

# Synthesis, stability and protonation studies of a self-complementary dodecamer containing the modified nucleoside 2'-deoxyzebularine

M. Vives <sup>1</sup>, R. Eritja <sup>2</sup>, R. Tauler <sup>1</sup>, V. E. Marquez <sup>3</sup>, R. Gargallo <sup>1\*</sup>

<sup>1</sup> *Department of Analytical Chemistry. University of Barcelona. Martí i Franqués, 1 – 11, E-08028 Barcelona, Spain*

<sup>2</sup> *Institut de Biologia Molecular de Barcelona-CSIC, Jordi Girona 18-26, E-08034 Barcelona, Spain*

<sup>3</sup> *Laboratory of Medicinal Chemistry, Center for Cancer Research, National Cancer Institut, Frederick, MD 21702, USA.*

\* Correspondence to: R. Gargallo; e-mail: [raimon@apollo.qui.ub.es](mailto:raimon@apollo.qui.ub.es); fax: +34-934021233

**Abbreviations:** AcOEt: ethyl acetate, ACN: acetonitrile, bzI: benzoyl, CD: circular dichroism, DCM: dichloromethane, dmf: dimethylaminomethylidene, DMT: 4,4'-dimethoxytrityl, Et<sub>3</sub>N: triethylamine, LCAA-CPG: long chain amino alkyl-controlled pore glass. MCR-ALS: Multivariate Curve Resolution-Alternating Least Squares. OD: optical density measured at 260 nm, TEAA: triethylammonium acetate.

**Keywords:** 2'-deoxyzebularine, spectroscopy, DNA structure, Multivariate Curve Resolution, factor analysis

## ABSTRACT

The nucleoside 2'-deoxyzebularine (K) was incorporated into the self-complementary dodecamer 5'CGTACGKGTACG3' by solid-phase 2-cyanoethylphosphoramidite chemistry using dimethoxytrityl (DMT) as the 5'-hydroxyl protecting group. Standard synthesis cycles using trichloroacetic acid and short ammonia treatment (50 °C for 30 min) were found to be the optimal conditions to obtain the desired dodecamer with minimum acid and basic degradation of the acid- and base-sensitive 2-pyrimidinone residue.

The protonation equilibria of the K nucleoside and of the dodecamer at 37 °C were studied by means of spectroscopically monitored titrations. For the K nucleoside, a pKa value of  $3.13 \pm 0.09$  was obtained. For the dodecamer, four acid-base species were found in the pH range 2 – 12, with pKa values of  $9.60 \pm 0.07$ ,  $4.46 \pm 0.16$  and  $2.87 \pm 0.19$ . Melting experiments were carried out to confirm the proposed acid-base concentration profiles. Finally, kinetics experiments were also carried out at several pH values to evaluate the stability of the K nucleoside and of the dodecamer. An increased stability was shown by the K nucleoside when incorporated into the dodecamer. Multivariate methods based on both hard- and soft-modeling were applied for the analysis of spectroscopic data, allowing the estimation of concentration profiles and pure spectra.

## INTRODUCTION

The incorporation of modified bases into oligonucleotides is a useful approach for the study of recognition processes between proteins and nucleic acids. These processes take place between functional groups of the protein amino acid chains and the nucleobases, and the specificity of these interactions depends on the protein's ability to recognize a characteristic nucleic acid sequence. The present paper reports the substitution of 2'-deoxycytidine for 2'-deoxyzebularine (K), which chemically corresponds to the deletion of the exocyclic amino group from cytosine. Both 2'-deoxyzebularine, (1- $\beta$ -D-deoxyribofuranosyl-1,2-dihydropyrimidin-2-one) and the corresponding riboside were identified as comparable inhibitors of cytidine deaminase (1,2), but only the riboside (zebularine) showed cellular cytotoxic against L1210 leukemia cells *in vitro* and important antitumor activity in L1210 bearing mice (3). Because of the fluorescent nature of the aglycon, modified oligonucleotides where key cytosines have been replaced with 1,2-dihydropyrimidin-2-one have been synthesised to explore the role of ribozyme function (4). Additionally, based on the reduced stability of the glycosyl bond under mildly acidic conditions (pH 3), this substitution provides a useful approach to the generation of apurinic / apyrimidinic (abasic) sites. The generation of abasic sites is a common form of DNA damage (5,6). Contrary to its behavior under acidic conditions, based-catalyzed hydrolysis causes degradation of the 1,2-dihydropyrimidin-2-one ring leading to the generation of the known mutagen, 1,3-propanedialdehyde (malondialdehyde, MDA) (1,7). MDA has been shown to react rapidly at amino group of amino acids and nucleobases such as adenine and cytosine, causing important mutations (8,9). This chemical behavior is a concern since both acidic and basic conditions are encountered during the dodecamer synthesis based on the solid-phase method by 2-cyanoethylphosphoramidite chemistry (10). Several works have described improved conditions for oligonucleotide synthesis. Some of them substitute the more common 5'-

hydroxyl protector group employed in the dodecamer synthesis, 4,4'-dimethoxytrityl (DMT), by a more acid labile group such as 9-phenylxanthene or pixyl group (11) and / or modify the basic treatment for the cleavage of the dodecamers from the supports (12).

