Protonation Studies and Multivariate Curve Resolution on Oligodeoxynucleotides Carrying the Mutagenic Base 2-Aminopurine

R. Gargallo,* M. Vives,* R. Tauler,* and R. Eritja†

*Department of Analytical Chemistry, University of Barcelona, E-08028 Barcelona, and †Centro de Investigación y Desarrollo—Consejo Superior de Investigaciones Científicas, E-08034 Barcelona, Spain

ABSTRACT 2-Aminopurine (P) is a mutagen causing A·T to G·C transitions in prokaryotic systems. To study the base-pairing schemes between P and cytosine (C) or thymine (T), two self-complementary dodecamers containing P paired with either C or T were synthesized, and their protonation equilibria were studied by acid-base titrations and melting experiments. The mismatches were incorporated into the self-complementary sequence d(CGCPCCGGXGCG), where X was C or T. Spectroscopic data obtained from molecular absorption, circular dichroism (CD), and molecular fluorescence spectroscopy were analyzed by a factor-analysis-based method, multivariate curve resolution based on the alternating least squares optimization procedure (MCR-ALS). This procedure allows determination of the number of acid-base species or conformations present in an acid-base or melting experiment and the resolution of the concentration profiles and pure spectra for each of them. Acid-base experiments have shown that at pH 7, 150 mM ionic strength, and 37°C, both C and P are deprotonated. At pH near 4, the majority of species shows C protonated and P deprotonated. Finally, at pH values near 3, the majority of species shows both protonated C and P. These results are in agreement with NMR studies showing a wobble geometry for the P·C base pair and a Watson-Crick geometry for the P·T base pair at neutral pH. Melting experiments were carried out to confirm the proposed acid-base distribution profile. For the sequence including the P·T mismatch, only one transition was observed at neutral pH. However, for the sequence including the P·C mismatch, two transitions were detected by CD but only one by molecular absorption. This behavior agrees with that observed by other authors for oligonucleotides of similar sequence and suggests the following sequence of conformational changes during melting: duplex → hairpin → random coil.

INTRODUCTION

2-Aminopurine (P), a base analog of adenine, is a powerful mutagen causing A·T to G·C transitions in prokaryotic systems (Freese, 1959; Ronen, 1979). The proposed pathway for such transitions requires that P mispair with both thymine and cytosine. Although P usually forms base pairs with T during replication (Watanabe and Goodman, 1981) it can also mispair with C (Goodman and Ratliff, 1983). The formation and persistence of P·C mispairs is the molecular basis of 2-aminopurine mutagenesis.

Suggested base-pairing schemes between P and C involve formation of rare tautomers (Freese 1959), a protonated base pair and a wobble base pair (Sowers et al., 1986) (Scheme 1). NMR studies showed that a P·C base pair at neutral (Fagan et al., 1996; Sowers et al., 2000) and high (Sowers et al., 1986) pH is in a wobble geometry. On the other hand, NMR studies have shown that 2-aminopurine forms a two-hydrogen-bond base pair with thymine in Watson-Crick geometry (Sowers et al., 1986) (Scheme 1 *c*).

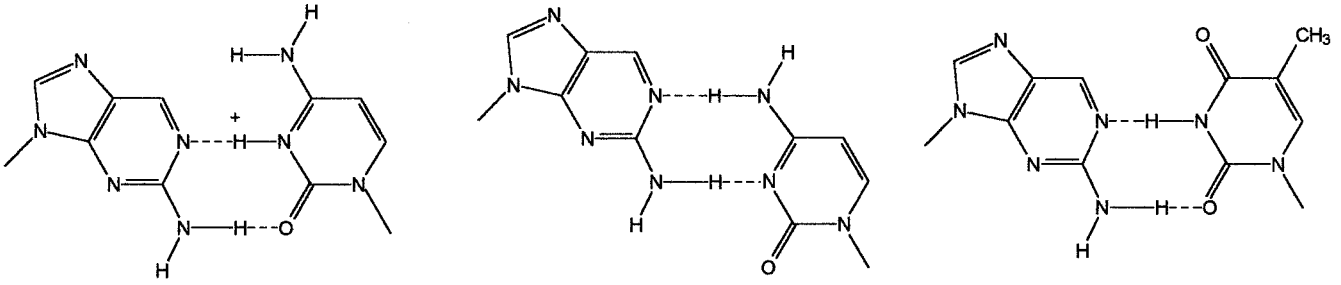
Owing to its fluorescence properties, the P base analog has been extensively used as a real-time reporter molecule to study the dynamics of DNA duplexes (Guest et al., 1991). Currently, 2-aminopurine is being used as a local fluorescent probe to study a large number of enzymes that interact with DNA, such as DNA polymerases (Bloom et al., 1993), helicases (Raney et al., 1994), DNA methylases (Holz et al., 1998), and junction-resolving enzymes (Déclais and Lilley, 2000). The P·T and P·C in DNA have been studied by NMR (Fagan et al. 1996; Sowers et al. 1986, 1989, 2000) and melting and calorimetric studies (Eritja et al., 1986; Law et al., 1996).

Univariate measurements are usually carried out in most biophysical and biochemical studies. For example, melting studies are usually monitored at one or several wavelengths, and appropriate models must be applied for data analysis (Breslauer, 1995). Multivariate measurements and factor analysis methods have been shown to be a powerful tool, which can improve the results of univariate analysis (Henry, 1997; Ren et al., 1999; Zimanyi et al., 1999; Malinowski and Howery, 1991). In some cases, the univariate approach is enough and gives similar results to those obtained from multivariate analysis. However, when selective wavelengths or temperature regions are not available, i.e., when there is a high overlap of the signals related to the different conformations, multivariate analysis is more adequate for the study of such systems. In the analysis of melting experiments, as well as in the study of protonation equilibria of oligonucleotides, factor-analysis-based methods provide a powerful tool for exploratory analysis because they do not need the previous postulation of any chemical model before calculations. Among these methods, multivariate curve resolution based on alternating least squares (MCR-ALS) has been widely applied for the study of acid-base and confor-

Received for publication 8 May 2001 and in final form 17 July 2001.

Address reprint requests to Dr. R. Gargallo, University of Barcelona, Department of Analytical Chemistry, Marti i Franques, 1–11, E-08028 Barcelona, Spain. Tel.: 34-934034445; Fax: 34-934021233; E-mail: raimon@apolo.qui.ub.es.

0006-3495/01/11/2886/11 $2.00



SCHEME 1. Suggested base pairing schemes for P·C and P·T mispairs. (*left*) protonated Watson–Crick P·C Sowers *et al*., 1986, (*center*) wobble P·C Sowers *et al*., 1986, (*right*) Tautomer P·T Fresse, 1959.

mational transitions of polynucleotides (de Juan et al., 1997; Gargallo et al., 1997; Vives et al., 1999). This method can also handle multiple experiments simultaneously, improving resolution and reducing inherent ambiguities concerning single-data experiment factor-analysis methods (Tauler et al., 1995).

The objective of the work described here is to determine the protonation state of both P and C in two dodecamers incorporating the P·T and P·C mispairs at the experimental conditions of neutral pH, 37°C, and 150 mM ionic strength. For this purpose, acid-base titrations of oligonucleotides containing the P·T and P·C base pairs were carried out and monitored by molecular absorption, fluorescence, and circular dichroism (CD). Experimental spectroscopic data were analyzed by means of MCR-ALS to determine the number of acid-base species present along the titration and to resolve their individual concentration profiles as a function of pH and their individual pure spectra. From the calculated concentration profiles, it is possible to know the relative proportion of the different protonation states of both P and C in any point inside the pH range studied. Melting experiments were also carried out to check the validity of the concentration profiles finally proposed. To our knowledge, this is the first attempt to study the protonation state of mispairs by applying this approach.

