic monophosphate sodium salt (cCMP) (Sigma, a.r.), and dioxane (Carlo Erba, a.r.) have been used without further purification.

Stock solutions of poly(C) and cCMP have been prepared by weighing and dissolving a known amount of the solid reagent in the hydroorganic mixture and have been stored at 4°C. The concentration of the polynucleotide solutions is expressed as the concentration of the monomeric unit in the polynucleotide chain. CO_2-free sodium hydroxide (Merck, a.r.) solutions have been prepared following Kolthoff's procedure (13) and standardized with potassium hydrogenphtalate.

The ionic strength of all the solutions has been adjusted to 0.15 $mol.l^{-1}$ by adding the appropriate amount of sodium chloride.

Apparatus

Ultraviolet 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.

EMF measurements have been performed with an ORION 720 pH meter (±0.1 mV precision). Titrant additions have been done with a Metrohm Dosimat 655 autoburette (±0.005 ml volume 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-(KClsat in water–dioxane), where GE is the glass electrode, WS is the working solution, and RE is the Ag/AgCl reference electrode, whose inner solution contains the working hydroorganic mixture in order to minimize the liquid junction potential.

Neutral or slightly basic solutions with different concentrations of poly(C) have been used as the initial solution to be titrated in the uv and CD spectrometric titrations. For all the uv titrations, the solution in the titration vessel is continuously pumped through the flow cell in the spectrophotometer. CD titrations are not performed continuously because of the spectropolarimeter working procedure. 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 230–330 nm with a $\Delta\lambda$ = 1 nm between consecutive absorbance readings.

All titrations have been performed under nitrogen atmosphere, placing the working solution in a double-walled titration cell, thermostatted at 25°C. Gran's method is always used to calibrate the electrodic system (14). This "*in situ*" calibration method standardizes the electrode system without the need of using buffer solutions prepared in the solvent mixture. The pH values measured after calibration are values related to the pH scale of the hydroorganic mixture. These values are defined as pH = $-\log [H^+]$ and not as pH = $-\log a_H$, as usual. Such a simplification can be accepted because all the titrations have been carried out at a constant ionic strength and therefore the activity coefficients do not show variations during the whole reaction process.

Data Treatment

Potentiometric Titrations

Point-by-point log K calculation has been applied to the poly(C) titrations in order to detect the existence of polyelectrolytic effect (i.e., changes of the log K value depending on the protonation degree). 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 macromolecular titrations.

Numerical analysis of the monomer cCMP EMF data has been carried out using the SUPERQUAD program (15), a least-squares curve-fitting method, whose working procedure consists of the refinement of the parameters (e.g., log K) related to a previously proposed chemical model. Figures of merit provided by the program 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 EMF readings.

Spectrometric Titrations

The measurements coming from a spectrometric acid–base titration can be stored in a data matrix **D**, whose rows are the spectra recorded at the different pH values of the equilibrium process. Both directions of this original matrix are chemically meaningful; thus, the information in the rows is related to the spectral changes of the absorbing species during the titration and the information in the columns concerns the evolution of the species concentration with pH.

The goal of all resolution methods is to unmix the information in the original data set by decomposing matrix **D** into the product of two smaller matrices, **C** and

S, containing the concentration profiles and spectra, respectively, related to each absorbing species in the solution (see Fig. 2). Such a decomposition inherently assumes a linear model formed by additive contributions of the individual species in the data matrix. This makes sense for a spectrometric titration because of the linearity associated with the fulfillment of the Beer–Lambert law or the analogous CD spectrometric equation.

The ALS method works following the next steps.

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

The alternating least-squares method allows the treatment of a single matrix and of several matrices together, the so-called three-way data sets. 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 the other. The resulting row-wise **D** matrix is sized (*npht* × *nw*), *npht* being the total number of spectra coming from all the appended $\mathbf{D}_i$ matrices.

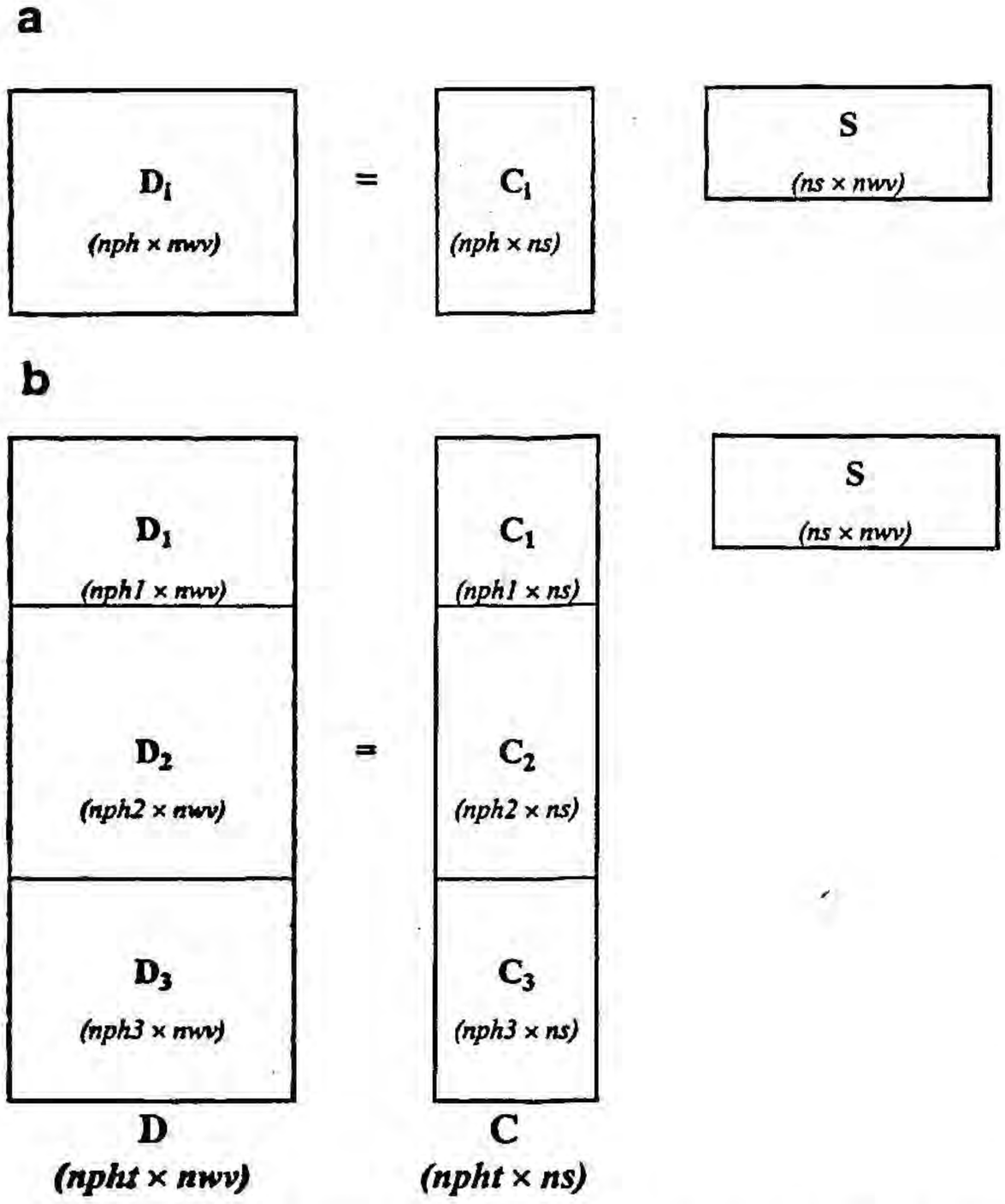


FIG. 2. Alternating least-squares working procedure for the resolution of (a) a single spectrometric titration and (b) several spectrometric titrations (three-way data set).

TABLE 1

Solvatochromic Parameters of the Solvents Used

Solvent	α	β	π^*
Water	1.17	0.47	1.09
30% (v/v) dioxane–water	0.74	0.38	1.08
50% (v/v) dioxane–water	0.61	0.474	0.96
Dioxane	0.0	0.36	0.54

*(b) Determination of the number of absorbing species in **D**, ns.* The number of species, *ns,* present in the data matrix is usually determined by using principal component analysis (PCA) or some other derived technique focused on the analysis of matrices related to evolutionary processes, such as evolving factor analysis (EFA) or fixed-size moving window evolving factor analysis (FSMW-EFA) (16, 17). A sign of the correct selection of the number of absorbing species is the similar magnitude between the estimated experimental error and the residual standard deviation coming from the difference between **D** and the matrix reproduced by using the first *ns* principal components, **D***.

(c) Building a matrix of initial estimates. To begin the least-squares optimization step, a matrix of initial estimates, containing either concentration profiles or spectra, is necessary. Whatever the matrix direction of this initial guess, the use of random estimates has been avoided. Techniques, such as the above-mentioned EFA and FSMW-EFA, provide initial estimates based on the structure of the data matrix and are therefore much closer to the true solutions. The use of these sound estimates also minimizes the risk that the final solutions converge to local minima.

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

(d) Application of the ALS optimization to obtain the definitive concentration and spectra matrices. The generalized Beer law equation in matrix form, $\mathbf{D} = \mathbf{CS}^T$, is solved iteratively by using a constrained least-squares procedure. Working with individual titrations or treating simultaneously several experiments, the iterative 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 constraints related to the internal structure and to the chemical features of the data set. Selectivity, belonging to the first group of constraints, comes out as the key to the resolution procedure. Indeed, the presence of selective zones (i.e., those where only one species is present) for all species ensures recovery of the true solutions. According to the chemical features of the

TABLE 2

Description of the Experiments Performed

System	Technique	Solvent	[L] (M)	pH range
cCMP-H^+	Potentiometry	30% (v/v) D/W	1.83×10^{-3}	3.4–7.2
			1.41×10^{-3}	2.8–7.1
			1.74×10^{-3}	3.0–7.5
		50% (v/v) D/W	1.98×10^{-3}	3.4–4.6
			1.39×10^{-3}	3.1–4.5
			1.06×10^{-3}	3.0–4.2
Poly(C)-H^+	Potentiometry	30% (v/v) D/W	6.09×10^{-4}	6.2–4.0
			4.86×10^{-4}	7.1–4.0
			7.02×10^{-4}	8.0–4.1
	UV	30% (v/v) D/W	9.95×10^{-5}	6.8–4.2
			1.27×10^{-4}	7.7–4.2
			6.93×10^{-5}	7.3–4.0
	CD	30% (v/v) D/W	9.95×10^{-5}	7.5–4.0
			6.60×10^{-5}	6.4–4.0
			1.25×10^{-4}	7.4–4.3

spectrometric titrations, all the concentration profiles have been forced to be positive and unimodal and each row in the $\mathbf{C}_i$ matrix forced to fulfill the closure constraint (i.e., the sum of the concentrations of all the species at each point of the titration is constant). Non-negativity has also been applied to the uv spectra, whereas CD spectra have not been subject to this constraint because of the possibility of having negative measurements with this technique. In the simultaneous treatment of several titrations, no constraints forcing the shape of the concentration profile of an individual species to be the same in all the different titrations have been applied. In contrast, only one pure spectrum per species has been accepted because all the titrations analyzed together have been performed under 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:

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

where d_{ij} are the experimental data and d_{ij}^* the reproduced data by using the ALS method. Subscripts i and

TABLE 3

Results of the Experiments Performed

System	Solvent	Technique	Data treatment	Log K	σ^a	χ^{2a}	Lack of fit (%)[b]	Fit log $K = f(\alpha_p)$[c] Model	r^2
cCMP-H	30% D/W	Potentiometry	CV[d]	4.010 (5)[e]	1.29	30.55	—	—	—
	50% D/W	Potentiometry	CV	3.766 (3)	2.41	34.03	—	—	—
Poly(C)-H	Water[f]	uv	ALS[g]	4.21 (5)[h]	—	—	1.3	$\text{Log } K = 4.21 + 4.6\alpha - 4.5\alpha^2$	0.980
		CD	ALS	4.16 (4)	—	—	3.6	$\text{Log } K = 4.16 + 3.9\alpha - 3.1\alpha^2$	0.985
	30% D/W	uv	ALS	4.04 (2)	—	—	1.6	$\text{Log } K = 4.04 + 17\alpha - 110\alpha^2 + 350\alpha^3 - 400\alpha^4$	0.996
		CD	ALS	3.96 (2)	—	—	3.5	$\text{Log } K = 3.96 + 21\alpha - 140\alpha^2 + 480\alpha^3 - 600\alpha^4$	0.996

[a] Figures of merit only related to the curve-fitting program SUPERQUAD (see text for definitions).
[b] Figure of merit only related to ALS (see text for definition).
[c] The fit only makes sense for macromolecular systems with possible polyelectrolytic effect.
[d] CV, classical least-squares curve-fitting procedure (SUPERQUAD program).
[e] Numbers in parentheses are the errors associated with the last figure.
[f] The experimental data in aqueous solution from Ref. 3.
[g] Alternating least-squares curve-resolution method.
[h] The numerical values for the poly(C)-H system are those extrapolated from the appropriate fitted model for $\alpha = 0$.

AID AB 2175 / 6m30$$$121 05-12-97 09:49:17 abl

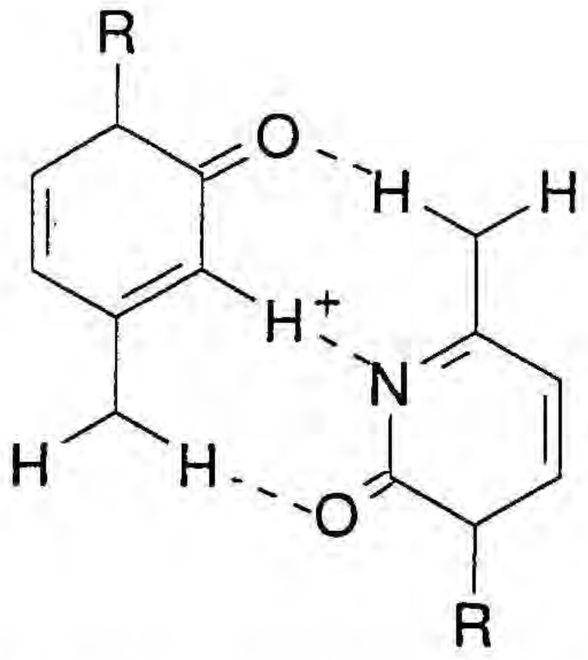


FIG. 3. Cytosine base-pair arrangement in the double-stranded conformation [poly(C)·poly(CH^+)].

j refer to rows and columns of the original data matrix, respectively.

RESULTS AND DISCUSSION

Solvent effects on any solute property can be better interpreted when the solvent properties in the solvation sphere of the solutes are known. The characterization of this particular environment is often done by using empirical microscopic parameters, whose scales are based on the measurements coming from a solvent-dependent process in a reference solute. Previous studies by the authors (18, 19) have been devoted to the microscopic characterization of dioxane–water mixtures by using the α, β, and π^* solvatochromic parameters proposed by Kamlet *et al.* related to the hydrogen-bonding acidity, hydrogen-bonding basicity, and polarity–polarizability, respectively (20–22). Table 1 contains these parameters for pure water and pure dioxane and for the solvent mixtures used in this work. The general decrease of all the solvatochromic parameters as the percentage of organic cosolvent increases confirms that dioxane causes mainly a decrease of the activity developed by water without introducing any kind of specific interaction. Such an interaction would be, if existing, far from what occurs naturally in biological systems.

Dioxane–water mixtures [30 and 50% (v/v)] have been widely used in many biocoordination studies (23, 24). In this work, all the experiments involving the system cCMP-H have been performed in both media. The poly(C)-H system was only studied at 30% (v/v) dioxane–water because of the low solubility of the polynucleotide in solvent mixtures with higher percentage of dioxane. Table 2 describes in detail the conditions of all the experiments carried out.

The N(3) in the nitrogen base moiety is the only protonation site to be taken into account in the pH range studied. The results concerning the potentiometric and spectrometric experiments related to both cCMP-H and poly(C)-H systems are listed in Table 3.

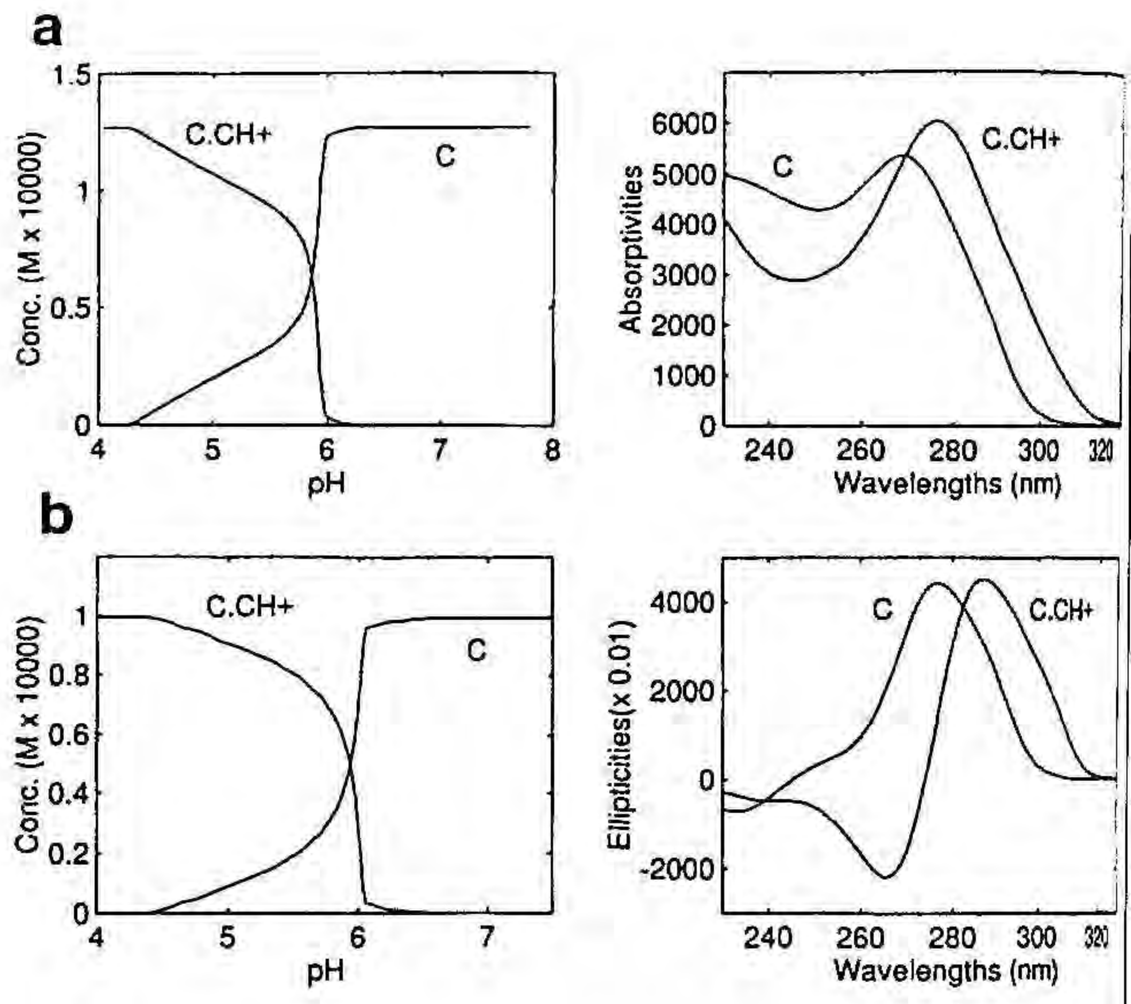


FIG. 4. Concentration profiles and spectra related to the poly(C) protonation process for (a) a uv spectrometric titration and (b) a CD spectrometric titration.

The potentiometric study of the polycytidylic acid has only given qualitative results owing to the partial solubility of the polynucleotide at the concentration levels suitable for this technique. These results show the existence of the polyelectrolytic effect and its pattern of variation as the degree of protonation increases: thus, log K shows a positive or increasing tendency that turns negative as the protonation process goes further. Such behavior, also observed in aqueous solution (3), is likely due to changes in the electrostatic surroundings of the macromolecule and/or to conformational transitions associated with the poly(C) protonation. To know more about the origin of the polyelectrolytic effect, the results coming from the spectrometric experiments must be taken into account.

Literature about poly(C) structure in aqueous solution usually proposes three different conformations depending on the protonation form of the molecule. Thus, deprotonated poly(C) appears to be a single-stranded helix, left- (25) or right-handed (26–32), whereas protonated poly(C) is reported to have two different conformations: a first formed by a half-protonated double-stranded helix, [poly(C)·poly(CH^+)], with the base-pair arrangement shown in Fig. 3, and a second coming from the strand splitting of the previous form, where the species poly(CH^+) adopts a random coil structure (26, 29–32). Little work has been done concerning the quantitative aspects related to the transitions between these structural forms (i.e., the evolution of the concentration profiles of the conformers with pH) in aqueous solution (3) and there is a complete lack of information about these processes occurring in less polar media.

A quantitative study of the structural and thermody-

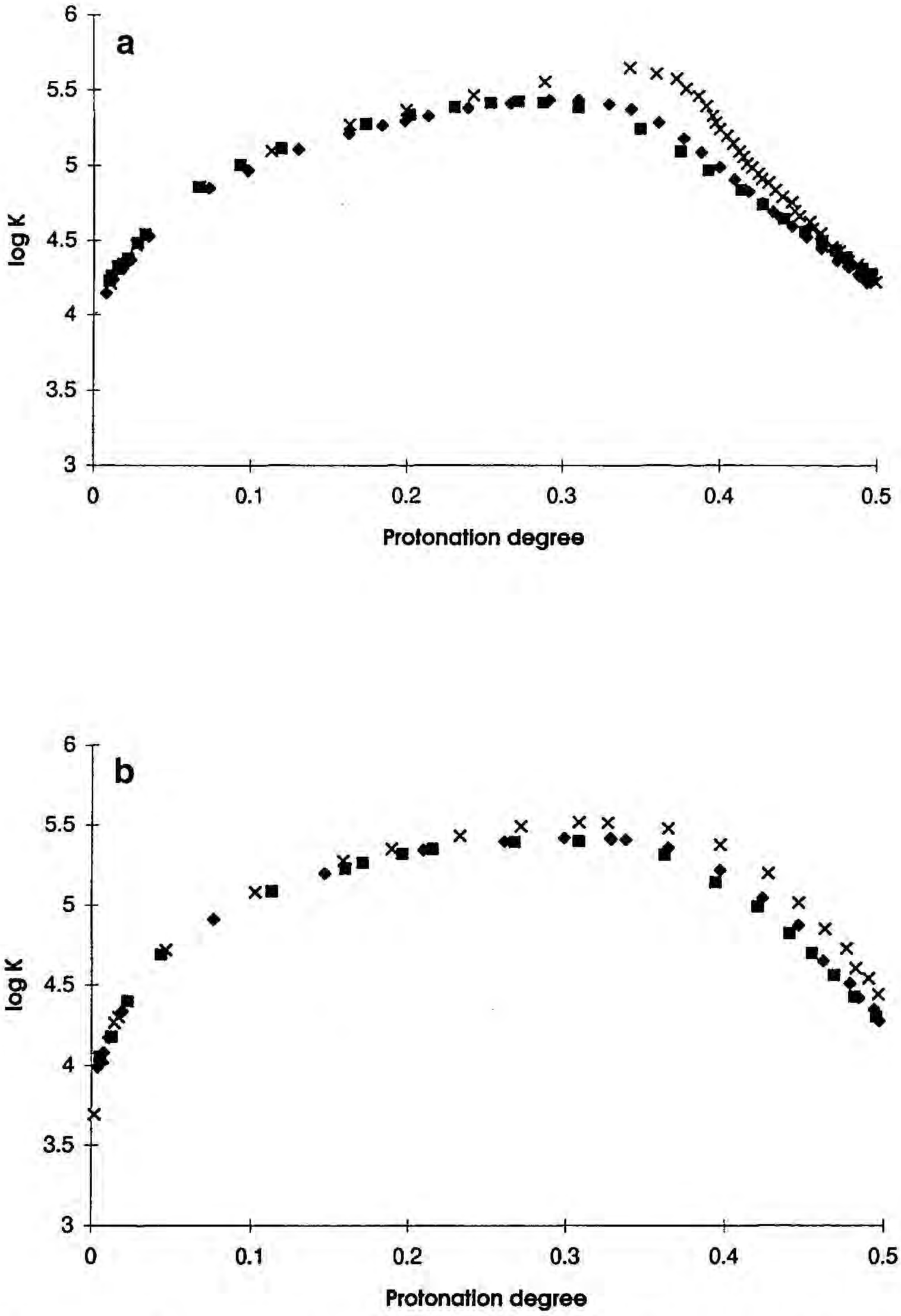


FIG. 5. Plots of the experimental log K values vs α_p from (a) uv titrations and (b) CD titrations. Different symbols belong to different titrations.

namic species related to the poly(C) protonation process in 30% (v/v) dioxane–water has been carried out spectrometrically. No problems related to the poly(C) solubility affected this study becuse of the low concentrations used in both uv and CD studies. Nevertheless, no spectra could be recorded for pH values lower than 4 because of the precipitation of the charged polynucleotide.

PCA and EFA detected the presence of two absorbing species for both uv and CD titrations in the pH range studied. Initial estimates of the concentration profiles were made using EFA. All the $\mathbf{D}_i$ data matrices coming from titrations performed with the same technique as well as their initial estimates $\mathbf{C}_i$ have been appended to form the row-wise data matrices **D** and **C** to be input in the ALS optimization step. Figure 4 shows the response profiles obtained for the uv and CD data sets once the resolution procedure is complete. The results in the plots include the spectra obtained for the two absorbing species, common to all the titrations followed by the same spectrometric technique, and the concentration profiles related to one of the titrations of each

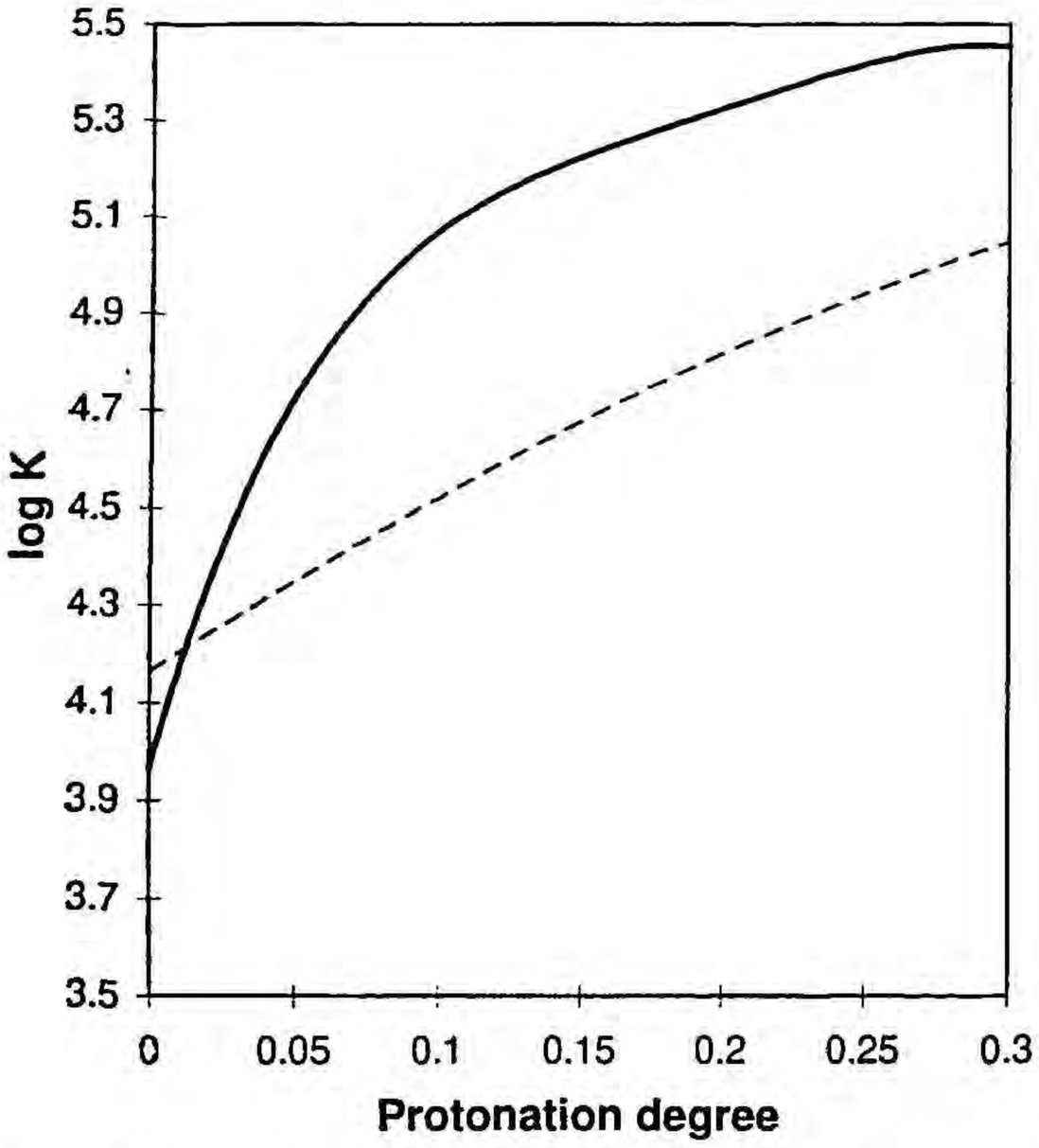


FIG. 6. Theoretical models of the polyelectrolytic effect related to the poly(C) protonation process in 30% (v/v) dioxane–water (solid line) and in aqueous solution (dashed line), according to the fits obtained from CD data.

data set. The concentration profiles coming from other titrations are chemically in agreement with those shown in Fig. 4.

According to the ALS results, there is a one-to-one correspondence between chemical and conformational species. This means that the deprotonated poly(C) and the protonated poly(C) each present one conformation and, therefore, the structural transition between them is associated with the proton uptake process. Because of the pH range studied and the similar shapes between the pure spectra in Fig. 4 and those found in aqueous solution (3), the most reasonable hypothesis is the identification of the chemical species in 30% (v/v) dioxane–water as deprotonated poly(C) in a single-stranded helical form and half-protonated [poly(C)·poly(CH^+)] in the double-stranded helical form. Additional sound reasons to support the existence of [poly(C)·poly(CH^+)] instead of the random-coil poly(CH^+) as the protonated species in water–dioxane are the absence of decrease in the CD ellipticities when going from the deprotonated to the protonated species in contrast with what would happen if a helix → random-coil transition occurred and the more favorable situation of a charged species partially stabilized by the formation of a hydrogen bond N-H^+···N in comparison with a species with a neat charge NH^+ in a low polar solution. Indeed, the poly(CH^+) species, present in aqueous solution, is probably formed in water–dioxane at pH values lower than 4, but its low stability in such a solvent mixture causes the immediate precipitation of the polynucleotide.