The work described here is divided into two parts. First, the synthesis of a self-complementary dodecamer containing the K nucleoside, 5'CGTACGKGTACG3', is described. For this, and due to the 2-pyrimidinone basic catalyzed hydrolysis previously described, two different treatments for the cleavage of the dodecamer from the supports were evaluated. The objective of the second part is to characterize the protonation equilibria of the K nucleoside and of the dodecamer at 37 °C and 150 mM ionic strength. For this purpose, acid-base titrations were carried out and monitored by molecular absorption, fluorescence and circular dichroism (CD) spectroscopies. Melting experiments at different pH values were also carried out to check the validity of the concentration profiles proposed for the dodecamer. Kinetics studies of the K nucleoside and of the dodecamer at several pH values were also carried out to evaluate the acid and base catalyzed hydrolysis of 2-pyrimidinone residues.

## EXPERIMENTAL SECTION

### Reagents and solutions

*Synthesis:* HPLC grade solvents were from Merck (Germany). Standard 2-cyanoethyl phosphoramidites were obtained from Cruachem Ltd. (United Kingdom) and dry solvents employed in dodecamer synthesis from SDS (France). Analytical TLC was run on aluminum sheets coated with silica gel 60 F254 from Merck (Germany). Column chromatography was performed on silica gel 60 Å from SDS (France). The rest of reagents were purchased from Aldrich (USA) and Fluka (Switzerland). 2'-Deoxyzebularine was prepared as described in Barchi *et. al* work (2).

*Analytical studies:* samples were prepared in Ultrapure water (Millipore, USA) with the appropriate buffer compounds: sodium monohydrogenphosphate, potassium dihydrogenphosphate, NaOH, acetic acid, sodium acetate, hydrochloric acid. NaCl was added to adjust the ionic strength to 150 mM. Reagents were analytical reagent grade and were from Panreac (Spain), Probus (Spain) and Merck (Germany).

### Synthesis

*Synthesis of 5'-O-DMT-2'-deoxyzebularine:* 2'-Deoxyzebularine (2) (294 mg, 1.38 mmol) was dried after three successive evaporations of anhydrous pyridine. The residue was dissolved in 3 ml of pyridine and treated with 610 mg (1.79 mmol) of 4,4'-dimethoxytrityl chloride. After 3h of magnetic stirring at room temperature 1 ml of methanol was added to stop the reaction. The yellow solution obtained was evaporated to eliminate the solvents. The oily residue was dissolved in 20 ml of DCM and the solution washed twice in 5% NaHCO<sub>3</sub> and saturated NaCl. The organic phase was dried (Na<sub>2</sub>SO<sub>4</sub>) and concentrated to dryness. The residue was purified by column chromatography on silica gel packed with 1% pyridine /

DCM and eluted with a 0 - 8% methanol gradient in DCM. The collected fractions containing the pure product were concentrated to dryness to obtain a white foam. Yield 51%, TLC (7% methanol / DCM) Rf= 0.34. <sup>1</sup>H NMR (CDCl<sub>3</sub>) δ(ppm): 8.5 (m, 2H, H<sub>6</sub>,H<sub>4</sub>), 7.2 -7.4 (9H, phenyl DMT), 6.8 (4H, phenyl DMT), 6.2 (t, 1H, H<sub>1</sub>' ), 6.0 (m, 1H, H<sub>5</sub>), 4.5 (m, 1H, H<sub>3</sub>' ), 4.1 (m, 1H, H<sub>4</sub>' ), 3.8 (s, 6H, -OCH<sub>3</sub>(DMT)), 3.5(m, 2H, H<sub>5</sub>' ), 2.8 -2.35 (m, 2H, H<sub>2</sub>' ). <sup>13</sup>C NMR (CDCl<sub>3</sub>) δ(ppm): 41.9 (C<sub>2</sub>' ), 55.2 (CH<sub>3</sub>-O DMT), 61.8 (C<sub>5</sub>' ), 72.3 (C<sub>3</sub>' ), 86.5 (C q DMT), 87.0 (C<sub>4</sub>' ), 87.5 (C<sub>1</sub>' ), 113.3 – 144.0 (C DMT), 156.2 (C<sub>6</sub>), 158.6 (C<sub>2</sub>), 165.5 (C<sub>4</sub>).

*Synthesis of 5'-O-DMT-2'-deoxyzebularine-3'-O-2-cyanoethyl-N,N-*

*diisopropylphosphoramidite:* 5'-O-DMT-2'-deoxyzebularine (360 mg, 0.7 mmol) was dissolved in 5 ml of anhydrous DCM and 0.50 ml (2.8 mmol) of N-ethyldiisopropylamine were added. The reaction mixture was cooled by ice, and 0.20 ml (0.91 mmol) of 2-cyanoethyl N,N-diisopropyl-chlorophosphoramidite was added dropwise with a syringe. After the addition the mixture was allowed to reach room temperature. After 30 min of magnetic stirring, the progress of the reaction was checked by TLC (AcOEt / DCM / Et<sub>3</sub>N (45:45:10)). The reaction was judged completed, solvents were evaporated and the oily residue was dissolved in 30 ml of ethyl acetate and 1.5 ml of Et<sub>3</sub>N. The solution was washed twice with 5% NaHCO<sub>3</sub> and saturated NaCl. The organic phase was dried (Na<sub>2</sub>SO<sub>4</sub>) and concentrated to dryness. The residue was purified by column chromatography on silica gel packed with a AcOEt / DCM / Et<sub>3</sub>N (45:45:10) solution and eluted with AcOEt / DCM (1:1). The collected fractions containing the pure product were concentrated to dryness to obtain a white foam. Yield 71%, TLC (AcOEt / DCM / Et<sub>3</sub>N (45:45:10)) Rf= 0.60. <sup>1</sup>H NMR (CDCl<sub>3</sub>) δ(ppm): 8.5 (m, 2H, H<sub>6</sub>,H<sub>4</sub>), 8.46 – 8.39 (m,4H, 2-cyanoethyl), 7.2 -7.4 (9H, phenyl DMT), 6.8 (4H, phenyl DMT), 6.2 (t, 1H, H<sub>1</sub>' ), 6.0 (m, 1H, H<sub>5</sub>), 4.5 (m, 1H, H<sub>3</sub>' ), 4.1 (m, 1H, H<sub>4</sub>' ), 3.8 (s, 6H, -OCH<sub>3</sub>(DMT)), 3.4(m, 2H, H<sub>5</sub>' ), 2.65 -2.43 (m, 2H, H<sub>2</sub>' ), 1.0 – 1.3 (14H, diisopropyl).