MATERIALS AND METHODS

Reagents and solutions

Sequences 5′CGCXCCGGYGCG3′, where X = P and Y = C or T were prepared on a 1-μmol scale using standard 2-cyanoethyl phosphoramidites and the 2-cyanoethyl phosphoramidite of 2-aminopurine protected with the dimethylaminomethylidene group (Glen Research, Sterling, VA). Syntheses were performed using an automatic DNA synthesizer (Applied Biosystems model 392, Foster City, CA). Sequences were deprotected using standard protocols (concentrated ammonia at 55°C overnight). After deprotection, oligonucleotides were desalted using Sephadex G-25 columns. The DNA concentration was determined by UV absorbance measurements (260 nm) at 90°C, using for the DNA coil state the following extinction coefficients: 7500, 8500, 12,500, 15,000, and 4000 M^{-1} cm^{-1} for C, T, G, A, and P, respectively.

Solutions were prepared in Ultrapure water (Millipore, Bedford, MA) with the appropriate buffer compounds: sodium monohydrogenphosphate (Panreac, Montcada i Reixac, Spain), potassium dihydrogenphosphate (Panreac), NaOH (Merck, Darmstadt, Germany), glacial acetic acid (Merck), sodium acetate (Probus, Badalona, Spain), hydrochloric acid (Merck). NaCl (Merck) was added to adjust the ionic strength to 150 mM.

Apparatus

UV molecular absorption spectra (220–360 nm) were recorded on a Perkin-Elmer lambda-19 spectrophotometer. Fluorescence spectra were recorded on an Aminco-Bowman series 2 spectrofluorimeter (λ_{exc} = 302 nm; λ_{em} = 350–500 nm; slits, 4/4 nm). These instruments were located in the same laboratory, which enabled the simultaneous recording of molecular absorption and fluorescence data. CD spectra (220–360 nm) were recorded on a Jasco J-720 spectropolarimeter equipped with Neslab RET-110 temperature control unit. Instrumental control and data acquisition and analysis were carried out using personal computers. pH measurements were performed with an Orion model 701A pH meter (with a precision of ±0.1 mV) and a combined Ross pH electrode (Orion 9103). For all measurements, a Hellma quartz cuvette (path length of 1.0 cm) was used.

Procedure

Experimental on-line set-up for spectroscopically monitored acid-base titrations has already been described elsewhere (de Juan et al., 1997; Gargallo et al., 1997). Titrations were carried out at 37°C and 150 mM ionic strength in NaCl. Before each titration, a calibration of the electrode system was carried out according to the Gran's method (Gran, 1952) at the same experimental conditions of ionic strength and temperature. Titrations were carried out by adding increasing volumes of HCl stock solution to a slightly basic solution of oligonucleotide. At each pH value a complete spectrum (molecular absorption and fluorescence or CD) was recorded.

Melting experiments were carried out at 1°C increments with a temperature ramp of 0.5°C/min and monitored either by molecular absorption, fluorescence, or CD. At each temperature, a complete spectrum was recorded. Each sample was heated at 65°C for 5 min and allowed to equilibrate at the starting temperature for 30 min before the melting experiment was started. After each experiment, the sample was cooled to starting temperature and the final spectrum was compared with the initial one to confirm the reversibility of the process.

Data treatment

Fluorescence data obtained along the acid-base titration of P were analyzed by means of SQUAD (Leggett, 1977), a hard-modeling-based program that has been widely used for the calculation of pKa values of monomers. Hard-modeling programs are those based on the previous postulation of a set of chemical species defined by their stoichiometries and by an initial

estimation of protonation constants. Previous works have shown that this kind of hard-modeling program does not work well when applied to data analysis obtained from acid-base or complexation studies of nucleic acids due to their polymeric nature (Gargallo et al., 1996), which makes it difficult to postulate the set of stoichiometries and equilibrium constants valid for the whole experiment. Because of this, spectroscopic data obtained along acid-base titrations and melting experiments of oligonucleotides were analyzed by a method based on factor analysis, namely, MCR-ALS. The main steps of this procedure are as follows (Matlab codes are freely available at the electronic address http://www.ub.es/gesq/mcr/mcr.htm.).

1. Spectroscopic data from each experiment were collected in a table or matrix **D** (N_r rows $\times$ N_w columns), where N_r represents the spectra recorded at successive pH or temperature values and N_w is the number of wavelengths measured in every spectra.

2. The number of acid-base species or conformations, *N*, which can be detected from the spectroscopic monitoring of a titration or melting experiment, respectively, were determined. *N* was estimated by applying several methods such as singular value decomposition (SVD) (Malinowski and Howery, 1991), evolving factor analysis (EFA) (Maeder, 1987), or SIMPLISMA (Windig and Guilment, 1991). The chemical rank (rank in absence of experimental noise) of **D** calculated by any of these methods is assumed to be the number of chemical species in the system. If *N* is correctly estimated, the difference between the experimental data matrix **D** and that reproduced by principal component analysis (PCA) (**D***) (i.e., the residual data not explained by the PCA model) should be similar to the experimental error.

3. The Beer law in matrix form is solved iteratively by an alternating least-squares (ALS) algorithm to calculate the matrix of pure spectra $\mathbf{S}^T$ and the matrix of concentration profiles **C** that best fit the data in **D** for the *N* chemical species proposed. This model can be written as

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

where **E** is the matrix of the residual data not explained by the chemical species in **C** and $\mathbf{S}^T$, which should be close to the experimental error. This iterative method starts with an initial estimation of the **C** data matrix obtained from EFA for the considered *N* species. The optimization algorithm consists of two steps.

In the first step an estimation of the species spectra $\mathbf{S}^T$ is obtained from the solution of the least-squares optimization problem:

$$\min \|\mathbf{D}^* - \mathbf{C}\mathbf{S}^T\| \quad \mathbf{C}, \text{ constraints}, \tag{2}$$

where **D*** is the reproduced data matrix for the *N* species considered. For molecular absorption and fluorescence data, the species spectra must be positive. This constraint is applied accordingly during the least-squares optimization.

In the second step, a new estimation of the species concentrations **C** is obtained by solving the least-squares optimization problem:

$$\min \|\mathbf{D}^* - \mathbf{C}\mathbf{S}^T\| \quad \mathbf{S}^T, \text{ constraints} \tag{3}$$

In this case the concentrations are constrained not only to be positive but also to give unimodal concentration profiles; i.e., it can show only one maximum. Because the total concentration of oligonucleotide is constant throughout the experiment, the constraint of closure for the concentration profiles is also applied at this stage; i.e., the sum of the different species concentrations at each pH or temperature value is forced to be constant.

The first two steps are repeated until the data matrix **D*** is well explained within experimental error. Convergence is usually achieved in a few iterations of the optimization procedure.

The results obtained by factor-analysis-based methods can show some mathematical ambiguities and rank-deficiency problems, which can produce results different from the true ones (Tauler et al., 1995). This difference depends greatly on data selectivity and on local rank conditions, i.e., the degree of overlap of the pH or temperature regions where the different species coexist, or the spectral overlap. A way to overcome these difficulties is the simultaneous analysis of several data matrices corresponding to different experiments. It has been shown that when this procedure is applied, the number of possible solutions of Eq. 1 is substantially reduced and converges to a unique solution (de Juan et al., 1997; Tauler et al., 1995). In this work, the simultaneous analysis of several experiments using the MCR-ALS procedure is carried out in the following way.