Once poly(C) and [poly(C)·poly(CH^+)] have been identified as the chemical species involved in the poly(C) protonation, the value of the equilibrium constant in each point of the titration can be calculated. Bearing in mind that the concentration of the polynucleotide has been expressed as the concentration of monomeric units in solution, the protonation degree (α_p) is defined as the concentration of protonated nitrogen bases divided by the total concentration of bases:

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

Please note that the [poly(C)·poly(CH^+)] concentration in the numerator has been divided by 2 due to the ratio of one proton per two nitrogen bases present in this chemical species. The numerical values used in these calculations are those coming from the concentration

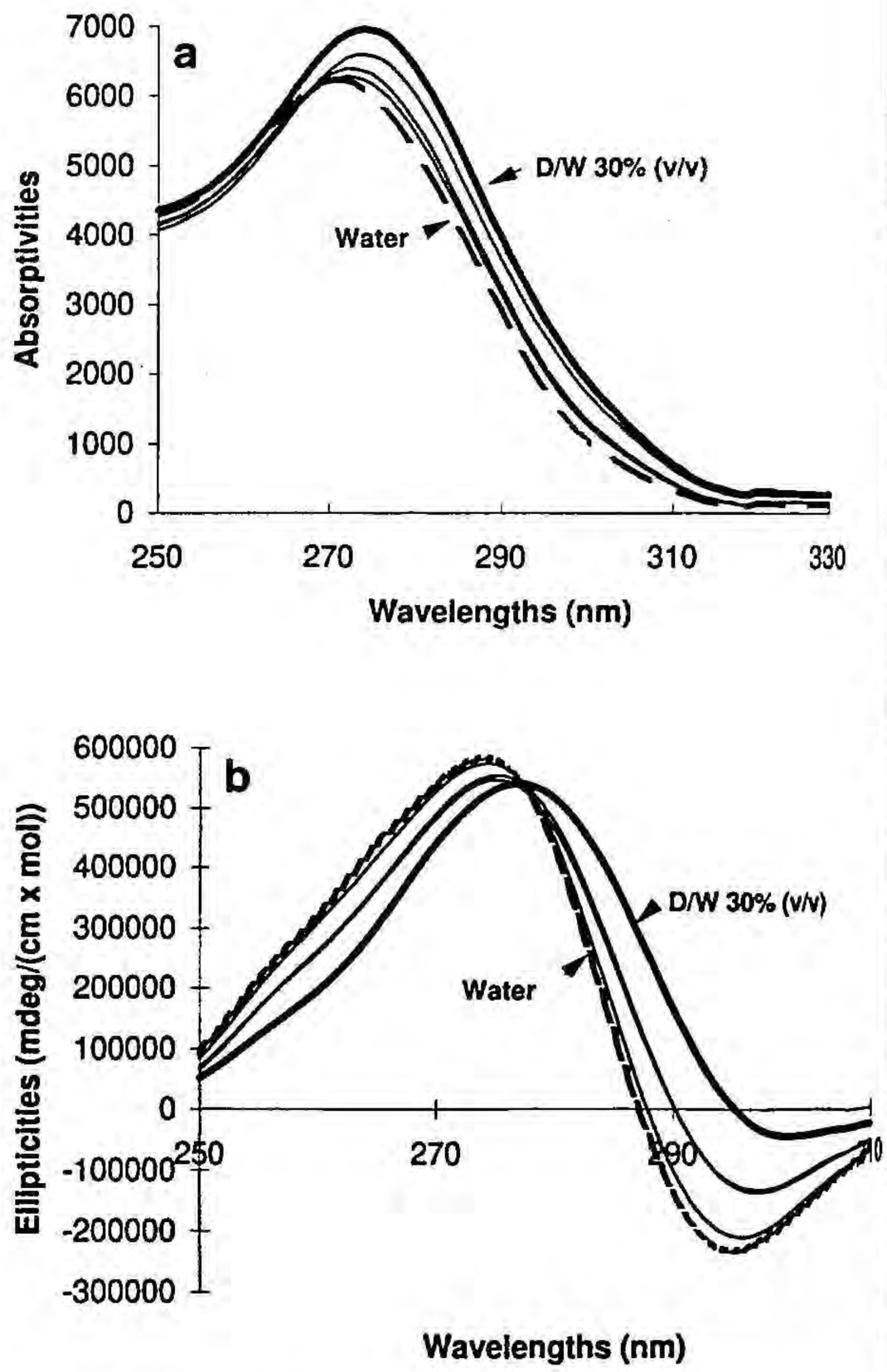


FIG. 7. Evolution of the deprotonated poly(C) spectra with the composition of the dioxane–water mixtures: (a) uv spectra, (b) CD spectra

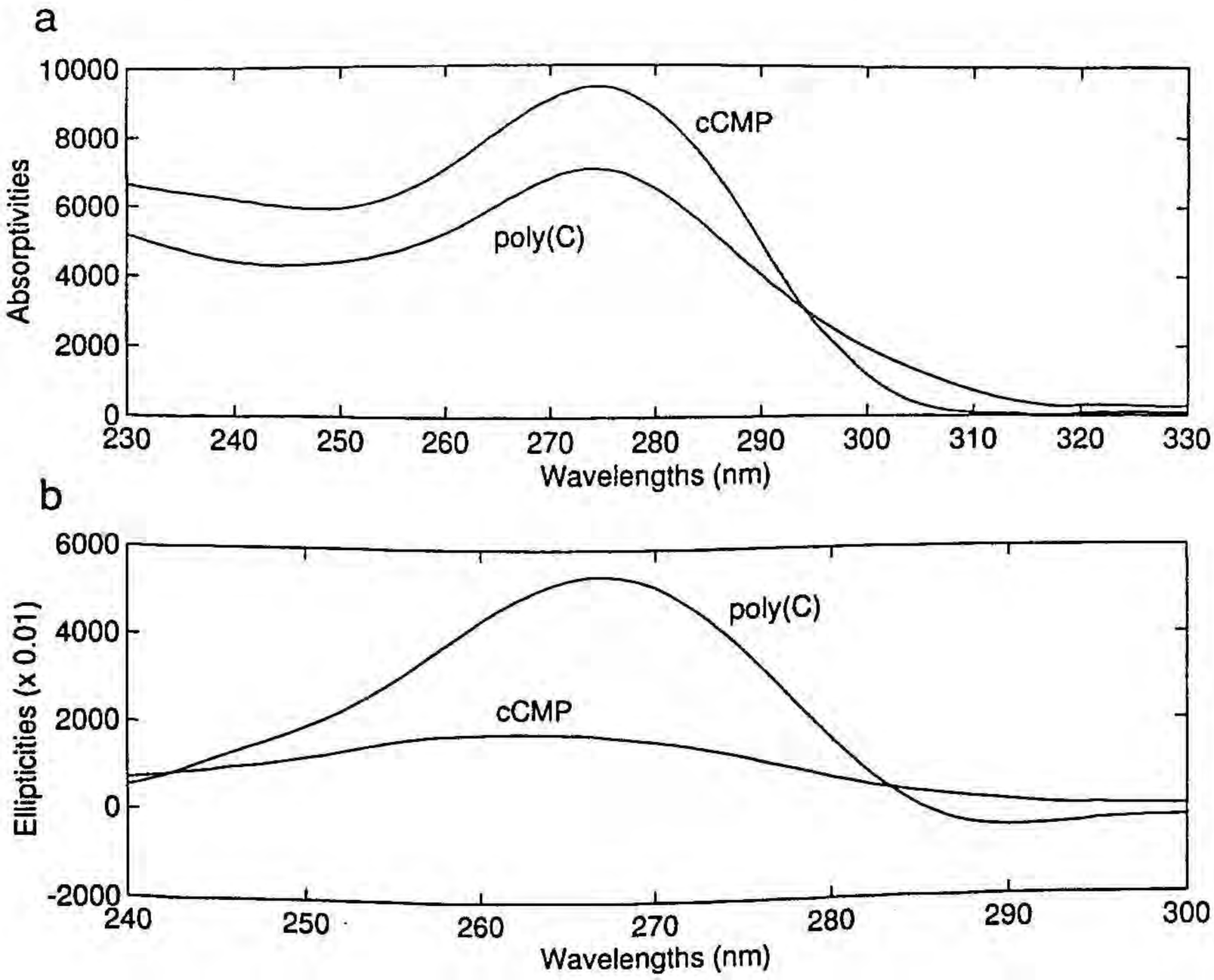


FIG. 8. Comparison of the poly(C) and cCMP spectra in 30% (v/v) dioxane–water. (a) uv spectra, (b) CD spectra.

profiles obtained with the ALS method. The α_p values determined for each pH value are subsequently used to evaluate the protonation constant of poly(C) as

$$K = \frac{\alpha_p}{[H^+](1 - \alpha_p)}.$$

The existence of a polyelectrolytic effect associated with the poly(C) protonation process is confirmed by the spectrometric studies. The evolution of this effect is shown in Fig. 5, where log K is plotted vs the protonation degree. As mentioned before, there is a nonlinear pattern in the polyelectrolytic effect with opposite tendencies depending on the degree of protonation, i.e., a positive polyelectrolytic effect for $\alpha_p < 0.3$ and a negative effect as the protonation process goes further. The first positive tendency is due 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 hydrogen bond N-$H^+ \cdots$N and, at the same time, this base-pair arrangement is responsible for the growth and stabilization of the helical structure. The polyelectrolytic effect becomes negative for α_p values between 0.3 and 0.5 because of the increase of charge density in the macromolecular structure, i.e., over a certain α_p value, the repulsive effect between the protonated sites becomes more important than the stabilization caused by the formation of the double-helical structure. No experimental results are available for $\alpha_p > 0.5$ because of the precipitation of the charged poly(CH^+) species. Though the polyelectrolytic effect in aqueous solution also shows a change in sign from positive to negative, some differences arise between the evolution of this effect in aqueous solution and that in water–dioxane. The first difference is related to the ranges of existence of the positive and the negative polyelectrolytic effects. Whereas the change of sign appears in α_p values around 0.3 in the hydroorganic solution, the negative tendency is not seen till α_p values approximately equal 0.5 in water. The higher stability of charged structures in aqueous solution explains why the negative polyelectrolytic effect does not appear until the strands of the half-protonated double helix are split (i.e., when more than half of the protonation sites have reacted). Then, the breakdown of the stable helical structure and the stronger electrostatic repulsion caused by the increasing concentration of charged protonated sites are responsible for the hindrance of the protonation process. The second difference concerns the pattern of the positive effect in both media: in the hydroorganic mixture, the fit of log K as a function of α_p needs a fourth order polynomial, whereas a second order polynomial is enough to explain the same data in water, as shown in Table 3 [please note the good agreement between the extrapolated value of log $K_{poly(C)\text{-}H}$ for $\alpha_p = 0$, i.e., without presence of macromolecular effects

caused by protonated sites, and the log K corresponding to the cCMP, the reference monomeric unit of poly(C)]. The fitted models for the polyelectrolytic effect in both media are plotted in Fig. 6. A much steeper effect for low α_p values is detected in water–dioxane due to the easier formation of the hydrogen bonds N-H$^+$···N in this solvent mixture due to reduced competition by the solvent molecules. As the degree of protonation increases, this tendency is inverted and the polyelectrolytic effect in water–dioxane becomes less and less pronounced, being close to a plateau for α_p values next to 0.3. Such a smoothing in the evolution of the polyelectrolytic effect is due to the gradual balance between the favorable tendency to form the hydrogen bond N-H$^+$···N and the destabilizing influence associated with the increase of charge density in the macromolecular structure.

In addition to the commented solvent effect on the thermodynamic aspects of the poly(C) protonation process, a series of deprotonated poly(C) spectra have been recorded in hydroorganic mixtures ranging from 0 to 30% (v/v) dioxane to learn more about the solvent effect on the structure of this conformer. Figure 7 shows the evolution of the series of spectra recorded by using uv and CD techniques. The simultaneous hyperchromicity in the uv spectra and hypochromicity in the CD spectra when going from water to 30% (v/v) dioxane–water confirms the expected weakening of the base-stacking interactions associated with decreases of polarity around polynucleotides (26). The most accepted theory about base stacking maintains the main role of hydrophobic forces in the establishment of such interactions. Therefore, in less polar solvents, less base stacking can be expected. Generally speaking, the weakening of this kind of interaction leads to less-ordered structures. Nevertheless, the introduction of 30% dioxane is not enough to induce the helix → random-coil transition, i.e., the complete loss of the initial poly(C) structure, but only to widen the helical pitch of the single-stranded helix, giving more relaxed structures. To support the existence of this partially ordered structure in 30% (v/v) dioxane–water, the spectrum of poly(C) in this solvent mixture is compared with the spectrum of the cCMP molecule in Fig. 8. When no base stacking exists, e.g., in the poly(U) random coil conformation (12), the spectra of the macromolecule and the monomer have comparable absorptivities. Figure 8 shows clearly the phenomena of hyperchromicity in uv and hypochromicity in CD when comparing cCMP and poly(C) spectra. These variations in the absorptivity magnitudes are much larger than those observed when poly(C) spectra recorded in different solvent mixtures are compared, this fact confirming the impossible attribution of the solvent effect on the macromolecular structure to a helix → random-coil transition.

REFERENCES

1. Thang, M. N., and Guschlbauer, W. (1992) *Pathol. Biol.* 10, 1
2. Doukas, J., Cutler, A. H., and Mordes, J. P. (1994) *Am. J. Pa* **145,** 137.
3. Casassas, E., Gargallo, R., Izquierdo-Ridorsa, A., and Taule (1995) *Reactive Polymers* **27,** 1.
4. Gargallo, R., Cuesta-Sánchez, F., Izquierdo-Ridorsa, A., a Massart, D. L. (1996) *Anal. Chem.* **68,** 2241.
5. Tauler, R., Izquierdo-Ridorsa, A., Gargallo, R., and Casassas (1995) *Chemom. Intel. Lab. Systems* **27,** 163.
6. Casassas, E., Tauler, R., and Marqués, I. (1994) *Macromol* **27,** 1729.
7. Sigel, H., Martin R. B., Tribolet, R., Häring, U. K., and Mal Balakrishnan, R. (1985) *Eur. J. Biochem.* **152,** 187.
8. Kanski, R., and Murray, C. J. (1993) *Tetrahedron Lett.* 34, 2
9. Li, Y. K., Kuliopulos, A., Mildvan, A. S., and Talalay, P. (19 *Biochemistry* **32,** 1816.
10. Tauler, R., Smilde, A. K., and Kowalski, B. R. (1995) *J. Chem* **9,** 31.
11. Tauler, R. (1995) *Chemom. Intell. Lab. Systems* **30,** 133.
12. de Juan, A., Fonrodona, G., Gargallo, R., Izquierdo-Ridorsa, Tauler, R., and Casassas, E. (1996) *J. Inorg. Biochem.* 63, 1
13. Kolthoff, I. M., Sandell, E. B., Meehan, E. J., and Bruckenste S. (1969) Quantitative Chemical Analysis, Collier-MacMill Canada, Toronto.
14. Gran, G. (1952) *Analyst* **77,** 661.
15. Gans, P., Sabatini, A., and Vacca, A. (1985) *J. Chem. Soc., Dalt Trans.,* 1195.
16. Gampp, H., Maeder, M., Meyer, C. J., and Zuberbühler, A. (1986) *Talanta* **33,** 943.
17. Keller, H. S., and Massart, D. L. (1991) *Anal. Chim. Acta* 2 379.
18. Casassas, E., Fonrodona, G., and de Juan, A. (1992) *J. Sol. Ch* **21,** 147.
19. Casassas, E., Domínguez, N., Fonrodona, G., and de Juan, (1993) *Anal. Chim. Acta* **283,** 548.
20. Kamlet, M. J., Abboud, J. L., and Taft, R. W. (1977) *J. Am Ch Soc.* **99,** 6027.
21. Taft, R. W., and Kamlet, M. J. (1976) *J. Am. Chem. Soc.* 98, 28
22. Kamlet, M. J., and Taft, R. W. (1976) *J. Am. Chem. Soc.* 98, 3
23. Liogang, G., Tribolet, R., and Sigel, H. (1988) *Inorg. Chem.* 2877.
24. Sigel, H. (1993) *Chem. Soc. Rev.,* 255.
25. Broido, M. S., and Kearns, D. R. (1982) *J. Am. Chem. Soc.* 1 5207.
26. Saenger, W. (1988) Principles of Nucleic Acid Structu Springer-Verlang, New York.
27. Adler, A., Grossman, L., and Fasman, G. D. (1967) *Proc. N Acad. Sci. USA* **57,** 423.
28. Arnott, S., Chandrasekaran, R., and Leslie, A. G. W. (1976) *Mol. Biol.* **106,** 735.
29. Hartman, K. A., and Rich, A. (1965) *J. Am. Chem. Soc.* 87, 20
30. Garriga, P., García-Quintana, D., and Manyosa, J. (1992) *J. Biochem.* **210,** 205.
31. Antao, V. P., and Gray, D. M. (1993) *J. Biomol. Struct. Dyn* **5,** 819.
32. Tanigawa, M., and Yamaoka, K. (1995) *Bull. Chem. Soc. J* **68,** 481.

A soft-modelling approach to interpret thermodynamic and conformational transitions of polynucleotides.

Anna de Juan[1], Anna Izquierdo-Ridorsa, Romà Tauler, Gemma Fonrodona and Enric Casassas.

Departament de Química Analítica. Universitat de Barcelona. 08028 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-modelling 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 due to 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 polyelectrolytic effects and the presence of conformational transitions associated or not with the proton uptake process is presented.

Keywords: polyU, polyA, polyC, curve resolution, polyelectrolytic effect, solvent effect.

1 Corresponding author.

Introduction

Nowadays, 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, T, 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 Figure 1).

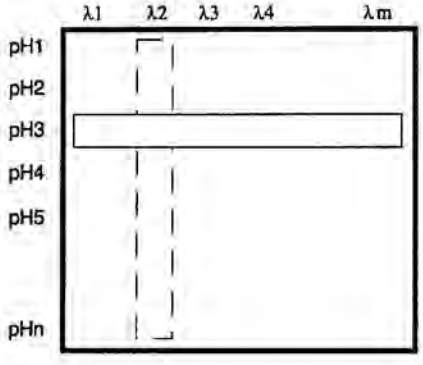


Figure 1. Experimental output from the multivariate monitoring of a pH-dependent process (thick line), an experiment performed at fixed pH conditions (solid hairline) and a univariate monitoring of a pH-dependent process (dashed line).

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: ambiguous structural information is obtained since 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 measurements. In the second approach, a column of the data matrix described in Figure 1 is employed: the ambiguity related to univariate measurements is again apparent and, though 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 (Legget, 1977). This approach is often applied to the data coming from the spectrometric monitoring of chemical equilibria. The clear understanding of these reactions allows the following assumptions in the process of model building:

a) *the fulfilment of the mass action law* (i.e. the validity of a fixed equilibrium constant throughout the reaction) and,

b) *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.) that causes a more complex evolution of the processes. The inapplicability of the classical methods can be explained by:

a) *the existence of polyelectrolytic 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 which might model this effect cannot be devised since no information concerning its existence and its pattern (i.e., linear or non-linear) for an unknown macromolecular process is available *'a priori'*.

b) *the 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 analysing multivariate data sets, the self-modelling curve resolution methods are the most appropriate for the analysis of evolutionary chemical processes when neither models nor prior information about the number and identity of the species involved are available, as it is the case with biomacromolecular equilibria (Kvalheim and Liang, 1992); (Malinowski, 1996); (Vandeginste et al., 1985); (Tauler et al., 1995); (Tauler, 1996). 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 fulfil the Beer-Lambert law.

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.

To illustrate the usefulness of this soft-modelling approach in the interpretation of biomacromolecular processes, the spectrometric (UV and CD) study of the acid-base behaviour of the homopolynucleotides polyuridylic acid (polyU) (de Juan et al., 1996), polycytidylic acid (polyC) (de Juan et al., 1997b) 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 polyelectrolytic 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.

Experimental

The experimental protocol followed in the study of the acid-base behaviour 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 only explains in detail 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); (de Juan et al., 1997b).

Reagents and solutions

Hydrochloric acid and sodium chloride (Merck, a.r.), polyadenylic acid sodium salt (polyA) and adenine-3',5'-cyclic monophosphate sodium salt (cAMP) (Sigma, a.r.) and dioxane (Carlo Erba, a.r.) 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 monomeric unit in the polynucleotide chain. CO_2-free sodium hydroxide (Merck, a.r.) solutions have been prepared following the Kolthoff's procedure and standardized with potassium hydrogenphtalate (Kolthoff et al., 1969).

The ionic strength of all the solutions has been adjusted at 0.15 $mol \cdot l^{-1}$ by adding the appropriate amount of sodium chloride. The solutions have been 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 e.m.f. measurements. Titrant additions have been carried out with a Metrohm Dosimat 655 autoburette (±0.005 ml volume 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 electrodic system consists of:

G.E. / W.S., I=0.15 M, water-dioxane / R.E.(KClsat in water-dioxane)

where G.E. is the glass electrode, W.S. is the working solution and R.E. is the Ag/AgCl reference electrode, whose inner solution contains the working hydroorganic mixture in order 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 a $\Delta\lambda$=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 has always been used to calibrate the electrodic 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 in order to detect the presence of polyelectrolytic 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 fulfilment 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 the cyclic nucleotides has been carried out using the SUPERQUAD program, a least-squares curve fitting method whose working procedure 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 while 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 Figure 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 non-iterative and obtain the solutions directly from the abstract analysis of the data (Kvalheim and Liang, 1992); (Malinowski, 1996); while some others use this abstract information as the starting point of an iterative process which 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., 1997). This method has been chosen due to its great flexibility in the modelling 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

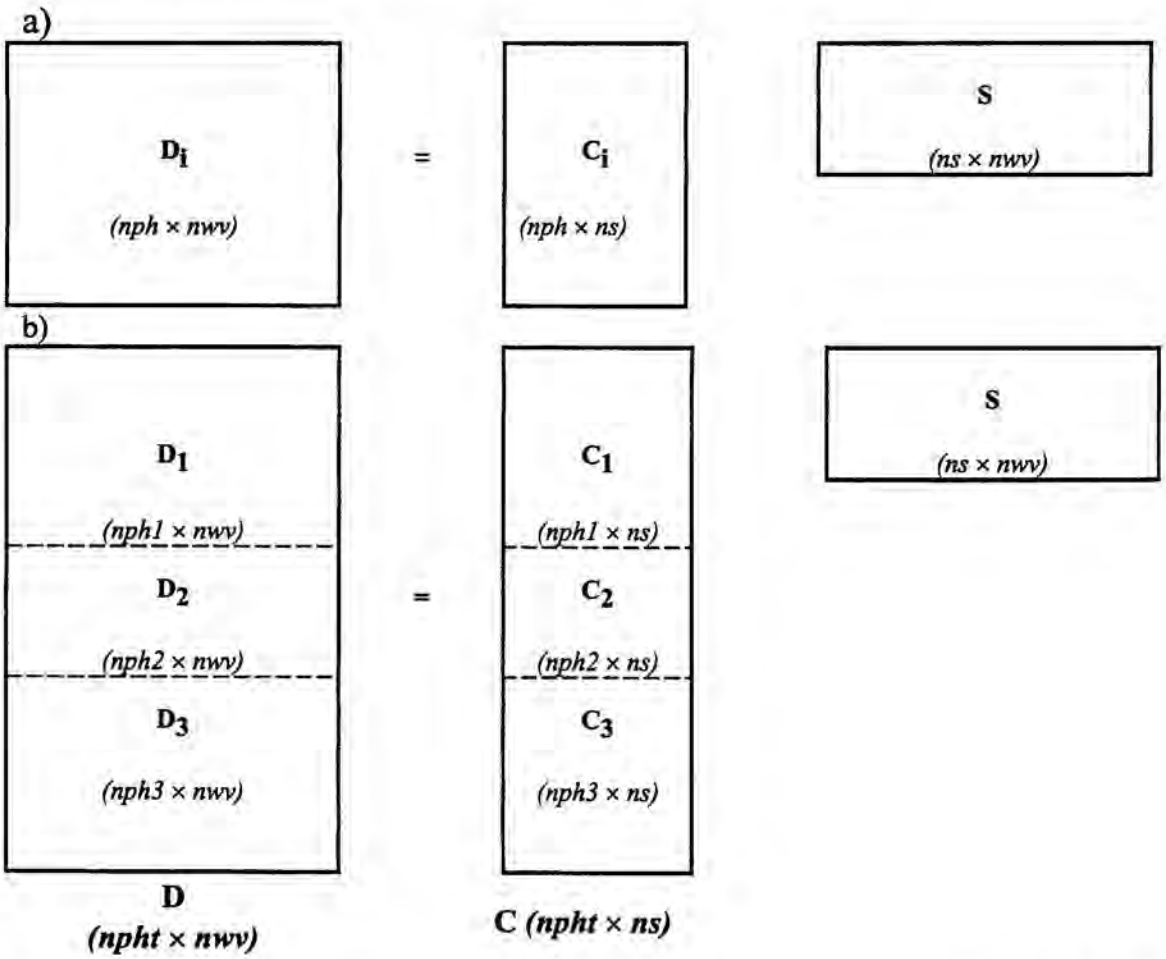


Figure 2. Alternating Least Squares working procedure for the resolution of a) a single pH-dependent experiment and b) several pH-dependent experiments.

The steps followed by the ALS method are summarized below:

*a) 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* × *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 on top of each other. The resulting column-wise augmented **D** matrix has a size of (*npht* × *nw*), where *npht* is the total number of spectra from all the appended $\mathbf{D}_i$ matrices.

*b) 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), though certain derived techniques which 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 using the first *ns* principal components, $\mathbf{D}^*$, has a magnitude similar to the estimated experimental error.

c) 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 as well as others yield initial estimates based on the structure of the data matrix which 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, **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 **D**.

d) Application of the Alternating Least Squares (ALS) optimization to obtain the definitive concentration and spectra matrices.

The equation **D** = **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., 1997). 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 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 the concentration profiles have been constrained so as to be positive and unimodal (i.e. with only one maximum) while each row in the $\mathbf{C}_i$ matrix is constrained so as to fulfil 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). Non-negativity 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 which force the shape of the concentration profile of an individual species to be the same

in all the different titrations are applied. In contrast, a single unit spectrum per species is accepted because all 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 (Math Works Inc., 1994).

Results and discussion

All 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.

Table 1. Description of the experiments performed in dioxane-water 30% (v/v).

System	Nr. experiments	Technique	[L] range ($M \times 10^3$)	pH range
cUMP-H	3	Potentiometry	0.5 - 0.7	5 - 10.5
cCMP-H	3	"	1.4 - 1.8	2.8 - 7.2
cAMP-H	3	"	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

* Potentiometric experiments, though performed, have not been listed as no quantitative results were obtained.

ALS has been applied to the UV and CD acid-base titrations of polyA, polyC and polyU. For each homopolynucleotide, all 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 on in detail below.

a) 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. Figures 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.

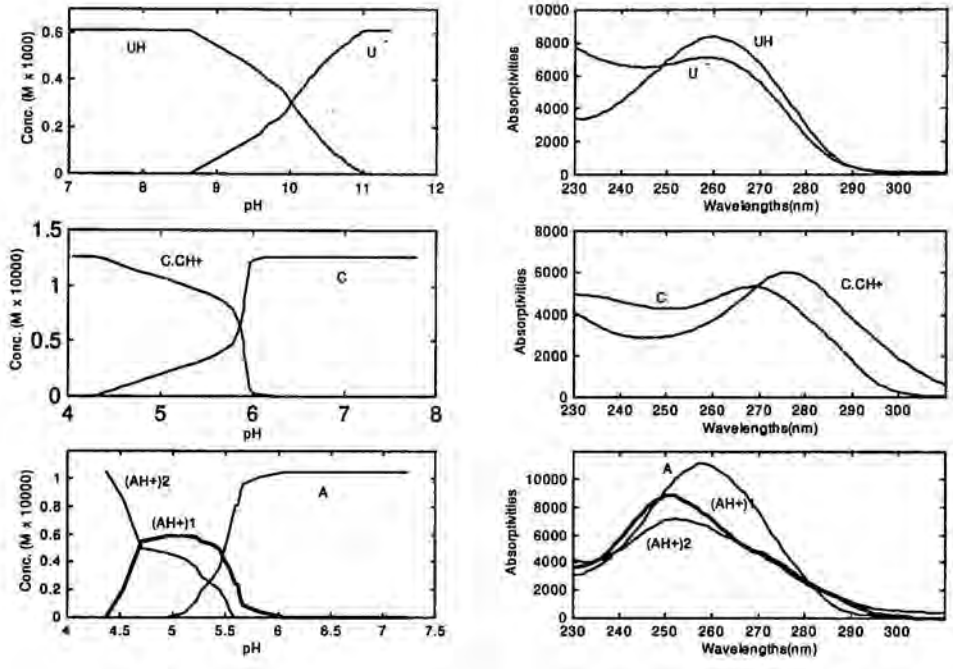


Figure 3. ALS results from the UV titrations in dioxane/water 30% (v/v): concentration profiles and UV spectra related to the polyU-H, polyC-H and polyA-H systems.

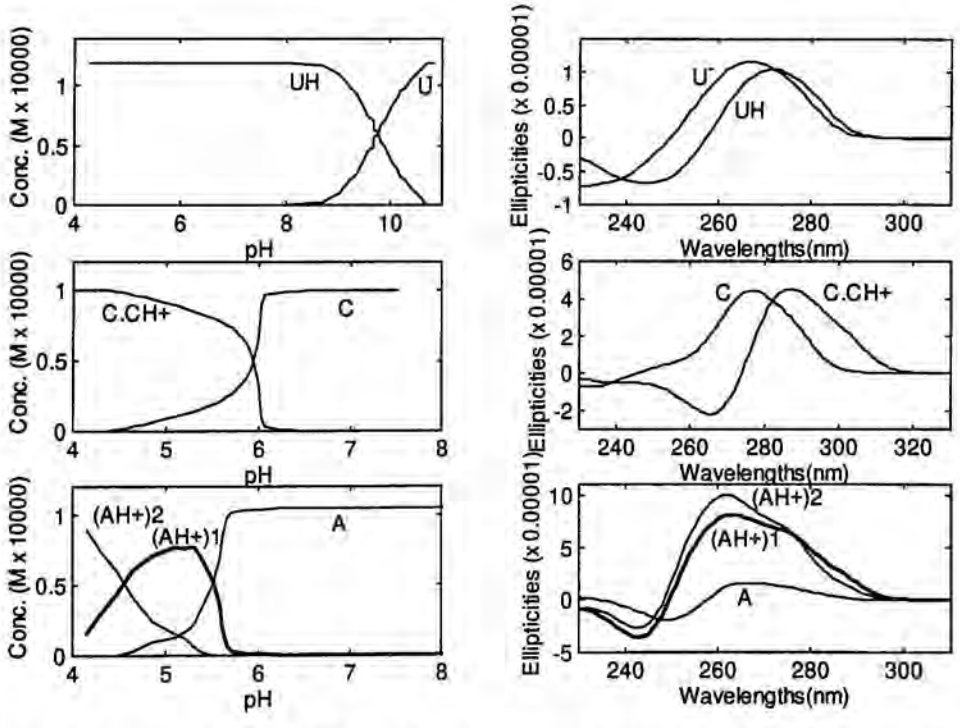


Figure 4. ALS results from the CD titrations in dioxane/water 30% (v/v): concentration profiles and CD spectra related to the polyU-H, polyC-H and polyA-H systems.

b) Detection and description of conformational transitions.

The concentration profiles diagram obtained with ALS provides complete information about the evolution of all the species present in the acid-base equilibria. This includes transitions between chemical species which 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 which 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); (Broido and Kearns, 1982); (Garriga et al., 1992); (Antao and Gray, 1993); (Tanigawa and Yamaoka, 1965). 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 figures 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 favourable 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 neat 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 which is somewhat more complex than that of the two previous polynucleotides. Though 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 been proposed (Antao and Gray, 1993); (Maggini et al., 1994); (Casassas et al., 1994). Figures 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 seem 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→ 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 neighbouring 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 Figure 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.

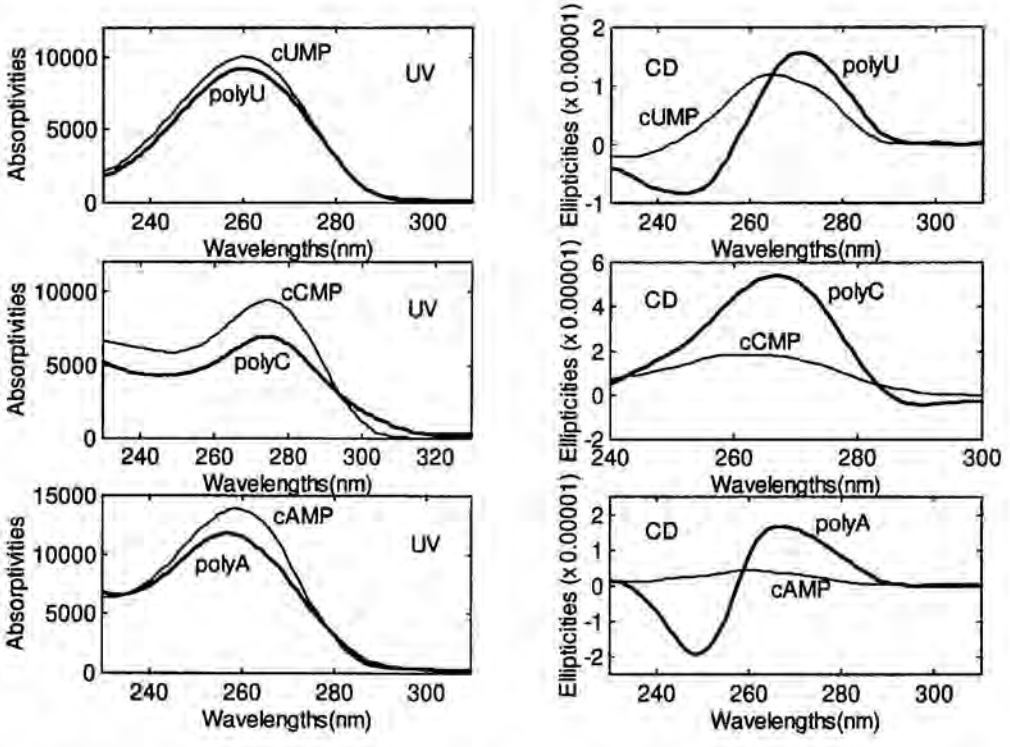


Figure 5. UV and CD spectra of the polymer and the respective cyclic nucleotide (polyU|cUMP, polyC|cCMP and polyA|cAMP) in dioxane-water 30% (v/v).

c) Existence and pattern of polyelectrolytic 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 polyelectrolytic effect exists. Plotting the log K values vs. their corresponding protonation degrees is a graphical way of studying both the existence and the pattern of a polyelectrolytic 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 neighbouring protonated sites are present.

Figure 6 shows in thick lines the log K vs. α plots for polyA, polyC and polyU in dioxane/water 30% (v/v). The plots include the lines fitted to the experimental data.

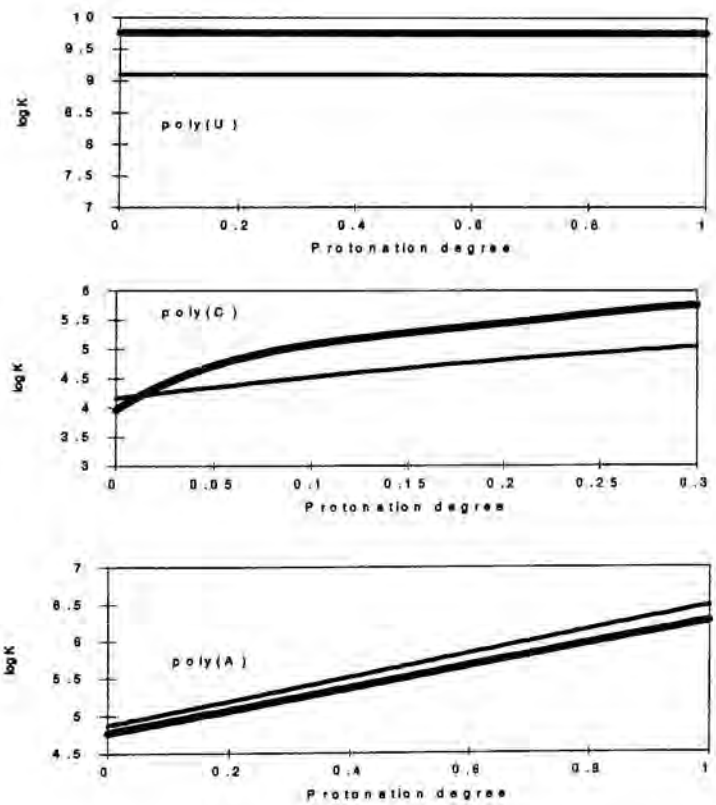


Figure 6. Theoretical models of the polyelectrolytic effect related to the polyU, polyC and polyA protonation processes in dioxane-water 30% (v/v) (thick line) and in aqueous solution (hairline).