$^{31}\text{P}$  NMR ( $\text{CDCl}_3$ , and 85%  $\text{H}_3\text{PO}_4$  as external reference.  $\delta(\text{ppm})$ : 149.5 and 150.0 (two diastereoisomers).

*Synthesis of 5'CGTACGKGTACG3' dodecamer:* The dodecamer was prepared in 1  $\mu\text{mol}$  scale on an Applied Biosystems 392 automatic DNA synthesizer using commercially available 2-cyanoethyl phosphoramidites of the natural bases ( $\text{A}^{\text{bzI}}$ ,  $\text{C}^{\text{bzI}}$ , T,  $\text{G}^{\text{dmf}}$ ) and the 5'-O-DMT-2'-deoxyzebularine-3'-O-2-cyanoethyl-N,N-diisopropylphosphoramidite previously synthesized. The last DMT group at the 5' end was not removed to facilitate purification. Removal of base (dmf on N2-dG and bzI on N6-dA and N4-dC) and phosphate protecting groups and cleavage from the LCAA-CPG support was evaluated with two treatments using concentrated ammonia : 1) at room temperature overnight and 2) at 55 °C for 30 min. Ammonia solutions were filtrated, centrifuged, concentrated to dryness and dissolved in Ultrapure water (Millipore). Dodecamer was purified by reverse phase HPLC. The major peak was collected and analyzed by MS. (The collected fractions of pure product were concentrated by liophilization. Yield: 83 OD<sub>260</sub> (from 5 synthesis of 1  $\mu\text{mol}$ ), MS (MALDI-TOF):  $m/z$  3645.42 [ $\text{M}+\text{Na}^+$ ], (expected  $m/z$  3646 [ $\text{M}+\text{Na}^+$ ]).

## Procedure

The experimental set-up used for spectroscopically monitored acid-base titrations has already been described elsewhere (13,14). Melting experiments were carried out at 1 °C increments with a temperature ramp of 0.5 °C / minute and monitored either simultaneously by molecular absorption and CD, or individually by fluorescence. At each temperature, a complete spectrum was recorded. Each sample was heated at 65 °C for 5 minutes and allowed to equilibrate at the starting temperature for 30 minutes before the melting experiment was started. After each experiment, the sample was cooled to starting temperature and the final

spectrum was compared to the initial one in order to confirm the reversibility of the process. Kinetics experiments were carried out at 37 °C and monitored by molecular absorption and fluorescence. The processes were considered finished when no more spectral changes were observed.

## **Apparatus**

*Synthesis:* HPLC and mass spectrometry were used for the dodecamer purification and characterization. HPLC was performed on a Shimadzu instrument as follows: solvent A was 0.1M TEAA / ACN (95:5) and solvent B was 0.1M TEAA / ACN (3:7); column: a polystyrene column, Hamilton PRP-1, 250 x 10 mm; flow rate: 3ml/min. A 0 – 80 % B linear gradient in 20 min and detection at 260 nm was applied for DMT-oligonucleotides and, a 0 – 50 % B linear gradient in 20 min and detection at 260 nm was applied for oligonucleotides after the removal of the DMT group. MALDI-TOF mass spectrometry measurements were performed in a Voyager DE-RP Applied Biosystems instrument using 3-hydroxypicolinic acid as the matrix.

*Analytical studies:* Molecular absorption spectra were recorded on a Perkin-Elmer lambda-19 spectrophotometer. Fluorescence spectra were recorded on an Aminco-Bowman series 2 spectrofluorimeter ( $\lambda_{exc}$ : 303 nm,  $\lambda_{em}$ : 325 - 500 nm). These instruments were located in the same laboratory, which enabled the simultaneous recording of molecular absorption and fluorescence spectra. CD spectra were recorded on a Jasco J-720 spectropolarimeter equipped with Neslab RET-110 temperature control unit. This spectropolarimeter allowed the calculation of molecular absorption spectra by means of the mathematical treatment of the output signal included in the software, allowing the simultaneous recording of molecular absorption and CD spectra. Hellma quartz cuvettes (path length of 1.0 cm) were used. pH



measurements were performed with an Orion model 701A pH meter (precision of  $\pm 0.1$  mV) and a combined Ross pH electrode (Orion 9103).