1. An acid-base titration is monitored simultaneously with molecular absorption and fluorescence

2. Spectroscopic data are arranged in a row-wise augmented data matrix $[\mathbf{D}_{abs}, \mathbf{D}_{fl}]$, whose dimensions are N_r pH values by N_w wavelengths measured. N_w is now the sum of the wavelengths in the molecular absorption spectrum plus the wavelengths measured in the fluorescence spectrum; i.e., the dimensions of the augmented data matrix are $N_r \times (N_{w,abs} + N_{w,fl})$.

3. *N* is estimated as in the case of the analysis of an individual data matrix.

4. The Beer law is solved iteratively by applying the following model:

$$[\mathbf{D}_{abs}, \mathbf{D}_{fl}] = \mathbf{C}\,[\mathbf{S}_{abs}, \mathbf{S}_{fl}]^T + [\mathbf{E}_{abs}, \mathbf{E}_{fl}], \tag{4}$$

where **C** is the matrix giving the changes of concentration of the *N* species detected in the considered experiment, and it has dimensions $N_r \times N$, the same dimensions as in Eq. 1. $[S_{abs}, S_{fl}]^T$ is the row-wise augmented matrix of species spectra, where each row has for each species its pure spectrum in the two spectroscopies, molecular absorption and fluorescence. This matrix has dimensions $N \times (N_{w,abs} + N_{w,fl})$. $[\mathbf{E}_{abs}, \mathbf{E}_{fl}]$ is the row-wise augmented error matrix and it has the dimensions $N_r \times (N_{w,abs} + N_{w,fl})$. In this new data arrangement, the two individual matrices, $\mathbf{D}_{abs}$ and $\mathbf{D}_{fl}$ share their row space, defined by concentration matrix **C**; i.e., the species concentration profiles are exactly the same in the two experiments. In the analysis of the augmented matrices some new constraints may be applied during the above-mentioned ALS optimization steps: 1) correspondence between common species in the different data matrices and 2) pure spectra of common species in different experiments are equal.

With this simultaneous data matrix treatment, selectivity resolution conditions may be achieved due to the absence of signal in some spectroscopies for some species. Hence, for example, the deprotonated form of P shows a high fluorescence signal, but not C and G. The simultaneous analysis of molecular absorption and fluorescence data will help resolve the pure molecular absorption spectra, which are very overlapped, because of the selective signal of species containing deprotonated P.

The explained simultaneous analysis of molecular absorption and fluorescence data matrices could not be extended to the analysis of CD data because CD experiments were carried out independently using an instrument located in a different building.

RESULTS AND DISCUSSION

Protonation studies

First, a fluorescence-monitored acid-base titration of 2-aminopurine-2′-deoxyriboside was carried out to determine the pKa value of the nucleoside at the experimental conditions and to characterize the fluorescence signal before the study of the protonation of oligonucleotides (Fig. 1). The most interesting trend was the dramatic decrease of the fluorescence intensity upon protonation of P. Analysis of fluorescence data by means of SQUAD allowed the calculation of the pKa value for the N1 position at the experimental conditions (3.3 $\pm$ 0.1), which agreed well with

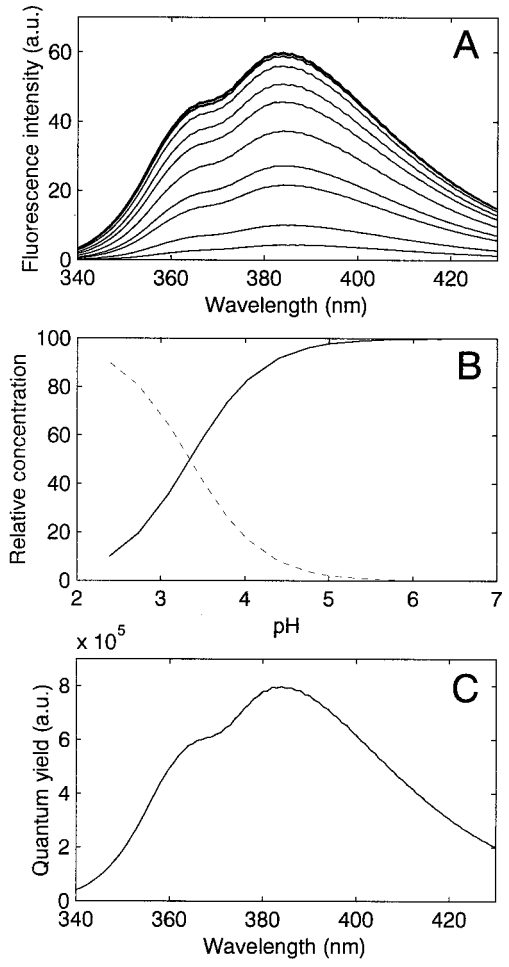


FIGURE 1 Acid-base titration of P monitored by fluorescence. (*A*) Experimental data (λ_{exc} = 311 nm; slits, 4/4 nm; 14 spectra); (*B*) Concentration profiles recovered by SQUAD (pKa = 3.3 ± 0.1); (*C*) Pure fluorescence spectra.

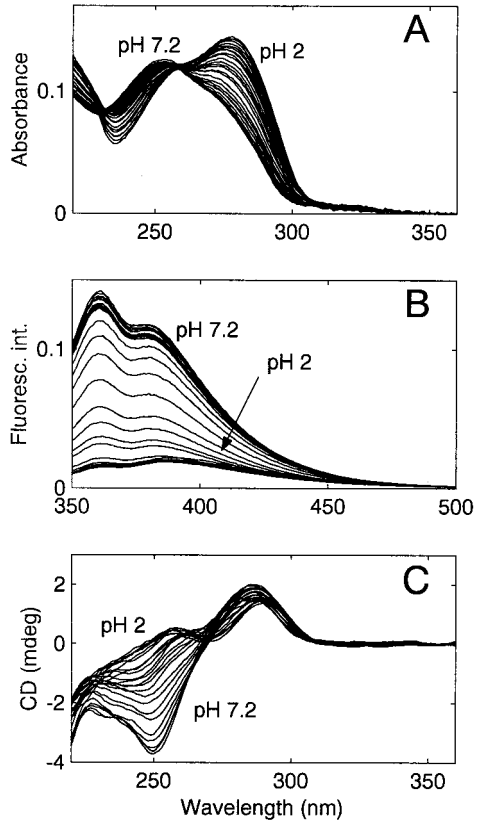


FIGURE 2 (*A* and *B*) Experimental data (32 spectra) obtained along one of the acid-base titrations of PC dodecamer monitored simultaneously by molecular absorption, data matrix $\mathbf{D}_{abs}$ (*A*), and fluorescence, data matrix $\mathbf{D}_{fl}$ (*B*); (*C*) Experimental data (23 spectra) obtained along one of the acid-base titrations of PC dodecamer monitored by CD, data matrix $\mathbf{D}_{CD}$. In both titrations, sample was 3 μM PC and 150 mM NaCl.

values found in the literature for the 2-aminopurine nucleoside: 3.40 ± 0.05 at 23°C and 50 mM ionic strength (Fox et al., 1958) or 3.2 at 22°C (Nair and Buenger, 1990). From the calculated value it was clear that for the isolated nitrogenated base at neutral pH values, only the deprotonated species of P was present. On the other hand, it should be pointed out that pKa for P was lower than that described for cytosine (pKa 4.5, in 35°C and 150 mM ionic strength) (Izatt et al., 1971).