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

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

No polyelectrolytic effect has been detected and, therefore, a thermodynamic protonation constant can be given for all 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 next equation:

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

please note that the [poly(C)·poly(CH^+)] concentration is always divided by 2 when included as protonated and deprotonated form due to the existence of one protonated base and one deprotonated base per base-pair. There is a non-linear pattern in the polyelectrolytic 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 polyelectrolytic effect decreases becoming negative for α_p values higher than 0.3 (de Juan et al., 1997b) because the increase of charge density in the macromolecular structure makes that the repulsive effect between the protonated sites 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:

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

A linear positive polyelectrolytic effect caused by the stabilizing electrostatic interactions between the negatively charged phosphate groups and the protonated adenine sites is shown. The difference between the polyelectrolytic 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, though 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 non-linear pattern of the polyelectrolytic 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.

Table 2. Results of the experiments performed.

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

* 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 come from references (Casassas, 1993) for polyU, (Casassas, 1995) for polyC and (Casassas, 1994) for polyA.

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

d) Solvent effect on the acid-base equilibria of polynucleotides.

The acid-base equilibria of polyA, polyC and polyU have been studied in aqueous solution (Casassas et al., 1994); (Casassas et al., 1995); (Casassas et al., 1993). The comparison of the results obtained in water with those obtained in dioxane/water 30% (v/v) will allow the inference of the solvent effects on the protonation processes of the polynucleotides.

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^+), whose formation involves the appearance of two neat 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 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 neat charges.

All 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 figure 7. 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.

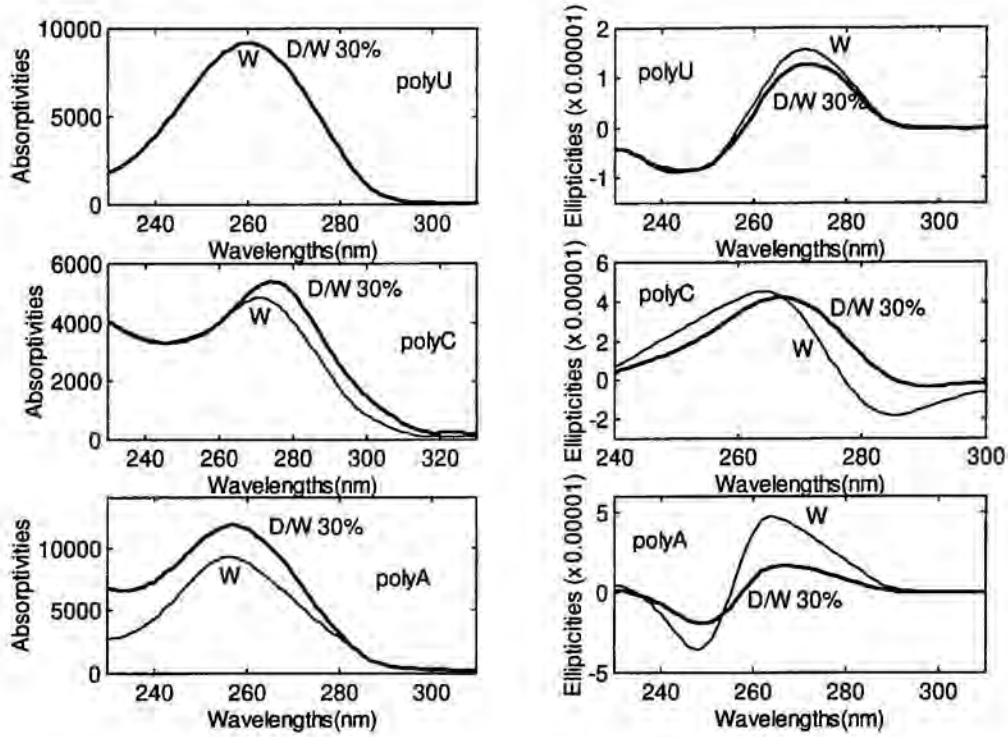


Figure 7. Evolution of the UV and CD spectra of protonated polyU, deprotonated polyC and deprotonated polyC with the composition of the dioxane-water mixtures.

The polyelectrolytic effect, if absent or linear, maintains the same behaviour in both solvents studied, as shown in Table 2 and in Figure 6. Thus, the polyU acid-base behaviour does not present a polyelectrolytic 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 polyelectrolytic 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 polyelectrolytic 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 polyelectrolytic effect when a cooperative mechanism is involved in the polynucleotide protonation process, as can be seen in the polyC protonation process. Though not visible in Figure 6, where only the mathematically fitted positive polyelectrolytic effect is shown, there is a change in the sign of the polyelectrolytic effect associated with the polyC protonation process in both water and water/dioxane solutions (de Juan, 1997b). Solvent effect is first noticed in the ranges of existence of the positive and the negative polyelectrolytic effects. Whereas the change of sign appears in α_p values around 0.3 in the hydroorganic solution, this negative behaviour is not seen till α_p higher than 0.5 in water. The greater stability of charged structures in water explains that the negative polyelectrolytic 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 polyelectrolytic 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_p, 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_p 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 polyelectrolytic effect in water/dioxane becomes less and less pronounced, being close to a plateau for α_p values next to 0.3. The smoothing in the evolution of the polyelectrolytic effect comes from the gradual balance between the favourable tendency to form the hydrogen bond N-H^+····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 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 whose instrumental response follows an additive model.

References

Adler, A., Grossman, L. and Fasman, G.D. 1967. Single-stranded oligomers and polymers of cytidylic and 2'-deoxicytidylic acids: comparative optical rotatory studies. *Proc. Natl. Acad. Sci.* **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. of 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.

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.

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. of Inorg. Biochem.* 63:155-173.

de Juan, A., Vander Heyden, Y., Tauler, R. and Massart, D.L. 1997. Assessment of new constraints applied to the Alternating Least Squares (ALS) method. *Anal. Chim. Acta.* In press.

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. Acid-base and copper(II) complexation equilibria of poly(inosinic)-poly(cytidylic) acid. *Biopolymers.* In press.

Garriga, P., Garcia-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. *The 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.

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. 1. 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.

Liang, G., Tribolet, R. and Sigel, H. 1988. Ternary complexes in solution. 50. Dependence of intra,olecular 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.

MATLAB v. 4.2. math Works Inc. Cochituate Place, MA. US.

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. *Chemical Society Reviews* 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. and 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. of Chemometrics.* 9:31-58.

Vandeginste, B.G.M., Derks, W. And Kateman, G. 1985. Multicomponent self-modelling curve resolution in high-performance liquid chromatography by iterative Target Transformation Analysis. *Anal. Chim. Acta.* 173:253-264.

III. General discussion and conclusions.

CHAPTER 7.

GENERAL DISCUSSION.

On the microscopic description of solvent space.

Understanding the evolution from the classical macroscopic perspective to the microscopic outlook in the solvent description requires a picture of the solvent space in the microscopic domain. The knowledge of the solvent space structure in terms of microscopic solvent similarities is the first step in this project and the main concern of article I.

Article I (*Solvent classification based on solvatochromic parameters...)* is focused on the establishment of a solvent classification based on the similarity in strength and nature of the solvent interactions with solutes. To this end, a large set of representative solvents described by the solvatochromic parameters α, β and π^*, related to the solvent hydrogen bond acidity, solvent hydrogen bond basicity and solvent polarity, respectively, are structured in groups by using several multivariate clustering techniques.

The microscopic solvent space has been defined by the α, β and π^* parameters because of their independent chemical meaning and their lack of mathematical correlation. No other microscopic descriptors have been used since most of them can be expressed as linear combinations of the three aforementioned parameters. Introduction of bias in the formation of solvent groups (coming from the chemist or from the mathematical features of a particuylar chemometrical technique) has been prevented by using several multivariate clustering techniques whose grouping criteria, albeit of diverse mathematical backgrounds,

are always based on either properties of the whole clusters or on characteristics of the space regions.

The solvent classification obtained is formed by six solvent groups, namely:

- **I:** slightly basic solvents with low polarity (aliphatic ethers and substituted aliphatic amines).
- **II:** aprotic polar solvents relatively basic and moderately or highly polar (aliphatic cyclic ethers, esters, ketones and nitriles).
- **III:** strongly basic and strongly polar solvents (pyridines, small amides, sulfoxides, ureas and phosphoramides).
- **IV:** relatively polar solvents with a general low tendency to form hydrogen bonds. The next separation level splits the moderate hydrogen-bond donor solvents from the rest, yielding subgroups A (aromatic ethers, hydrocarbons and halogenated species, and aliphatic apolar halogenated hydrocarbons) and B (polyhalogenated polar aliphatic hydrocarbons, whose heteroatoms cause the inductive effect responsible for the hydrogen-bond donor interaction).
- **V:** amphiprotic solvents (alcohols and water) with marked hydrogen-bond properties Successive splitting separates glycol and water from the other alcohols due to their greater association ability.
- **VI:** aliphatic hydrocarbons without solute/solvent interactions, i.e., with α, β and π^* equal to zero.

The hydrogen-bond basicity and the overall solvent interaction ability are seen to be the most discriminating features among solvents in this classification scheme. The internal diversity in terms of size, functional groups and spatial structure of the solvents included in each group is the most remarkable characteristic of this solvent classification since it reveals the similar hidden behaviour in terms of solute/solvent interactions among solvents traditionally classified further apart. This fact opens a wide range of possibilities in the process of solvent selection, since classical limitations caused by steric hindrance or specific interactions with a certain functional group can be overcome due to the variety of compounds in the same solvent group that behave similarly with the solutes.

This solvent classification has also been compared with Snyder's approach, a previous solvent classification applied extensively in chromatography which uses solvent descriptors analogous to those in the classification proposed in this project. The Snyder's microscopic descriptors are the appropriately corrected logarithms of certain gas-liquid distribution constants of some reference solutes, namely dioxane ($\log(K_g'')_d$), ethanol ($\log(K_g'')_e$) and nitromethane ($\log(K_g'')_n$), which are meant to represent solvent contributions due to hydrogen-bond acidity, hydrogen-bond basicity and polarity, respectively. However, in contrast to the structured spatial distribution of solvents in the 3D space defined by α, β and π^*, the solvents placed in the Snyder space form a continuous swarm of points. The lack of structure of the 3D Snyder space, together with the mixed character of the Snyder hydrogen-bond solvent descriptors ($\log(K_g'')_d$ and $\log(K_g'')_e$ are linear combinations of α and π^* and of α, β and π^*, respectively) recommend the use of the solvatochromic scheme when the solvent is to be selected according to its ability to develop interactions with the solutes.

Going from pure solvents to solvent mixtures: the water-dioxane example.

The structure of the microscopic space of pure solvents is far from being a continuous distribution; therefore, the solvent selection process would be somewhat limited if solvent mixtures were not considered. Actually, these mixed solvents, which form a network of infinite lines connecting the pure solvents among them, fill the gaps present in the space of pure solvents and provide the chemist with the suitable solvent for each purpose.

Microscopic characterization of the mixture and determination of acid-base equilibria.

As mentioned in chapter 2, the water-dioxane combination is especially interesting because of the different nature of its solvent partners and the consequent diversity of its solvent mixtures. This variability is evident in terms of macroscopic features, such as the dielectric constant, and has also been confirmed at microscopic level by the studies carried out in this project.

Table 7.1 includes the experimental values of the microscopic parameters $E_T(30)$, α, β and π^* for water-dioxane mixtures covering the composition range from 0 to 100% (v/v) of dioxane. These values are reported in articles II, III, IV and VI.

Table 7.1. Microscopic parameters related to water-dioxane mixtures.

% dioxane (v/v)	$x_{dioxane}$	α	β_1	β_2	π^*	$E_T(30)$
10	0.023	0.94	0.217	0.432	1.143	61.1
20	0.050	0.81	0.296	0.474	1.124	58.6
30	0.083	0.74	0.379	0.494	1.088	57.1
40	0.123	0.67	0.433	0.525	1.049	55.6
50	0.174	0.61	0.474	0.575	0.989	53.6
60	0.241	0.57	0.486	0.596	0.920	52.3
65	0.282	0.55	0.465	0.610	0.885	51.5
70	0.330	0.54	0.469	0.628	0.849	50.9
75	0.388	0.51	0.501	0.613	0.823	50.2
80	0.458	0.48	0.532	0.638	0.781	49.0
85	0.545	0.45	0.552	0.640	0.748	48.0
90	0.655	0.40	0.551	0.636	0.704	46.7
95	0.801	0.35	0.518	0.632	0.636	44.8
100	1.000	-0.07	0.360	0.419	0.54	36.0

The π^* values shown in table 7.1. come from the average of the results obtained for this parameter with the solvatochromic probes 2-nitroanisole, 4-nitroanisole and 4-ethylnitrobenzene, all recommended for use in amphiprotic solvents or solvent mixtures. The convergent values obtained with these three reference solutes indicate the absence of significant probe-to-probe solvation differences and confirm the validity of the π^* parameter to measure the polarity of water-dioxane mixtures. The π^* values coming from N-methyl-2-nitroaniline were not included in the average due to the significant difference with the results obtained with the three aforementioned indicators. However, this disagreement

should be attributed to differences in the solvation around this indicator, but to the greater ability of this reference solute to establish hydrogen-bonding interactions with the basic groups of the solvent.

The hydrogen-bond basicity of the water-dioxane mixtures was measured with the parameters β_1, determined by using the solute pair 4-nitroaniline and N,N-diethyl-4-nitroaniline, and β_2, determined with the solute pair 4-nitrophenol and 4-nitroanisole. These two measurements were not averaged because of the scattering in the results. However, despite the numerical differences between the two series of β values, which are always larger than those observed among π^* values in any solvent system, there is a clear correlation between them, i.e. they show the same kind of variation with the composition of the mixture. This close relationship makes that both βi series lead to the same conclusions about the significance of the solvent basicity on a solute process when included in an LSER and, therefore, the use of either as a measure of the solvent basicity for water-dioxane mixtures is equally valid. The values of α and $E_T(30)$ were obtained using one pair of solutes (4-nitroanisole and Reichardt's betaine) and one indicator, respectively (Reichardt's betaine), but due to the lesser extent of preferential solvation phenomena in hydroorganic mixtures they may be taken, at least, as good approximations of the solvent hydrogen-bond acidity (α) and the overall solvation ability ($E_T(30)$).

The values of the parameters α, π^* and $E_T(30)$ decrease as the proportion of dioxane in the hydroorganic mixture increases. This behaviour could be expected because of the higher polarity of water and the null hydrogen-bond acidity of dioxane. A different behaviour is observed in both β_i series, where the values of hydrogen-bond basicity in the mixed solvents are higher than those shown by pure water and pure dioxane. This synergetic effect, which shows a maximum value at $x_w = 0.5$, may be due to the higher basic power of the water:dioxane complexes compared to pure solvents (Koppel,1983b), to the stronger basic interaction of the water molecules released from their network structure with the solute (Migron,1991), or to a combination of both.

The non-linear relationship between $E_T(30)$ values and π^* values for water-dioxane mixtures confirms the difference between Reichardt's concept of solvent polarity (defined as the overall solvation ability of the solvent) and that proposed by Kamlet et al., which involves the microscopic interactions governed exclusively by dipole affinities between the solvent and the solute without including any hydrogen-bonding phenomenon. The mixed nature of the $E_T(30)$ parameter, sensitive to solvent polarity/polarizability and to solvent hydrogen-bond acidity, is shown in these hydroorganic mixtures by a model that defines the $E_T(30)$ parameter as a linear combination of α and π^*. Equations related to this model present good fits when obtained using the averaged π^* values in table 7.1. or any of the π^* series related to one of the π^*-sensitive indicators (see article III). The relative weights of α and π^* in the model built for water-dioxane mixtures agree with those calculated in analogous expressions derived for other hydroorganic mixtures (Cheong,1988) and for pure solvents.

The phenomenon of preferential solvation in water-dioxane has been studied by examining the changes of the $E_T(30)$ values with the mixture composition. According to this empirical approach, any deviation of the experimental $E_T(30)$ values from the expected additive behaviour $E_T(30)_{theoret} = x_{water}\, E_T^o(30)_{water} + x_{dioxane}\, E_T^o(30)_{dioxane}$ can be attributed to the presence of preferential solvation. For a given bulk x_{water}, $E_T(30)$ values higher than $E_T(30)_{theoret}$ indicate that water is preferentially placed around the solute and vice versa when $E_T(30)$ values are lower than $E_T(30)_{theoret}$. In agreement with classical studies devoted to the solvation of water-dioxane mixtures, the plot of $E_T(30)$ vs. x_{water} shows a double behaviour. For water mole fractions lower than 0.45 (% dioxane (v/v) > 85%), water is preferentially placed around the solutes, whereas dioxane is the preferred solvent when the water content exceeds this threshold value. The differences between the water-rich mixtures and the dioxane-rich mixtures are also revealed when the Langhals equation is used to fit the experimental $E_T(30)$ values. Two equations are needed to model the evolution of this polarity parameter throughout the whole composition range. Langhals interprets this discontinuity in behaviour as a consequence of the change in the water structure in the mixture promoted by the introduction of dioxane molecules. Thus, for low dioxane mole fractions, the three-dimensional structure of water is kept and the cosolvent is preferentially

placed around the solutes; once a threshold value of $x_{dioxane}$ is exceeded, the water network is broken and the isolated water molecules interact more easily with the solutes. The dioxane operates as a "solvent-structure breaker" and the different behaviour shown by water-rich mixtures and dioxane-rich mixtures is partly explained since the polar interactions of water molecules in molecular associates are clearly stronger than those performed by isolated water molecules.

When water-dioxane mixtures have been described microscopically, the interpretation of the acid-base behaviour of single solutes in these mixed systems becomes easy. Autoprotolysis constants and acidity constants were determined potentiometrically for salicylic acid (carboxylic and phenol group) and propionic acid in mixtures covering the range from 10 to 70%(v/v) dioxane. No determinations were made at higher percentages of dioxane because of the limited solubility of the background electrolyte at the working conditions. The experimental data obtained in this project were also used by other authors as a reference set to test rank analysis methods (Faber, 1997) and chemometric procedures devoted to the establishment of LSER (González-Arjona,1994). In order to reach more general conclusions about the solvent effect on the acid-base equilibria of single solutes in water-dioxane mixtures, acidity constants taken from the literature from substances with different functional groups have also been used in the studies carried out.

The differences between the macroscopic and the microscopic properties of the water-dioxane mixtures are seen when analogous macroscopic and microscopic parameters, such as the polarity measures $1/\varepsilon$ and π^*, are related to each other. No linear relationship is obtained showing that the variation of polarity in the bulk solvent and in the solvation shell of the solutes is not the same. This conclusion is also supported by the non-linear relationship found between the variation of the pK values, which are mainly affected by the solvent properties in the solvation shell of the solute, and some typical bulk properties, such as the inverse of the dielectric constant or the mole fraction of one of the components in the mixture. Departures from the linearity are especially clear in the dioxane-rich mixtures. It is in these more organic media that the differences between the bulk solvent and the solvent around the solutes become more evident due to the breakdown of the water network and to

the increasing importance of solute/solvent interactions which cannot be properly included in classical macroscopic models based only on electrostatic interactions. In consequence, the interpretation of the acid-base behaviour of solutes in water-dioxane mixtures needs the use of microscopic solvent descriptors to be appropriately described.

Simple univariate relationships between each series of pK values and each of the microscopic parameters determined (α, β, π^* and $E_T(30)$) were studied as a first step in the establishment of LSERs, which explain the solvent effect on the acid-base behaviour of water-dioxane mixtures. The change in the internal structure of the solvent mixture promoted by the introduction of dioxane molecules is observed in the plots of pK vs. $E_T(30)$, where two zones described by different linear models are obtained. The same conclusion can be reached when the acidity constants are related to other microscopic parameters involving hydrogen-bonding contributions, such as α or β. In contrast, one linear model is enough to describe the evolution of the acidity constants with the polarity parameter π^* in the whole range of compositions of water-dioxane mixtures. From these relationships it can be concluded that the breakdown of the water structure affects more severely the mixture ability to interact via hydrogen-bonding with the solute than the overall tendency of the mixture to establish interactions based solely on the dipole affinities between solvent and solute.

In articles II and IV, the linear model $pK_i = pK_i^o + s\pi^*$, where pK_i^o is the pK value for the ith solute in a hypothetical inert solvent, was taken as the basic expression for the establishment of LSERs valid for all the studied composition range of water-dioxane mixtures. No significant improvement was found when other solvatochromic parameters, such as α and β, were included in this univariate model; only the addition of the β parameter can explain the evolution of the acidity constants of ethylenediamine, but even in this case, the contribution of this parameter is doubtful due to the slight variation of pK values in the composition range studied.

Modelling of solvent-dependent processes in water-dioxane mixtures: proposals for the establishment of Linear Solvation Energy Relationships (LSER).

According to articles II and IV, the solvent effect on the acid-base equilibrium constants in water-dioxane mixtures seems to be essentially a π^*-dependent process. Nevertheless, the need to confirm in a more rigorous way the validity of the models obtained has led to the proposal of two general procedures devoted to helping in the correct establishment of LSERs, namely: a hard-modelling approach based on the fitting of chemically meaningful models to the experimental data, and a soft-modelling approach, which establishes these models without the support of any external expression. The theoretical background of these methods is explained in detail in sections 1.4.1. and 1.4.2. and both are applied to the experimental data in articles V and VI.

In different ways, both procedures are focused on finding the right size of the LSER model (i.e. the correct number of terms needed) and on identifying the chemical nature of the terms to be included in it. The identification of the solvent descriptors is an obvious task in order to interpret the solvent-dependent process. Less evident, but equally important, is the right determination of the model size. Models missing essential terms do not explain accurately the variations of the solvent-dependent property, whereas models with too many terms are also dangerous because they hinder the interpretability of the model (the chemist must be aware that introduction of superfluous terms does not imply only a slight change in the overall model fit but the inherent assumption of the significance of the solvent effect related to those extra terms on the solute property under study) and decrease its predictive ability (the small terms usually fit specific variations that relate the solvent and solute measurements used to build the model and that might not exist among other solutes and solvents).

Compared with the classical regression procedures, the main contributions of the hard-modelling approach proposed are the prior selection of the size and nature of the parameters to be included in the LSER model by using a stepwise regression and the prevention of the negative influence of outlying observations in the definitive model through

the application of a robust regression method. Thus, taking as potential solvent descriptors the solvatochromic parameters of Kamlet et al., the influence of each solvent parameter on the explanatory ability of the model is assessed and the inclusion of terms which do not add significant information to the description of the solvent-dependent property is avoided. Once the number and identity of the solvent descriptors to be present in the model have been decided, the detection and removal of outliers is performed and the definitive LSER is built by using only the observations accepted according to the diagnostic methods for outlier detection.

The soft-modelling approach provides other benefits when compared to the classical regression approaches and to the hard-modelling approach. It shares with the hard-modelling approach the ability to determine the size of the LSER model and to identify the nature of the terms included in it. However, no external chemical models are needed for either task. The size of the model is determined by factor analysis (FA) of the data set formed by a matrix whose columns contain the arrays of solvent-dependent solute properties. FA provides the simplest abstract model, i.e. the smallest in size, formed by uncorrelated additive contributions. Target Factor Analysis (TFA) is then applied to transform this abstract model into a chemically meaningful LSER. In contrast to the hard-modelling approaches that use the limited group of solvent descriptors included in a basic expression, the main advantage of TFA is that any solvent descriptor, macroscopic or microscopic, can potentially be used to establish the LSER model. All of them are tested individually and the results of these individual tests sort the analyzed solvent descriptors according to their suitability to describe the data set under study. In some situations, the results of the test can reveal that combinations formed by solvent descriptors coming from different microscopic approaches, or formed by macroscopic and microscopic descriptors, can be more useful to describe a specific solvent-dependent property than some fixed multiparametric expressions used in hard-modelling approaches. Thus, TFA is able to discover the validity of unexpected combinations of solvent descriptors without requiring a prior knowledge about the chemical problem as sound as hard-modelling approaches do.

Another difference between the hard-modelling approach proposed and the soft-modelling procedure is that the first is univariate, i.e. makes only one LSER model related to an array of solvent-dependent measured properties every time, and the latter is multivariate, i.e., all the arrays of solvent-dependent data form a data matrix and their respective LSER models are calculated simultaneously.

Both hard-modelling and soft-modelling approaches have positive aspects. If the main virtues of the soft-modelling approach are the complete freedom in the selection of potential descriptors to be included in the LSER models and the saving in calculation time due to the simultaneous calculation of all the LSER models, on the other hand, the hard-modelling approach proposed presents a greater resistance to the influence of outlier observations in the building of LSERs. Therefore, the two approaches provide complementary benefits and the agreement of the results obtained by using both of them is an excellent validation of the LSERs established.

The LSERs built in articles V and VI by using the aforementioned soft- and hard-modelling approaches are related to the study of the spectral shifts of the probes used in the determination of microscopic parameters and to the description of the acid-base behaviour of several solutes in water-dioxane mixtures.

According to Kamlet et al., the solvatochromic shift of the probes used to determine the α, β and π^* parameters can be described by the general expression:

$$\nu_{max} = \nu_o + a\alpha + b\beta + s\pi^*$$

where ν_{max} is the wavenumber related to the maximum absorbance of the probe and ν_o is the wavenumber that would be found in a hypothetical inert solvent. This general expression can be reduced if the probe under study is only sensitive to one or two of the solvent properties represented by the terms $a\alpha$, $b\beta$ and $s\pi^*$.

Article V raises two questions about this subject:

a) Are these three solvent properties enough to describe the spectral shift of all the probes analyzed? and,

b) Is the spectral shift of each probe only sensitive to the solvent properties that is supposed to measure?

The soft-modelling approach answers these two questions. The data set to be treated in this case was formed by a matrix whose columns contained the ν_{max} values of each probe at several water-dioxane compositions. The probes analyzed were 2-nitroanisole, 4-nitroanisole, 4-ethylnitrobenzene and N-methyl-2-nitroaniline (proposed as solutes sensitive to solvent polarity), Reichardt's betaine (proposed as solute sensitive to solvent polarity and solvent hydrogen-bonding acidity) and 4-nitrophenol and 4-nitroaniline (proposed as solutes sensitive to solvent polarity and hydrogen-bonding basicity). FA of this data matrix gave a model with three abstract factors; therefore, no more than three contributions were necessary to describe the spectral shift of the probes analyzed. The subsequent steps were the confirmation of α, β and π^* as the three real contributions related to the spectral shift and the individual exam of the terms needed to describe the spectral shift of each of the probes analyzed.

Not surprisingly, after the application of TFA α, β and π^* were accepted as the most suitable parameters in the description of the spectral shifts. This behaviour could be expected because the three solvatochromic parameters are calculated as linear combinations of the ν_{max} of different probes. However, it is not clear which of these parameters has greater weight in the description of the ν_{max} variations related to each probe. The solvatochromic indicators 2-nitroanisole, 4-nitroanisole and 4-ethylnitrobenzene showed only large loadings (i.e. weight coefficients) related to the π^* parameter, as expected. In contrast, N-methyl-2-nitroaniline, which was found to be an inadequate indicator for the determination of the π^* parameter in work III, was again rejected because of the large loadings associated with solvent hydrogen-bonding interactions. In agreement with the conclusions of Kamlet et al., Reichardt's betaine was found to be sensitive only to π^* and α and 4-nitrophenol only to π^* and β. 4-nitroaniline gives an unusually large loading related

to α because the β values used came from the pair formed by 4-nitroanisole and 4-nitrophenol.

The establishment of LSERs to describe the pK variation of several solutes in water-dioxane mixtures is first treated in article V by applying the soft-modelling approach and is studied more extensively in article VI, where both hard- and soft-modelling approaches are used.

Each column in the data set described in article V contains a series of pK values related to the dissociation of the carboxylic group of a solute determined in several water-dioxane mixtures covering a composition range from 10 to 70 % dioxane (v/v). FA showed that the solvent-dependent variation of the pK data in these water-dioxane mixtures could be described by a two-factor model. Solvent macroscopic properties (dioxane mole fraction, n_2, and $1/\varepsilon$) and solvent microscopic descriptors (α, β, π^* and E_T^N) were tested as potential parameters to be included in the LSER. The best targets were n_2 and π^*. The definite LSERs could be built by using either a combination of n_2 or of π^* with the unity target, which represents the offset of the model. Although the fit of both models is similar, the goodness of the LSER involving the macroscopic dioxane mole fraction may be due to the limited range of mixture solvent composition in the data set.

To clarify the doubts cast upon the similar validity of the macroscopic and the microscopic models obtained to define the variation of pK values in dioxane-water mixtures, article VI presents a more general study involving data sets with a larger variety of functional groups in the solutes and with a wider composition range covered. Thus, A.1. matrix included pK of solutes with carboxylic groups, phenolic groups and amino groups in a composition range from 20 to 70 % dioxane (v/v), whereas A.2. matrix was formed by the pK of a smaller number of solutes determined in a composition range between 20 to 80 % dioxane (v/v). The better quality of the π^*-dependent model was demonstrated when more functional groups and solvent compositions were studied. For the A.1 data matrix, the π^* target was found to be better than n_2 according to the TFA results (e.g. SPOIL function,...),

but this situation became evident when the A.2 matrix was analyzed, because the n_2 target did not pass the test and the only accepted targets were π^* and unity.

The hard-modelling approach gives essentially the same results as the soft-modelling procedure, thus validating the LSER model $pK = pK^o + s\pi^*$ that describes the variation of the acid-base behaviour of solutes in water-dioxane mixtures as a solvent polarity-dependent process. The slight variation among the LSERs coefficients obtained from the FA-TFA approach (loadings) and from the application of the hard-modelling procedure stems from the different number of experimental observations used in both of them. Since the number of data used to establish the models is generally larger in the hard-modelling approach, the LSERs so obtained are preferably taken as definitive models in this study.

A tentative study on the effect of the nature of the background electrolyte in the variation of the pK values in dioxane-water mixtures is also included in article VI. The data in matrices A.1 and A.2 are transposed and some of them are taken to form matrices B.1 and B.2. B.2 contains solutes whose pK values have been determined using a common background electrolyte, KNO_3. B.1. matrix is enlarged by including more pK values determined using $NaClO_4$ as inert salt. Though the amount of data analyzed is not sufficient to justify definitive conclusions, there is no apparent effect related to the use of different electrolytes in the pK determination because the number of factors needed to describe the variation of the data sets does not increase when going from B.2. to B.1 matrices.

The acid-base equilibria of polynucleotides in water-dioxane mixtures.

The knowledge of the acid-base behaviour of single solutes in water-dioxane mixtures is a good starting point for the study of more complex proton transfer processes, such as those related to polynucleotides, in these hydroorganic media. Water-dioxane mixtures were chosen to emulate biological microenvironments of low polarity because they contain water as all biological systems and because low polarity is easily obtained with the introduction of the organic cosolvent.

The main concern of this project lies in determining to which extent a decrease in polarity can affect the pH-dependent transitions of polynucleotides. To monitor these pH-dependent transitions, potentiometry, when possible, and UV and circular dichroism spectrometries were used. The potentiometric technique provides information about the thermodynamic transitions related to the equilibria, whereas the spectrometric techniques confirm the potentiometric results and provide additional information concerning the identity (spectrum) and evolution of the concentration profiles of all the absorbing species involved in the proton transfer process.

As mentioned in section 3.5.3, success is not guaranteed when spectrometric data from macromolecular equilibria are treated with the same classical least-squares procedures as those applied to processes associated with single molecules. The failure of these procedures stems from the particular features that some macromolecular equilibria may present, namely:

a) the non-fulfilment of the mass action law and,

b) the lack of correspondence between the number of chemical species and the number of conformations.

The presence of these particularities, which are contrary to the chemical laws that control the equilibria of single solutes cannot be predicted beforehand and, therefore, the use of classical data treatments that assume the absence of such features is overtly dangerous. In this case, the use of soft-modelling approaches, such as self-modelling curve resolution, that do not take into account any chemical information and that rely on the systematic features of the experimental data (e.g., non-negativity, unimodality,...) to model the concentration profiles and the spectra of the species in solution is the most suitable and safest option.

Use of curve resolution techniques to interpret the multivariate monitoring of biochemical processes: improvement and understanding of the chemometric procedures.

Before applying curve resolution methods to the spectrometric data linked to the pH-dependent transitions of polynucleotides, a detailed study about the performance of these chemometric tools work was carried out. Special attention was paid to the following two points:

a) improvement of the currently used Alternating Least Squares (ALS) algorithm by the introduction of new constraints (article VII) and,

b) comparison of the performance of two three-way resolution methods: ALS and the Trilinear Decomposition (TLD) (article VIII).