## DATA TREATMENT

The molecular absorption, fluorescence or CD spectra recorded along each experiment were collected in a table or matrix **D**. The dimensions of **D** were  $Nr$  rows x  $Nm$  columns, where  $Nr$  were the spectra recorded at successive pH, temperature or time values and  $Nm$  was the number of wavelengths measured in every spectra (Scheme 1a). Spectroscopic data arranged in matrices were analyzed to obtain information about the thermodynamic behavior of the system. Hence, for an acid-base titration, the information to be recovered consisted on the number of acid-base species present along the titration, their concentration profiles and the pure spectrum for each one of those species. For melting experiments, the information consisted of the number of conformations, their concentration profiles and the pure spectrum for each one of them. Finally, for kinetics experiments, the information consisted of the number of species, their kinetics profiles, and the pure spectra. Mathematically, the experimental data model can be written as:

$$\mathbf{D} = \mathbf{C} \mathbf{S}^T + \mathbf{E} \quad (\text{equation 1})$$

where **C**, **S<sup>T</sup>** and **E** are, respectively, data matrices containing the concentration profiles, the pure spectra and the residual noise not explained by the species or conformations in **C** and **S<sup>T</sup>**.

Two different approaches have been used to decompose the experimental data matrix **D** according to equation 1: the hard-modeling-based EQUISPEC procedure (15) and the soft-modeling-based Multivariate Curve Resolution based on Alternating Least Squares procedure (16). EQUISPEC is based on the postulation of an initial chemical model defined by: (1) the stoichiometries of the proposed species, (2) by approximate values of their equilibrium

constants and (3) on the compliance of the mass-action law. This procedure is especially useful when applied to the study of chemical equilibria of monomers or oligomers, where the mass-action law is fulfilled and no secondary effects related to large polymeric structures, like polyelectrolytic or polyfunctional effects or conformational changes are present. On the other hand, for large polymeric systems or when analyzing data from melting experiments, application of hard-modeling procedures is rather difficult or even impossible since it is difficult to postulate a physical and chemical model describing the observed data variance. In these cases, however, data analysis is still possible by applying soft-modeling factor analysis-based methods because these do not require the previous postulation of a physical or chemical model. Multivariate Curve Resolution based on Alternating Least Squares (MCR-ALS) has been widely applied for the study of acid-base and conformational transitions of polynucleotides (13, 14, 17-19). The mathematical bases of EQUISPEC and MCR-ALS have been previously described elsewhere and not given here for the sake of brevity (16, 19, 20).

Equation 1 can be applied to the analysis of a single data matrix or to the simultaneous analysis of several data matrices (14, 17). In this work, a simultaneous analysis was applied several times (Scheme 1b - 1d). To carry out such an analysis, an augmented data matrix is first built up by joining several data matrices before applying equation 1. These data matrices correspond either to spectra recorded with different instrumental techniques along a single experiment (row-wise augmented data matrix), to spectra recorded with a single instrumental technique in different experiments (column-wise augmented data matrix), or to a combination of the two previous data matrix arrangements, *i.e.*, to spectra recorded with different instrumental techniques throughout a single experiment combined with spectra recorded in different experiments (row- and column-wise augmented data matrix). In all these cases, this simultaneous analysis was shown to improve dramatically the reliability of the obtained results (16, 17), solving difficulties appearing in the analysis of a single data matrix such as

rank deficiency (17, 20-21). This is a difficulty commonly present in chemical reaction based systems when the number of independent reactions plus one is lower than the number of chemical species involved in these reactions (22). This fact produces that concentration profiles obtained will not be linearly independent and the number of detected species will be less than the number of species really present in the system. Deeper investigation of the analysis of augmented data matrices have also been published recently (14, 17, 23).

## RESULTS AND DISCUSSION

### Dodecamer synthesis

Dodecamer was prepared using solid-phase 2-cyanoethylphosphoramidite chemistry (10). The required synthon to incorporate 2'-deoxyzebularine into oligonucleotides was prepared starting from the corresponding nucleoside that was synthesized as previously described (2). The DMT group was selected for protection of the 5'-end. Reaction of 2'-deoxyzebularine with DMT chloride in pyridine gave the desired DMT-nucleoside that was reacted with 2-cyanoethyl-N,N-diisopropylchlorophosphine to yield the desired phosphoramidite. The 2'-deoxyzebularine phosphoramidite was purified by silica gel chromatography and characterized by proton and phosphorous NMR. Solid phase synthesis study using this phosphoramidite revealed a step-coupling yield similar to standard phosphoramidites (>98%).

On the other hand, to investigate the dodecamer stability under the basic treatment of cleavage from the support, LCAA-CPG, two syntheses at small scale were carried out. The last DMT was not removed to facilitate purification using reversed phase HPLC. One of the oligonucleotides was treated with concentrated ammonia at 55 °C for 30 min and another with concentrated ammonia at room temperature overnight. In both cases (Figure 1a) and in the area corresponding to DMT-containing oligonucleotides a major eluting peak (at 12.18 min) was found with a smaller peak near the major peak (at 11.62 min). From both syntheses these two eluting peaks were collected and, after the DMT group was removed, analyzed by means of HPLC. Chromatograms showed a single eluting peak from the collected previously eluted peak at 12.18 min (Figure 1b) and a still mixture of the major and minor components from the collected previously eluted peak at 11.62 min (Figure 1c). Molecular absorption analyses of major component showed an absorption band maximum at 260 nm characteristic of the four natural nucleosides absorption with a shoulder at 303 nm characteristic of the 2'-