Fig. 2, *A* and *B,* shows experimental data obtained along an acid-base titration of dodecamer carrying P·C base pairs (sequence PC: 5′-CGCPCCGGCGCG-3′) simultaneously monitored by molecular absorption (data matrix $\mathbf{D}_{abs}$) and fluorescence (data matrix $\mathbf{D}_{fl}$). When protonation occurred, the maximum of absorption spectra shifted gradually from 253 nm to 276 nm. Fluorescence signal showed a different behavior. First, a small increase of fluorescence intensity was observed in the pH range 7.8–4. Finally, the fluorescence intensity dramatically decreased in the pH range 4–2.

Molecular absorption and fluorescence spectra were arranged in an augmented data matrix $[\mathbf{D}_{abs}, \mathbf{D}_{fl}]$ obtained by joining the two data matrices, $\mathbf{D}_{abs}$ and $\mathbf{D}_{fl}$. This data matrix had the dimensions N_r = 32 pH values by N_w = 292 (141 wavelengths measured in molecular absorption and 151 wavelengths measured in molecular fluorescence). Preliminary analysis performed by SVD, EFA, and SIMPLISMA showed the presence of three main components that could be related to chemical species in the pH range 2–8; i.e., N = 3. Taking into account the number of components observed, an initial estimate of the pure spectra was obtained by SIMPLISMA. Fig. 3, *A–C*, shows the concentration profiles $\mathbf{C}$ and pure spectra $[\mathbf{S}_{abs}, \mathbf{S}_{fl}]^T$, which were obtained from the initial estimate of pure spectra by applying the MCR-ALS optimization procedure according to Eq. 4. From the pKa values for P and C and from the intensity of the resolved pure fluorescence spectra, the three chemical species were interpreted as follows: 1) PC dodecamer at neutral pH values, showing all the nitrogenated bases deprotonated (P at N1 and C at N3); 2) PC dodecamer between pH 2.5 and pH 6, approximately, showing only P deproto-

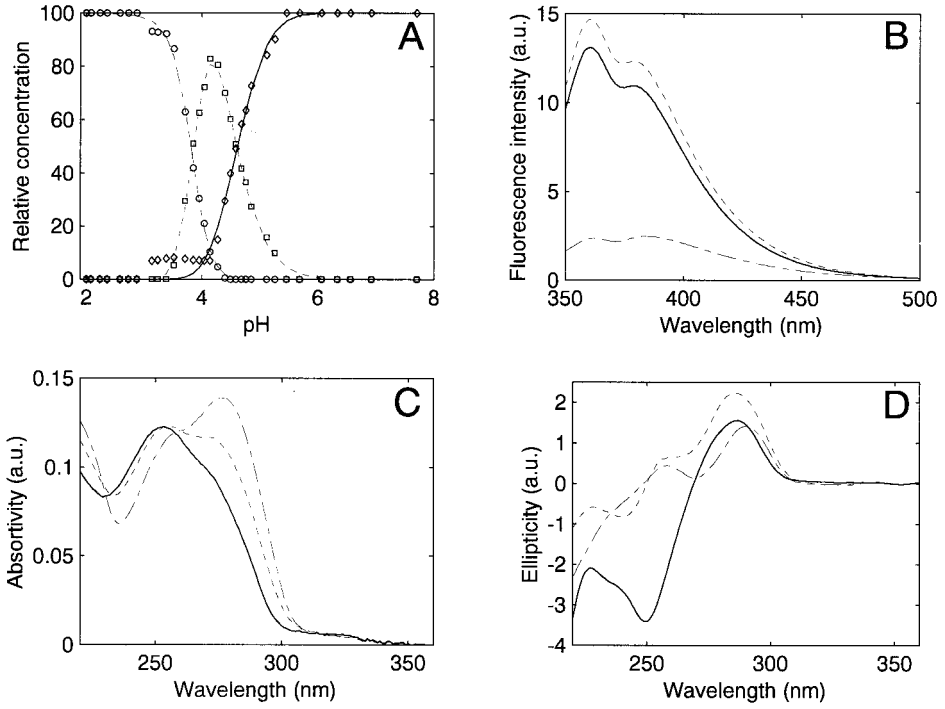


FIGURE 3 MCR-ALS results obtained for the acid-base titrations of PC dodecamer showed in Fig. 2. (*A*) Concentration profiles **C**; (*B*) Pure fluorescence spectra $\mathbf{S}_{fl}$; (*C*) Pure molecular absorption spectra $\mathbf{S}_{abs}$; (*D*) Pure CD spectra $\mathbf{S}_{CD}$. The solid line represents PC dodecamer showing both deprotonated C and P. The dotted line represents PC dodecamer with protonated C and deprotonated P. The dashed line represents PC dodecamer with both protonated C and P. **C**, $\mathbf{S}_{fl}$, and $\mathbf{S}_{abs}$ were obtained from the analysis of $[\mathbf{D}_{abs}, \mathbf{D}_{fl}]$ in Fig. 1 according to Eq. 4. $\mathbf{S}_{CD}$ values were obtained from the analysis of CD data according to Eq. 1.

nated, whereas C was protonated; and 3) PC dodecamer at pH below 4, showing both P and C protonated. For all three species, G is neutral (pKa near 9).

Pure fluorescence spectra obtained for each one of the species (Fig. 3 *B*) show the trend described in the fluorescence titration of 2-aminopurine 2′-deoxyriboside. Hence, pure spectra for the first and second species, where P is proposed to be deprotonated, show higher fluorescence intensity than the pure spectra for the third species, where P is proposed to be protonated. Moreover, the fluorescence intensity for the second species is slightly higher than for the neutral species, a fact that could reflect the partial disruption of base stacking upon protonation. This behavior has been previously described by other authors when monitoring the denaturation of oligonucleotides containing P by molecular fluorescence (Law et al., 1996).

Pure resolved molecular absorption spectra are depicted in Fig. 3 *C*. Whereas the spectra for the first and third species have been resolved without ambiguities, the pure spectrum for the second species should be considered critically because of the high overlap of the signal in molecular absorption. It should be stressed that analysis of data matrix $\mathbf{D}_{abs}$ alone by using Eq. 1 would have made it very difficult to ascertain the presence of three acid-base species because of the above-mentioned spectral overlap. However, the simultaneous analysis of $\mathbf{D}_{abs}$ and $\mathbf{D}_{fl}$ allowed the resolution because of the high selectivity of fluorescence, where only P environment is observed (Law et al., 1996).

CD-monitored acid-base titrations were also performed to confirm the results obtained from molecular absorption and fluorescence-monitored titrations of dodecamer PC. In this case, CD data were analyzed by applying Eq. 1. MCR-ALS-resolved concentration profiles **C** obtained agreed with those obtained previously. Fig. 3 *D* shows the resolved MCR-ALS pure spectra $\mathbf{S}_{CD}$ for each one of the postulated species explained above. The spectrum for the first species shows a characteristic negative band at 250 nm, typical of the B-DNA, which is lost in the CD spectra of the second and third species. That behavior could reflect a dramatic change in the PC dodecamer conformation upon protonation of cytosines, in agreement with the loss of stacking observed at 310–330 nm in molecular absorption spectra.