The strategies followed in articles VII and VIII include the combined use of simulated and real data to assess how the inclusion of a new constraint (article VII) or the selection of a certain resolution method (article VIII) might affect the performance of the resolution procedure. Indicators of this performance include the recovery of the qualitative information and the overall data fit (for two- and three-way data sets) and the recovery of the quantitative information (only for three-way data sets). The use of simulated data sets in these studies is essential because, in a strict sense, it is only for these systems that all the information to be recovered is perfectly known. Simulations also offer the possibility of obtaining easily a pool of representative data sets, which prevents the statement of biased conclusions that might arise if only a limited set of real examples was used. However, the exclusive use of simulations is not recommended either. Conclusions inferred by using simulations (e.g., the inclusion of a certain constraint improves the overall data fit) must be confirmed when real data sets are employed. Disagreements might indicate that some important features present in real data sets have been missed in the simulated data and this fact would suggest the need to generate new simulations which match the real situation more closely.

Article VII is focused on helping to solve the problem associated with the resolution of multicomponent systems with severe overlap among the profiles of the species present.

When this occurs, selectivity is partially or completely lost. The successful resolution of the system relies then on the tailoring of the domain of possible solutions by the application of suitable constraints that introduce the information related to the systematic features of the concentration profiles and/or instrumental responses of the experimental data. The correct selection and application of these constraints are responsible for the decrease in the ambiguity associated with the decomposition of bilinear matrices and, therefore, finding new effective constraints is a positive contribution to the resolution of complex data sets.

In this study, two new implementations of the unimodality constraint (named hereafter horizontal unimodality and localized unimodality) and the so-called symmetry constraint are proposed. The first of these constraints is applicable to any kind of profile, whereas the second and the third are only restricted to peak-shaped profiles.

Horizontal unimodality is proposed as a means of correcting the drastic trimming of non-unimodal profiles performed by the classical implementation of unimodality. Both modalities of the constraint share the same structure, that is, a first step for seeking the maximum in the profile and a second step for detecting and eliminating the non-unimodal regions of the profile. The difference between classical and horizontal unimodality lies in the way these secondary maxima are suppressed; whereas in classical unimodality sets all the non-unimodal elements in a profile equal to zero, in the horizontal unimodality sets these unacceptable elements equal to the closest element that fulfils the unimodal condition in the profile. From a graphical point of view, the classical unimodality cuts through the original profiles *vertically* and this often yields very narrow and steep constrained profiles, while the horizontal unimodality cuts the original profiles *horizontally* and, as a result, the constrained profiles look softer and closer to the shape of a real profile (see Fig. 1, article VII).

The localized unimodality includes an additional restriction to the plain unimodality constraint, whatever implementation applied. Apart from the existence of only one maximum in the profile, the position of this maximum is also controlled. The reference positions for the different profiles have been determined externally by using other methods, such as the needle algorithm (de Juan, 1996), OPA (Cuesta Sánchez, 1996) or

SIMPLISMA (Windig, 1991). The original profile is constrained by relocating the profile maximum to the reference position and by removing the non-unimodal regions.

The symmetry constraint is specific for chromatographic profiles and suppresses the front-tailed peaks by making the back half of the peak symmetrical to the front half. Despite its name, this constraint does not transform all the profiles into symmetrical peaks, but only those that present the fronting phenomenon.

To test the effectiveness of these constraints, a basic chromatographic system with two components was chosen. No spectral selectivity was included, since this feature would have masked the effect of the constraints to be tested. A two-level factorial design was used to generate representative simulated data sets. The factors taken into account in the design and the respective levels are included in Table 7.2.

Table 7.2. Factors and levels of the two-level full factorial design.

	Levels	
Factors	**(-)**	**(+)**
Constraint (**A**)	Absence	Presence
Resolution (**B**)	0.2	0.8
Ratio minor/major compound (**C**)	1:100	1:10
Noise pattern (**D**)	Homoscedastic	Heteroscedastic
S/N ratio minor compound (**E**)	20	50

The constraint to be tested is introduced as a factor, whose levels are presence (+) and absence (-), depending on its inclusion or not in the resolution process. The remaining factors are related to certain properties of the data set that are also thought to influence the quality of the resolution results. Two factors (**C** and **E**) are linked to the presence of minor compounds. The effect of constraints other than selectivity in the modelling of these minor species is critical because their signals can hardly be differentiated from the other major species and from the noise and, therefore, the external information coming from constraints is essential in recovering the correct profiles.

ALS was applied to the whole pool of simulated data to check the influence of each of the constraints proposed on the resolution results. The responses determined after each

ALS run were the dissimilarities between the true and the recovered concentration profiles for the major and the minor compound, the standard deviation of the residuals and the lack of fit. Normal probability plots were used to assess the significance of the effect associated with each factor and with the interactions established among the individual factors on the resolution results.

The effect of factors other than the constraints confirms what might be expected from any resolution procedure. In the recovery of qualitative information, the overlap between profiles is the factor that has most influence on the recovery of the concentration profile of the major compound, the signal-to-noise ratio is more crucial for the recovery of the profile of the minor compound, while heteroscedasticity and high noise levels hinder the recovery of any profile whatsoever. The effect of these factors goes in the same direction when parameters related to the error associated with the results are analyzed, though the intensity of these effects is not nearly so great. This lower sensitivity of the error parameters to the effect of the different factors is a consequence of the rotational ambiguity associated with the decomposition of bilinear matrices (i.e., the original data matrix can be accurately reproduced by using the product of either the matrix of true concentration profiles and the matrix of spectra or matrices containing linear combinations of both).

Of the three constraints assessed, only the horizontal unimodality led to an improvement in the quality of the resolution results. The effect of this constraint was found as being significant in the recovery of the profile of the minor compound and it had a clearly positive influence on the decrease in the lack of fit. The goodness of this constraint was confirmed by the results obtained in a number of real data sets, where ALS including horizontal unimodality has always yielded a better data fit than ALS with the former implementation of this constraint. Localized unimodality and symmetry did not cause any noticeable difference in the resolution results; however, since the real information included in these constraints is potentially useful, future research could be oriented towards finding alternative implementations that might increase the effectiveness of these constraints.

Updating the classical implementation of the unimodality constraint with the horizontal unimodality in the ALS method improved the resolution results of various data sets, including those related to polynucleotide equilibria. However, the improvement of these results could also be due to the use of three-way resolution methods other than ALS, which might be more suitable to the treatment of the data under study. To test this possibility, ALS was compared with a three-way resolution method with a different background, the Trilinear Decomposition (TLD). An exhaustive study of the differences in performance of both methods according to the features of the data sets was carried out in article VIII.

Trilinear Decomposition is a clear example of a non-iterative resolution method and its solutions are based on the resolution of a generalized eigenvalue/eigenvector problem, whereas ALS is representative of the iterative methods that optimize initial estimates by using constraints related to the internal data structure of the data sets and to the features of their experimental profiles. Given the working procedures of both methods, certain general advantages and drawbacks can be identified in each of these methods. TLD is fast, user-friendly and furnishes unique solutions. However, trilinear data structure is always assumed (i.e., common concentration profiles and spectra for the same species in all the appended matrices) and no information about data structure and features of the profiles can be included to prevent meaningless solutions. ALS can work with non-trilinear matrices having only one order in common (either rows or columns) because the trilinear condition is optionally selected and meaningless solutions are easily avoided through the introduction of constraints in the optimization step. Despite these advantages, unique solutions are not always guaranteed, the iterative optimization slows down the resolution process and more user intervention is required.

The performance of TLD and ALS were tested on a representative group of three-way simulated data sets showing common sources of variation in real data, such as changes in shape and position of the profiles from matrix to matrix and shape distortions caused by noise. The basic trilinear data sets are formed by pairs of matrices that contain a binary chromatographic system. Varying degrees of chromatographic and spectral overlap were

represented. The concentration profiles of one or both matrices were modified according to the source of data variation under study, i.e., changes in signal position, in signal shape or noise addition. The indicators used to assess the effect of each of these sources on the resolution results are related to the recovery of qualitative and quantitative information and to the quality of the overall data description. Conclusions inferred from this theoretical study are used afterwards to suggest general guidelines for selecting the most suitable method according to the features of the data set to be analyzed.

Both TLD and ALS were run assuming the existence of two compounds in all the three-way data sets. ALS method was always used by introducing the non-negativity constraint in both chromatographic and spectral directions and the unimodal constraint in the chromatographic profiles. ALS method was run twice for each example: once forcing trilinear structure in the data set and once without doing so (the acronyms ALSf and ALS have been used to represent both modalities, respectively).

Any conclusions drawn from the comparison of the Trilinear Decomposition and Alternating Least Squares methods are mainly dependent on the existence of selectivity and on the internal structure of the three-way data set. Linked to the latter feature, those variations introduced in the three-way data sets which induce a loss of the trilinear data structure (i.e., changes in shape and position of the signal) must be distinguished from those which do not (i.e., signal distortions produced by noise).

As a general rule, ALS works much better than TLD and ALSf when there is no trilinearity as it has greater flexibility in the modelling of the response profiles. The better performance of ALS becomes more evident for non-trilinear systems with selectivity since the presence of this feature suppresses ambiguities in the decomposition of the augmented matrix (in such cases, the minimal dissimilarities between the true and the calculated profiles by ALS come from the noise added to all data sets). When there is no selectivity in any of the orders of the data set, ALS still performs better in most of the cases. This tendency fails to hold in but a few cases, for example when the departures of trilinearity are extremely small or when these departures are associated with an increase in the overlap between

compounds. In these unusual situations, the negative effects on the resolution results caused by the increase in correlation among profiles on ALS and by the loss of trilinearity on TLD and ALSf are comparable. However, the similar performance of TLD or ALSf and ALS in these particular cases is limited to the recovery of response profiles and to the recovery of quantitative information; the overall data description remains much better when ALS is used. Since TLD and ALSf only include the trilinear information present in the first two compounds to reproduce the data matrix, the non-trilinear contribution excluded from the model is responsible for the deterioration in the overall data description, i.e., for the increase in the lack of fit; on the contrary, the larger amount of information input by ALS in the data reproduction, though using profiles that may be linear combinations of the real ones, yields a description of the data which is better than that obtained by using methods with trilinear structure imposed. This behaviour and the possibility of calculating the lack of fit for real data provides an additional method for checking the presence of trilinear structure in any three-way data sets because non-trilinear systems will always have a significantly greater lack of fit for TLD than for ALS when the same number of compounds is used in the resolution process.

Differences in performance between TLD and ALSf for non-trilinear systems stem from the distinct way in which the trilinear condition is included in their resolution processes. TLD works on the whole data set to obtain the trilinear combination of response profiles that best fits the original data, whereas ALSf operates on each column of the augmented concentration matrix individually to get the best constrained profile for each compound. This means that TLD gives priority to overall data description while ALSf focuses more on the recovery of the individual profiles. These explanations match the results found for non-trilinear systems, in which ALSf provided better profiles and smaller errors in quantitation, whereas TLD showed lower lack of fit. The greater lack of fit in the ALSf models was due to the removal of both the non-trilinear contributions and the trilinear information that do not respect the constraints imposed (e.g. non-unimodal and negative parts of the concentration profile,...). The worse response profiles obtained by TLD can be attributed to the acceptance of any kind of profile shape, provided that it keeps the trilinear condition.

Noise (different levels and patterns) was added to the data sets without modifying their trilinear structure. All three-way resolution methods were more affected by increases in the noise level than by variations in the noise pattern. Generally speaking, both TLD and ALSf provided more satisfactory results than ALS for these noisy trilinear sets. Their superiority was obvious when there was no selectivity in the three-way data set because the internal structure of the data sets led both methods to attain the real unique solutions. In contrast, ALS suffered from the ambiguity associated with the decomposition of the bilinear matrices lacking selective information. Then, the profiles obtained, though they correctly reproduced the original data, were linear combinations of the true response profiles. For the sake of speed and ease of application, TLD is more recommendable to treat trilinear three-way data sets with no selective information than ALSf. When selectivity was present, TLD, ALS and ALSf all obtained unique solutions in very good agreement with the true profiles. Nevertheless, TLD and ALSf usually showed a subtle superiority with respect to ALS, especially in systems with high noise levels. This difference is probably due to the generally less noisy profiles recovered by TLD and ALSf. These two methods have a smoothing effect on the profiles obtained since the noise contribution differs from matrix to matrix, i.e., noise is seldom a trilinear contribution. Consequently, the profiles recovered in this way are more similar to the true noise-free signals than those obtained by using ALS, where the noise in the final profiles cannot be suppressed unless it distorts the profiles in such a way that the constraints imposed in the optimization procedure are not respected.

From the study performed on simulated data sets, it can be concluded that before selecting a resolution method the kind of data set to be analyzed must be known. In classifying it, the key features are the presence of selectivity and the existence of trilinear structure.

The data associated with polynucleotide equilibria are non-trilinear and can or cannot show selectivity. ALS was found to be the best method to deal with non-trilinear data sets, though systems lacking selectivity can sometimes show solutions of similar quality with ALS and TLD. However, ALS appears to be more robust in the modelling of the response profiles and since this is the main goal in the resolution of polynucleotide

equilibria, ALS was adopted as the most suitable resolution method in all subsequent studies.

Application of curve resolution techniques to the study of pH-dependent transitions of some homopolynucleotides in water-dioxane mixtures.

The work carried out up to this point in this project has been performed so as to obtain the necessary information for studying of the pH-dependent transitions of polynucleotides in environments of low polarity. Thus, on the one hand, the water-dioxane mixtures used to emulate these special environments have been characterized and the solvent-dependency of the acid-base equilibria of single solutes has been described with suitable LSERs while, on the other hand, the selection and improvement of chemometric tools suitable for the interpretation of macromolecular equilibria have also been a subject of research.

Three homopolynucleotides, polyuridylic acid (polyU), polycytidylic acid (polyC) and polyadenylic acid (polyA), were chosen for study because of the diversity of their features (see section 3.3). Articles IX and X include the results obtained for the systems polyU-H and polyC-H, respectively, whereas article XI explains the general strategy used in tackling the problem of polynucleotide equilibria (i.e., the combination of experimental multivariate monitoring and application of curve resolution methods) and reports systematically all the different kinds of information that can be obtained when this approach is selected. The potential of this approach is proved by showing the results in articles IX and X and those obtained for the polyA-H system.

The study of the acid-base behaviour of the homopolynucleotides was carried out by using UV and circular dichroism spectrometries to monitor the pH-dependent transition of polyA, polyC and polyU and by applying additionally the potentiometric technique to those compounds that are soluble at the concentration levels suitable for this technique. In the pH working ranges, only one protonation site has to be considered in each monomeric unit of the polynucleotides. Thus, polyA protonates the N(1) nitrogen in the adenine ring and

polyC the N(3) nitrogen in the cytosine ring, respectively, giving positively charged polymers, whereas polyU deprotonates the uracil N(3) amide-like nitrogen and becomes negatively charged at basic pH values. 30 and 50 %(v/v) dioxane-water mixtures were used in performing all the experiments; larger amounts of the organic cosolvent induced the precipitation of the polynucleotides.

The information gathered from the experiments described above allow the straightforward interpretation of the acid-base behaviour of each of the polynucleotides studied. Additionally, either the sole combination of these results or their comparison with those from other kinds of experiment can contribute to the understanding of certain interesting effects, namely:

a) *the effect of the molecular structure on the acid-base behaviour of the polynucleotides.* This can be interpreted by comparing the differences between the behaviour of the three polynucleotides studied.

b) *the presence of macromolecular effects on the evolution of the polynucleotide equilibria.* This can be assessed directly from the results of the polynucleotide experiments or by comparing these results with those obtained for their respective cyclic nucleotides.

c) *the solvent effect on the polynucleotide equilibria.* This information comes from the comparison of the experiments performed in water-dioxane mixtures with those previously reported in aqueous solution.

The potentiometric technique was applied to polyA, polyC and polyU and to their respective cyclic nucleotides adenine-3',5'-cyclic monophosphate (cAMP), cytosine-3',5'-cyclic monophosphate (cCMP) and uracil-3',5'-cyclic monophosphate (cUMP). Equilibrium constants were determined for all the monomeric units at 30% (v/v) dioxane-water mixture. Among the polynucleotides studied, only quantitative results could be obtained for the polyU-H system, because both polyC and polyA precipitate at the concentration levels required for the potentiometric technique. The presence of polyelectrolytic effect was

checked by examining the variation of the pK value with the degree of protonation of the macromolecule. No evidence of this effect was found for the polyuridylic acid, since the pK value remained constant during the whole protonation process. The information obtained from the potentiometric titrations of polyA and polyC seemed to indicate the presence of the polyelectrolytic effect in both systems, though this statement needed to be confirmed with the spectrometric studies due to the qualitative nature of these experiments.

The use of spectrometric techniques, such as UV and circular dichroism, in monitoring polynucleotide acid-base equilibria provided quantitative results for the three systems studied due to the lower concentration levels needed for the correct performance of these techniques. Furthermore, the information related to the structure of all the absorbing species in solution (conformers or chemical species) was obtained and therefore, the description of all the pH-dependent transitions (associated with a thermodynamic process or otherwise) was possible.

For each polynucleotide, all the titrations performed using the same technique are treated together with the ALS method. In the case of UV titrations, the constraints applied are selectivity, when present, non-negativity for concentration profiles and spectra, unimodality for concentration profiles and closure. CD titrations are treated similarly, the only difference being that the CD spectra are not subject to the non-negativity constraint. The results directly provided by the ALS method and the information that can be obtained from them are detailed in the sections below:

a) ALS results: concentration profiles and pure unit spectra of the absorbing species involved in the acid-base equilibrium.

ALS yields the number of absorbing species involved in each acid-base equilibrium and the concentration profile and spectrum related to each of them. In the polyC-H and polyU-H systems, two species explain the protonation process, whereas for the polyA-H system three species are needed. The agreement between the concentration profiles obtained with both spectrometric techniques for the same polynucleotide protonation process

validate the quality of the experimental work and the correct performance of the ALS method on this kind of biochemical data.

b) Detection and description of conformational transitions.

The distribution plot of the species involved in an acid-base equilibrium is built by plotting the concentration profiles obtained by ALS vs. the pH values of the experiment. This plot includes transitions between chemical species without structural changes, conformational changes associated with the protonation process and spatial rearrangements of the macromolecule that are not linked to alterations of its protonation state. The task of identifying which kinds of transition take place and which conformations are involved in them is beyond the ALS method and depends on independent evidence and on the knowledge of the chemist about the process being analyzed. PolyU, polyC and polyA are good examples of the three transitions mentioned above.

According to previous studies performed in aqueous solution, polyU enters the group of polynucleotides that do not show a conformational change when going from the protonated to the deprotonated form. In water-dioxane mixtures, the similar shape of the UV and CD spectra related to the poly(UH) and poly(U^-), where only a small shift between their absorption maxima is observed, seems to confirm this hypothesis. The conformation implied has been identified as a random coil due to the slight differences in the absorption intensity between the polyU and cUMP spectra.

The distribution plot of the species involved in the polyC protonation in water-dioxane mixtures shows two species with clearly different spectra. These two species are identified with two differently protonated species and, therefore, this polynucleotide provides an example of conformational transition associated with the protonation process of the polynucleotide. Of the three possible polyC species suggested for this protonation equilibria in aqueous solution, i.e., the deprotonated polyC, the half-protonated [poly(C)·poly(CH^+)] and the completely protonated poly(CH^+), the first and the second appear to be the most probable species in hydroorganic solution. Such an identification is

supported by the similar shape of the spectra and the similar pH region of existence of these two species in these media and in aqueous solution. Additionally, in a solution of low polarity, the situation of a charged species stabilized by the formation of an interstrand hydrogen bond N-H^+····N, as in [poly(C)·poly(CH^+)], is more favourable than that of a species with a neat charge NH^+, as in poly(CH^+). Both poly(C) and [poly(C)·poly(CH^+)] present ordered conformations, as can be inferred from the marked hyperchromism in the CD spectra and hypochromism in the UV spectra of the polyC species with respect to the cCMP spectrum. The conformations of the two absorbing species of polyC are a single helix for the deprotonated form of polyC and a double-stranded helix for the half-protonated form of this polynucleotide, as in aqueous solution.

The pH-dependent transitions related to polyA require the description of three absorbing species. Since there is only one protonation site to be taken into account and three absorbing species, one of the transitions of this system must be associated with the protonation process and the other must be a spatial rearrangement of the macromolecule which is not linked to any reaction process. When looking at the plot of concentration profiles, the transition performed in the most basic pH region includes species whose pH regions of existence and profile shape are highly reproducible from experiment to experiment, whereas the transition occurring at more acidic pH values involves species with more irregular concentration profiles that are slightly shifted in terms of pH regions of existence between the different experiments. Such behaviour seems to associate the former transition with the proton uptake process and the second with a spatial rearrangement of the protonated macromolecule. The shape of the spectra also confirms this correspondence of transitions because the spectrum of the most basic species, i.e., the deprotonated polyA, looks completely different from the spectra of the other two species, which are very similar to each other and can reasonably be attributed to different conformers of the polyA protonated species (poly(AH^+)1 and poly(AH^+)2). As in polyC, the comparison between the spectra of the three absorbing species of polyA and the cAMP spectrum confirms the ordered structure of all the absorbing species in the polynucleotide. The spectra of the polyA species in water-dioxane are quite similar to those in aqueous solution and are in agreement with the most widely accepted theories which assume a single helical

conformation for the polyA species and diverse varieties of double-stranded helices for the protonated forms of the polynucleotide. The CD and UV spectra of both protonated conformations show hyperchromism and hypochromism, respectively, when going from poly(AH^+)1 to poly(AH^+)2. This is due to the minimization of the electrostatic repulsion of neighbouring phosphate groups induced by the stabilizing electrostatic interactions between the negative charges of these groups and the protonated sites of the bases. This increasing stabilizing effect is probably responsible for the transition from a double helix with a larger helical pitch to a more compact structure with the phosphate groups closer to each other and with stronger base stacking.

c) Existence and pattern of polyelectrolytic effect.

After identifying all the species involved in a macromolecular protonation, the equilibrium constant related to this reaction can easily be calculated. Table 7.3. shows the absorbing species found for the systems polyA-H, polyC-H and polyU-H.

Table 7.3. Identity and conformations of the species involved in the protonation of polyA, polyC and polyU.

System	Species	Conformations
polyU-H	poly(UH) → poly(U^-)	random coil → random coil
polyC-H	poly(C) → poly(C)·poly(CH^+)	single helix → double helix
polyA-H	poly(A) → poly(AH^+)1 → poly(AH^+)2	single helix → double $helix_1$ → double $helix_2$

The information in the distribution plot includes the concentration of all the deprotonated and protonated absorbing species at each point of the acid-base titration. Therefore, for each pH value, an apparent value of the protonation constant can be calculated. Plotting the log K values obtained vs. their corresponding protonation degrees (α_p) allows the detection of the presence and the pattern of polyelectrolytic effect. If this effect is present, the apparent constant (K_{app}) of analogous sites of the polynucleotide changes along the protonation process and the log K value reported for the system is the intrinsic constant (K_{int}), defined as the extrapolated K_{app} value for a protonation degree

equal to zero. If this effect is absent, the K_{app} values remain constant throughout the whole reaction process and a thermodynamic protonation constant is obtained.

The general equation used to calculate the apparent constant of a polynucleotide is:

$$K = \frac{\sum[protonated\ conformers]}{\left(\sum[deprotonated\ conformers]\right)[H^+]}$$

According to the species identification in the ALS distribution plots, the expression above adopts the following forms for the polyU-H system:

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

for the polyC system:

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

and, for the polyA system:

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

The presence of polyelectrolytic effect was detected for polyA and polyC systems. In contrast, PolyU showed constant values of log K in the whole protonation process. This different behaviour of polyU is probably due to the more flexible and disordered structures of the protonated and deprotonated species, which easily allow spatial rearrangements of the macromolecule that minimize the between-sites effect during the protonation process.

PolyA and polyC show polyelectrolytic effects with different patterns. The non-linear pattern of the polyelectrolytic effect of polyC is caused by the cooperative action between the protonation process and the formation of the double-stranded helical structure (i.e., the double helix stabilizes the charged protonated site by the interstrand hydrogen bond $N\text{-}H^{+}\cdots N$ and, at the same time, this base-pair arrangement is directly responsible for the growth of the helical structure). As the protonation process advances, the positive polyelectrolytic effect becomes negative because the repulsive effect between protonated sites becomes more intense than the stabilization caused by the formation of the double-helical structure. For polyA, the positive polyelectrolytic effect has a linear pattern due to the stabilizing electrostatic interactions between the negatively charged phosphate groups and the protonated adenine sites.

There seems to be a certain relation between the pattern of the polyelectrolytic effect and the mechanism of stabilization of the protonated sites. Thus, a non-linear pattern of the polyelectrolytic effect could be linked to the presence of a cooperative mechanism associated with the protonation process (polyC protonated sites are responsible for the formation of one of the interstrand hydrogen bonds of the double helix of the polynucleotide), whereas the more common linear pattern could be mainly caused by electrostatic interactions (polyA protonated sites stabilize the double helical structure minimizing the electrostatic repulsion of neighbouring phosphate groups, but do not participate directly in the interstrand hydrogen-bonding).

Protonation constants of cCMP and cUMP have numerical values similar to the intrinsic protonation constant of polyC and the protonation constant of polyU, respectively, the slight differences between the values of K_{polyU} and K_{cUMP} being due to hindrance effects. However, a clear disagreement is found when the cAMP protonation constant and the polyA intrinsic protonation constant are compared. The much lower value of the cAMP protonation constant could be caused by the absence of the positive influence of the surrounding phosphate groups of other monomeric units in the sites of polyA on the protonation sites of the cyclic nucleotide, cAMP.

d) Effect of the molecular structure of the polynucleotide on the acid-base equilibria.

PolyA, polyC and polyU can be grouped according to two different criteria: the functional group of their protonation sites and the nature of the ring containing the rotonation site in the homopolymer. The former criterion distinguishes polyC and polyA, both with a nitrogen in the aromatic ring as a protonation site, from polyU, whose protonation site is an amide-like nitrogen. The latter criterion puts polyC and polyU in the same group, both with pyrimidinic bases, and polyA in a different class, with a purinic base.

The effect of the functional group of the protonation site on the acid-base equilibria seems to be much stronger than the nature of the ring. According to the similarities of their acid-base behaviours, polyC and polyA belong to the same group, whereas polyU is included in a clearly different class. PolyA and polyC both present ordered structures (single- and double-stranded helices) and polyelectrolytic effect, whereas all polyU species have random coiled structures and no polyelectrolytic effect is detected in the protonation process of the polynucleotide.

The differences in the acid-base behaviour caused by the base ring are less intense and are apparent when polyA and polyC are compared. The different spatial location of the protonation sites in both polynucleotides is responsible for the different roles that these sites play in the macromolecule. Whereas the polyC protonation site is oriented to the core of the double-stranded helix and participates in one of the interstrand bonds, the polyA protonation site is placed in the outer part of the double helix and is not directly responsible for the interstrand union. As commented above, these different roles explain the different pattern of the polyelectrolytic effect detected for both polynucleotides. The different properties of the rings also affect the extent of the stacking in the macromolecule. Purinic rings have a greater stacking ability than pyrimidinic rings. In agreement with this concept, the CD spectra of the double helices of polyA present much higher ellipticities than the double helix of polyC, this hyperchromism of polyA helices being a sign of more ordered and, therefore, more stacked, structures.

e) Solvent effect on the acid-base equilibria of polynucleotides.

Solvent effect on the acid-base equilibria of polynucleotides is studied by comparing the results previously reported in water and those obtained in dioxane/water 30% (v/v) and 50% (v/v) for these reaction processes. Table 7.4. lists the species detected in the polyU, polyA and polyC protonation processes in the different media.

Table 7.4. Species involved in the protonation equilibria of polynucleotides.

System	Solvent	Species
polyU-H	Water	poly(UH), poly(U^-)
	D/W 30% (v/v)	poly(UH), poly(U^-)
	D/W 50% (v/v)	poly(UH), poly(U^-)(1)
polyC-H	Water	poly(C), poly(C)·poly(CH^+), poly(CH^+)
	D/W 30% (v/v)	poly(C), poly(C)·poly(CH^+)
polyA-H	Water	poly(A), poly(AH^+)1, poly(AH^+)2, poly(AH^+)3(2),...
	D/W 30% (v/v)	poly(A), poly(AH^+)1, poly(AH^+)2

(1) The deprotonated species precipitates at pH > 10.

(2) The existence of poly(AH^+)3 and more acidic species is not accepted by all authors.

Hydroorganic media decrease the stability of charged macromolecular structures. Indeed, some species detected in aqueous solutions are not formed in these lower polarity media because the macromolecule precipitates at the suitable pH values. This situation is found for the fully protonated poly(CH^+), for some conformations of protonated polyA and for the deprotonated polyU (only at very basic pH values in the 50% (v/v) dioxane-water mixture). Generally speaking, 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 neat charges.

The species present in dioxane/water mixtures and in water have the same conformation, the only difference being the less intense stacking interactions in hydroorganic media that lead to the formation of more relaxed configurations in these media than in aqueous solution. Such an effect is confirmed by the hypochromism in UV spectra and hyperchromism in CD spectra of the aqueous species with respect to those found in the solvent mixtures (e.g., in the single helical structures of polyC and polyA).

The comments in the paragraphs above concern the solvent effect on the qualitative information related to the acid-base equilibria of polynucleotides. Table 7.6. summarizes the quantitative information associated with these processes.

Table 7.6. Quantitative information related to the acid-base equilibria of polynucleotides.

System	Solvent	log K	Model log K = f(α_p)
cUMP-H	D/W 30% (v/v)	9.250(4)[1]	-
"	D/W 50% (v/v)	9.761(3)	-
cCMP-H	D/W 30% (v/v)	3.766(3)	-
cAMP-H	D/W 30% (v/v)	3.195(8)	-
polyU-H	Water[2]	9.1(2)	log K = 9.1
	D/W 30% (v/v)[4]	9.75(6)	log K = 9.75
	D/W 50% (v/v)[5]	9.914(2)	log K = 9.914
polyC-H[3]	Water[2]	4.16(4)	$\log K = 4.16+3.9\alpha-3.1\alpha^2$
	D/W 30% (v/v)[4]	3.96(2)	$\log K = 3.96+21\alpha-140\alpha^2+480\alpha^3-600\alpha^4$
polyA-H[3]	Water[2]	4.87(4)	$\log K = 1.61\alpha + 4.87$
	D/W 30% (v/v)[4]	4.78(4)	$\log K = 1.46\alpha + 4.78$

(1) Numbers in parentheses are the errors associated with the last figure.
(2) Results in aqueous solution come from references (Casassas, 1993) for polyU, (Casassas, 1995) for polyC and (Casassas, 1994) for polyA.
(3) log K values for polyC and polyA are always intrinsic constants.
(4) Values obtained with CD spectrometry.
(5) Value obtained with potentiometrically.

The lower stability of charged structures in water-dioxane systems with respect to aqueous solutions is noticed in the variations of the numerical values of the equilibrium constants. Thus, the polyU protonation process, that goes from a negatively charged

structure, poly(U^-), to a neutral structure, poly(UH), yields higher constants in hydroorganic solutions than in water; in contrast, polyA and polyC protonations, which go from neutral deprotonated structures to positively charged structures, show higher constants in aqueous solution.

The presence of polyelectrolytic effect depends on the nature of the polynucleotide and follows the same tendency in aqueous and in hydroorganic solutions, i.e., in all media, polyA and polyC protonation processes are affected by polyelectrolytic effects, whereas polyU protonation is not.

In contrast, the pattern of the polyelectrolytic effect may change from solvent to solvent depending on the nature of this effect. In the case of a polyelectrolytic effect controlled by electrostatic interactions, as in polyA, linear models with quite similar slopes describe the variation of log K with the protonation degree in all the media used. However, these small differences between slopes may be due either to the intrinsic similarity of the polyelectrolytic effects in the solvents used or to the effect of the inert salt counterions, which shield the charged phosphates and hide the real solvent effect on the polyA protonation process. Working at lower ionic strengths would be necessary to clarify this point. The pattern of a polyelectrolytic effect caused by a cooperative protonation mechanism, as in polyC, is clearly solvent-dependent. Solvent effect is first noticed in the ranges of existence of the positive and the negative polyelectrolytic effects. Whereas the change of sign appears in α_p values around 0.3 in the hydroorganic solution, the negative behaviour is not seen till α_p higher than 0.5 in water. The negative polyelectrolytic effect does not appear in water 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 polyelectrolytic effect in less polar media. A second aspect of solvent effect concerns the mathematical expression used to model the positive polyelectrolytic effect in both media: in the hydroorganic mixture, a fourth-order polynomial is needed to explain the variation of log K with α_p, whereas a second-order polynomial is enough to explain the same data in aqueous solution. The much steeper effect for low α_p values detected in water/dioxane is due to the easier formation of the interstrand hydrogen

bond N-H^+····N in these hydroorganic media, where the competition of the solvent molecules in the development of these interactions is much weaker than in water. As the protonation proceeds, this tendency is inverted and the polyelectrolytic effect in water/dioxane becomes less and less pronounced, being close to a plateau for α_p values next to 0.3. Such a smoothing in the evolution of the polyelectrolytic effect comes from the increasing weight of the destabilizing influence associated with the augmentation of charge density in the macromolecular structure that gradually balances the favourable tendency to form the interstrand hydrogen bond N-H^+····N.

CHAPTER 8.

CONCLUSIONS.

1. A solvent classification based on the similarities of microscopic solvent properties has been established. The data used have been a set of representative solvents described by the solvatochromic parameters α, β and π^*, related to the solvent hydrogen bond acidity, solvent hydrogen bond basicity and solvent polarity.