deoxyzebularine absorption. Mass-spectrometry analysis of this material confirmed the mass expected for the dodecamer ( $m/z$  3645.42  $[M+Na^+]$ ). Mass-spectrometry analysis of the minor component showed a  $m/z$  3614.28  $[M+Na^+]$  peak not in agreement with the dodecamer structure without the 2-pyrimidinone ring ( $m/z$  expected for dodecamer without 2-pyrimidinone base 3552  $[M+Na^+]$ ), indicating no abasic site formation (6) under the experimental conditions of the acidic 5'-DMT group deprotection steps. This material could be an impurity or a degradation product from the basic catalyzed hydrolysis of 2'-deoxyzebularine produced by basic cleavage from the LCAA-CPG support. It was noticeable from the chromatograms that the area of the minor component was higher after treatment with concentrated ammonia at room temperature overnight than after 30 min at 50 °C. Moreover, HPLC analysis showed some small eluting peaks at higher retention times, and these were assigned to sequences still carrying some protecting groups. These small peaks did not appear in the sample treated with a 30 min treatment at 50 °C. From these results, all syntheses were finished with a concentrated ammonia treatment at 50 °C for 30 min.

### **Protonation studies**

Acid-base titrations of K nucleoside monitored using molecular absorption (Figure 2a, corresponding to data matrix  $D_{abs}$ ) and fluorescence (Figure 2b, matrix  $D_{fl}$ ) were carried out at 37 °C and 0.15 M ionic strength to determine the pKa value of the nucleoside and to characterize the fluorescence signal before its incorporation to the oligonucleotide. The main trends observed in molecular absorption spectra upon protonation were a red shift of the maximum from 303 to 308 nm and a slight decrease of intensity. Fluorescence spectra obtained at neutral pH values showed a double band with maxima at 360 and 382 nm. Upon

protonation the intensity of fluorescence increased dramatically and finally a maximum at 357 nm and a shoulder over 380 nm were measured.

Matrices  $\mathbf{D}_{\text{abs}}$  and  $\mathbf{D}_{\text{fl}}$  were grouped in a row-wise augmented data matrix with dimensions number of rows,  $Nr$ , = 18 pH values and number of columns,  $Nm$ , = 297 wavelengths (i.e., 121 wavelengths measured in molecular absorption from 240 to 360 nm and 176 wavelengths measured in fluorescence measurements from 325 to 500 nm). This augmented matrix is given in Scheme 1b and written in a shorter way (using Matlab® notation (23)) as  $[\mathbf{D}_{\text{abs}}, \mathbf{D}_{\text{fl}}]$ . The analysis of this augmented matrix by EQUISPEC (14) according to equation 1 allowed the calculation of the pKa value for the N3 position at the experimental conditions ( $3.13 \pm 0.09$ ). This value is rather close to the pKa value determined for 2-pyrimidinone base at different experimental conditions of 20 °C (pKa = 2.5, (24)). Figure 2c-d show the calculated pure spectra  $\mathbf{S}_{\text{abs}}^T$  and  $\mathbf{S}_{\text{fl}}^T$ , which agree with the experimental observations, as commented on above. Figure 2e shows the resolved concentration profiles  $\mathbf{C}$  which shows that at neutral pH the K nucleoside is fully deprotonated.

Once the protonation equilibrium of the K nucleoside was studied, a similar procedure was applied to study the acid-base behavior of the dodecamer. From basic to neutral pH the maximum of the molecular absorption spectra shifted from 261 to 256 nm (Figure 3a, matrix  $\mathbf{D}_{\text{abs}}$ ). From neutral to acidic pH, a slight shift from 256 to 261 nm and a hyperchromic effect was also observed, a fact that could reflect the base stacking disruption when protonation takes place. CD spectra recorded at basic pH values showed a negative band at 240 nm (Figure 3b, matrix  $\mathbf{D}_{\text{CD}}$ ). Upon protonation the intensity increased and the position shifted to 250 nm, which would reflect a B-helical form at neutral pH values. At more acidic pH values, the CD negative band shifted to 244 nm, indicating the partial disruption of the B-helical form due to the protonation of cytosine and adenine bases. At pH values below 2, the complete breaking of the double-stranded structure was reflected in the low signal-to-noise ratio

observed. Finally, fluorescence spectra for the pH range 7.7 – 2.2 (Figure 3c, matrix  $\mathbf{D}_f$ ) reflect a dramatic change in the fluorescence of the fully protonated dodecamer species due to the protonation of the K nucleoside. No precipitate was observed along the titrations either at acidic or basic pH values.