Data analysis was also carried out considering two and four components. Therefore, an initial estimation of the purest spectra was obtained with SIMPLISMA in both cases, and MCR-ALS was applied. When considering two components, i.e., only the totally protonated and deprotonated species, very large residuals in the pH region 3–6 were obtained, an indication that, at least, a third component was necessary to fit the data into the expected experimental

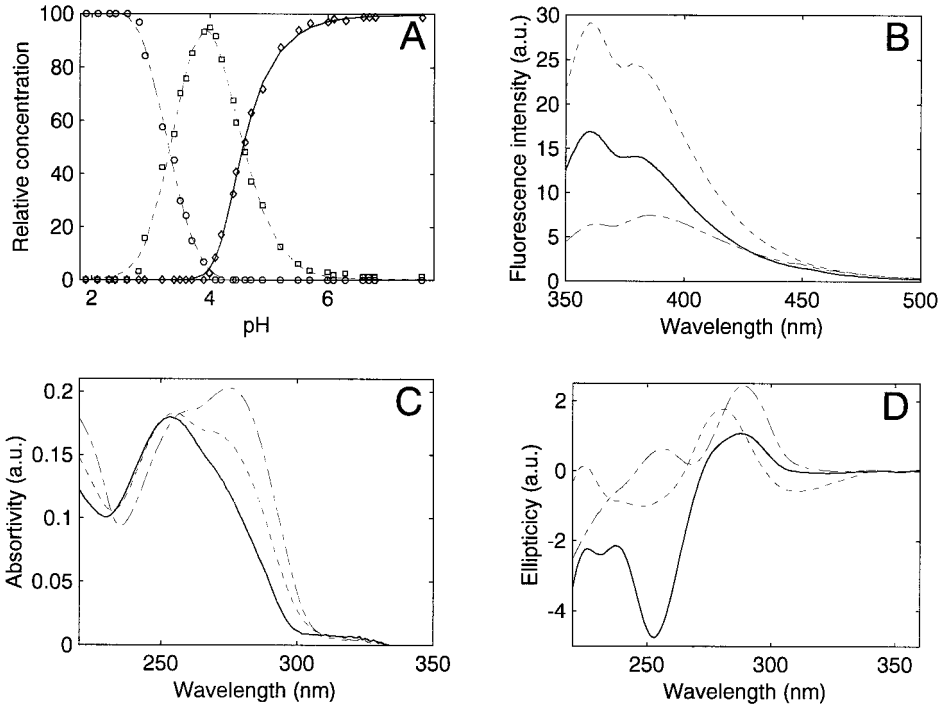


FIGURE 4 MCR-ALS results for two representative acid-base titrations of PT dodecamer. (*A*) Concentration profiles **C**; (*B*) Pure fluorescence spectra $\mathbf{S}_{fl}$; (*C*) Pure molecular absorption spectra $\mathbf{S}_{abs}$; (*D*) Pure CD spectra $\mathbf{S}_{CD}$. The solid line represents PT dodecamer showing both deprotonated C and P. The dotted line represents PT dodecamer with protonated C and deprotonated P. The dashed line represents PT dodecamer with both protonated C and P.

uncertainty. When considering four components in the data analysis of independent titrations, two trends were observed. First, the concentration profile and pure spectra for the fourth component were not reproducible. Second, both lacked chemical meaning; i.e., they described only part of the random noise contribution.

The protonation equilibria of dodecamer carrying P·T base pairs (sequence PT: 5′-CGCPCCGGTGCG-3′) showed similar trends to the above described for PC. The MCR-ALS-resolved concentration profiles (Fig. 4 *A*) also showed the presence of three main species, which were interpreted as follows: 1) PT dodecamer at neutral pH values, showing all the nitrogenated bases deprotonated (P at N1 and C at N3); 2) PT dodecamer between pH 2.5 and pH 6, approximately, showing P deprotonated and C protonated; and 3) PT dodecamer at pH below 4, showing both P and C protonated. For all three species, G and T were totally neutral (pKa near 9). The presence of T instead of C seems not to affect greatly the concentration profiles for PT dodecamer. As in the case of PC sequence, two more titrations monitored either by molecular absorption and fluorescence or by CD were carried out, and in all cases, MCR-ALS final results were similar.

It is also worthy mentioning the difference observed for the resolved pure fluorescence spectra of the main species at pH 4. For both sequences, the most intense fluorescence spectrum is that of the second species. However, for the PT sequence, this difference is greater than for the fluorescence spectra of PC dodecamer. The presence of cytosines in the P·C base pair, absent in the P·T base pair, is probably the reason for this difference.

The resolved MCR-ALS CD spectra for the first species of PC and PT, which are present at neutral pH values, are quite similar, with both showing a negative band at 250 nm. Something similar can be observed for the resolved spectra of the fully protonated species. On the contrary, the resolved MCR-ALS CD spectra for the PT species present in the pH

TABLE 1 Tm values obtained from MCR-ALS-resolved concentration profiles

Sequence	Technique	pH 4.0	pH 5.0	pH 6.0	pH 7.2
PT	Molecular absorption	37.2	51.7	56.7	57.1
	CD	23.8, 48.2	54.7	55.7	58.6
PC	Molecular absorption	45.1	46.7	46.8	47.4
	CD	25.7, 48.0	41.5, 64.0	39.0, 64.5	39.4, 63.8

range 3–6 showed a negative band at 310 nm that is not present for the corresponding species in PC.

From these results, it was proposed that the geometry of the P·C and P·T base pairs at neutral pH in PC and PT dodecamers do not involve the protonation of either of these nitrogenated bases because the protonation of cytosines starts at pH below 6 and the protonation of P starts at pH values below 4, approximately (Scheme 1). Our results did not agree with the studies of Sowers et al. (1986) obtained by NMR. These authors pointed out that the second hydrogen bond between P and C was formed by protonation of the base pair, and that was the dominant species at pH 7 at 18°C and 150 mM ionic strength. The calculated pKa value for P and the resolved concentration profiles, including the changes in the fluorescence spectrum of PC dodecamer, showed that for this sequence, under physiological conditions of temperature and ionic strength, the base pair P·C is not protonated. This conclusion is in agreement with the previous work of Fagan et al. (1996), where the wobble geometry for the P·C mispair at neutral pH was proposed. Recently, Sowers et al. (2000) proposed that the wobble geometry is the predominant species at neutral pH and 9°C in equilibrium with a small quantity of the protonated species (1–4%, approximately). Our results confirm the wobble base pair as the majority species but disagrees with the pKa of the protonation of the P·C mispair. Those authors propose a pKa 5.9 at the experimental conditions studied, whereas our results reflect a pKa near 4.7. This disagreement could be related to the different experimental conditions used in both studies and to the fact that pKa usually increases when temperature decreases (Izatt et al., 1971).

Melting experiments

Melting experiments at different pH values were carried out to confirm the conclusions obtained from protonation studies (Table 1). For every melting experiment, monitored either by molecular absorption, CD, or fluorescence, spectral data were arranged in a data matrix and analyzed by the MCR-ALS procedure according to Eq. 1. Fig. 5 shows the experimental molecular absorption data obtained along a melting experiment of PT dodecamer at pH 7.2. The loss of the initial ordered structure is reflected in an increase of the absorbance at 260–280 nm and a decrease at 320 nm. Spectroscopic data were arranged in a data matrix (29 spectra by 141 wavelengths) and analyzed by SVD, EFA, and SIMPLISMA to get the number of main factors that could be related to oligonucleotide conformations. From this preliminary analysis, only two main components were detected. An initial estimation of their pure spectra was obtained by SIMPLISMA and used in the later MCR-ALS optimization. Fig. 6 *A* shows the MCR-ALS-resolved concentration profiles for these two conformations, which were related respectively to the initial double-stranded ordered conformation and to the final random-coil conformation.