2. The six solvent groups obtained are: **I** → slightly basic solvents with low polarity (aliphatic ethers and substituted aliphatic amines), **II** → aprotic polar solvents relatively basic and moderately or highly polar (aliphatic cyclic ethers, esters, ketones and nitriles), **III** → strongly basic and strongly polar solvents (pyridines, small amides, sulfoxides, ureas and phosphoramides), **IV** → relatively polar solvents with a generally low tendency to form hydrogen bonds. The next separation level splits the moderate hydrogen-bond donor solvents from the rest, yielding subgroups A (aromatic ethers, hydrocarbons and halogenated species, and aliphatic apolar halogenated hydrocarbons) and B (polyhalogenated polar aliphatic hydrocarbons, whose heteroatoms cause the inductive effect responsible for the hydrogen-bond donor interaction), **V** → amphiprotic solvents (alcohols and water) with marked hydrogen-bond properties Successive splittings separate glycol and water from the other alcohols due to their greater association ability and, **VI** → aliphatic hydrocarbons with null solute/solvent interactions, i.e., with α, β and π^* equal to zero.

3. The new solvatochromic scheme has been compared with the Snyder scheme, which uses other microscopic descriptors related to solvent hydrogen-bond acidity, $\log(K_g'')_d$, solvent hydrogen-bond basicity, $\log(K_g'')_e$, and solvent polarity, $\log(K_g'')_n$. The lack of structure of the 3D Snyder solvent space and the mixed character of the Snyder hydrogen-bond solvent descriptors ($\log(K_g'')_d$ and $\log(K_g'')_e$ are actually linear combinations of α and π^* and of α, β and π^*, respectively) recommend preferably the solvatochromic scheme for solvent selection purposes.

4. Water-dioxane mixtures covering the whole range of compositions have been studied. The solvent parameters α, β and π^*, proposed by Kamlet et al., and the polarity parameter $E_T(30)$, proposed by Reichardt, have been determined for these mixtures. π^* values were determined as the average of the results obtained with the solvatochromic indicators 2-nitroanisole, 4-nitroanisole and 4-ethylnitrobenzene. Two independent series of β were determined: β_1, using the solute pair 4-nitroaniline and N,N-diethyl-4-nitroaniline, and β_2, using the solute pair 4-nitrophenol and 4-nitroanisole. The values of α and $E_T(30)$ have been obtained with one pair of solutes (4-nitroanisole and Reichardt's betaine) and one indicator, respectively (Reichardt's betaine).

5. The correlation between the Reichardt parameter $E_T(30)$ and the parameters of Kamlet et al. in water-dioxane mixtures confirms the mixed nature of the $E_T(30)$ parameter, sensitive to solvent polarity/polarizability and to solvent hydrogen-bond acidity, through a model that describes the $E_T(30)$ parameter as a linear combination of α and π^*. Equations related to this model present good fits either using both the averaged π^* values or any of the π^* series related to each of the π^*-sensitive indicators.

6. Preferential solvation of solutes has been detected in water-dioxane mixtures by studying the evolution of the $E_T(30)$ parameter with solvent composition. A double behaviour has been found in the plot of $E_T(30)$ vs. x_{water}. For water mole fractions lower than 0.45 (% dioxane (v/v) > 85%), water is preferentially placed around the solutes, whereas dioxane is preferred when the water content exceeds this threshold value. This behaviour is attributed to the change of water structure in the mixtures: thus, for low dioxane

contents, the three-dimensional structure of water is kept and the cosolvent is preferentially placed around the solutes; for high dioxane contents, the water network is broken and the isolated water molecules interact more easily with the solutes.

7. The solvent-dependency of certain solute properties has been described using LSERs. Two approaches are proposed to help in the correct establishment of these behaviour models: a hard-modelling approach, that fits a chemical model to the experimental data, and a soft-modelling approach, that does not need the support of any external expression. Both approaches work determining the size and identity of the terms to be included in the model. The former does this by performing a stepwise selection of the suitable solvent properties to be included in the model, a robust regression to eliminate the experimental outliers and a LS model using only the observations accepted. The soft-modelling approach combines a first Factor Analysis of the data to identify the size of the model and a Target Factor Analysis to find out the nature of the solvent properties to be included in the model. Both approaches provide complementary benefits and the agreement of the results obtained by using the two different procedures is an excellent validation of the LSERs established.

8. The soft-modelling approach is used to assess the sources of variation of the spectral shifts of the solvatochromic indicators. In agreement with the Kamlet et al. postulations, the shifts of all the indicators used to determine α, β and π^* parameters may be described with three factors that have been identified as solvent hydrogen bond acidity, solvent hydrogen bond basicity and solvent polarity. Furthermore, each individual indicator accepted in the determination of these parameters shifts only according to the solvent features that it is supposed to describe, thus confirming the quality of the solvatochromic descriptors proposed.

9. Autoprotolysis constants for the solvent and acidity equilibrium constants for several single solutes have been determined in dioxane-water mixtures covering the range from 10 to 70% (v/v) dioxane. The determinations have been performed at 25°C and at constant ionic strength. The potentiometric cell used has been: G.E. / working solution,

n% dioxane / R.E. (KCl_{sat}, *n*% dioxane), where G.E. is the glass electrode, R.E. is the reference electrode and *n* is the percentage (v/v) of dioxane in the working solution and in the inner solution of the reference electrode. The Gran method has been used to calibrate *in situ* the potentiometric cell.

10. LSERs which describe the solvent-dependency of the acid-base equilibria of single solutes in water-dioxane mixtures have been established by using both hard- and soft-modelling approaches. These two approaches have given the same general model, $pK = pK^{o} + s\pi^{*}$. Thus, the solvent effect on the acid-base behaviour of solutes in water-dioxane mixtures is controlled by the microscopic solvent polarity, as described by the π^{*} parameter, despite the functionality of the protonation sites in the solutes and the composition range studied. No other univariate expression describes the solvent-dependent pK variation better than the model above in the whole range of solvent mixtures; thus, models using macroscopic solvent properties (e.g., mole fraction or relative permittivity) lose linearity for dioxane-rich mixtures and other models based on microscopic polarity parameters, such as $E_T(30)$, obey a linear expression in water-rich mixtures and a different expression in dioxane-rich mixtures.

11. Once the solvent effect on the acid-base processes of simple solutes in water-dioxane mixtures has been established, the pH-dependent transitions of the polynucleotides polyadenylic acid (polyA), polycytidylic acid (polyC) and polyuridylic acid (polyU) and of their respective cyclic nucleotides adenine-3',5'-cyclic monophosphate (cAMP), cytosine-3',5'-cyclic monophosphate (cCMP) and uracil-3',5'-cyclic monophosphate (cUMP) have been studied. The acid-base behaviour of these compounds has been investigated by using spectrometric (UV and circular dichroism) and potentiometric techniques in order to obtain a qualitative and quantitative description of all the transitions (thermodynamic and conformational) associated with the polynucleotides. All the studies have been performed at 30 or 50% (v/v) of dioxane because at higher dioxane percentages, precipitation phenomena occurred.

12. Curve resolution methods (CR) have been used to treat the output from the spectrometric experiments. In contrast to classical data treatments, CR does not assume the fulfilment of the mass action law or the one-to-one correspondence between chemical species and absorbing species. This lack of assumptions allows the modelling of the system to take into account the presence of possible macromolecular side effects, such as polyelectrolytic effect or the presence of more than one conformer associated with one chemical species.

13. The performances of two three-way resolution methods, Alternating Least Squares (ALS) and Trilinear Decomposition (TLD), have been compared to select the most appropriate procedure to deal with the spectrometric data of polynucleotides. Choosing ALS or TLD depends mainly on the presence of selectivity and the presence of trilinear structure in the data set. Comparison of the fits obtained by TLD and ALS is proposed as a tool to diagnose the internal structure of the data: thus, ALS always shows a better fit than TLD for non-trilinear data and vice versa. ALS works better than TLD for non-trilinear systems, though for those situations with null selectivity and slight departures from trilinearity, TLD and ALS can sometimes recover the qualitative and the quantitative information in a similar manner. TLD and ALS forcing trilinear structure are preferred for trilinear systems without selectivity; if selectivity is present, ALS performs as well as the methods forcing trilinear structure, though a slight superiority of the former is detected in systems with high noise levels. Since the spectrometric data of polynucleotides are non-trilinear and may or may not have selectivity, the use of ALS has been preferred because of the higher robustness in the modelling of concentration and spectral profiles.

14. Three new constraints have been proposed to improve the optimization step of the ALS method. The horizontal unimodality shares the same structure of the classical unimodality, but the suppression of secondary maxima is performed by setting the non-unimodal elements in the profile equal to the closest element keeping the unimodal condition; the localized unimodality forces the presence of one maximum per profile and restricts the position of this maximum and, finally, the symmetry constraint eliminates the

front-tailed profiles by making the back half of the peak symmetrical to the front half. The usefulness of these constraints is checked on a pool of representative data sets simulated by using a two-level full factorial design that takes into account the factors that seem to affect the ALS results the most. In their present implementations, only the constraint of horizontal unimodality caused a significant improvement in the resolution results, particularly in terms of recovery of profiles related to minor species. There were no significant effects associated with the inclusion of the other two constraints, though this could be due to the implementation used.

15. The polyU-H system has been studied potentiometrically and spectrometrically in 30% and 50% (v/v) dioxane-water mixtures. In all the experiments performed, the pH-dependent transitions can be explained with two species, the protonated poly(UH) and the deprotonated poly(U^-). In the working solvent mixtures, all the techniques used showed the absence of polyelectrolytic effect, and, therefore, thermodynamic constants related to the polyU protonation could be calculated. From the structural point of view, both protonated and deprotonated polyU share the same random-coiled structure and, therefore, the polyU-H system shows a pH-dependent thermodynamic transition without structural alterations of the macromolecule. The existence of these random-coiled structures has been confirmed through the similar absorptivities of the polynucleotide and the cUMP spectra.

16. The polyC-H system has been studied potentiometrically and spectrometrically in a 30% dioxane-water mixture. No experiments have been performed at higher dioxane contents because precipitation occurred. The pH-dependent transitions have been explained with two species, the deprotonated poly(C) and the half-protonated [poly(C)·poly(CH^+)]. In the working solvent mixtures, both the potentiometric titrations (qualitatively, because precipitation was present) and the spectrometric experiments detected the presence of a polyelectrolytic effect with a non-linear pattern. Such an effect is positive for low protonation degrees due to the cooperative mechanism between the protonation process and the formation of the double-helix of the half-protonated polyC and it becomes negative once the effect of the increase of charge density of the structure overcomes the

positive influence of the cooperative mechanism in the protonation process. From the structural point of view, the deprotonated polyC has a single-helical conformation, whereas the half-protonated polyC presents a double-stranded helical structure. The existence of these ordered structures has been confirmed through the hypochromicity in UV and hyperchromicity in CD of polyC spectra with respect to the cCMP spectrum. The polyC-H system presents an example of pH-dependent thermodynamic transition with an associated conformational change.

17. The polyA-H system has been studied potentiometrically and spectrometrically in a 30% dioxane-water mixture. No experiments have been performed at higher dioxane contents because precipitation occurred. The pH-dependent transitions have been explained with three species, the deprotonated polyA and the protonated conformations poly(AH^+)1 and poly(AH^+)2. In the dioxane-water mixture, the presence of a polyelectrolytic effect with a linear pattern was detected. Such a positive effect is explained by the minimization of the electrostatic repulsion of neighbouring phosphate groups through the interaction between the negative charges of these groups and the positive charges of the protonated sites. From the structural point of view, deprotonated polyA presents a single-helical conformation, whereas the two protonated forms of the polynucleotide are defined as different double-stranded helical conformations. The existence of these ordered structures has been confirmed through the hypochromicity in UV and hyperchromicity in CD of polyA spectra with respect to the cAMP spectrum. The polyA-H system shows one conformational transition associated with the protonation process and another which can be described as a spatial rearrangement of the macromolecule which is not linked to any chemical reaction.

18. The effect of the functional group of the protonation site on the polynucleotide is much stronger than the nature of the base ring on the protonation processes. Thus, polyA and polyC, which have a nitrogen in the aromatic ring as protonation site, both present ordered structures (single- and double-stranded helices) and polyelectrolytic effect, whereas polyU, whose protonation site is an amide-like nitrogen, have random coiled structures and no polyelectrolytic effect is detected. Differences in the acid-base

behaviour caused by the base ring are less intense and concern mainly the degree of base stacking of the ordered structures (higher in purinic rings than in pyrimidinic rings) and the role that the protonation site plays on the macromolecular structure (directly responsible for the interstrand union in double helices, as in polyC, or not, as in polyA).

19. Solvent effect on the qualitative information related to acid-base equilibria of polynucleotides has been studied by comparing the results previously reported in water and those obtained in dioxane/water 30% (v/v) and 50% (v/v) for these reaction processes. Hydroorganic media decrease the stability of charged macromolecular structures and some charged species found in aqueous solution, such as the fully protonated poly(CH^+), some conformations of protonated polyA and the deprotonated polyU (only at very basic pH values in the 50% (v/v) dioxane-water mixture), are not formed in water-dioxane mixtures. No variations in conformation have been detected between the species present in water and in dioxane/water mixtures, the only difference being the formation of more relaxed configurations in dioxane-water mixtures due to weaker base stacking.

20. Solvent effect on the quantitative aspects of the acid-base equilibria involves variations in the numerical values of the thermodynamic or apparent constants related to the protonation processes and, in some situations, modifications of the pattern of the polyelectrolytic effect. Owing to the low stability of charged structures in water-dioxane systems, protonation processes involving the creation of charged species, such as those related to polyC and polyA, show lower constants in these systems, whereas processes involving the transition from charged species to neutral, as the polyU protonation, are more favourable in hydroorganic media. The presence of polyelectrolytic effect follows the same tendency in aqueous and in hydroorganic solutions, i.e., in all media, polyA and polyC protonation processes are affected by polyelectrolytic effects, whereas polyU protonation is not. The pattern of the polyelectrolytic effect remains unchanged when this effect is absent or shows a linear pattern, as in polyU and polyA, respectively, and is clearly solvent-dependent when this effect is associated with a cooperative mechanism, as in polyC. Solvent effect on polyC influences the ranges of existence of the positive and the negative polyelectrolytic effects. Thus, the change of sign from positive to

negative appears at lower α_p values in hydroorganic mixtures than in aqueous solution. The mathematical expression used to model the positive polyelectrolytic effect in both media also reveals differences: whereas a fourth-order polynomial is needed to explain the variation of log K with α_p in water-dioxane mixtures, a second-order polynomial is enough to explain the same effect in aqueous solution. The much steeper effect for low α_p values detected in water/dioxane stems from the weaker competition of the solvent molecules in the development of the interstrand hydrogen bond N-H^+····N in these low polarity media. As the protonation proceeds, the positive polyelectrolytic effect in water-dioxane media becomes weaker than in water due to the destabilizing influence caused by the augmentation of charge density in the macromolecular structure that gradually balances the favourable effect due to the formation of the interstrand hydrogen bond N-H^+····N.

IV. REFERENCES

Abraham, 1993.
M.H. Abraham. Scales of Solute Hydrogen-Bonding - Their Construction and Application to Physicochemical and Biochemical Processes. *Chem. Soc. Rev.* 22 (1993) 73.

Adams, 1992.
R.L.P. Adams, J.T. Knowler and D.P. Leader. *The biochemistry of the nucleic acids.* 11th. ed. (1992) Ed. Chapman & Hall.

Åkerlöf, 1936.
G. Åkerlöf and O.A. Short. The dielectric constant of dioxane-water mixtures between 0 and 80 °C. *J. Am. Chem. Soc.*, 53 (1936) 1241.

Albert, 1996.
I.D.L. Albert, T.J. Marks and M.A. Ratner. Rational design of molecules with large hyperpolarizabilities - electric-field, solvent polarity, and bond-length alternation effects on merocyanine dye linear and nonlinear-optical properties. *J. Phys. Chem.*, 100 (1996) 9714.

Arnett, 1962
E.M. Arnett and C.Y. Wu. Base strengths of some saturated cyclic ethers in aqueous sulfuric acid. *J. Am. Chem. Soc.*, 84 (1962) 1684.

Balakrishnan, 1981.
S. Balakrishnan and A.J. Easteal. Empirical polarity parameters for some binary solvent mixtures. *Aust. J. Chem.*, 34 (1981) 933.

Berthélot, 1862.
M. Berthélot and L. Péan de Saint-Gilles. *Ann. Chim. et Phys.,3. Ser.*, 65 (1862) 385.

Bloch, 1975.
A. Bloch. Chemistry, Biology and Clinical Uses of Nucleosides. *Analogs. Ann. NY Acad. Sci.*, (1975) 225.

Boerhaave, 1733.
H. Boerhaave. *Elementa Chemiae.* Editio Altera (1733) G. Cavelier, Paris.

Boggetti, 1994.
H. Boggetti, J.D. Anunziata, R. Cattana and J.J. Silber. Solvatochromic study on nitroanilines - preferential solvation vs dielectric enrichment in binary solvent mixtures. *Spectrochim. Acta Part A - Molec. Spectrosc.*, 50 (1994) 719.

Bolton, 1979.
P.H. Bolton and D.R. Kearns. Intramolecular water bridge between 2'-OH and phosphate groups in RNA. Cyclic nucleotides as a model system. *J. Amer. Chem. Soc.*, 101 (1979) 479.

Borer, 1974.
P.N. Borer, B. Dengler, I. Tinoco Jr. and O.C. Uhlenbeck. Stability of ribonucleic double-stranded helices. *J. Mol. Biol.*, 86 (1974) 843.

Bosch, 1992.
E. Bosch and M. Rosés. Relationships between E_T polarity and composition in binary solvent mixtures. *J. Chem. Soc. Faraday Trans.*, 88 (1992) 3541.

Bosch, 1995.
E. Bosch, C. Ràfols and M. Rosés. Variation of acidity constants and pH values of some organic acids in water/2-propanol mixtures with solvent composition. Effect of preferential solvation. *Anal. Chim. Acta*, 302 (1995) 109.

Brady,1982.
J.E. Brady and P.W. Carr. Perfluorinated solvents as non-polar test systems for generalized models of solvatochromic measures of solvent strength. *Anal. Chem.*, 54 (1982) 1751.

Brahms, 1966.
J. Brahms and C. Sadron. Ionic and hydrogen-bond interactions contributing to the conformational stability of polynucleotides. *Nature*, (1966) 1309.

Bro,1996.
R. Bro and H. Heimdal. Enzymatic Browning of Vegetables. Calibration and Analysis of Variance by Multiway Methods. *Chemom. Intel. Lab. Sys.*, 34 (1996) 85.

Brooker, 1951.
L.G.S. Brooker, G.H. Keyes and D.W. Heseltine. Colour and constitution. XI. Anhydronium bases of p-hydroxystyryl dyes as solvent polarity indicators. *J. Am. Chem. Soc.*, 73 (1951) 5350.

Brooker, 1965.
L.G.S. Brooker, A.C. Craig, D.W. Heseltine, P.W. Jenkins and L.L. Lincoln. Color and constitution. XIII. Merocyanines as solvent property indicators. *J. Am. Chem. Soc.*, 87 (1965) 2443.

Buffle, 1988.
J. Buffle. *Complexation reactions in aquatic systems: an analytical approach.* (1988) Ellis Horwood Ltd. Wiley & Sons, NY.

Bugg, 1971.
C.E. Bugg, J.M. Thomas, M. Sundaralingam and S.T. Rao. Stereochemistry of nucleic acids and their constituents. X. Solid-state base-stacking patterns in nucleic acid constituents and polynucleotides. *Biopolymers*, 10 (1971) 175.

Burger, 1983.
K. Burger. *Solvation, ionic and complex formation reactions in non-aqueous solvents. Experimental methods for their investigation.* (1983) Elsevier.

Burger, 1990.
K. Burger. *Biocoordination chemistry: coordination equilibria in biologically active systems.* (1990) Ellis Horwood. NY.

Burgess, 1970.
J. Burgess. Solvent effects on visible absorption spectra of bis-(2,2'-bipyridil) biscyanoiron(II),bis-(1,10-phenantroline)biscyanoiron(II), and related conpounds. *Spectrochim. Acta, Part A*, 26 (1970) 1369.

Calvin,1945.
M. Calvin and K.W. Wilson. Stability of chelate compounds. *J. Am. Chem. Soc.*, 67 (1945) 2003.

Carr, 1993.
P.W. Carr. Solvatochromism, Linear Solvation Energy Relationships and Chromatography. *Microchem. J.*, 48 (1993) 4

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

Casassas, 1995.
E. Casassas, R. Gargallo, A. Izquierdo-Ridorsa and R. Tauler. 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 (1995) 1.

Catalán, 1995.
J. Catalán, V. López, P. Pérez R. Martin Villamil and J.G. Rodriguez. Progress Towards a Generalized Solvent Polarity Scale - The Solvatochromism of 2-(Dimethylamino)-7-Nitrofluorene and Its Homomorph 2-Fluoro-7-Nitrofluorene. *Liebigs Ann.*, 2 (1995) 241

Catalán,1996.
J. Catalán, C. Díaz, V. López, P. Pérez, J.L. G. de Paz and J.G. Rodríguez. A generalized solvent basicity scale: the solvatochromism of 5-nitroindoline and its homomorph 1-methyl-5-nitroindoline. *Liebigs Ann.*, (1996) 1785.

Causley, 1982.
G.C. Causley and W.C. Johnson Jr. Polynucleotide conformation from flow dichroism studies. *Biopolymers*, 21 (1982) 1763.

Claisen, 1896.
L. Claisen. *Liebigs Ann. Chem.*, 291 (1896) 25.

Covington, 1974.
A.K. Covington, I.R. Lantzke and J.M. Thain. Nuclear magnetic resonance studies of preferential solvation. Part 3. thermodynamic treatment involving change of solvation number, and application to dymethilsulphoxide-containing solvents. *J. Chem. Soc. Faraday Trans. I*, (1974) 1869.

Covington, 1976.
A. K. Covington. Thermodynamics of preferential solvation of electrolytes in binary solvent mixtures. *Adv. Chem. Ser.*, 155 (1976) 153.

Covington, 1989.
A.K. Covington and M. Dunn. Nuclear magnetic resonance studies of preferential solvation. Part 6. Application of Blander's coordinated cluster theory to the methanol-water solvent system. *J. Chem. Soc. Faraday Trans. I*, 85 (1989) 2827.

Critchfield, 1953.
F.E. Critchfield, J.A. Gibson, Jr. and J.L. Hall. Dielectric constant for the dioxane-water system from 20 to 35 °C. *J. Am. Chem. Soc.*, 75 (1953) 1991.

Cheong, 1988.
W.J. Cheong and P.W. Carr. Kamlet Taft π^* polarizability/dipolarity of mixtures of water with various organic-solvents. *Anal. Chem.*, 60 (1988) 820.

Cuesta-Sánchez, 1996.
F. Cuesta Sanchez, J. Toft, B. Van den Bogaert and D.L. Massart. Orthogonal projection approach applied to peak purity assessment. *Anal. Chem.* 68 (1996) 79

Dähne, 1975.
S. Dähne, F. Shob, K.D. Nolte and R. Radeglia. New parametres of solvent polarity. *Ukr. Khim. Zh.*, 41 (1975) 1170.

Das, 1980.
R.C. Das, U.N. Dash and K.N. Panda. Thermodynamics of dissociation of DL-malic, maleic and fumaric acids in water and water + dioxane mixtures. *J. Chem. Soc. Faraday Trans. I*, 76 (1980) 2152.

Dawber, 1988.
J.G. Dawber, J. Ward and R.A. Williams. A study in preferential solvation using a solvatochromic pyridinium betaine and its relationship with reaction rates in mixed solvents. *J. Chem. Soc. Faraday Trans.*, 84 (1988) 713.

Dawber, 1990.
J.G. Dawber. Relationship between solvatochromic solvent polarity and various thermodynamic and kinetic data in mixed solvent systems. *J. Chem. Soc. Faraday Trans.*, 86 (1990) 287.

de Juan, 1996.
A. de Juan, B. Van den Bogaert F. Cuesta Sánchez and D.L. Massart. Application of the needle algorithm for exploratory analysis and resolution of HPLC-dad data. *Chemom. Intell. Lab. Sys.*, 33 (1996) 133

de Leeuw, 1980.
H.P.M. de Leeuw, C.A.G. Haasnoot and C. Altona. Empirical correlations between conformational parameters in β-D-furanoside fragments derived from a statistical survey of crystal structures of nucleic acid constituents. Full description of nucleoside molecular geometries in terms of four parameters. *Isr. J. Chem.*, 20 (1980) 108.

Dealencastro, 1994.
R.B. Dealencastro, J.D.D. Neto, M.C. Zerner. Solvent Effects on the Electronic-Spectrum of Reichardts Dye. *Int. J. Quant. Chem.*, 28 (1994) 361

Delauzon, 1994.
S. Delauzon, K.M. Rajkowski and N. Cittanova. Investigation of a 17-Beta-Estradiol-Monoclonal Antiestradiol Antibody-Binding Mechanism Using Dilute-Solutions of Organic-Solvents. *J. Ster. Biochem. Mol. Biol.*, 48 (1994) 225.

DeVoe, 1962.
H. DeVoe and I. Tinoco Jr. The stability of helical polynucleotides: base contributions. *J. Mol. Biol.*, 4 (1962) 500.

Dimroth, 1963.
K. Dimroth, C. Reichardt, T. Siepmann and F. Bohlman. Pyridinium N-phenolbetaines and their use for the characterization of the polarity of solvents. *Liebigs Ann. Chem.*, 661 (1963) 1.

Dong,1982.
D.C. Dong and M.A. Winnik. The Py scale of solvent polarities. solvent effects on the vibronic fine-structure of pyrene fluorescence and empirical correlations with E_T value and Y value. *Photochem. and Photobiol.*, 25 (1982) 17.

Drago, 1965.
R.S. Drago and W.W. Wayland. Double-scale equation for correlating enthalpies of Lewis acid-base interactions. *J. Am. Chem. Soc.*, 87 (1965) 3571.

Drago, 1992.
R.S. Drago. A Unified Scale for Understanding and Predicting Nonspecific Solvent Polarity - A Dynamic Cavity Model. *J. Chem. Soc. Perkin Trans.*, 2 (1992) 1827.

Drago, 1994.
R.S. Drago, M.S. Hirsch, D.C. Ferris and C.W. Chronister. A unified scale of solvent polarities for specific and nonspecific interactions. *J. Chem. Soc. Perkin Trans.*, 2 (1994) 219.

Drummond, 1987.
C.J. Drummond, F. Grieser and T.W. Healy. Effect of Electrolyte on the Mean Interfacial Solvent and Electrostatic Characteristics of Cationic Micelles *Chem. Phys. Lett.*, 140 (1987) 493.

Dubois, 1966.
J.E. Dubois and J.E . Bienvenue. Ambivalent character of the solvation effects on certain electronic transitions. *Tetrahedron Lett.*, 8 (1966) 1809.

Effenberger, 1995.
F. Effenberger, F. Wurthner and F. Steybe. Synthesis and solvatochromic properties of donor-acceptor-substituted oligothiophenes. *J. Org. Chem.*, 60 (1995) 2082.

Erdey-Grúz, 1973.
T. Erdey-Grúz and B. Levay. Self-diffusion coefficient of iodine in methanol-water and dioxane-water mixtures. *Acta Chim. Acad. Sci. Hung.*, 79 (1973) 401.

Faber, 1997.
K. Faber-K and B.R. Kowalski. Critical-Evaluation of 2 F-Tests for Selecting the Number of Factors in Abstract Factor-Analysis. *Anal. Chim. Acta,* 337 (1997) 57.

Famini, 1989.
G.R. Famini. Using theoretical descriptors in QSAR. V. CRDEC-TR-085. *US Army Chem. Res., Dev. Eng. Cent.*, Aberdeen Proving Ground, MD, 1989.

Famini, 1992.
G.R. Famini, W.P. Ashman, A.P. Mickiewicz and L.Y. Wilson, Using Theoretical Descriptors in Quantitative Structure - Activity-Relationships - Some Physicochemical Properties. *J. Phys. Org. Chem.*, 5 (1992) 395.

Finch, 1969.
J.T. Finch and A. Klug. Two double helical forms of polyriboadenylic acid and the pH-dependent transition between them. *J. Mol. Biol.*, 46 (1969) 597.

Forina, 1988.
M. Forina, R. Leardi, C. Armanino and S. Lanteri. *PARVUS, an extendable package of programs for data exploration, classification and correlation.* (1988), Elsevier, Amsterdam.

Freifelder, 1987.
D. Freifelder. *Molecular Biology.* 2nd. ed. (1987) Ed. Jones and Bartlett.

Friedman, 1995.
R.A. Friedman and B. Honig. A free-energy analysis of nucleic acid base stacking in aqueous solution. *Biophys. J.*, 69 (1995) 1528.

Gans, 1985.
P. Gans, A. Vacca and A. Sabatini. Superquad - an improved general program for computation of formation-constants from potentiometric data. *J. Chem. Soc. Dalton Trans.*, 6 (1985) 1195.

Gemperline, 1984.
P.J. Gemperline. A priori estimates of the elution profiles of the pure components in overlapped liquid chromatography peaks using Target Factor Analysis. *J. Chem. Inf. Comput. Sci.*, 24 (1984) 206.

Gemperline, 1986.
P.J. Gemperline. Target Transformation Factor Analysis with linear inequality constraints applied to spectroscopic-chromatographic data. *Anal. Chem.*, 58 (1986) 2856.

Gielen, 1963.
M. Gielen and J. Nasielski. Electrophilic substitution at a saturated carbon atom. III. Competition between steric and inductive effects as a function of the polarity of solvents in the reaction of halogens with tetraalkyl compounds. *J. Organomet. Chem.*, 1 (1963) 173.

Glover,1965.
D.J. Glover. Equilibria in solution. I. Ion solvation and mixed solvent interaction. *J. Am. Chem. Soc.*, 87 (1965) 5275.

González-Arjona, 1994.
D. Gonzalez-Arjona, J.A. Mejias and A.G. Gonzalez. Holmes - A Program for Target Factor-Analysis. *Anal. Chim. Acta*, 295 (1994) 119.

Gran, 1952.
G. Gran. Determination of the equivalence point in potentiometric titrations. Part II. *Analyst 77*, (1952) 661.

Grunwald, 1948.
E. Grunwald and S. Winstein. Correlation of solvolysis rates. *J. Am. Chem. Soc.*, 70 (1948) 846.

Gutmann, 1966.
V. Gutmann and E. Wychera. Coordination reactions in nonaqueous solutions. Role of the donor strength. *Inorg. Nucl. Chem. Lett.*, 2 (1966) 257.

Haak, 1986.
J.R. Haak and J.B.F.N. Engberts. Solvent polarity and solvation effects in highly aqueous mixed solvents. Application of the Dimroth-Reichardt $E_T(30)$ parameter. *Rec. Trav.Chim. Pays-Bas*, 105 (1986) 307.

Hall, 1971.
R.H. Hall. *The modified nucleosides in nucleic acids* (1971) Columbia Univ. Press, NY.

Hamilton,1990.
J.C. Hamilton and P.J. Gemperline. Mixture analysis using factor analysis. II: self-modeling curve resolution. *J. of Chemom.*, 4 (1990) 1.

Hanlon, 1966.
S. Hanlon. The importance of London dispersion forces in the maintenance ofthe deoxyribonucleic acid double helix. *Biochem. Biophys. Res. Commun.*, 23 (1966) 861.

Harmon, 1978.
R.E. Harmon. *Chemistry and biology of nucleotides and nucleosides.* (1978) R.E. Harmon, R.K. Robins and L.B. Townsend ed. Academic Press, NY.

Henry, 1997.
E.R. Henry. The use of matrix methods in the modeling of spectroscopic data sets. *Biophys. J.*, (1997) (in press).

Hildebrand, 1979.
J. H. Hildebrand. Is there a "hydrophobic effect"? *Proc. Natl. Acad. Sci.* 76 (1979) 194.

Izquierdo, 1986.
A. Izquierdo and J.L. Beltrán. Miniglass, an interactive program for the evaluation of stability-constants of metal-ligand complexes from potentiometric data. *Anal. Chim. Acta*, 181 (1986) 87.

Jano, 1992.
I. Jano. Study of the Solvents Effect on the Electronic-Spectra of Betaine and Azo-Merocyanine Dyes. *J. Chim. Phys.* 89 (1992) 1951.

Johnson,1986.
B.P. Johnson, J.G. Dorsey, B. Gabrielsen, M. Matulenko and C. Reichardt. Solvatochromic solvent polarity measurements in analytical chemistry. Synthesis and applications of $E_T(30)$. *Anal. Lett.*, 19 (1986) 939.

Kamlet,1976.
M.J. Kamlet and R.W. Taft. 1. The solvatochromic comparison method. 1. The β-scale of solvent hydrogen bond acceptor (HBA) basicities. *J. Am. Chem. Soc.*, 98 (1976) 377.

Kamlet,1977.
M.J. Kamlet, J.L. Abboud and R.W. Taft. The solvatochromic comparison method. 6. The π* scale of solvent polarities. *J. Am. Chem. Soc.*, 99 (1977) 6027.

Kamlet,1981.
M.J. Kamlet, J.L.M. Abboud and R.W. Taft. An examination of linear solvation energy relationships. *Progress in Physical Chemistry*. 13 (1981) 485 Interscience, NY.