As molecular absorption and CD data were simultaneously recorded, an augmented [ $\mathbf{D}_{\text{abs}}$ ,  $\mathbf{D}_{\text{CD}}$ ] matrix was built up (Scheme 1b). On the other hand, as fluorescence data were obtained in an independent experiment,  $\mathbf{D}_f$  matrix had to be analyzed individually (Scheme 1a). In all the cases, data were analyzed by EQUISPEC considering the presence of three, four or five acid-base species. The reliability on those models was analyzed according to the final results obtained with each one of them, i.e., according to the chemical sense of the calculated pKa values, concentration profiles, the pure spectra and to the residual magnitude of the variance not explained by each model. The system including four species, i.e., three pKa values, was finally considered to be the most chemically meaningful. For this system, the calculated pKa values were:  $9.60 \pm 0.07$ ,  $4.46 \pm 0.16$  and  $2.87 \pm 0.19$ . Figure 3d-f show, respectively, the pure molecular absorption spectra  $\mathbf{S}_{\text{abs}}^T$ , the pure CD spectra  $\mathbf{S}_{\text{CD}}^T$ , and the pure fluorescence spectra  $\mathbf{S}_f^T$  for each one of the four proposed species. Figure 3g shows the corresponding calculated concentration profiles  $\mathbf{C}$ . The explanation for these species is as follows: at pH values around 10, the dodecamer shows all nitrogenated bases deprotonated, with neutral cytosine, adenine and 2-pyrimidinone but with anionic thymine and guanine at N3 (N(3)-T) and N1 (N(1)-G) positions, respectively. At neutral pH, the protonation of N(3)-T and N(1)-G shows the dodecamer with neutral bases. At pH around 3.5, protonation of adenine at N1 and cytosine at N3 is observed. Finally, at lower pH values, the dodecamer shows protonation of 2-pyrimidinone at N3 and guanine at N7. Concentration profiles  $\mathbf{C}$  show that double-stranded B-dodecamer is only present as a major species in a very narrow region around pH 7. The calculated pKa values for the four natural bases into the dodecamer are quite similar to the

pKa values found in literature for the free bases (25). As expected, the calculated pKa value for 2-pyrimidinone when incorporated into the dodecamer is slightly lower than that calculated for the K nucleoside, i.e., protonation is a little more difficult when the base forms hydrogen bonds with guanine bases in the dodecamer.

The calculated pure spectra (Figure 3d-f) agree well with the observed experimental trends. The protonation of cytosine and adenine bases is more clearly reflected in the calculated molecular absorption and CD pure spectra ( $S_{\text{abs}}^{\text{T}}$ ,  $S_{\text{CD}}^{\text{T}}$ ) than in the calculated fluorescence pure spectra ( $S_{\text{fl}}^{\text{T}}$ ).  $S_{\text{fl}}^{\text{T}}$  for the neutral (labeled as 3) and intermediate (labeled as 2) species are very similar. However, the protonation of the 2-pyrimidinone base at the K nucleoside (see above) is clearly reflected in the dramatic increase of fluorescence for the pure spectrum of the fully protonated species (labeled as 1).

Cooperative effects related to the polymeric structure of polynucleotides can usually be denoted in protonation equilibria by the presence of pronounced slopes in species distribution concentration profiles (13, 26). When these effects are present, the acid-base equilibria for each protonation site in the polynucleotide cannot be characterized by a single constant pKa value. However, in the present case, from the slope of the acid-base concentration profiles **C** (Figure 3g), no cooperative behavior related to the polymeric structure could be deduced, i.e. the pKa value for each one of the nitrogenated bases was not affected by the protonation / deprotonation of neighboring bases. This conclusion is reinforced by the results obtained when applying EQUISPEC, which uses a hard-modeling approach, for the study of protonation equilibria. From the results obtained in this previous research (13, 14, 18), hard-modeling methods can be used without problems for the study of protonation equilibria for oligonucleotides like dodecamers.



## Melting studies

Melting experiments at pH values 3.1, 3.9 and 6.7 were carried out to confirm the conclusions derived from the previous protonation studies. The reversibility of the unfolding processes at pH 3.1 and 3.9 were checked by recording one spectrum after a long stabilization time at the initial temperature. Reversibility of the melting process at pH 6.7 was studied in detail with an annealing experiment monitored by molecular absorption and CD. While the melting experiments at pH 3.9 and 6.7 were reversible, melting at pH 3.1 was irreversible, probably because of the labile behavior of 2-pyrimidinone residues in acidic medium and high temperature. Kinetic experiments were carried out to confirm this point (see below). Due to the irreversibility of the melting process at pH 3.1, experimental data were not analyzed in this case. At pH 6.7, the loss of the initial structure is reflected in: a) the increase of absorbance with a shoulder formation over 270 nm; b) the blue shift and decrease of the intensity of the CD band at 250 nm, typical of B-DNA; c) the fluorescence intensity decrease with the differentiation of the band over 350 nm, characteristic of the K nucleoside. Spectroscopic data were analyzed by means of the soft-modeling procedure MCR-ALS. For this purpose molecular absorption and CD spectra recorded along melting experiments at pH 3.9 and 6.7 and annealing experiment at pH 6.7 were arranged in a row-wise augmented data matrix together with the molecular absorption and the CD pure spectra resolved by EQUISPEC from the previous protonation studies of the dodecamer (Figure 3d-e). The augmented matrix used in these analyses is given in Scheme 1c and written in a shorter way as  $[[\mathbf{S}_{\text{abs}}^T, \mathbf{S}_{\text{CD}}^T]^{\text{acid-base}}, [\mathbf{D}_{\text{abs}}, \mathbf{D}_{\text{CD}}]^{\text{melting pH 3.9}}, [\mathbf{D}_{\text{abs}}, \mathbf{D}_{\text{CD}}]^{\text{melting pH 6.7}}, [\mathbf{D}_{\text{abs}}, \mathbf{D}_{\text{CD}}]^{\text{annealing pH 6.7}}]$ . The dimensions of this augmented matrix are  $Nr = 95$  (3 estimated EQUISPEC spectra plus 28 spectra measured in the melting experiment at pH 3.9 plus 31 spectra measured in the melting experiment at pH 6.7 plus 33 spectra measured in the annealing experiment at pH 6.7) and  $Nm = 244$  (122 wavelengths recorded in every technique from 235 to 356 nm).