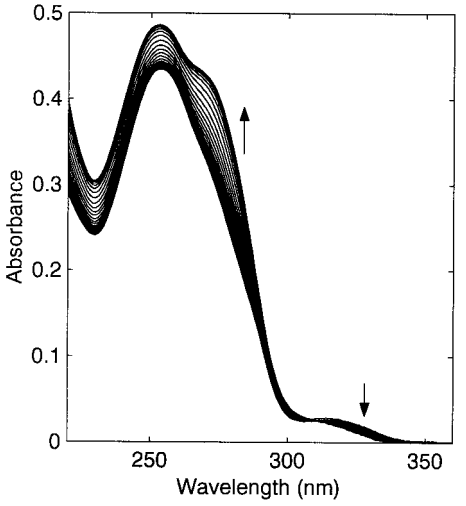


FIGURE 5 Experimental data (29 spectra) obtained along a representative melting experiment of PT dodecamer at pH 7.2 monitored by molecular absorption. Sample was 5 μM in PT, in 10 mM Na_2HPO_4, 10 mM KH_2PO_4, 150 mM NaCl. The arrows show the absorbance changes upon heating.

From the crossing point of the MCR-ALS-resolved concentration profiles for both components, a value for T_m = 57.1 ± 0.6°C (±SD) was calculated. This value was in the expected range, according to the relatively high G·C content of the oligonucleotide. Resolved molecular absorption spectra for each component (Fig. 6 *B*) also gave the expected shapes and intensities according to the observed experimental data trends.

Melting experiments for PT sequence at pH 7.2 monitored by CD and molecular fluorescence also showed the presence of only one transition. When CD experimental data were analyzed, the calculated T_m value (58.6 ± 0.6°C) was similar to the T_m value obtained from molecular absorption, which indicated that both molecular absorption and CD offered similar descriptions of the melting process for PT sequence. Whereas a molecular absorption melting curve monitors a global duplex dissociation, a fluorescence melting profile monitors the local duplex perturbations in the vicinity of the fluorophore, in this case the P residues. From fluorescence studies, a slightly lower T_m value was observed (51.5 ± 0.7°C), which was related to the breaking of the neighborhood of the P·T base pair. This observation could be consistent with the fact that local perturbations of the helix around the P·T base pairs and global disruption of the duplex are not totally coincident equilibrium events. This would be explained by the relative weakness of the P·T base pair hydrogen bonding in comparison with the G·C base pair hydrogen bonding.

At pH values 6.0 and 5.0, the calculated T_m values are slightly lower than at pH 7.2. This fact can be related to the presence of a relative increase of the species involving protonated cytosines, according to the concentration profile

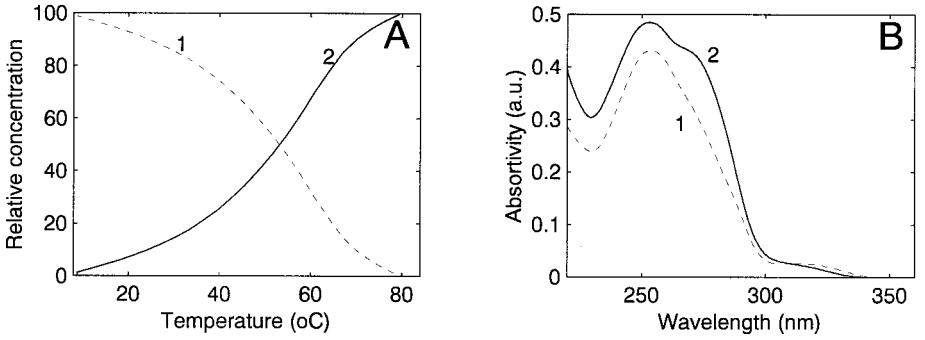


FIGURE 6 MCR-ALS results for the analysis of melting data showed in Fig. 5 by using Eq. 1. (*A*) Concentration profiles C; (*B*) Pure molecular absorption spectra $\mathbf{S}_{abs}$. The dashed line represents the initial ordered duplex, and the solid line represents the final disordered oligonucleotide.

in Fig. 4 *A*. Accordingly, cytosine bases would be protonated in such species, which in turn produced the weakness of the Watson-Crick G·C+ base pairs or the formation of Hoogsteen G·C+ base pairs and, consequently, decreasing T_m values. This trend is also observed at pH 4.0. It is worth mentioning that two transitions were observed at this pH when the melting was monitored by CD and only when monitored by molecular absorption. A similar behavior was observed for PC dodecamer and is commented on below.

Fig. 7 shows the MCR-ALS-resolved concentration profiles and pure spectra for molecular absorption and for CD-monitored melting experiments of the PC dodecamer at pH 7.2. Two transitions were observed when CD was used to monitor the melting, but only one transition was observed when molecular absorption was used. Fig. 8 shows the experimental molecular absorption and CD melting curves monitored at only one wavelength. Whereas the CD curve at 280 nm clearly reflects the presence of two transitions, it is most difficult to detect the presence of two transitions from the melting curve monitored by molecular absorption. Pre-

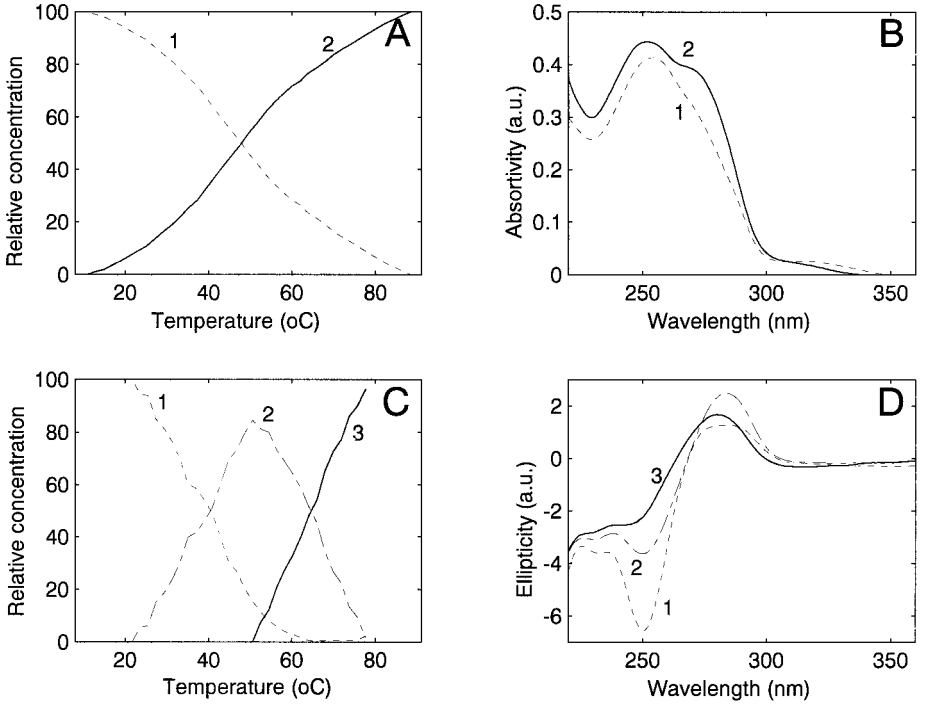


FIGURE 7 MCR-ALS results for data obtained in two melting experiments of PC dodecamer at pH 7.2. (*A* and *B*) Concentration profiles (*A*) and pure spectra (*B*) resolved from the molecular absorption monitored melting experiment; (*C* and *D*) Concentration profiles (*C*) and pure spectra (*D*) resolved from the CD monitored melting experiment. The dotted line represents the initial ordered duplex, the solid line represents the final disordered oligonucleotide, and the dashed line represents the intermediate state observed in the CD study. Sample was 4 μM in PT, in 10 mM Na_2HPO_4, 10 mM KH_2PO_4, 150 mM NaCl.