Kamlet,1983.
M.J. Kamlet, J.L.M. Abboud, M.H. Abraham and R.W. Taft. Linear Solvation Energy Relationships .23. A comprehensive collection of the solvatochromic parameters, π^*, α and β, and some methods for simplifying the generalized solvatochromic equation. *J. Org. Chem.*, 48 (1983) 2877.

Kamlet,1985.
M.J. Kamlet and R.W. Taft. Linear Solvation Energy Relationships. Local empirical rules or fundamental laws of chemistry? A reply to the chemometricians. *Acta Chem. Scand. Ser. B.* 39 (1985) 611.

Kanski,1993.
R. Kanski and C.J. Murray. Enzymichromism: determination of the dielectric properties of an enzyme active site. *Tetrahedron Lett.* 34 (1993) 2263.

Karjalainen,1989.
E.J. Karjalainen. The spectrum reconstruction problem. Use of Alternating Regression for unexpected spectral components in two-dimensional spectroscopies. *Chemom. and Intell. Lab. Sys.* 7 (1989) 31.

Keller, 1991.
H.R. Keller and D.L. Massart. Peak Purity Control in Liquid-Chromatography with Photodiode-Array Detection by a Fixed Size Moving Window Evolving Factor-Analysis *Anal. Chim. Acta*, 246 (1991) 379.

Kistenmacher, 1978.
T.J. Kistenmacher, M. Rossi and L.G. Marzilli. A model for the interrelationship between asymmetric interbase hydrogen bonding and base-base stacking in hemiprotonated polyribocytidylic acid: crystal structure of 1-methylcytosine hemihydroiodide hemihydrate. *Biopolymers*, 17 (1978) 2581.

Knorr, 1896.
L. Knorr. *Liebigs Ann. Chem.*, 293 (1896) 70.

Koppel, 1971.
I.A. Koppel and V.A. Palm. General equation for solvent effects. *Reakts. Sposobnost Org. Soedin.*, 8 (1971) 291.

Koppel, 1983a.
I.A. Koppel and J.B. Koppel. E_T parameters of some binary mixtures of hydroxylic and aprotic solvents. *Organic Reactivity*, 4 (1983) 547.

Koppel, 1983b.
I.A. Koppel and J.B. Koppel E_T parameters of binary mixtures of alcohols with DMSO and MeCN. Synergetic solvent effect of high intensity. *Organic Reactivity*, 20 (1983) 523.

Koppel, 1994.
Koppel-IA Anvia-F Taft-RW. Title: Rechecking of the Equilibrium Gas-Phase Basicity Scale for Low-Basicity Compounds Using Fourier-Transform Ion Cyclotron Resonance Spectrometry. *J. Phys. Org. Chem.*, 7 (1994) 717.

Kosower, 1958a
E.M. Kosower. Effect of solvent on spectra. I. Empirical measure of solvent polarity.- Z values. *J. Am. Chem. Soc.*, 80 (1958) 3253.

Kosower, 1958b.
E.M. Kosower. Effect of solvent on spectra. II. Correlation of spectral absorption data with Z values. *J. Am. Chem. Soc.*, 80 (1958) 3261.
Kosower, 1958c.
E.M. Kosower. Effect of solvent on spectra. III. Use of Z values with kinetic data. *J. Am. Chem. Soc.*, 80 (1958) 3267.

Krigowski, 1975.
T.M. Krigowski and W.R. Fawcett. Complementary Lewis acid-base description of solvent effects. I. Ion.ion and ion-dipole interactions. *J. Am. Chem. Soc.* 97 (1975) 2143.

Kvalheim,1993.
Y.Z. Liang and O.M. Kvalheim. Heuristic evolving latent projections: resolving hyphenated chromatographic profiles by component stripping. *Chemom. and Intell. Lab. Sys.*, 20 (1993) 115.

Lacorte,1995.
S. Lacorte, D. Barceló and R. Tauler. Determination of traces of herbicide mixtures in water by on-line solid-phase extraction followed by liquid chromatography with diode-array detection and multivariate self-modelling curve resolution. *J. of Chromatog. A*, 697 (1995) 345.

Langhals, 1982.
H. Langhals. Polarity of binary liquid mixtures. *Angew. Chem. Int. Ed. Engl.*, 21 (1982) 724.

Laurence,1986.
C. Laurence, M. Helbert and P. Nicolet. Polarity and basicity of solvents .2. Solvatochromic hydrogen-bonding shifts as basicity parameters. *J. Chem. Soc. Perkin Trans. II.* 7 (1986) 1081.

Laurence,1986.
C. Laurence, P. Nicolet and M. Helbert. Polarity and basicity of solvents. Part 2. Solvatochromic hydrogen-bonding shifts as basicity parameters. *J. Chem. Soc. Perkin Trans.* II (1986) 1081.

Laurence,1994.
C. Laurence, P. Nicolet, M. Tawfik Dalati, J.L.M. Abboud and R. Notario. The empirical treatment of solvent-solute interactions. 15 years of π^*. *J. Phys. Chem.*, 98 (1994) 5807.

Lawson,1974.
C.L. Lawson and R.J. Hanson. *Solving Least Squares Problem.* Prentice-Hall, Inc., Englewood Cliffs, N.J., (1974).

Lawton,1971.
W.H. Lawton and E.A. Sylvestre. Self modeling curve resolution. *Technometrics* 13 (1971) 617.

Legget,1977.
D.J. Legget. Numerical analysis of multicomponent spectra. *Anal. Chem.* 49 (1977) 276.

Legget,1983.
D.J. Legget, S.L. Kelly, L.R. Shiue, Y.T. Wu, D. Chang and K.M. Kadish. A computational approach to the spectrophotometric determination of stability constants. II. Application to metallophorphyrin-axial ligand interactios in non-aqueous solvents. *Talanta*, 30 (1983) 579.

Li, 1991.
J. Li, Y. Zhang, A.J. Dallas and P.W. Carr. Measurement of Solute Dipolarity Polarizability and Hydrogen-Bond Acidity by Inverse Gas-Chromatography *J. Chromatogr.* 550 (1991) 101.

Li, 1995.
Li-ZB Rutan-SC. Title: Solvatochromic Studies of the Surface Polarity of Silica Under Normal-Phase Conditions. *Anal. Chim. Acta*, 312 (1995) 127.

Lowrey,1995.
A.H. Lowrey, C.J. Cramer, J.J. Urban and G.R. Famini. Quantum chemical descriptors for linear solvation energy relationships. *Computers Chem.* 19 (1995) 209.

Lu, 1996.
Lu-HY Rutan-SC. Title: Solvatochromic Studies on Reversed-Phase Liquid-Chromatographic Phases .1. Synthesis and Solvent Effects of a New Merocyanine Dye, 4,6-Dichloro-2-(2-(1-Methyl-4-Pyridinio)Vinyl)-Phenolate. *Anal. Chem.*, 68 (1996) 1381.

Maeder, 1986.
M. Maeder and A.D. Zuberbühler. The Resolution of Overlapping Chromatographic Peaks by Evolving Factor-Analysis. *Anal. Chim. Acta* 181 (1986) 287.

Maeder, 1987.
M. Maeder. Evolving Factor-Analysis for the Resolution of Overlapping Chromatographic Peaks. *Anal. Chem.*, 59 (1987) 527.

Maksimovic, 1974.
Z.B. Maksimovic, C. Reichardt and A. Miksa-Spiric. Determination of empirical parameters of solvent polarity E_T in binary mixtures by solvatochromic N-hydroxy-phenylpiridinium betaine dyes. Fresenius Z. *Anal. Chem.* 270 (1974) 100.

Malinowski,1991.
E.R. Malinowski. *Factor Analysis in Chemistry* (1991) 2nd. Ed. Wiley-Interscience, NY.

Malinowski,1992.
E.R. Malinowski. Window Factor Analysis: theoretical derivation and application to flow injection analysis data. *J. of Chemom.* 6 (1992) 29.

Manne, 1995.
R. Manne. On the Resolution Problem in Hyphenated Chromatography. *Chemom. and Intell. Lab. Sys.* 27 (1995) 89.

Marcus, 1989.
Y. Marcus. Preferential solvation. Part 3. Binary solvent mixtures. *J. Chem. Soc. Faraday Trans.* I, 85 (1989) 381.

Marcus, 1990.
Y. Marcus. Preferential solvation in mixed solvents. Part 5. Binary mixtures of water and organic solvents. *J. Chem. Soc. Faraday Trans.* 86 (1990) 2215.

Marcus, 1994a.
Y. Marcus. Use of chemical probes for the characterization of solvent mixtures. Part 1. Completely non-aqueous mixtures. *J. Chem. Soc. Faraday Trans.*, 2 (1994) 1015.

Maria,1985.
P.C. Maria and J.F. Gal. A Lewis basicity scale for nonprotogenic solvents enthalpies of complex formation with boron trifluoride in dichloromethane. *J. Phys. Chem.* 89 (1985) 1296.

Mentschutkin, 1890.
N. Mentschutkin. *Z. Phys. Chem.* 5 (1890) 589.

Migron, 1991.
Y. Migron and Y. Marcus. Polarity and hydrogen-bonding ability of some binary aqueous-organic mixtures. *J. Chem. Soc. Faraday Trans.* 87 (1991) 1339.

Mohr, 1965.
S.C. Mohr, W.D. Wilk and G.H. Barrow. The association of water with bases and anions in an inert solvent. *J. Am. Chem. Soc.* 87 (1965) 3048.

Mui,1974.
K.K. Mui, W.A.E. McBryde and E. Nieboer. The stability of some metal complexes in mixed solvents. *Can. J. Chem.*, 52 (1974) 1821.

Murray, 1994.
J.S. Murray, T. Brinck, P. Lane, K. Paulsen and P. Politzer. Statistically-Based Interaction Indexes Derived from Molecular-Surface Electrostatic Potentials - A General Interaction Properties Function (Gipf). *J. Mol. Struct.* (Theochem) 307 (1994) 55.

Mussini, 1984.
T. Mussini, M. Cicognini, A.K. Covington P. Longhi and S. Rondinini. Standard pH Values for Potassium Hydrogenphthalate Reference Buffer Solutions in 10-Percent, 30-Percent and 50-Percent (W/W) 1,4-Dioxane Water Mixed-Solvents at Temperatures from 288.15-K to 318.15-K. *Anal. Chim. Acta*, 162 (1984) 103.

Narang, 1996.
Narang-U Zhao-CF Bhawalkar-JD Bright-FV Prasad-PN. Title: Characterization of a New Solvent-Sensitive 2-PhotonInduced Fluorescent (Aminostyryl)Pyridinium Salt Dye. *J. Phys. Chem.*, 100 (1996) 4521.

Nicolet,1986.
P. Nicolet and C. Laurence. Polarity and basicity of solvents .1. A thermosolvatochromic comparison method. *J. Chem. Soc. Perkin Trans. II*, 7 (1986) 1071.

O'Neill, 1993.
M.L. Oneill, P. Kruus and R.C. Burk. Solvatochromic Parameters and Solubilities in Supercritical-Fluid Systems. *Can. J. Chem.*, 71 (1993) 1834.

Park, 1994.
Park-JH Chae-JJ Nah-TH Jang-MD. Title: Characterization of Some Silica-Based Reversed-Phase Liquid Chromatographic Columns Based on Linear Solvation Energy Relationships. *J. Chromatog.* A, 664 (1994) 149.

Patnaik, 1994.
Patnaik-LN Sahu-B. Title: Solvent Effects on the Absorption-Spectra of Merocyanine Dyes *J. Sol. Chem.*, 23 (1994) 1317.

Pethig,1992.
R. Pethig. Protein-water interactions determined by dielectric methods. *Annu. Rev. Phys. Chem.*, 43 (1992) 177.

Politzer, 1994.
P. Politzer, J.S. Murray, Ed. *Quantitative treatments of solute/sovent interactions.* (1994) Elsevier.

Pullman, 1966.
B. Pullman, P. Claverie and J. Caillet. Van der Waals-London interactions and the configuration of hydrogen-bonded purine and pyrimidine pairs. *Proc. Nat. Acad. Sci.* 55 (1966) 904.

Pytela,1988.
O. Pytela. Empirical approach to description of solvent effect on processes in solutions. A review. *Coll. Czech. Chem. Comm.* 53 (1988) 1333.

Ramirez, 1995.
Ramirez-CB Carrasco-N Rezende-MC. Title: Halochromism and Solvatochromism of a Pi-Asterisk Probe in Binary Solvent Mixtures. *J. Chem. Soc. Faraday Trans.*, 91 (1995) 3839.

Reichardt, 1995.
Reichardt-C Harms-K Kinzel-M Schafer-G Stein-J Wocadlo-S. Title: Chiral Polymethine Dyes .4. Synthesis, AbsoluteConfiguration, Spectroscopic and Chiroptical Properties of Chiral Trimethinium and Pentamethinium Cyanine Dyes with 1,2,3,4-Tetrahydro-6-Methylquinolyl End-Groups. *Liebigs Ann. Chem.*, 2 (1995) 317.

Reichardt,1982.
C. Reichardt. Solvent effects on chemical-reactivity. *Pure and Appl. Chem.* 54 (1982) 1867.

Reichardt,1983.
C. Reichardt and E. Harbusch-Görnert. Pyridinium N-phenoxide betaines and their application for the characterization of solvent polarities .10. Extension, correction, and new definition of the e_t solvent polarity scale by application of a lipophilic penta-tert-butyl-substituted pyridinium N-phenoxide betaine dye. *Lieb. Ann. Chem.* 5 (1983) 721.

Reichardt, 1988.
C. Reichardt. *Solvent and solvent effects in organic chemistry.* VCH Weinheim (Germany)

Reichardt, 1993.
C. Reichardt, S. Asharin-Fard, A. Blum, M. Eschner, A.M. Mehranpour, P. Milart, T. Niem, G. Schäfer and M. Wilk. Solute Solvent Interactions and Their Empirical Determination by Means of Solvatochromic Dyes. *Pure and Appl. Chem.* 65 (1993) 2593.

Reichardt,1994.
C. Reichardt. Solvatochromic dyes as solvent polarity indicators. *Chem. Rev.* 94 (1994) 2319.

Richert, 1993.
R. Richert and A. Wagener. Spectroscopic Line-Shapes in Polar Supercooled Liquids. *J. Phys. Chem.* 97 (1993) 3146.

Rifkind, 1976.
J.M. Rifkind, Y.A. Shin, J.M. Heim and G.L. Eichhorn. Cooperative disordering of single-stranded polynucleotides through copper crosslinking. *Biopolymers*, 15 (1976) 1879.

Rodley, 1976.
G.A. Rodley, R.S. Scobie, R.H. Bates and R.M. Lewitt. A possible conformation for double-stranded polynucleotides. *Proc. Natl. Acad. Sci.* 73 (1976) 2959.

Rosés 1993.
M. Rosés and E. Bosch. Linear Solvation Energy Relationships in Reversed-Phase Liquid-Chromatography - Prediction of Retention from a Single Solvent and a Single Solute Parameter. *Anal. Chim. Acta*, 274 (1993) 147.

Rousseeuw, 1981.
P.J. Rousseeuw and A.M. Leroy. *Robust regression and outlier detection*, (1981) Wiley, NY.

Saenger, 1984.
W. Saenger. *Principles of nucleic acid structure.* C.R. Cantor ed. (1984) Ed. Springer-Verlag, NY.

Sánchez,1986.
E. Sánchez and B.R. Kowalski. Generalized Rank Annihilation Factor-Analysis. *Anal. Chem.* 58 (1986) 496.

Sánchez,1990.
E. Sánchez. Tensorial resolution: a direct trilinear decomposition. *J. of Chemom.* 4 (1990) 29.

Saurina,1995.
J. Saurina, S. Hernández-Cassou and R. Tauler. Multivariate Curve Resolution applied to continuous-flow spectrophotometric titrations. Reaction between amino acids and 1,2-naphtoquinone-4-sulfonic acid. *Anal. Chem.* 67 (1995) 3722.

Scremin, 1994.
Scremin-M Zanotto-SP Machado-VG Rezende-MC. Title: Preferential Solvation of a Beta-Sensitive Dye in BinaryMixtures of a Non-Protic and a Hydroxylic Solvent. *J. Chem. Soc. Faraday Trans.*, 90 (1994) 865.

Scheraga, 1978.
H.A. Scheraga. Interactions in aqueous solution. *Acc. Chem. Res.* 12 (1978) 7.

Schulte, 1995.
Schulte-RD Kauffman-JF. Title: Fluorescence from the Twisted Intramolecular Charge-Transfer Compound bis(4,4'-Dimethylaminophenyl)Sulfone in Ethanol CO2 - A Probe of Local Solvent composition. *App. Spectrosc.*, 49 (1995) 31.

Schweizer, 1971.
M.P. Schweizer, J.T. Witkowski and R.K. Robins. Nuclear magnetic resonance determination of *syn* and *anti* conformation in pyrimidine nucleosides. *J. Amer. Chem. Soc.*, 93 (1971) 277.

Shin, 1973.
Y.A. Shin. Interaction of metal ions with polynucleotides and related compounds. XXII. Effect of divalent metal ions on the conformational changes of polynucleotides. *Biopolymers* 12 (1973) 2459.

Sigel,1985.
H. Sigel, R.B. Martin, R. Tribolet, U.K. Häring and R. Malini-Balakrishnan. An estimation of the equivalent solution dielectric constant in the active-site cavity of metalloenzymes. *Eur. J. Biochem.*, 152 (1985) 187.

Sigel,1993.
H. Sigel. Interactions of metal ions with nucleotides and nucleic acids and their constituents. *Chem. Soc. Rev.* (1993) 255.

Sjöström,1981.
M. Sjöström and S. Wold.Linear free-energy relationships. Local empirical rules or fundamental laws of chemistry? *Acta Chem. Scand. Ser. B* 35 (1981) 537.

Skwierczynski, 1994.
R.D. Skwierczynski and K.A. Connors. Solvent effects on chemical processes. Part 7. Quantitative description of the composition dependence of the solvent polarity measure $E_T(30)$ in binary aqueous-organic solvent mixtures. *J. Chem. Soc. Perkin Trans.* 2 (1994) 467.

Smilde, 1994. A.K. Smilde, R. Tauler, H.M. Henshaw, L.W. Burgess and B.R. Kowalski. Multicomponent determination of chlorinated hydrocarbons using a reaction-based chemical sensor. 3. Medium rank second order calibration with restricted Tucker models. *Anal. Chem.*, 66 (1994) 3345.

Spange, 1996.
Spange-S Reuter-A Vilsmeier-E. Title: On the Determination of Polarity Parameters of Silica by Means of Solvatochromic Probe Dyes. *Coll. Polym. Sci.*, 274 (1996) 59.

Suhaldonik, 1970.
R.J. Suhaldonik. *Nucleoside antibiotics.* (1970) Wiley, NY.

Suhaldonik, 1979.
R.J Suhaldonik. *Nucleosides as biological probes.*(1979) Wiley, NY.

Sun, 1995.
Sun-YP Bunker-CE Title: Solute and Solvent Dependencies of Intermolecular Interactions in Different Density Regions in Supercritical Fluids - A Generalization of the 3-Density-Region Solvation Mechanism. *Phys. Chem.*, 99 (1995) 976.

Sutton, 1965.
Tables of Interatomic distances and configuration in molecules and ions. L.E. Sutton ed. (1965) Chem. Soc., Burlington House, London.

Swain, 1955.
C.G. Swain, R.B. Mosely and D.E. Bown. Correlation of rates of solvolysis with a four-parameter equation. *J. Am. Chem. Soc.* 77 (1955) 3731.

Swain,1983.
C.G. Swain, S. Alunni, A.L. Powell and M.S. Swain. Solvent effects on chemical reactivity. evaluation of anion and cation solvation components. *J. Am. Chem. Soc.* 105 (1983) 502.

Taft,1976.
R.W. Taft and M.J. Kamlet. The solvatochromic comparison method. 2. The α-scale of solvent hydrogen bond donor (HBD) acidities. *J. Am. Chem. Soc.*, 98 (1976) 2886.

Taft,1985.
R.W. Taft, M.J. Kamlet, J.L.M. Abboud and M.H. Abraham. Linear Solvation Energy Relationships. *J. Sol. Chem.*, 14 (1985) 153.

Tanford, 1978.
C. Tanford. The hydrophobic effect and the organization of living matter. *Science*, 200 (1978) 1012.

Tauler,1993.
R. Tauler, B.R. Kowalski and S. Fleming. Multivariate curve resolution applied to spectral data from multiple runs of an industrial process. *Anal. Chem.,*. 65 (1993) 2040.

Tauler,1995.
R. Tauler, A. Smilde and B.R. Kowalski. Selectivity, local rank, three-way data analysis and ambiguity in multivariate curve resolution. *J. of Chemom.* 9 (1995) 31.

Tauler,1995b.
R. Tauler. Multivariate curve resolution applied to second order data. *Chemom. and Intell. Lab. Sys.* 30 (1995) 133.

Taylor, 1982.
R. Taylor and O. Kennard. The molecular structures of nucleosides and nucleotides. 1. The influence of protonation on the geometries of nucleic acid constituents. *J. Mol. Struct.* 78 (1982) 1.

Thang, 1992.
M.N. Thang and W. Guschlbauer. Biologic Actions of Double-Stranded Polyribonucleotides and Potential Therapeutic Uses. A Reappraisal. *Pathologie Biologie,*.40 (1992) 1006.

Toft, 1993 .
J. Toft and O.M. Kvalheim. Eigenstructure Tracking Analysis for Revealing Noise Pattern and Local Rank in Instrumental Profiles - Application to Transmittance and Absorbency Ir Spectroscopy. *Chem. Intell. Lab. Sys.*, 19 (1993) 65.

Ts'o, 1963.
P.O.P. Ts'o, I.S. Melvin and A.C. Olson. Interaction and association of bases and nucleosides in aqueous solution. *J. Amer. Chem. Soc.* 85 (1963) 1289.

Tunuli,1984.
M.S. Tunuli and M.A. Farhataziz Rauf. Dimroth $E_T(30)$ as parameter of solvent polarity. A caveat. *J. Photochem.* 24 (1984) 411.

Van Uitert,1952.
L.G. Van Uitert and C.G. Haas. Studies on coordination compounds. I. A method for determining thermodynamic equilibrium constants in mixed solvents. *J. Am. Chem. Soc.* 75 (1952) 451.

Vandeginste,1985.
B.G.M. Vandeginste, W. Derks and G. Kateman. Multicomponent self-modelling curve resolution in high-performance liquid chromatography by iterative transformation factor analysis. *Anal. Chim. Acta* 173 (1985) 253.

Walker, 1979.
Nucleoside Analogues, Chemistry, Biology and Medical Applications. R.T. Walker, E. de Clerck and F. Eckstein ed. (1979) Plenum Press, NY.

Weast, 1976.
Handbook of Chemistry and Physics. R.C. Weast ed. (1976) F-215. CRC Press Cleveland, Ohio.

Windig,1988.
W. Windig. Mixture analysis of spectral data by multivariate methods. *Chemom. and Intell. Lab. Sys.* 4 (1988) 201.

Windig,1991.
Interactive self-modeling mixture analysis. *Anal. Chem.* 63 (1991) 1425.

Wislicenus, 1896.
W. Wislicenus. *Liebigs Ann. Chem.* 291 (1896) 147.

Wrona, 1991.
P.K. Wrona, T.M. Krygowski and Z. Galus. Correlation between empirical Lewis acid-base solvent parameters and the thermodinamic parameters of ion solvation. 2. Acidity parameters of cations and basicity parameters of anions. *J. Phys. Org. Chem.* 4 (1991) 439.

Yathindra, 1974.
N. Yathindra and M. Sundaralingam. Backbone conformations in secondary and tertiary structural units of nucleic acids. Constraint in the phosphodiester conformation. *Proc. Nat. Acad. Sci.*, 71 (1974) 3325.

V. SUMMARY IN CATALAN

Objectius

L'objectiu últim d'aquesta memòria és la descripció qualitativa i quantitativa de totes les transicions termodinàmiques i conformacionals lligades als equilibris àcid-base d'alguns polinucleòtids en ambients biològics de polaritat baixa. L'emulació d'aquests ambient especials s'ha dut a terme fent ús de mescles aigua-dioxà que mantenen la natura aquosa dels medis biològics i al mateix temps presenten la baixa polaritat desitjada gràcies a les propietats del cosolvent orgànic.

La complexitat associada al caràcter macromolecular dels polinucleòtids i al fet d'utilitzar dissolvents mixtos ha aconsellat la realització de certs estudis fonamentals abans de tractar específicament el problema dels equilibris àcid-base dels polinucleòtids en mescles aigua-dioxà. Aquests estudis han inclòs, d'una banda, la caracterització detallada de les mescles aigua-dioxà i la interpretació de llur efecte sobre els equilibris àcid-base de soluts senzills i, d'altra banda, l'estudi de les unitats monòmeres dels polinucleòtids en aquestes mescles.

Un punt constant d'interès en tot el treball que s'ha realitzat ha estat l'assoliment de la màxima qualitat en el tractament de les dades experimentals. En aquest sentit, han merescut una atenció especial els problemes associats a l'establiment de les relacions lineals d'energia de solvatació (LSER), que són els models que descriuen l'efecte del solvent sobre el comportament del solut, i la interpretació de les dades multivariants procedents del seguiment dels equilibris dels polinucleòtids. El primer problema s'ha solucionat amb l'aplicació de diversos mètodes de modelatge dur (hard-modelling) i de modelatge tou (soft-modelling), mentre per al segon s'ha optat per l'aplicació de mètodes de resolució de corbes, que no necessiten la postulació de cap model químic per a interpretar la variació de les diferents espècies en solució.

Estructura de la Tesi Doctoral

La sèrie d'articles inclosa en aquesta memòria mostra el procés seguit en l'estudi dels equilibris àcid-base dels polinucleòtids en ambients de polaritat baixa. Aquest procés comença amb la selecció i avaluació dels descriptors del solvent que s'utilitzaran posteriorment en la caracterització de les mescles aigua-dioxà (article I), continua amb la caracterització microscòpica d'aquestes mescles hidroorgàniques i amb la descripció de llur efecte sobre els equilibris àcid-base de soluts senzills (articles II-VI) i acaba amb el seguiment multivariant i la interpretació del comportament àcid-base dels polinucleòtids en aquestes mescles (articles VII-XI).

L'article I presenta una visió general de l'evolució que ha seguit la descripció del solvent, que es podria resumir com la transició de les propietats macroscòpiques (que caracteritzen adequadament el si del solvent) al de les propietats microscòpiques (més adients per a la descripció de les propietats en la zona cibotàctica dels soluts). Els descriptors microscòpics proposats per Kamlet i col. per a quantificar les interaccions solut/solvent de polaritat i la capacitat de formació de ponts d'hidrogen han estat utilitzats per a estructurar l'espai dels solvents en grups de dissolvents construïts en funció de la similaritat de les propietats microscòpiques de llurs membres.

Els espais lliures entre els grups de solvents purs poden ésser omplerts amb la infinitud de possibles mescles de solvents. El sistema aigua-dioxà cobreix una àmplia zona de l'espai de solvents i ha estat seleccionat per a tots els estudis posteriors duts a terme en aquesta memòria. Els articles II, III i IV inclouen la determinació experimental de les propietats microscòpiques de mescles hidroorgàniques que cobreixen tot el rang de composicions que va del 0 al 100% (v/v) de dioxà i la determinació de les constants de protonació de certs soluts senzills dissolts en les mescles ja caracteritzades. Es presenten i s'interpreten algunes correlacions simples entre sèries de paràmetres microscòpics del solvent, entre aquests paràmetres i la composició de les mescles i entre les constants de protonació i els descriptors dels solvents. Els articles V i VI estan orientats a l'establiment de models de comportament més acurats per tal d'explicar l'efecte del solvent sobre la

propietat del solut en estudi. S'han proposat dues possibles alternatives per a la construcció dels models LSER: la primera inclou una estratègia robusta que ajusta el model químic postulat a les dades experimentals (modelatge dur) i la segona, que combina l'ús de l'anàlisi de Factors (FA) i el Target Factor Analysis (TFA), transforma un model abstracte (FA) en una LSER (TFA) sense l'aplicació de cap expressió química externa (modelatge tou).

Després de la recerca fonamental descrita en els articles I-VI, ja es pot procedir a l'estudi dels equilibris àcid-base dels polinucleòtids en mescles aigua-dioxà. El tractament de les dades multivariants provinents d'aquests experiments no és una tasca senzilla i per aquest motiu s'han realitzat alguns estudis dedicats al coneixement i millora del mètode de resolució de corbes més comunament utilitzat en treballs anàlegs duts a terme en solució aquosa, el mètode de mínims quadrats alternats (ALS). La millora d'aquest mètode s'ha centrat en la proposta i avaluació de noves restriccions incloses en el procés de resolució iteratiu (article VII). El funcionament d'aquest mètode s'ha comparat amb el d'un altre mètode de resolució de dades tridireccionals, la Descomposició Trilineal (TLD), de fonaments teòrics clarament diferents (article VIII). Dades simulades i dades reals han estat utilitzades en ambdós treballs. Un cop s'ha confirmat la idoneïtat del mètode ALS per al tractament de les dades procedents dels estudis de polinucleòtids, els articles IX-XI descriuen detalladament les transicions termodinàmiques i conformacionals relacionades amb els equilibris àcid-base de l'àcid poliuridílic (polyU), de l'àcid policitidílic (polyC) i de l'àcid poliadenílic (polyA), i comparen aquests processos amb llurs anàlegs en solució aquosa. Els articles IX i X s'orienten específicament a l'estudi dels sistemes polyU-H i polyC-H, respectivament, i al de llurs nucleòtids cíclics. L'article XI compila els resultats que es mostren als articles IX i X i els compara amb els obtinguts per al sistema polyA-H. Aquest darrer article posa èmfasi en el gran potencial que té l'ús combinat del seguiment multivariant experimental i de l'aplicació de mètodes de resolució de corbes per a l'estudi dels equilibris biomacromoleculars i exposa sistemàticament tots els tipus d'informació que proporciona l'ús d'aquesta estratègia de recerca.

Resum general

El concepte de solvent ha anat evolucionant amb el decurs del temps. Si en un principi era considerat com un ens continu que podia ésser descrit amb propietats macroscòpiques, posteriorment es va comprovar que l'estructura que adoptava el solvent al voltant dels soluts, amb els quals establia certes interaccions específiques, era força diferenciada de l'estructura que posseïa en el si de la solució. Aquesta constatació va mostrar que les propietats macroscòpiques no eren suficients per a fer una correcta descripció del solvent i calia incloure descriptors de natura microscòpica, que donessin informació sobre les característiques d'aquests composts al voltant dels soluts.

La major part de paràmetres microscòpics són de natura empírica i s'han construït a partir de mesures procedents de processos que experimenten soluts de referència i que depenen del solvent. La variació d'aquestes mesures, que poden ésser termodinàmiques, cinètiques o espectroscòpiques, en modificar el solvent que envolta el solut, es fa servir com a base per a construir les escales empíriques de paràmetres microscòpics del solvent. Segons la natura del solut de referència (sovint anomenat indicador), la propietat que se'n mesura pot ésser sensible a un o a diversos tipus d'interacció solut/solvent. Alguns autors, entre ells Reichardt, han preferit seleccionar un solut de referència (el 2,6-difenil-4-(2,4,6-trifenil-1-piridini)fenolat) capaç d'establir diversos tipus d'interacció amb el solut. L'escala que se'n deriva, $E_T(30)$, es construeix a partir dels desplaçaments espectrals que sofreix el màxim de la banda d'absorció del solut en funció del solvent i és proposada pel seu autor com una mesura de polaritat (en aquest context, la polaritat es defineix com la capacitat global del solvent d'interaccionar amb el solut). D'altres autors, com Kamlet i col., han preferit proposar una sèrie de paràmetres, en què cadascun d'ells sigui sensible a un únic tipus d'interacció específica. D'aquesta manera, per a la descripció microscòpica del solvent han proposat els paràmetres α, β i π^*, que descriuen l'acidesa per formació de ponts d'hidrogen, la basicitat per formació de ponts d'hidrogen i la polaritat-polaritzabilitat, respectivament.

Per tal d'entendre més profundament el que suposa l'evolució des de la perspectiva macroscòpica a la microscòpica en la descripció dels solvents, el primer punt d'aquest treball ha consistit en fer un estudi de l'estructura de l'espai global de solvents a partir de les similituds que aquests composts mostren quant a les seves característiques microscòpiques.

A l'article I es proposa un esquema de classificació de solvents basat en la semblança de la intensitat i la natura de les interaccions que poden exercir aquests composts amb els soluts. Per tal de dur a terme aquesta tasca, un conjunt representatiu de solvents, descrits amb llurs corresponents paràmetres α, β i π^*, s'ha estructurat en grups fent ús de diferents tècniques d'anàlisi d'agrupacions (cluster analysis).

L'espai microscòpic dels solvents ha estat definit amb els paràmetres α, β i π^* degut a la independència de llurs significats químics i a la manca de correlació matemàtica que presenten entre ells. No s'han utilitzat altres descriptors microscòpics, ja que la majoria poden ésser expressats com a combinacions lineals dels paràmetres prèviament esmentats. La introducció de biaix en la formació dels grups s'ha evitat utilitzant diversos procediments d'agrupació que, tot i tenir diferents fonaments teòrics, comparteixen el fet de basar llurs criteris d'agrupació en característiques dels grups o de les regions de l'espai i no en propietats associades a objectes (solvents) individuals.