Simultaneous MCR-ALS analysis of all these data sets resolved four components in the temperature range under study with the molecular absorption and CD pure spectra,  $S_{abs}^T$ ,  $S_{CD}^T$  (Figure 4a and 4b). Two of the calculated pure spectra of these four components were coincident with the pure spectra previously obtained for the dodecamer with neutral nucleobases and for the dodecamer with protonated cytosine and adenine bases. The other two components were assigned to unfolded random-coil conformations. Calculated concentration profiles for the melting experiment at pH 3.9 and pH 6.7 are shown in Figure 4d-e, respectively. Calculated concentration profiles for the annealing experiment at pH 6.7 are shown in Figure 4f. Figure 4d shows the presence at initial temperature values of a mixture of the dodecamer with deprotonated nucleobases and, as a main component, of the dodecamer with protonated adenine and cytosine. This agrees with the protonation species previously obtained at this pH (Figure 3g). When temperature is increased, the transition of these two species to random coil-conformations is observed. From Figure 4d, a  $T_m$  value = 42 °C was estimated from the crossing point of the two resolved concentration profiles (curves 2 and 5) for the dodecamer with protonated adenine and cytosine. A  $T_m$  value = 44 °C was determined from the crossing point of resolved concentration profiles (curves 3 and 6) for the deprotonated dodecamer at pH 3.9. Calculated concentration profiles obtained for the melting experiment at pH 6.7 (Figure 4e) showed the transition between the initial double-stranded conformation with all nucleobases deprotonated (B-form) and the final random-coil conformation. For this melting process, a value for  $T_m$  = 50.5 °C from the resolved concentration profiles C. For the annealing process at pH 6.7, a value for  $T_m$  = 50.9 °C was determined. From the lack of significant differences in the calculated  $T_m$  values for the melting and the annealing experiments, it was concluded the reversibility of the process at pH 6.7.

Fluorescence data obtained in the melting experiment at pH 6.7 were analyzed individually (Scheme 1a). In Figure 4c the resolved fluorescence species normalized spectra  $S_n^T$  are given. The value of  $T_m = 50.3$  °C was determined from the resolved concentration profiles  $C$ . The concentration profiles obtained from the fluorescence experiment at pH 6.7 were not plotted because they were very similar to Figure 4d-f. To study the possible existence of non-hyperchromic transitions assigned to intermediate hairpin-shaped structures (27), MCR-ALS analysis of experimental CD data was also performed. No intermediate structures were detected in this case.

From the results obtained, it was deduced that  $T_m$  values increase with the pH. This fact was related to the relatively more stable double-stranded structure at neutral pH values. Upon protonation, a disruption of the structure was observed and, consequently, the corresponding  $T_m$  value decreased.  $T_m$  values estimated for dodecamer at pH 3.9 and 6.7 were quite different, which confirmed that one species was predominant in each region, as depicted in Figure 3g. Simultaneous analysis of melting experiments augmented with estimated pure spectra obtained in protonation studies allowed the resolution of the melting process of the mixture of species at pH 3.9 and confirmed that at pH 6.7 only one species was predominant.  $T_m$  values calculated from molecular absorption-CD or fluorescence monitored melting experiments at pH 6.7 showed good agreement, indicating that both approaches gave a similar description of the melting process for the dodecamer sequence. Whereas the melting experiment monitored by molecular absorption showed the global perturbation of the duplex, the same experiment monitored by fluorescence showed local perturbations in the neighborhood of the fluorophore (in this case, the 2-pyrimidinone base). Because  $T_m$  values obtained from molecular absorption and fluorescence experiments were quite similar, it was concluded that the global duplex disruption is coincident with the local perturbation of the

helix around of the K·G base pair. This would be related to a similar strength of the K·G base pair in relation to the global duplex stability.

### **Kinetic studies**

Due to the acidic and basic labile behavior of the residues containing the 2-pyrimidinone base, it was decided to study the stability of the K nucleoside and of the dodecamer as a function of time during the acid-base studies. Hence, kinetics experiments of the K-nucleoside at 37 °C and at pH values 2.1, 3.1, 4.1, 9.2 and 11.9 were carried out monitored by molecular absorption and fluorescence. Finally, the labile behavior of the K nucleoside when incorporated into the dodecamer was studied at pH 3.1 at 37 °C.

The molecular absorption spectra of the K nucleoside at pH 2.1, 3.1 and 4.1 showed a progressive decrease of absorbance with a blue shift of the maximum (from 310 to 307 nm for experiment at pH 2.1, from 304 to 301 nm for experiment at pH 3.1 and from 303 to 300 nm for experiment at pH 4.1) along time. Individual analysis of these experiments by means of MCR-ALS revealed the presence of only two components at each pH value, which were assigned to the K nucleoside and to a mixture of the ribose ring and of the 2-pyrimidinone base, obtained after a N-glycosyl bond cleavage without base structure breakage (5).