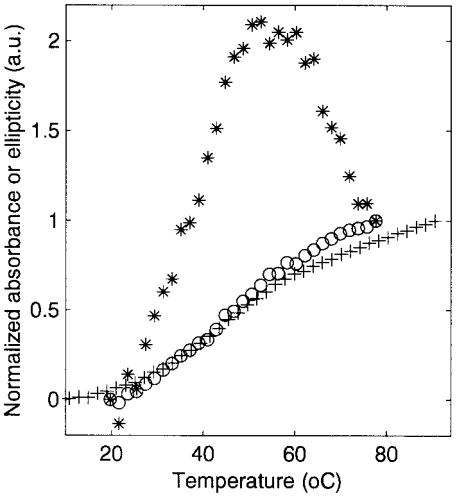


FIGURE 8 Melting curves for the PC dodecamer at pH 7.2 monitored by CD at 250 nm (○), CD at 280 nm (*), and molecular absorption at 280 nm (+).

vious reports by other authors on the existence of transitions involving very low or even non-hyperchromic transitions in molecular absorption have shown that CD melting monitored at 263 nm could provide a useful tool for the study of such non-hyperchromic transitions (Davis et al., 1998). In that work, oligonucleotides of sequence d(CGCAXXXX-CGCG) where XXXX = TTTT, ATTT, and TATT were studied. The PC sequence studied here, d(CGCPCCG-GCGCG), is identical in the first and last four base pairs, apart from P, which is similar to A, but it is different for the four central base pairs, CCGG. Davis et al. (1998) described these transitions as the formation of a hairpin instead of a duplex structure due to the extensive number of possible mismatches in the duplex state of the oligonucleotides studied. A similar case can occur for the PC oligonucleotide studied here at low temperature, where the mismatches are located on the P·C base pairs in the duplex state. When temperature was increased, hairpin formation was observed.

Isaksson et al. (1999) also described the formation of intermediate structures for the sequence 5′-CGCGAAT-TCGCG-3′ and its carboxylic-nucleotide analogs in which the pentose sugar in 2′-dA residue was replaced with 2′-deoxyaristeromycin. Based on NMR data, they proposed the existence of a hairpin structure in the temperatures range from 40°C to 60°C, which finally melts to a single-strand structure above 67°C (Isaksson et al., 1999). Also, the sequence studied by these authors has the same four first bases as our sequence. We believe that what we are monitoring by CD is a similar melting behavior: duplex → hairpin → random coil (Fig. 7 *C*).

The T_m value calculated from fluorescence data for PC dodecamer at pH 7.2 was 39.6 ± 0.7°C. This value corresponds to the first transition observed by CD; i.e., this first transition describes the breaking of the neighborhood of the

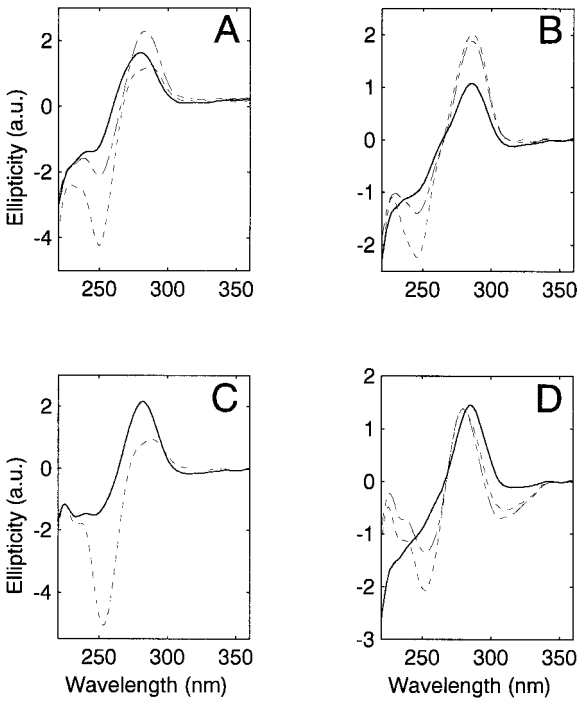


FIGURE 9 MCR-ALS-resolved CD spectra for PC dodecamer at pH 6.0 (*A*), PC dodecamer at pH 4.0 (*B*), PT dodecamer at pH 6.0 (*C*), and PT dodecamer at pH 4.0 (*D*). The dotted line represents the initial low-temperature conformation, the dashed line represents the intermediate conformation, and the solid line represents the final high-temperature random-coil conformation. Solutions were 5 μM in PT and 4 μM in PC; 10 mM KH_2PO_4, NaOH, 150 mM NaCl, pH 6.0; 10 mM sodium acetate, acetic acid, 150 mM NaCl, pH 4.0.

P·C base pair. This value is lower than the T_m calculated for the neighborhood of the P·T base pair, which should be related to the weakness of the P·C base pair in comparison with the P·T base pair, as has been pointed out from calorimetric data (Law et al., 1996).

Table 1 shows that for PC, T_m values calculated from molecular absorption or CD data did not change significantly in the whole pH range 5.0–7.2, which confirmed that only one species was predominant in that region, as depicted in Fig. 3 *A*.

Comparison of the MCR-ALS-resolved CD spectra obtained for the different conformations of P·C and P·T show some significant trends (Fig. 9). First, the band at 253 nm is very intense at pH 6 for the ordered structure of both oligonucleotides, but less intense for pH 4. Looking at the shape of the resolved CD spectra for the initial ordered structure, both for PC and PT sequences, it is clear that a new species is formed at pH 4.0, which agrees with the proposed protonation model. Finally, the CD spectrum of the disordered structure at higher temperatures is observed as showing a similar shape in all the cases, which indicates that the final structure is quite similar.

In conclusion, MCR is proposed for the analysis of spectroscopic data in the study of acid-base equilibrium and

melting processes of oligonucleotides. This multivariate approach gives more accurate results because it takes advantage of the information contained in all the measured wavelengths, not only at a single wavelength. Moreover, simultaneous analysis of multiple data sets corresponding to different experiments (three-way data analysis) can eliminate some of the ambiguities related to factor-analysis methods in the analysis of a single data set.

We analyzed the behavior of oligonucleotides containing the mutagenic base 2-aminopurine using this multivariate approach. The presence of protonated species in neutral solutions at the experimental conditions is clearly ruled out, giving new evidence for the wobble geometry for the P·C base pair. The existence of an intermediate form between duplex and random coil during the melting process at pH 7.2, probably a hairpin structure, has been demonstrated for the PC dodecamer whereas the perfectly matched PT dodecamer does not show this behavior. These results could not be clearly observed by simple univariate measurements, showing the usefulness of the multivariate approach for the analysis of conformational changes on synthetic oligonucleotides.

Petr Taborsky (Masaryk University, Brno, Czech Republic) is thanked for his help.

This study was supported by the Ministerio de Educación y Cultura (grant DGICYT PB96–0377) and the Generalitat de Catalunya (grant 1999SGR-00048). M. V. also thanks the University of Barcelona for a Ph.D. grant.