La classificació proposada estructura els solvents en els sis grups següents:

- **I:** solvents lleugerament bàsics de baixa polaritat (èters alifàtics i amines alifàtiques substituïdes).
- **II:** solvents polars apròtics relativament bàsics i moderadament o altament polars (èters cíclics alifàtics, esters, cetones i nitrils).
- **III:** solvents fortament bàsics i fortament polars (piridines, amines petites, sulfòxids, urees i fosforamides).
- **IV:** solvents relativament polars amb tendència general baixa a formar ponts d'hidrogen. Un següent nivell de separació separa aquests solvents entre els que tenen una capacitat de donació de ponts d'hidrogen moderada i la resta, i així s'obtenen els subgrups A (èters aromàtics, hidrocarburs i altres espècies halogenades i hidrocarburs alifàtics

halogenats poc polars) i B (hidrocarburs alifàtics polars polihalogenats, els heteroàtoms dels quals exerceixen l'efecte inductiu responsable de la formació de ponts d'hidrogen per donació de protons).

- **V:** solvents amfipròtics (aigua i alcohols) amb notable capacitat de formació de ponts d'hidrogen. Separacions posteriors separen aigua i glicol de la resta d'alcohols degut a la seva major capacitat d'autoassociació.
- **VI:** hidrocarburs alifàtics que no exerceixen interaccions específiques amb els soluts; és a dir, composts amb α, β i π^* iguals a zero.

La basicitat de ponts d'hidrogen i la capacitat global d'interacció del solvent són les dues característiques més diferenciadores en l'esquema de solvents proposat. La diversitat interna dels grups quant a la mida, grups funcionals i estructura espacial dels solvents que els composen és probablement una de les propietats més interessants d'aquesta classificació. Aquesta diversitat permet descobrir analogies entre solvents relacionades amb el comportament microscòpic que han estat classificats tradicionalment en grups diferents. Aquest fet obre un nou camí en el procés de selecció de solvents, ja que problemes que es presenten en l'ús de determinats solvents a causa d'impediments estèrics o d'interaccions específiques desenvolupades per certs grups funcionals poden ésser superats degut a la varietat de solvents dins un mateix grup que interaccionen similarment amb els soluts.

Aquesta classificació de solvents s'ha comparat amb l'aproximació de Snyder, una classificació de solvents anterior a la proposada que s'ha utilitzat àmpliament en cromatografia i que treballa amb descriptors de solvents anàlegs als emprats en l'esquema d'aquesta memòria. Els descriptors microscòpics proposats per Snyder són els logaritmes corregits de les constants de distribució gas-líquid del dioxà, $(\log(K_g'')_d)$, de l'etanol, $(\log(K_g'')_e)$, i del nitrometà, $(\log(K_g'')_n)$, que representen les interaccions del solvent degudes a donació de ponts d'hidrogen, a acceptació de ponts d'hidrogen i les degudes a polaritat, respectivament. Malgrat l'aparent similitud entre els dos esquemes, l'estructura existent en l'espai tridimensional definit per α, β i π^* no s'aprecia en l'espai de Snyder, on els solvents es disposen formant un núvol continu de punts. La manca d'estructura de l'espai tridimensional de Snyder, unida al caràcter mixt dels descriptors de Snyder

relacionats amb les contribucions de formació de ponts d'hidrogen (en realitat, $(\log(K_g'')_d)$ i $(\log(K_g'')_e)$ són combinacions lineals de α i π^* i de α, β i π^*, respectivament) recomanen la utilització de l'esquema de classificació proposat en aquesta memòria quan la selecció del solvent s'ha de realitzar d'acord amb la intensitat i tipus d'interaccions que el solvent pot exercir sobre els soluts.

Dels solvents purs a les mescles de solvents: l'exemple aigua-dioxà.

L'estructura de l'espai microscòpic de solvents purs no és contínua. Per tant, el procés de selecció de solvents restaria certament limitat si no existissin les mescles de solvents. De fet, són aquests solvents mixts els qui omplen els buits presents a l'espai de solvents purs i els qui proporcionen al químic el medi adient per al procés en estudi.

Caracterització microscòpica del sistema aigua-dioxà i estudi de processos àcid-base.

La combinació aigua-dioxà és especialment interessant degut a la natura extremadament diferent dels dos solvents que la formen i a la gran varietat de mescles que se'n poden obtenir en variar la proporció dels constituents. Aquesta varietat es reflecteix en els valors de les propietats macroscòpiques de les mescles i també en els valors de les propietats microscòpiques, determinats en aquesta tesi (veure articles II, III, IV i VI).

Els valors del paràmetre π^* vénen de la mitjana dels resultats obtinguts per a aquest paràmetre amb els indicadors solvatocròmics 2-nitroanisol, 4-nitroanisol i 4-etilnitrobenzè, tots ells recomanats per a la determinació del paràmetre π^* en solvents amfipròtics. Els valors concordants obtinguts amb aquests soluts de referència indiquen l'absència de diferències significatives de solvatació degudes a la natura dels indicadors i confirmen la validesa del paràmetre π^* per a mesurar la polaritat de les mescles aigua-dioxà. Els valors de π^* procedents de la N-metil-2-nitroanilina no s'han inclòs en la mitjana perquè el desplaçament espectral d'aquest indicador és causat, en part, per interaccions bàsiques del solvent.

La basicitat per ponts d'hidrogen de les mescles aigua-dioxà s'ha mesurat amb els paràmetres β_1, determinat amb el parell de soluts format per la 4-nitroanilina i la N,N-dietil-4-nitroanilina, i β_2, determinat amb el parell de soluts 4-nitoanisol i 4-nitrofenol. No s'han calculat valors promig del paràmetre β degut a la dispersió entre els resultats de les dues sèries. Malgrat aquesta dispersió, que és sempre superior a la que es troba en la determinació del paràmetre π^* en qualsevol sistema de solvents, hi ha una clara correlació entre els valors de les dues sèries, és a dir, ambdues mostren el mateix tipus de variació amb la composició de la mescla. Aquesta relació fa que ambdues sèries duguin a la mateixa conclusió pel que fa a la significació de la basicitat del solvent sobre el procés d'un solut i, per tant, qualsevulla d'elles és vàlida com a mesura de la basicitat de les mescles aigua-dioxà.

Els valors de α i $E_T(30)$ han estat obtinguts amb un únic parell de soluts (4-nitroanisol i la betaïna de Reichardt) i un únic solut (la betaïna proposada per Reichardt), respectivament. De tota manera, gràcies a la menor intensitat dels fenòmens de solvatació preferencial en mescles hidroorgàniques, aquests paràmetres poden ésser considerats, al menys, com a bones aproximacions de l'acidesa per ponts d'hidrogen (α) i la capacitat global de solvatació ($E_T(30)$).

Els valors dels paràmetres α, π^* i $E_T(30)$ segueixen una tendència descendent quan la proporció de dioxà augmenta. Aquest comportament és lògic, ja que l'aigua és clarament més polar que el dioxà i, a més, aquest solvent orgànic no té capacitat per a formar ponts d'hidrogen per donació de protons. Una evolució radicalment diferent s'observa en ambdues sèries β_i, en què els valors de basicitat del solvent en les mescles arriben a ésser superiors als valors d'aquesta propietat en aigua pura o en dioxà pur. Aquest efecte sinèrgic podria ésser explicat per la major basicitat dels complexos aigua-dioxà comparada amb la dels solvents purs, per la interacció bàsica més forta exercida per les molècules d'aigua alliberades de la xarxa tridimensional pròpia d'aquest solvent, o per una combinació dels dos factors anteriors.

La relació no lineal entre els valors de $E_T(30)$ i π^* en mescles aigua-dioxà confirma la diferència entre el concepte de polaritat de Reichardt, definit com la capacitat global d'interacció del solvent, i el de Kamlet i col., que considera únicament les interaccions microscòpiques governades per afinitat de dipols entre solvent i solut i exclou tot fenomen de formació de ponts d'hidrogen. La natura mixta del paràmetre $E_T(30)$, sensible a la polaritat/polaritzabilitat del solvent i a l'acidesa per ponts d'hidrogen, es confirma clarament amb el model que defineix el paràmetre $E_T(30)$ com a combinació lineal de α i π^*. Les equacions relacionades amb aquest model presenten ajusts bons tant si s'obtenen amb el valor promig del paràmetre π^* com si es construeixen fent ús de qualsevol de les sèries de valors que vénen d'un dels indicadors acceptats. Les contribucions de α i π^* en el model establert en mescles aigua-dioxà concorden amb expressions anàlogues construïdes a partir de valors d'aquests paràmetres en altres solvents purs i mixtos.

El fenomen de solvatació preferencial en aigua-dioxà s'ha estudiat examinant els canvis dels valors d'$E_T(30)$ amb la composició de la mescla. Segons aquest procediment empíric, qualsevol desviació dels valors experimentals d'$E_T(30)$ del model de comportament additiu $E_T(30)_{teòric} = x_{aigua}\, E_T^o(30)_{aigua} + x_{dioxà}\, E_T^o(30)_{dioxà}$ pot ésser atribuïda a la presència de solvatació preferencial. Donada una fracció molar x_{aigua}, valors d'$E_T(30)$ superiors a l'$E_T(30)_{teòric}$ indiquen que l'aigua se situa preferencialment al voltant del solut i viceversa quan els valors d'$E_T(30)$ són inferiors a l'$E_T(30)_{teòric}$. El gràfic d'$E_T(30)$ enfront de x_{aigua} mostra un doble comportament. Per a fraccions molars d'aigua inferiors a 0.45 (% dioxà (v/v) > 85), l'aigua se situa preferencialment al voltant del solut, mentre per a fraccions superiors, la proporció de dioxà al voltant del solut ultrapassa el valor teòric que li correspondria en cas de complir-se el model additiu de comportament. Les diferències entre les mescles riques en aigua i les mescles riques en dioxà ja van ésser notades per Langhals, que va necessitar dues equacions diferents per tal de modelar el comportament químic relatiu a aquests dos grups de solvents. Langhals interpreta aquesta discontinuïtat com una conseqüència del canvi d'estructura de l'aigua causat per la introducció de molècules de dioxà. Així, per a fraccions molars de dioxà baixes, l'estructura tridimensional de l'aigua es manté i és el cosolvent orgànic qui es disposa preferencialment al voltant del solut; quan s'excedeix un cert valor llindar de $x_{dioxà}$, la xarxa de molècules d'aigua es trenca i les

molècules d'aigua alliberades interaccionen més fàcilment amb els soluts. El dioxà funciona com un solvent trencador d'estructures i el comportament diferent mostrat per les mescles riques en aigua i les riques en dioxà s'explica en part perquè les interaccions polars de les molècules d'aigua que formen part d'agregats moleculars són clarament més fortes que les que exerceixen les molècules d'aigua aïllades.

Després de la caracterització de les mescles aigua-dioxà, la interpretació del comportament àcid-base de soluts senzills en aquests sistemes mixtos esdevé simple. Les constants d'autoprotòlisi del solvent i algunes constants d'acidesa de soluts (grups carboxílic i fenòlic de l'àcid salicílic i grup carboxílic de l'àcid propiònic) s'han determinat potenciomètricament en el rang de composició de les mescles que va del 10 al 70% de dioxà (v/v). No s'han efectuat determinacions a composicions més riques en dioxà degut a la precipitació de l'electròlit suport en les condicions de treball. Per tal d'obtenir conclusions més generals pel que fa a la interpretació de l'efecte del solvent en els equilibris àcid-base, s'ha treballat també amb dades procedents de la bibliografia que fan referència a composts que tenen grups funcionals diferents als de les molècules estudiades experimentalment.

Les diferències entre els paràmetres macroscòpics i microscòpics de les mescles aigua-dioxà es palesen quan es comparen paràmetres anàlegs, com són ara la mesura de polaritat macroscòpica, $1/\varepsilon$, i la mesura de polaritat microscòpica, π^*. La relació que existeix entre aquests dos paràmetres no és lineal, amb la qual cosa es demostra que la variació de la polaritat amb la composició de la mescla en el si de la solució i a l'esfera de solvatació dels soluts és clarament diferent. Aquesta conclusió es confirma en constatar la relació no lineal entre la variació dels valors de pK, principalment afectats per les propietats del solvent al voltant dels soluts, i algunes propietats típicament macroscòpiques, com la inversa de la constant dielèctrica o la fracció molar d'un component de la mescla. Les desviacions de la linealitat són especialment evidents en les mescles riques en dioxà. És en aquests medis on les diferències entre propietats macroscòpiques i microscòpiques es fan més evidents a causa de la importància creixent de les interaccions específiques solut/solvent que no poden ésser descrites de forma adient amb un model clàssic que només consideri fenòmens de tipus electrostàtic. En conseqüència, la interpretació del

comportament àcid-base dels soluts en mescles aigua-dioxà necessita la introducció de paràmetres microscòpics per a ésser descrit adequadament.

Han estat estudiades algunes relacions univariants simples entre cadascuna de les sèries de pK i cadascun dels paràmetres microscòpics determinats (α, β, π^* i $E_T(30)$), com a pas previ per a l'establiment de LSERs que descriguin l'efecte del solvent sobre els equilibris àcid-base de les mescles aigua-dioxà. El canvi de l'estructura interna del solvent provocat per la introducció de molècules de dioxà s'observa en els gràfics de pK enfront d'$E_T(30)$, on s'obtenen dues zones amb comportaments lineals diferents. Aquesta mateixa conclusió s'obté quan els valors de pK es relacionen amb altres paràmetres microscòpics relacionats amb la formació de ponts d'hidrogen, com són ara α i β. Contràriament, un únic model lineal és necessari per tal d'explicar la relació entre la variació de pK i el paràmetre π^* de polaritat en tot el rang de composicions de les mescles. D'aquestes relacions univariants, es pot concloure que el trencament de l'estructura de l'aigua afecta més fortament la capacitat d'interacció del solvent per ponts d'hidrogen amb el solut que no pas la capacitat de la mescla d'establir interaccions basades únicament entre l'afinitat de dipols entre solvent i solut.

Als articles II i IV, el model lineal $pK_i = pK_i^o + s\pi^*$, on pK_i^o és el valor del pK per al solut *i* en un solvent inert hipotètic, s'ha pres com a model bàsic per a l'establiment de LSERs vàlides en tot el rang de composicions de les mescles. No s'han detectat millores importants quan s'han inclòs al model altres paràmetres solvatocròmics, com α i β.

Modelatge de processos dependents del solvent en mescles aigua-dioxà: propostes per a l'establiment de relacions lineals d'energia de solvatació (LSER).

Les conclusions dels articles II i IV semblen interpretar l'efecte del solvent sobre la variació de les constants d'acidesa dels soluts en aigua-dioxà com un procés controlat per les interaccions polars entre solvent i solut. No obstant això, la necessitat de confirmar d'una manera més rigorosa la validesa dels models de comportament obtinguts ha conduït a proposar dos procediments generals orientats a ajudar a l'establiment correcte de LSERs:

un procediment de modelatge dur que treballa ajustant un model químic a les dades experimentals, i un procediment de modelatge tou, que estableix els models de comportament sense el suport de cap expressió externa. Els articles V i VI descriuen l'aplicació d'aquests dos mètodes a dades experimentals.

De diferents maneres, ambdós procediments determinen primerament la mida correcta del model LSER, és a dir, el nombre de termes que ha d'incloure, i posteriorment identifiquen la natura química dels termes que hi han d'ésser presents. La necessitat d'identificar les propietats dels solvent que afecten la propietat del solut en estudi és una tasca evident per tal d'interpretar el procés dependent del solvent. Menys evident, però igualment important, és la determinació prèvia de la mida del model LSER. Models amb un nombre de termes menor del necessari no expliquen de forma acurada les variacions de la propietat del solut, mentre que models amb un nombre excessiu de termes dificulten la interpretació del model (la inclusió de qualsevol terme en una LSER no implica únicament la variació de l'ajust matemàtic del model, sinó l'acceptació implícita de la significació d'aquella propietat del solvent en el procés del solut estudiat) i disminueixen la seva capacitat predictiva (els termes poc importants d'un model sovint descriuen relacions específiques entre els soluts i els solvents concrets que s'han utilitzat per a construir el model i que poden no existir entre altres soluts i solvents).

Comparat amb altres procediments de regressió clàssica, les principals contribucions del procediment de modelatge dur proposat són la selecció prèvia de la mida i la natura dels paràmetres que s'inclouran en el model LSER mitjançant una regressió seqüencial (stepwise regression) i la supressió de la influència negativa de les observacions aberrants amb l'aplicació d'un procediment de regressió robusta. D'aquesta manera, prenent com a descriptors del solvent els proposats per Kamlet i col., és avaluada la influència de cadascun d'aquests paràmetres en la capacitat explanatòria del model i són rebutjats els termes que no afegeixen informació significativa a la descripció de la propietat del solut dependent del solvent. Quan s'ha determinat l'expressió química del model, es procedeix a la detecció i eliminació d'aberrants i es construeix la LSER definitiva utilitzant només les dades experimentals que han superat els tests de detecció d'aberrants.

El procediment de modelatge tou proporciona beneficis distints dels associats als mètodes de regressió clàssics i a l'aproximació de modelatge dur proposada anteriorment. Comparteix amb aquesta darrera aproximació la capacitat de determinar la mida del model LSER i d'identificar la natura dels termes que hi han d'ésser presents. A diferència d'aquesta, però, no li calen models químics externs per tal de dur a terme cap d'aquestes dues tasques. La mida del model és determinada amb l'anàlisi de factors (FA) de la matriu de dades formada per columnes que contenen les sèries de valors de les propietats del solut dependents del solvent. FA determina quin és el model abstracte més simple per a descriure les dades; és a dir, el de mida menor format per contribucions additives descorrelacionades. Un cop sabuda la dimensió del model, s'aplica Target Factor Analysis (TFA) per tal de transformar el model abstracte en una LSER dotada de significat químic. Contràriament al procediment de modelatge dur que treballa només amb el grup de descriptors del solvent inclosos en l'expressió bàsica que s'ajusta a les dades experimentals, el principal avantatge de TFA és que qualsevol descriptor del solvent, macroscòpic o microscòpic, pot ésser utilitzat en l'establiment del model LSER. Tots ells són assajats separadament i els resultats d'aquests tests individuals ordenen els descriptors del solvent segons la seva capacitat per a modelar el conjunt de dades en estudi. En alguns casos, els tests poden descobrir que models construïts amb paràmetres microscòpics del dissolvent proposats per diferents autors, o amb combinacions de paràmetres macroscòpics i microscòpics, poden ésser més útils per a descriure les variacions d'una propietat concreta d'un solut que expressions multiparamètriques fixes que s'utilitzen en molts procediments de modelatge dur. TFA pot, doncs, descobrir la validesa de combinacions poc freqüents de paràmetres del solvent i no requereix un coneixement del problema químic tan profund com els procediments de modelatge dur.

Tant el procediment de modelatge dur com el de modelatge tou proposats tenen punts positius. Si els principals avantatges del modelatge tou són la completa llibertat en la selecció de descriptors potencials del solvent per a construir el model LSER i l'estalvi de temps de càlcul associat al càlcul simultani dels models LSER de tots els soluts inclosos en la matriu de dades, el procediment de modelatge dur proposat té una major resistència a la

influència d'observacions aberrants a l'hora d'establir les LSERs. En conseqüència, els dos procediments gaudeixen de beneficis complementaris i la concordança entre els resultats obtinguts utilitzant-los ambdós és una validació excel·lent del model LSER establert.

Les LSERs presentades en els articles V i VI estan relacionades amb l'estudi del desplaçament espectral dels indicadors utilitzats en la determinació de paràmetres microscòpics i amb la descripció del comportament àcid-base d'alguns soluts en mescles aigua-dioxà. Segons Kamlet i col., el desplaçament solvatocròmic dels indicadors utilitzats en la determinació dels paràmetres α, β i π^* pot ésser descrit amb l'expressió general:

$$\nu_{max} = \nu_o + a\alpha + b\beta + s\pi^*$$

on ν_{max} és el nombre d'ona relacionat amb el màxim d'absorció de l'indicador i ν_o és el nombre d'ona que podria ésser trobat en un hipotètic solvent inert. Aquesta expressió pot ésser reduïda si l'indicador en estudi és sensible només a una o dues de les propietats del solvent α, β i π^*.

A l'article V sorgeixen dues qüestions relacionades amb aquest tema:

a) Són les propietats α, β i π^* suficients per a descriure el desplaçament espectral de tots els indicadors analitzats? i,

b) El desplaçament espectral de cadascun dels indicadors analitzats és sensible només a les propietats del solvent que se suposa que mesura?

S'ha aplicat el procediment de modelatge tou s'ha aplicat per tal de respondre aquestes dues qüestions. El conjunt de dades analitzat està format per una matriu, les columnes de la qual contenen els valors de ν_{max} de cadascun dels indicadors a diferents composicions de la mescla aigua-dioxà. Els indicadors analitzats han estat 2-nitroanisol, 4-nitroanisol, 4-etilnitrobenzè i N-metil-2-nitroanilina (proposats per a mesurar la polaritat del solvent), la betaïna de Reichardt (suposadament sensible a la polaritat del solvent i a l'acidesa per formació de ponts d'hidrogen) i 4-nitrofenol i 4-nitroanilina (proposats com a soluts sensibles a la polaritat del solvent i a la basicitat per formació de ponts d'hidrogen).

L'anàlisi de factors de la matriu de dades anterior dóna com a resultat un model amb tres factors abstractes; per tant, no calen més de tres contribucions per tal d'explicar el desplaçament espectral dels soluts analitzats. Les etapes següents consisteixen en confirmar α, β i π^* com les tres contribucions reals relacionades amb els desplaçaments espectrals dels indicadors i en l'examen individual dels termes necessaris per a descriure el desplaçament espectral de cadascun dels soluts analitzats. Com era d'esperar, després de l'aplicació de TFA, han estat acceptats α, β i π^* com els paràmetres més adequats per a la descripció dels desplaçaments espectrals. Aquest resultat és lògic, ja que els tres paràmetres solvatocròmics es calculen com a combinacions lineals de les ν_{max} dels diferents indicadors. No obstant això, el punt que continua sense ésser evident és quins d'aquests paràmetres tenen un pes més important en la descripció del ν_{max} relacionat amb cadascun dels indicadors. Els indicadors solvatocròmics 2-nitroanisol, 4-nitroanisol i 4-etilnitrobenzè només van mostrar loadings grans (és a dir, coeficients alts) associats al paràmetre π^*. En canvi, la N-metil-2-nitroanilina, que va ésser rebutjada per a la determinació del paràmetre π^*, té loadings significatius associats a les interaccions del solvent per formació de ponts d'hidrogen. En concordança amb els postulats de Kamlet i col., s'ha confirmat la sensibilitat de la betaïna de Reichardt a la polaritat del solvent i a l'acidesa per formació de ponts d'hidrogen i la del 4-nitrofenol a la polaritat del solvent i a la basicitat per formació de ponts d'hidrogen. La 4-nitroanilina presenta un loading excessivament gran associat a α perquè els valors de β utilitzats en aquest estudi eren els procedents del parell 4-nitroanisol i 4-nitrofenol.

L'establiment de LSERs per a descriure la variació de valors de pK de diversos soluts en mescles aigua-dioxà es tracta primerament a l'article V aplicant únicament el modelatge tou i de forma més extensa a l'article VI, on són utilitzats els dos tipus de modelatge proposats.

Cada columna del conjunt de dades de l'article V conté una sèrie de valors de pK relacionats amb la dissociació del grup carboxílic d'un solut determinat en diverses mescles aigua-dioxà que cobreixen un rang de composicions entre el 10 i el 70 % (v/v) de dioxà. FA mostra que la variació del pK en funció del solvent pot ésser descrita amb un model de dos factors. S'han provat propietats macroscòpiques del solvent (fracció molar del dioxà i $1/\varepsilon$) i

propietats microscòpiques (α, β, π^* i E_T^N) com a paràmetres potencials per a ésser inclosos dins del model LSER. Els millors paràmetres per a construir el model han estat la fracció molar del dioxà i π^*. Les LSERs definitives poden ésser construïdes utilitzant un model que inclogui qualsevol d'aquests dos paràmetres i la contribució constant deguda a la presència de l'ordenada a l'origen. Malgrat l'ajust similar d'aquests dos models possibles, la bondat del model que inclou la fracció molar de dioxà podria ésser deguda al limitat interval de composicions de la mescla utilitzat en aquest estudi.

Per tal d'aclarir els dubtes referents a la qualitat similar del model de tipus macroscòpic (que utilitza la fracció molar del dioxà) i el de tipus microscòpic (que utilitza el paràmetre π^*) que expliquen la variació dels equilibris àcid-base en mescles aigua-dioxà, l'article VI s'ha orientat cap a la realització d'un estudi més general que inclogui conjunts de dades més variats quant als grups funcionals dels soluts i a l'interval de composicions de la mescla analitzat. Així, es van dissenyar una matriu A.1. que inclou els pK de soluts amb grups funcionals carboxílics, fenòlics i amino, que han estat determinats en mescles que cobreixen l'interval de composicions que va del 20 al 70% de dioxà (v/v), i una matriu A.2. que conté els valors de pK d'un nombre menor de soluts determinats en mescles que cobreixen composicions des del 20 fins el 80 % (v/v) de dioxà. La major qualitat del model que inclou el paràmetre π^* es confirma en augmentar la varietat de composts i de composicions de les mescles. Per a la matriu A.1., el paràmetre π^* resulta més correcte que la fracció molar del dioxà, tot i que ambdós paràmetres superen el test de TFA; en canvi, quan és analitzada la matriu A.2., la fracció molar de dioxà ni tan sols passa el test TFA i és π^* l'únic paràmetre del solvent acceptat.

El modelatge dur dóna pràcticament els mateixos resultats que el modelatge tou, la qual cosa suposa la validació del model LSER $pK = pK^o + s\pi^*$. La lleugera variació que existeix entre els valors dels coeficients dels models LSER trobats amb els dos procediments de modelatge ve del diferent nombre d'observacions emprat en cadascun dels casos. Atès que el nombre de dades utilitzat en el modelatge dur acostuma a ésser superior, s'han adoptat les LSER que s'obtenen amb aquest procediment com a expressions definitives en aquest estudi.

L'article VI inclou un estudi breu de l'efecte de la natura de l'electròlit inert en la variació dels valors de pK en mescles aigua-dioxà. Les matrius de dades A.1. i A.2. es transposen i alguns dels seus elements s'organitzen per a formar les matrius B.1. i B.2. B.2. conté soluts, els valors de pK dels quals han estat determinats utilitzant com a electròlit suport el KNO_3. La matriu B.1. s'ha format incloent més valors de pK que han estat determinats amb $NaClO_4$ com a sal inert. Tot i que el nombre limitat de dades no és prou gran com per a extreure conclusions definitives, no s'aprecia un efecte significatiu relacionat amb l'ús d'electròlits diferents en la determinació dels valors de pK ja que el nombre de factors necessari per a descriure la variació de les dades no augmenta en passar de la matriu B.1. a la matriu B.2.

Els equilibris àcid-base dels polinucleòtids en mescles aigua-dioxà.

El coneixement dels equilibris àcid-base de soluts senzills en mescles aigua-dioxà és un bon punt de partida per a l'estudi de processos anàlegs de molècules més complexes, com són ara els polinucleòtids. Per tal d'emular els microambients biològics de polaritat baixa s'han triat les mescles aigua-dioxà perquè contenen aigua com tots els sistemes biològics, i perquè la introducció del cosolvent orgànic permet assolir fàcilment la baixa polaritat

L'interès principal d'aquest projecte s'orienta al coneixement de l'efecte d'una disminució de polaritat del medi en les transicions termodinàmiques i estructurals dels polinucleòtids que depenen del pH. El seguiment d'aquestes transicions s'ha dut a terme potenciomètricament (quan ha estat possible) i amb tècniques espectromètriques (UV i dicroisme circular). La potenciometria proporciona informació sobre les transicions termodinàmiques associades a processos d'equilibri, mentre que les tècniques espectromètriques confirmen els resultats potenciomètrics i afegeixen informació relacionada amb la identitat (espectre) i evolució dels perfils de concentració de totes les espècies absorbents involucrades en el procés de transferència de protó.

Tal com es discuteix a la secció 3.5.3. de la tesi, no existeix cap garantia d'obtenir resultats correctes quan les dades espectromètriques d'equilibris macromoleculars es tracten amb els procediments clàssics de mínims quadrats que s'apliquen als processos d'equilibri de molècules senzilles. Els problemes associats a l'ús d'aquests tractaments vénen de certes característiques particulars que poden presentar els equilibris macromoleculars, com per exemple:

a) l'incompliment de la llei d'acció de masses i
b) la falta de correspondència entre el nombre d'espècies químiques i el nombre d'espècies absorbents (conformers).

La presència d'aquests fenòmens que van contra les lleis químiques que controlen els equilibris dels soluts senzills no pot ésser coneguda per endavant i per tant, l'ús de tractaments de dades clàssics que assumeixen l'absència d'aquestes característiques especials és clarament perillós. En aquests casos, procediments de modelatge tou com els mètodes de resolució de corbes, que no tenen en compte un model químic de comportament, sinó les característiques sistemàtiques de les dades experimentals (per ex., no negativitat, unimodalitat,...) per a modelar els perfils de concentració i els espectres en solució són l'opció més adequada i segura.

Ús de mètodes de resolució de corbes per a la interpretació del seguiment multivariant de processos bioquímics: millora i comprensió dels procediments quimiomètrics.

Abans d'utilitzar els mètodes de resolució de corbes per a tractar les dades espectromètriques lligades a les transicions dels polinucleòtids, s'ha aprofundit en el coneixement d'aquestes eines quimiomètriques. Els dos punts fonamentals d'aquest estudi han estat:

a) millora del mètode de regressió per mínims quadrats alternats (ALS) amb la introducció de noves restriccions (article VII) i,
b) comparació del funcionament del mètode ALS amb un altre procediment de resolució de corbes per a dades tridireccionals, la descomposició trilineal (TLD).

L'estratègia seguida en els articles VII i VIII inclou l'ús combinat de dades simulades i exemples reals per tal d'avaluar de quina manera la inclusió d'una restricció nova (article VII) o la selecció d'un mètode de resolució (article VIII) pot afectar el funcionament del procés global de resolució. Indicadors d'aquest funcionament són la recuperació de la informació qualitativa i l'ajust global de les dades (per a dades bi- i tridireccionals) i la recuperació de la informació quantitativa (per a conjunts de dades tridireccionals). L'ús de dades simulades en aquests estudis és essencial perquè, en sentit estricte, aquests són els únics sistemes dels quals es pot conèixer tota la informació. Les simulacions també ofereixen la possibilitat d'obtenir fàcilment un ampli conjunt de dades representatives que evita l'obtenció de conclusions esbiaixades que poden sorgir en cas d'utilitzar un conjunt reduït d'exemples reals. No obstant això, l'ús exclusiu de simulacions tampoc és aconsellable. Les conclusions obtingudes amb els estudis realitzats amb simulacions s'han de veure confirmades quan s'empren dades reals. Discordances entre els resultats procedents de les simulacions i de les dades reals sovint indiquen que no s'han inclòs en les simulacions característiques importants de les dades reals. Quan això ocorre, s'ha de generar un nou conjunt de simulacions més properes a les dades reals.

L'article VII vol contribuir a la solució del problema associat a la resolució de sistemes multicomponent que presenten un grau de superposició important entre els perfils de llurs espècies. En aquests sistemes, la selectivitat es perd de forma parcial o completa. La resolució del sistema depèn en aquests casos de la delimitació del domini de solucions possibles mitjançant l'aplicació de restriccions adequades que introdueixen informació relacionada amb les característiques pròpies dels perfils de concentració i/o de les respostes instrumentals de les dades experimentals. La selecció i aplicació correctes d'aquestes restriccions disminueix l'ambigüitat associada a la descomposició de matrius bilineals i, per tant, la proposta de noves restriccions efectives constitueix una contribució positiva a la resolució de conjunts de dades especialment complexos.

En aquest treball, dues implementacions noves de la restricció d'unimodalitat (que seran anomenades unimodalitat horitzontal i unimodalitat localitzada) i la denominada restricció de simetria han estat proposades. La primera d'aquestes restriccions pot ésser

aplicada a qualsevol tipus de perfil, mentre que les dues darreres són aptes només per a perfils en forma de pic.

La unimodalitat horitzontal ha estat proposada per tal de corregir la modificació dràstica dels perfils no unimodals que té lloc en aplicar la implementació clàssica d'aquesta restricció. Ambdues implementacions d'aquesta restricció comparteixen el mateix *modus operandi* que inclou una primera etapa de cerca del màxim del perfil i una segona de detecció i eliminació de les seves zones no unimodals. La diferència entre la unimodalitat clàssica i l'horitzontal rau en la forma d'eliminar els màxims secundaris del perfil; mentre la unimodalitat clàssica iguala a zero tots els elements del perfil no unimodals, l'horitzontal iguala aquests elements als que els són més propers i que respecten aquesta restricció. Des d'un punt de vista gràfic, la unimodalitat clàssica talla el perfil verticalment, la qual cosa dóna sovint perfils anormalment estrets i rosts, mentre la unimodalitat horitzontal talla horitzontalment els perfils originals i dóna com a resultat perfils corregits més suaus, que presenten una forma més propera a la dels perfils reals (veure figura 1, article VII).

La unimodalitat localitzada inclou una limitació addicional a la restricció d'unimodalitat en qualsevol de les implementacions suara esmentades. A més de limitar el nombre de màxims del perfil a un de sol, també limita la posició d'aquest màxim dins del perfil. Les posicions de referència dels màxims dels diferents perfils es determinen externament amb mètodes com l'algoritme d'agulla, OPA, SIMPLISMA o d'altres. El perfil original es modifica resituant el seu màxim en cas necessari i eliminant després les seves zones no unimodals.