REFERENCES

Bloom, L. B., M. R. Otto, J. M. Beechem, and M. F. Goodman. 1993. Pre-steady-state kinetic analysis of sequence-dependent nucleotide excision by the 3′-exonuclease activity of bacteriophage T4 DNA polymerase. *Biochemistry.* 32:11247–11258.

Breslauer, K. M. 1995. Extracting thermodynamic data from equilibrium melting curves for oligonucleotide order-disorder transitions. *Methods Enzymol.* 259:221–242.

Davis, T. M., L. McFail-Isom, E. Keane, and L. D. Williams. 1998. Melting of a DNA hairpin without hyperchromism. *Biochemistry.* 37: 6975–6978.

Déclais, A. C., and D. M. Lilley. 2000. Extensive central disruption of a four-way junction on binding CCE1 resolving enzyme. *J. Mol. Biol.* 296:421–433.

de Juan, A., A. Izquierdo-Ridorsa, R. Gargallo, R. Tauler, G. Fonrodona, and E. Casassas. 1997. Three-way curve resolution applied to the study of solvent effect on the thermodynamic and conformational transitions related to the protonation of poly(C). *Anal. Biochem.* 249:174–184.

Eritja, R., B. E. Kaplan, D. Mhaskar, L. C. Sowers, J. Petruska, and M. F. Goodman. 1986. Synthesis and properties of defined DNA oligomers containing base mispairs involving 2-aminopurine. *Nucleic Acids Res.* 14:5869–5884.

Fagan, P. A., C. Fàbrega, R. Eritja, M. F. Goodman, and D. E. Wemmer. 1996. NMR study of the conformation of the 2-aminopurine:cytosine mismatch in DNA. *Biochemistry.* 35:4026–4033.

Fox, J. J., I. Wempen, A. Hampton, and I. L. Doerr. 1958. Thiation of nucleosides. I. Synthesis of 2-amino-6-mercpato-9-β-D-ribofuranosyl purine "Thioguanosine" and related purine nucleosides. *J. Am. Chem. Soc.* 80:1669–1675.

Freese, E. 1959. The specific mutagenic effect of base analogs on phage T4. *J. Mol. Biol.* 1:87–105.

Gargallo, R., R. Tauler, and A. Izquierdo-Ridorsa. 1996. Influence of selectivity and polyelectrolyte effects on the performance of soft-modelling and hard-modelling approaches applied to the study of acid-base equilibria of polyelectrolytes by spectrometric titrations. *Anal. Chim. Acta* 331:195–205.

Gargallo, R., R. Tauler, and A. Izquierdo-Ridorsa. 1997. Acid-base and copperII complexation equilibria of polyinosinic-polycytidylic acid. *Biopolymers.* 42:271–283.

Goodman, M. F., and R. L. Ratliff. 1983. Evidence of 2-aminopurine·cytosine base mispairs involving two hydrogen bonds. *J. Biol. Chem.* 258:12842–12846.

Gran, G. 1952. Determination of the equivalence point in potentiometric titrations. II. *Analyst.* 77:661–667.

Guest, C. R., R. A. Hochstrasser, L. C. Sowers, and D. P. Millar. 1991. Dynamics of mismatched base pairs in DNA. *Biochemistry.* 30: 3271–3279.

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

Holz, B., S. Klimasauskas, S. Serva, and E. Weinhold. 1998. 2-Aminopurine as a fluorescent probe for DNA-base flipping by methyltransferases. *Nucleic Acids Res.* 26:1076–1083.

Isaksson, J., T. Maltseva, P. Agback, X. Luo, A. Kumar, E. Zamaratski, and J. Chattopadhyaya. 1999. A single carboxylic nucleotide substitution in a 12mer DNA gives a Hoogsteen basepaired duplex till 38°C and a hairpin till 65°C: a 600 MHz NMR spectroscopic study. *Nucleosides Nucleotides.* 18:1593–1596.

Izatt, R. M., J. J. Christensen, and J. H. Rytting. 1971. Sites of thermodynamic quantities associated with proton and metal ions. *Chem. Rev.* 71:439–481.

Law, S. M., R. Eritja, M. F. Goodman, and K. J. Breslauer. 1996. Spectroscopic and calorimetric characterization of DNA duplexes containing 2-aminopurine. *Biochemistry.* 35:12329–12337.

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

Maeder, M. 1987. Evolving factor-analysis for the resolution of overlapping chromatographic peaks. *Anal. Chem.* 59:527–535.

Malinowski, E. R., and D. E. Howery. 1991. Factor Analysis in Chemistry, 2nd ed. Wiley, New York.

Nair, V., and G. S. Buenger. 1990. Hydrolysis of dideoxygenated purine nucleosides: effect of modification of the base moiety. *J. Org. Chem.* 55:3695–3697.

Raney, K. D., L. C. Sowers, D. P. Millar, and S. J. Benkovic. 1994. A fluorescence-based assay for monitoring helicase activity. *Proc. Natl. Acad. Sci. U.S.A.* 91:6644–6648.

Ren, J., X. Qu, J. B. Chaires, J. P. Trempe, S. S. Dignam, and J. D. Dignam. 1999. Spectral and physical characterization of the inverted terminal repeat DNA structure from adeno-associated virus 2. *Nucleic Acids Res.* 27:1985–1990.

Ronen. A. 1979. 2-Aminopurine. *Mutat. Res.* 75:1–47.

Sowers, L. C., Y. Boulard, and G. V. Fazakerley. 2000. Multiple structures for the 2-aminopurine-cytosine mispair. *Biochemistry.* 39:7613–7620.

Sowers, L. C., E. Eritja, F. M. Chen, T. Khwaja, B. E. Kaplan, M. F. Goodman, and V. Fazakerley. 1989. Characterization of the high pH wobble structure of the 2-aminopurine·cytosine mismatch by N-15 NMR spectroscopy. *Biochem. Biophys. Res. Commun.* 165:89–92.

Sowers, L. C., G. V. Fazakerley, R. Eritja, B. E. Kaplan, and M. F. Goodman. 1986. Base pairing and mutagenesis: observation of a protonated base pair between 2-aminopurine and cytosine in an oligonucleotide by proton NMR. *Proc. Natl. Acad. Sci. U.S.A.* 83:5434–5438.

Tauler, R., A. Smilde, and B. R. Kowalski. 1995. Selectivity, local rank, three-way data analysis and ambiguity in multivariate curve resolution. *J. Chemom.* 9:31–45.

Vives, M., R. Gargallo, and R. Tauler. 1999. Study of the intercalation equilibrium between the polynucleotide polyA-polyU and the ethidium bromide dye by means of MCR and the multivariate extension of the

continuous variation and mole ratio methods. *Anal. Chem.* 71: 4328–4337.

Watanabe, S. M., and M. F. Goodman. 1981. On the molecular biology basis of transition mutations: the frequencies of forming 2-aminpurine. cytosine and adenine.cytosine mispairs in vitro. *Proc. Natl. Acad. Sci. U.S.A.* 78:2864–2868.

Windig, W., and J. Guilment. 1991. Interactive self-modeling mixture analysis. *Anal. Chem.* 63:1425–1435.

Zimanyi, L., A. Kulcsar, J. K. Lanyi, D. F. Sears, Jr., and J. Saltiel. 1999. SVD with self-modeling applied to determine bacteriorhodopsin intermediate spectra: analysis of simulated data. *Proc. Natl. Acad. Sci. U.S.A.* 96:4408–4413.