La restricció de simetria és específica per a perfils cromatogràfics i modifica els pics que presenten cua en la seva part anterior fent la meitat posterior del perfil simètrica a l'altra. Cal tenir present que, malgrat el seu nom, aquesta restricció no transforma tots els pics en perfils simètrics, sinó només aquells que tenen una cua frontal.

Per tal d'avaluar la utilitat de les tres restriccions proposades, s'ha procedit a la resolució de diversos exemples de sistemes cromatogràfics binaris del tipus HPLC-DAD. En

cap cas s'ha inclòs selectivitat espectral, ja que aquesta característica hauria amagat l'efecte real de les restriccions proposades. La representativitat de les dades simulades s'ha assegurat fent ús d'un disseny factorial complet de dos nivells.

La restricció que s'avalua s'inclou com un factor, els nivells del qual són presència i absència, segons s'apliqui o no al procés de resolució. La resta de factors està relacionada amb propietats de les dades experimentals que se suposa que poden exercir algun efecte en els resultats de la resolució (resolució cromatogràfica, relació de concentracions component minoritari/component majoritari, model de soroll i relació senyal/soroll per al component minoritari). Hi ha dos factors relacionats amb la presència de components minoritaris. L'efecte de restriccions diferents de la selectivitat en el modelatge d'aquestes espècies és especialment important, ja que els seus senyals es diferencien molt poc dels d'altres espècies majoritàries i del nivell de soroll i, per tant, la informació externa procedent de les restriccions és essencial per a recuperar els perfils correctes.

El mètode ALS s'aplica al conjunt de simulacions per tal d'avaluar l'efecte de les tres restriccions proposades. Les respostes que es determinen després de cada execució del programa ALS són les dissimilituds entre els perfils de concentració reals i els recuperats per al component majoritari i per al component minoritari, la desviació estàndard dels residuals i la manca d'ajust. Els gràfics normals de probabilitat s'han utilitzat per tal d'avaluar la significació de l'efecte associat a cadascun dels factors i a les interaccions que s'estableixen entre ells.

L'efecte dels altres factors, apart de les restriccions, sobre les respostes no s'allunyen del que caldria esperar en qualsevol procés de resolució. En la recuperació de la informació qualitativa, la superposició entre perfils de concentració és el factor que influeix de forma més decisiva sobre la recuperació dels perfils del component majoritari, la relació senyal/soroll és més determinant en la recuperació del component minoritari i tant la presència d'heteroscedasticitat com els nivells alts de soroll afecten la recuperació de qualsevol perfil de concentració. L'efecte de tots aquests factors sobre els paràmetres associats a l'error és similar, encara que d'intensitat més baixa. La menor sensibilitat dels

paràmetres d'error a l'efecte dels diferents factors és conseqüència de l'ambigüitat rotacional associada a la descomposició de matrius bilineals (és a dir, la matriu original de dades pot ésser reproduïda acuradament utilitzant el producte de les matrius dels perfils de concentració i d'espectres reals o matrius que continguin combinacions lineals d'aquests perfils).

De les tres restriccions avaluades, només la unimodalitat horitzontal causa una millora de la qualitat dels resultats de la resolució. L'efecte d'aquesta restricció és significatiu en la recuperació del perfil de concentració del component minoritari i té una influència clara en la qualitat de l'ajust. La bondat d'aquesta restricció s'ha confirmat amb els resultats obtinguts amb un conjunt d'exemples reals, on ALS amb unimodalitat horitzontal sempre ha donat un ajust millor de les dades que ALS utilitzant la implementació clàssica d'aquesta restricció. Les restriccions d'unimodalitat localitzada i de simetria no han exercit cap millora apreciable en els resultats procedents d'ALS; no obstant això, atès que la informació inclosa en aquestes dues restriccions és potencialment útil, no es descarta que implementacions alternatives puguin augmentar l'efectivitat d'aquestes restriccions.

Encara que la introducció en el mètode ALS de la restricció d'unimodalitat horitzontal permet ja la millora dels resultats obtinguts en tractar conjunts de dades com els procedents dels equilibris dels polinucleòtids, una millora d'aquests resultats podria venir també de l'aplicació d'un altre mètode de resolució que fos més adequat per al tractament d'aquestes dades. Aquesta possibilitat s'ha contemplat amb l'estudi comparatiu del funcionament de dos mètodes de resolució de dades tridireccionals: l'ALS i la descomposició trilineal (TLD). L'article VIII recull de forma exhaustiva les diferències entre ambdós mètodes i proposa uns criteris de selecció basats en les característiques dels conjunts de dades que s'analitzen.

El mètode de descomposició trilineal (TLD) és un exemple clar de mètode de resolució no iteratiu i les seves solucions estan basades en la resolució del problema generalitzat de valors propis/vectors propis, mentre ALS és un representant dels mètodes iteratius que operen optimitzant unes estimacions inicials amb l'ajut de restriccions

relacionades amb l'estructura matemàtica del conjunt de dades i amb les característiques pròpies dels perfils experimentals. TLD és un mètode ràpid i simple que proporciona solucions úniques. Malgrat això, sempre presuposa estructura trilineal en les dades (és a dir, perfils de concentració i espectres comuns per a una mateixa espècie en totes les matrius analitzades) i no ofereix la possibilitat d'introduir informació relativa a les característiques dels perfils experimentals, la qual cosa fa que no es pugui evitar que apareguin solucions inacceptables des del punt de vista químic. ALS pot treballar amb matrius no trilineals que comparteixen un sol dels seus ordres (nombre de files o nombre de columnes) perquè la restricció de trilinealitat és opcional. A més, les solucions mancades de sentit químic es poden evitar introduint les restriccions adequades. Els aspectes menys favorables del mètode ALS són el major temps de càlcul necessari per a l'optimització iterativa, la major intervenció de l'usuari i el fet que les solucions que s'obtenen no sempre són úniques.

Els mètodes TLD i ALS s'han aplicat a un grup representatiu de conjunts de dades tridireccionals que mostren fonts de variació comunes en dades reals, com per exemple, canvis de posició i de forma del senyal d'una mateixa espècie en diferents experiments (és a dir, en diferents matrius) i distorsions dels senyals degudes a la presència de soroll. Els conjunts bàsics de dades tridireccionals estan formats per parelles de matrius amb estructura trilineal que contenen un mateix sistema binari cromatogràfic i que presenten diferents graus de superposició entre els perfils de concentració i els espectres. Els perfils de concentració d'una de les matrius del parell o de les dues matrius, segons el cas, es modifiquen segons la causa de variació que es vulgui simular (canvi de posició del senyal, canvi de forma o addició de soroll). Els paràmetres utilitzats per tal d'avaluar l'efecte d'aquestes causes de variació sobre els resultats dels dos mètodes de resolució comparats estan relacionats amb la recuperació de la informació qualitativa i quantitativa del sistema i amb la qualitat de l'ajust de les dades.

Tant ALS com TLD s'han executat admetent l'existència de dos components en tots els conjunts de dades tridireccionals. El mètode ALS s'ha executat de dues formes diferents: forçant la presència d'estructura trilineal en les dades (ALSf) i sense aplicar aquesta restricció (ALS).

Les conclusions extretes després de l'aplicació dels mètodes TLD i ALS a tots els conjunts de dades és que la qualitat dels resultats depèn principalment de l'existència de selectivitat i de l'estructura interna de les dades tridireccionals. En relació al darrer factor esmentat, cal distingir quines de les causes de variació simulades indueixen la pèrdua de l'estructura trilineal de les dades (canvis en la forma i posició dels perfils) i quines no ho fan (distorsions del senyal degudes al soroll).

Com a norma general, ALS funciona clarament millor que ALSf o TLD quan no hi ha trilinealitat, degut a la seva major flexibilitat en el modelatge dels perfils. Aquesta superioritat és més evident quan el sistema no trilineal presenta selectivitat, ja que la presència d'aquesta característica suprimeix les ambigüitats associades a la descomposició de matrius augmentades (en aquests casos, les petites dissimilaritats entre perfils reals i simulats vénen del soroll afegit a les dades). Quan no hi ha selectivitat a les dades, ALS continua essent millor en la major part de situacions. Només deixa d'ésser superior quan les desviacions de la trilinealitat són extremadament petites o quan vénen associades a un augment de la superposició dels perfils. En aquests casos particulars, els efectes negatius causats per la pèrdua de trilinealitat en el mètode TLD i per l'augment de correlació entre perfils en el mètode ALS són comparables. De tota manera, la similitud de resultats queda reduïda a la recuperació dels perfils i a la recuperació de la informació quantitativa; l'ajust de les dades continua essent clarament millor quan s'utilitza el mètode ALS. Atès que TLD i ALSf només inclouen la informació trilineal present en els dos primers components en el procés de reproducció de les dades originals, l'exclusió de les contribucions no trilineals del sistema ocasiona l'empitjorament de l'ajust; per contra, la major quantitat d'informació introduïda en la reproducció de les dades pel mètode ALS, encara que provingui de l'ús de perfils que són combinacions lineals dels perfils reals, proporciona un ajust de les dades millor que el que s'obté amb mètodes que imposen estructura trilineal. Aquest comportament i la capacitat de calcular l'ajust de les dades en sistemes reals suggereixen precisament l'ús de l'ajust per a dilucidar l'estructura interna de les dades, que serà no trilineal si l'ajust obtingut amb el mètode ALS és clarament millor que el que proporcionen

TLD o ALSf quan s'empra el mateix nombre de components en els processos de resolució de tots els mètodes.

Les diferències de comportament entre TLD i ALSf en l'anàlisi de dades no trilineals provenen de la forma distinta d'aplicar la trilinealitat dels dos procediments. TLD treballa amb tot el conjunt de dades per tal d'obtenir la combinació trilineal de perfils que permet una millor reproducció de les dades originals, mentre ALSf actua una per una en cadascuna de les columnes de la matriu augmentada de concentracions per tal d'obtenir el millor perfil trilineal corregit per a cadascun dels components de la matriu. Això significa que TLD dóna prioritat a la reproducció global de les dades i ALSf a la recuperació dels perfils individuals dels components. Aquesta explicació és consistent amb els resultats obtinguts per a sistemes no trilineals, per als quals ALSf sempre dóna millors perfils i quantificacions i TLD millors ajusts. L'ajust pitjor dels models ALSf es deu a la supressió de les contribucions no trilineals i de les trilineals que van contra les restriccions imposades (per exemple, parts negatives o no unimodals d'un perfil de concentració,...). La pitjor qualitat dels perfils obtinguts amb TLD s'explica perquè l'única limitació a la forma dels perfils és que han de mantenir la condició trilineal.

L'addició de soroll (diferents nivells i models) a les dades s'ha introduït sense modificar-ne l'estructura trilineal. Tots els mètodes de resolució són més afectats per augments dels nivells de soroll que per variacions en el tipus de soroll afegit. En general, TLD i ALSf proporcionen millors resultats que ALS per als sistemes amb soroll afegit. Aquesta superioritat és òbvia quan no hi ha selectivitat a les dades, ja que llavors ALS es veu afectat per problemes d'ambigüitat rotacional, mentre ALSf i TLD proporcionen solucions úniques i correctes degut a la trilinealitat de les dades. Per qüestions d'estalvi de temps de càlcul i de simplicitat, TLD és més recomanable que ALSf en sistemes trilineals mancats de selectivitat. Quan hi ha selectivitat, TLD, ALSf i ALS arriben a solucions úniques força similars als perfils reals. De tota manera, TLD i ALSf sovint mostren una subtil superioritat sobre ALS, especialment en sistemes amb un alt nivell de soroll. Aquesta diferència podria tenir el seu origen en el menor nivell de soroll present en els perfils recuperats per ALSf i TLD. Aquests dos mètodes suavitzen els perfils perquè la contribució

del soroll acostuma a ésser diferent de matriu a matriu, és a dir, el soroll rarament és trilineal. Per tant, els perfils que recuperen aquests dos mètodes són més semblants als senyals reals lliures de soroll que els perfils procedents d'ALS, on el soroll només es pot eliminar quan modifica el perfil de forma que deixa de complir les restriccions imposades en el procés de resolució.

De l'estudi efectuat se'n conclou que la selecció d'un o altre mètode de resolució vindrà determinada pel tipus de conjunt de dades que s'hagi d'analitzar. Per tal de classificar qualsevol conjunt de dades, les característiques clau són la presència de selectivitat i l'existència d'estructura trilineal.

Les dades associades als equilibris dels polinucleòtids són no trilineals i poden o no presentar selectivitat. S'ha comprovat que ALS és el millor mètode per a tractar sistemes no trilineals, tot i que si no es té selectivitat els resultats procedents de TLD poden tenir una qualitat similar. No obstant això, ALS sembla comportar-se de forma més robusta en el modelatge de perfils i com queaquest és el principal objectiu en la interpretació dels equilibris dels polinucleòtids, ALS ha estat el mètode de resolució adoptat per a tots els estudis posteriors.

Aplicació de tècniques de resolució de corbes a l'estudi de les transicions dependents del pH d'alguns homopolinucleòtids en aigua-dioxà.

Tres homopolinucleòtids, àcid poliuridílic (polyU), àcid policitidílic (polyC) i àcid poliadenílic (polyA) han estat seleccionats a causa de les seves diferents característiques. L'estudi d'aquests composts macromoleculars s'ha dut a terme utilitzant espectrometria UV i dicroisme circular i aplicant la tècnica potenciomètrica a aquells polinucleòtids solubles al nivell de concentració adequat per al bon funcionament d'aquesta tècnica. En l'interval de pH de treball, només cal considerar un seti de protonació per a cadascun dels tres composts analitzats. Així, polyA protona el N(1) de l'anell d'adenina i polyC el N(3) de l'anell de citosina i ambdós donen estructures amb càrregues positives, mentre polyU desprotona el N(3) nitrogen amídic de l'anell de l'uracil i dóna una estructura carregada negativament a

pH bàsic. S'han utilitzat mescles dioxà-aigua al 30 i 50% (v/v) en els diferents experiments. No s'han fet servir mescles més riques en solvent orgànic perquè induïen la precipitació de les estructures macromoleculars.

La informació que forneixen els experiments esmentats és suficient per a la interpretació directa del comportament àcid-base dels soluts estudiats. A més, la combinació interna d'aquests resultats o la seva comparació amb els procedents d'altres experiments pot ajudar a comprendre alguns efectes interessants, com són ara:

a) *l'efecte de l'estructura molecular en el comportament àcid-base del polinucleòtid.* Pot ésser interpretat comparant les diferències entre el comportament dels diferents polinucleòtids estudiats.

b) *la presència d'efectes de tipus macromolecular en l'evolució dels equilibris dels polinucleòtids.* Pot ésser avaluada amb els resultats dels experiments amb polinucleòtids o comparant els resultats d'aquests experiments amb els obtinguts amb els respectius nucleòtids cíclics.

c) *l'efecte del solvent en els equilibris dels polinucleòtids.* S'estudia comparant els experiments en aigua-dioxà amb d'altres anàlegs realitzats en solució aquosa.

La tècnica potenciomètrica va ésser aplicada a polyA, polyC i polyU i als seus respectius nucleòtids cíclics 3',5'-adenosinmonofosfat cíclic (cAMP), 3',5'-citidinmonofosfat cíclic (cCMP) i 3',5'-uridinmonofosfat cíclic (cUMP). Es van determinar constants d'equilibri per a tots els monòmers cíclics al 30% dioxà-aigua (v/v). Entre els polinucleòtids estudiats, l'obtenció de resultats quantitatius només va ésser possible per al sistema polyU-H, ja que polyA i polyC no eren completament solubles als nivells de concentració necessaris. La presència d'efecte polielectrolític es va estudiar examinant l'evolució dels valors de la constant d'equilibri amb el grau de protonació de la macromolècula. No es va detectar aquest efecte en l'equilibri associat a l'àcid poliuridílic, ja que els valors de pK van romandre constants durant tot el procés de protonació. De forma

qualitativa, es va apreciar aquest efecte en els sistemes polyC-H i polyA-H, tot i que aquesta conclusió va haver d'ésser confirmada amb els experiments espectromètrics, que van permetre un seguiment quantitatiu dels processos d'equilibri associats a aquests dos polinucleòtids.

L'ús de tècniques espectromètriques ha proporcionat informació quantitativa per a tots els polinucleòtids estudiats ja que el nivell de concentració necessari per al correcte funcionament d'aquestes tècniques és molt menor. A més, s'ha pogut obtenir la informació relacionada amb l'estructura de totes les espècies absorbents en solució (confòrmers o espècies químiques) i, en conseqüència, ha estat possible la descripció de totes les transicions dependents del pH (associades a processos termodinàmics o no).

Per a cada polinucleòtid, totes les valoracions realitzades amb la mateixa tècnica són tractades conjuntament amb el mètode ALS. En el cas de valoracions UV, les restriccions aplicades són: selectivitat, si existeix; no negativitat per a perfils de concentració i espectres; unimodalitat per als perfils de concentració i condició de sistema tancat. Les valoracions seguides amb dicroisme circular (CD) reben un tractament similar, que només es diferencia en que no s'aplica la restricció de no negativitat als espectres. Els resultats que proporciona directament el mètode ALS i la informació que es pot obtenir d'ells es detallen en els següents apartats:

a) *Resultats ALS: perfils de concentració i espectres unitaris de les espècies absorbents involucrades en l'equilibri àcid-base.*

ALS dóna el nombre d'espècies absorbents en cada equilibri àcid-base i el perfil de concentració i l'espectre relacionat amb cadascuna d'elles. Per als sistemes polyC-H i polyU-H, dues espècies han estat suficients per a explicar el procés de protonació, mentre el sistema polyA-H ha estat descrit amb tres espècies. La concordança entre els perfils de concentració obtinguts per a un mateix sistema amb les diferents tècniques espectromètriques valida la qualitat del treball experimental i el funcionament correcte del mètode ALS en el tractament d'aquestes dades bioquímiques.

b) *Detecció i descripció de les transicions conformacionals.*

El diagrama de distribució d'espècies de l'equilibri àcid-base en estudi es construeix dibuixant els perfils de concentració obtinguts amb ALS enfront els valors de pH de l'experiment. Aquest diagrama inclourà transicions entre espècies químiques sense canvis en l'estructura macromolecular, canvis conformacionals associats al procés de protonació i reorganitzacions espacials de la macromolècula que no suposen alteracions en el seu estat de protonació. La tasca d'identificar quins tipus de transició tenen lloc en cada polinucleòtid i quines conformacions hi prenen part és aliena a l'ALS i depèn d'evidències externes i del coneixement que es tingui del problema químic. PolyU, polyC i polyA són bons exemples de les transicions esmentades.

Segons estudis previs realitzats en medi aquós, polyU formaria part del grup de polinucleòtids que no presenten variacions conformacionals en passar de la forma protonada a la desprotonada. En mescles aigua-dioxà, la forma similar dels espectres corresponents a les espècies poly(UH) i poly(U^-), on el desplaçament espectral entre ells és gairebé l'única diferència que s'aprecia, sembla confirmar aquesta hipòtesi. La conformació implicada en aquest procés de protonació ha estat identificada com un cabdell aleatori degut a les lleugeres diferències d'intensitat d'absorció entre els espectres de polyU i de cUMP.

El diagrama de distribució d'espècies relacionades amb el procés de protonació del polyC mostra dues espècies amb espectres clarament diferents. Aquestes dues espècies s'han identificat com dues espècies diferentment protonades i per tant, polyC proporciona un exemple de transició conformacional associada a un procés de protonació. De les tres espècies suggerides per a aquest equilibri en solució aquosa (el polyC desprotonat, el [poly(C)·poly(CH^+)] semiprotonat i el poly(CH^+) totalment protonat), la primera i la segona semblen les més probables en medi hidroorgànic. Aquest identificació es basa en la similitud dels espectres i de la zona de pH d'existència d'aquestes dues espècies en aigua i en el medi estudiat. A més, en una solució de polaritat baixa, la situació d'una espècie carregada estabilitzada per la formació d'un pont d'hidrogen de tipus N-H^+····N, com [poly(C)·poly(CH^+)], és més favorable que la d'una espècie amb una càrrega neta NH^+, com

poly(CH^+). Tant polyC com [poly(C)·poly(CH^+)] presenten conformacions ordenades, com es pot observar a partir dels marcats hipercromisme en els espectres CD i hipocromisme en els espectres UV de les espècies de polyC respecte a l'espectre de cCMP. Les conformacions de les dues espècies de polyC, com en solució aquosa, deuen ésser una hèlix senzilla per a la forma desprotonada del polyC i una doble hèlix per a la forma semiprotonada d'aquest polinucleòtid.

Les transicions dependents del pH relacionades amb el polyA necessiten tres espècies per a ésser descrites adequadament. Com que només existeix un seti de protonació en l'interval de pH estudiat, una de les transicions d'aquest sistema estarà relacionada amb el procés de protonació i l'altra serà una reorganització espacial de la macromolècula que no altera el seu estat de protonació. En examinar el diagrama de distribució, s'observa que la transició que té lloc en la regió més bàsica inclou espècies que presenten perfils i zones d'existència molt reproduïbles en tots els experiments, mentre quela transició que ocorre a la zona més àcida involucra espècies amb perfils més irregulars, la zona de pH d'existència de les quals està lleugerament desplaçada en els diferents experiments. Aquest comportament sembla associar la primera transició amb el procés de protonació i la darrera amb la reorganització espacial de la macromolècula. La forma dels espectres també confirma aquesta hipòtesi, ja que l'espectre de la zona més bàsica, el polyA desprotonat, és completament diferent dels espectres de les altres dues espècies, que són força similars entre ells, i que podrien ésser atribuïts raonablement a dues conformacions de l'espècie protonada, poly(AH^+)1 i poly(AH^+)2. Com en el cas delper al polyC, la comparació entre els tres espectres del polyA i el del cAMP confirmen l'estructura ordenada de totes les espècies absorbents d'aquest polinucleòtid. Els espectres de les espècies de polyA en aigua-dioxà i en solució aquosa presenten formes força similars i concorden amb les teories generalment acceptades que postulen una conformació d'hèlix senzilla per al polyA desprotonat i diverses conformacions de doble hèlix per als confòrmers protons. Els espectres de CD i UV de les conformacions protonades presenten hipercromisme i hipocromisme, respectivament, en passar de poly(AH^+)1 a poly(AH^+)2. Aquest fet podria ésser degut a la minimització de la repulsió entre grups fosfat veïns induïda per les interaccions electrostàtiques estabilitzadores entre la càrrega negativa d'aquests grups i els

setis protonats de les bases. Aquest efecte estabilitzador creixent probablement és el responsable de la transició entre una doble hèlix amb un pas de rosca més gran a una estructura més compacta amb els grups fosfat més propers entre ells i un major grau d'apilament entre bases.

c) *Presència i tipus d'efecte polielectrolític.*

Després d'identificar les espècies involucrades en els diferents equilibris, es poden calcular fàcilment les constants associades als processos de protonació. La informació del diagrama de distribució inclou la concentració de totes les espècies absorbents protonades i desprotonades en cada punt de la valoració. Per tant, per a cada valor de pH mesurat, es pot calcular una constant. Dibuixar els valors obtinguts de log K enfront dels graus de protonació respectius, α_p, permet la detecció de la presència i del tipus d'efecte polielectrolític. Si aquest efecte és present, la constant aparent, K_{app}, dels setis anàlegs del polinucleòtid canvia durant el procés de protonació i s'ha de prendre com a valor de log K associat al sistema la constant intrínseca, K_{int}, definida com el valor extrapolat de K_{app} quan el grau de protonació és igual a zero. Si aquest efecte no existeix, els valors de K_{app} no varien en tot el procés i és possible definir una constant termodinàmica de protonació.

L'equació general que s'utilitza per a calcular la constant aparent d'un polinucleòtid és:

$$K = \frac{\sum[protonated\ conformers]}{\left(\sum[deprotonated\ conformers]\right)[H^+]}$$

Segons les espècies identificades en els diagrames de distribució dels diferents polinucleòtids, l'expressió anterior pren les formes

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

per al sistema polyU-H,

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

per al sistema polyC-H, i

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

per al sistema polyA-H.

S'ha detectat la presència d'efecte polielectrolític en els sistemes polyA-H i polyC-H. Contràriament, polyU mostra valors constants de log K durant tot el procés de protonació. El comportament diferent del polyU podria ésser probablement atribuït a les estructures més flexibles i desordenades de les espècies protonades i desprotonades, que permeten més fàcilment moviments de la macromolècula orientats a minimitzar l'efecte entre setis durant el procés de protonació.

PolyA i polyC presenten efectes polielectrolítics de diferents tipus. L'efecte polielectrolític no lineal del polyC és causat pel mecanisme cooperatiu entre el procés de protonació i la formació de l'estructura de doble hèlix (és a dir, la doble hèlix estabilitza el seti protonat mitjançant el pont d'hidrogen N-H^+····N establert entre les seves fibres i, al mateix temps, l'estructura adoptada pel parell de bases entre les quals es forma el pont d'hidrogen és directament responsable del creixement i l'estabilitat de l'estructura helicoidal). Quan el procés de protonació avança, l'efecte polielectrolític positiu evoluciona cap a negatiu degut a que la repulsió entre setis protonats esdevé més forta que l'efecte estabilitzant causat per la formació de l'estructura de doble hèlix. Per al polyA, l'efecte polielectrolític positiu segueix un model lineal degut a l'estabilització electrostàtica dels grups fosfats negatius amb els setis protonats de l'adenina.

Sembla ser que existeix una certa relació entre el tipus d'efecte polielectrolític i el mecanisme d'estabilització dels setis protonats. Així, un efecte polielectrolític no lineal

podria ésser relacionat a la presència d'un mecanisme cooperatiu associat al procés de protonació (els setis protonats del polyC són responsables de la formació d'un dels ponts d'hidrogen que uneixen les fibres de la doble hèlix del polinucleòtid), mentre que un efecte lineal seria causat principalment per interaccions electrostàtiques (els setis protonats del polyA estabilitzen l'estructura de doble hèlix minimitzant la repulsió electrostàtica entre grups fosfats propers, però no participen directament en la unió de les fibres de l'hèlix per ponts d'hidrogen).

Les constants de protonació del cCMP i del cUMP tenen valors numèrics similars a la constant intrínseca del polyC i a la constant de protonació del polyU, respectivament. Per contra, hi ha clares diferències entre la constant de protonació del cAMP i la constant intrínseca del polyA. El valor clarament inferior de la constant associada al cAMP podria ésser conseqüència de l'absència de la influència positiva dels grups fosfats situats en les unitats monòmeres adjacents als setis de protonació del polyA, que no és exercida sobre els setis de cAMP.

d) *Efecte de l'estructura molecular del polinucleòtid en els equilibris àcid-base.*

PolyA, polyC i polyU poden ésser agrupats segons dos criteris diferents: el grup funcional dels seus setis de protonació o la natura de l'anell que conté el seti de protonació del polinucleòtid. El primer criteri separa polyC i polyA, ambdós amb un nitrogen a l'anell aromàtic com a seti de protonació, de polyU, que presenta un nitrogen amídic com a seti de protonació. El darrer criteri posa al mateix grup polyC i polyU, ambdós amb bases pirimidíniques, i en una classe diferent situa el polyA, que conté una base purínica.

L'efecte del grup funcional del seti de protonació en els processos àcid-base sembla molt més important que la natura de l'anell de la base nitrogenada. Segons les similituds dels seus comportaments àcid-base, polyC i polyA pertanyen al mateix grup i polyU estaria formant part d'un grup diferent. Així, polyA i polyC presenten estructures ordenades (hèlix senzilles i dobles) i efecte polielectrolític, mentre totes les espècies del polyU tenen estructura de cabdell aleatori i no hi ha efecte polielectrolític associat a la protonació.

Les diferències en el comportament àcid-base degudes al tipus d'anell present en les bases nitrogenades són menys intenses i poden ésser apreciades quan es comparen polyA i polyC es comparen. La diferent situació espacial dels setis de protonació en ambdós polinucleòtids determina el paper que jugaran aquest setis en les seves macromolècules. Mentre el seti de protonació del polyC està orientat cap al nucli de l'hèlix doble i forma un dels enllaços interfibres, el seti de protonació del polyA mira cap a l'exterior de l'hèlix i no participa directament en la unió entre fibres. Com ja s'ha comentat, aquests diferents papers expliquen també els diferents tipus d'efectes polielectrolítics associats a aquests dos polinucleòtids. Les propietats diferents dels anells purínics i pirimidínics també afecten el grau d'apilament de la macromolècula. Els anells purínics presenten una capacitat d'apilament superior a la dels anells pirimidínics. Confirmant aquesta afirmació, els espectres de CD de les dobles hèlix del polyA presenten el.lipticitats clarament superiors a les de la doble hèlix del polyC. Aquest hipercromisme del polyA és un signe de major ordenació de l'estructura i, per tant, d'estructures amb més apilament.

e) *Efecte del solvent en els equilibris àcid-base dels polinucleòtids.*

L'efecte del solvent sobre els equilibris àcid-base dels polinucleòtids s'ha estudiat comparant resultats obtinguts en medi aquós (al nostre grup de recerca) amb els obtinguts en mescles aigua-dioxà. Del conjunt d'espècies químiques detectades en l'estudi d'aquests equilibris en els diferents medis, es conclou que l'estabilitat d'estructures macromoleculars carregades és menor en medis menys polars. De fet, algunes espècies detectades en aigua no apareixen en les mescles hidroorgàniques perquè la macromolècula precipita abans d'arribar als valors de pH adequats per a la formació d'aquestes espècies. Exemples d'espècies presents únicament en medi aquós són el polyC totalment protonat, algunes conformacions de polyA protonat i el polyU desprotonat (només a pH molt bàsics en mescles dioxà-aigua 50 % (v/v)). En general, estructures carregades desordenades, com el polyU desprotonat, són més estables que les estructures ordenades en aigua-dioxà perquè la macromolècula té més llibertat de moviments per tal d'acomodar de la forma més adequada les càrregues netes de la seva estructura.

Les espècies presents en aigua i en les mescles aigua-dioxà mostren la mateixa conformació i l'única diferència és el menor grau d'apilament entre bases en medis hidroorgànics, amb la qual cosa les estructures que s'obtenen són més relaxades que en solució aquosa. Aquest efecte es confirma amb l'hipocromisme dels espectres UV i l'hipercromisme dels de CD de les espècies mesurades en solució aquosa respecte a les mesurades en mescles aigua-dioxà, per exemple, en les hèlix senzilles del polyC i del polyA.

La informació quantitativa relativa a aquests equilibris es recull a la tesi. La menor estabilitat de les estructures carregades en medis hidroorgànics es reflecteix en els valors de les constants d'equilibri. Així, el procés de protonació del polyU, que va d'una espècie carregada negativament a una espècie neutra, dóna constants de protonació superiors en aigua-dioxà; en canvi, les protonacions de polyC i polyA, que impliquen el pas d'una estructura neutra a una carregada positivament, mostren valors superiors de constant en medi aquós.

La presència d'efecte polielectrolític depèn de la natura del polinucleòtid i no varia quan es canvia el medi de reacció. En tots els medis assajats en aquest estudi, polyU mai ha presentat efecte polielectrolític, mentre polyA i polyC sempre han sofert aquest tipus d'efecte.

El tipus d'efecte polielectrolític, en canvi, sí que és dependent del solvent. En el cas d'un efecte polielectrolític controlat per interaccions electrostàtiques, com en el polyA, models lineals amb pendents similars descriuen l'efecte polielectrolític en tots els medis assajats. No obstant això, aquestes petites diferències entre pendents poden ésser degudes a la gran similitud dels efectes dels diferents medis sobre aquest polinucleòtid, o bé als contraions de l'electròlit suport, que podrien estabilitzar els grups fosfat i obscurir l'efecte real del solvent sobre el procés de protonació. Treballar a forces iòniques més baixes podria aclarir aquest punt. El tipus d'efecte polielectrolític associat a protonacions afectades per mecanismes cooperatius, com la del polyC, és clarament dependent del solvent. L'efecte del medi es nota en les zones d'existència de l'efecte positiu i del negatiu. El canvi de signe té

lloc a graus de protonació al voltant de 0.3 en la solució hidroorgànica i al voltant de 0.5 en solució aquosa. L'efecte polielectrolític negatiu no apareix en aigua fins que té lloc el trencament de l'estructura de doble hèlix del polyC semiprotonat, mentre que un augment de la densitat de càrrega de l'estructura és suficient per a variar el signe d'aquest efecte en solucions menys polars. L'efecte del solvent també s'observa en l'expressió matemàtica que cal utilitzar per a modelar la zona d'efecte polielectrolític positiu en ambdós medis: en aigua-dioxà cal un polinomi d'ordre quart, mentre que un d'ordre dos és suficient per a descriure aquest mateix efecte en aigua. L'efecte brusc que s'aprecia en aigua-dioxà quan els graus de protonació són molt baixos s'explica per la fàcil formació de l'enllaç de pont d'hidrogen entre fibres en aquests medis poc polars, on la competició de molècules de solvent per a desenvolupar aquest tipus d'enllaç és molt menor que en solució aquosa. Quan la protonació avança, aquesta tendència s'inverteix i l'efecte polielectrolític en aigua/dioxà esdevé cada cop menys pronunciat fins que s'arriba a una zona gairebé plana per a graus de protonació al voltant de 0.3. Aquesta suavització de l'efecte polielectrolític ve provocada per la importància creixent de la influència desestabilitzadora associada a l'augment de densitat de càrrega en l'estructura, que gradualment compensa la tendència favorable a la protonació lligada a la formació de l'enllaç de pont d'hidrogen entre les dues fibres de l'hèlix.