However, from the concentration profile obtained along acid - base experiments (Figure 2e), it was clear that at the beginning of the kinetic experiment at pH 3.1 there should be a mixture (1:1) of protonated and deprotonated K nucleoside, being both hydrolyzed. This fact produces a rank deficiency problem (19), as has been previously shown in melting experiments (see above) and in previous works (13, 17, 18). This problem can be solved by MCR-ALS simultaneous analysis of the matrix obtained in different kinetic experiments at pH 2.1, 3.1 and 4.1 together with the molecular absorption pure spectra resolved by EQUISPEC in the previous protonation studies for protonated and deprotonated K nucleoside (Figure 2c). The

augmented matrix was built up according to Scheme 1d and written in a shorter way as  $[\mathbf{S}_{\text{abs}}^{\text{T acid-base}}, \mathbf{D}_{\text{abs}}^{\text{kinetic pH 2.1}}, \mathbf{D}_{\text{abs}}^{\text{kinetic pH 3.1}}, \mathbf{D}_{\text{abs}}^{\text{kinetic pH 4.1}}]$ . The dimensions of this augmented matrix are  $Nr = 363$  (2 estimated EQUISPEC spectra, plus 67 spectra measured in the kinetic experiment at pH 2.1, plus 110 spectra measured in the kinetic experiment at pH 3.1, plus 184 spectra measured in the kinetic experiment at pH 4.1) and  $Nm = 102$  (102 wavelengths recorded from 259 to 360 nm). The analysis revealed the presence of four components with the molecular absorption pure spectra shown in Figure 5a. Calculated molecular absorption pure spectra of two of these four components were practically identical to the pure spectra obtained for protonated and deprotonated K nucleoside (Figure 2c). The other two components were assigned to the species resulting from the hydrolysis process: a mixture of a ribose ring with a protonated nucleobase and a mixture of ribose ring and deprotonated nucleobase. Calculated concentration profiles obtained for the kinetic experiments at pH 2.1 and 4.1 (Figure 5b and 5d) showed the N-glycosyl bond cleavage with time for the protonated and deprotonated K nucleoside, respectively. From the crossing points of curves in these concentration profiles, it was observed that 50 % of the glycosyl bond hydrolysis took place after 0.23 and 6.5 hours, respectively. Calculated concentration profiles obtained for the kinetic experiment at pH 3.1 (Figure 5c) showed the glycosyl bond hydrolysis for a mixture of the protonated and of the deprotonated K nucleoside giving the corresponding mixture of ribose ring and protonated and deprotonated nucleobase. The resulting nucleobases obtained in these hydrolysis processes were also in acid-base equilibrium, which at this pH value is shifted to the deprotonated nucleobase ( $\text{pK}_a$  2.5 (24)). This equilibrium could explain the apparent absence, at the beginning of this experiment, of the expected 1:1 ratio for the mixture of deprotonated and protonated K nucleoside. From the crossing point of curves in concentration profile given in Figure 5c, it was observed that 50% of the depyrimidination process took place after 0.4 (curves 1' and 3') and 0.68 (curves 2' and 4') hours for

protonated and deprotonated K nucleoside, respectively. Concentration profiles (curve 1') given in Figure 5b-c for the hydrolysis of protonated K nucleoside at pH 2.1 and 3.1, respectively, were fitted to exponential curves and the obtained rate constant values were  $1.35 \times 10^{-3}$  and  $4.82 \times 10^{-4} \text{ s}^{-1}$ . In the same way, concentration profiles (curve 2') given in Figure 5c-d for the hydrolysis of deprotonated K nucleoside at pH 3.1 and 4.1, respectively, were fitted to exponential curves and the obtained rate constant values  $3.41 \times 10^{-4}$  and  $3.13 \times 10^{-5} \text{ s}^{-1}$  were obtained. A rate value of  $1.05 \times 10^{-4} \text{ s}^{-1}$  was previously reported (5) for the hydrolysis of K nucleoside at pH 3.0 and room temperature. This value agrees well with the values given in this work at pH 3.1, although in this previous work was not considered the acid - base equilibrium of K nucleoside. As expected, the glycosyl bond hydrolysis was pH dependent and followed pseudo first-order kinetics which can be expressed for the hydrolysis of protonated and deprotonated 2-pyrimidinone respectively, as:

$$d[\text{dKH}] / dt = -k_1' [\text{dKH}]^n \quad \text{and} \quad d[\text{dK}] / dt = -k_2' [\text{dK}]^n \quad (\text{equation 2})$$

where  $k_1'$  and  $k_2'$  are pseudoconstants which depends on the concentration of proton (i.e. depending on pH).

Molecular absorption spectra recorded in a kinetic experiment at pH 9.2 (72 hours) showed few changes, a finding which was related to the higher stability of the glycosyl bond at this pH value. It was only observed a small absorbance change at 270 nm. Whereas these changes at pH 9.2 were rather small, noticeable changes were however observed at pH 11.9. MCR-ALS analysis determined the presence of three main components along the kinetics at this pH 11.9. Figure 6 shows pure spectra  $S_{\text{abs}}^T$  and concentration profiles  $C$ . It was observed that the third species is the only species after 75 hours of reaction. This species was assigned to the breakage of the 2-pyrimidinone ring. The second species was related to the intermediate species formed during the nucleophilic attack by hydroxyl anions over the 2-pyrimidinone positions C5=C6. Finally, the first species is the initial K nucleoside (1, 9).

Kinetics of the dodecamer sample was also studied at pH 3.1 for 24 hours. During this time, no spectral changes were observed which could be related to the hydrolytic cleavage of the glycosyl bond in the dodecamer. From the comparison of these results with those obtained for the K nucleoside in the same experimental conditions, it was concluded that the stability of the K nucleoside increased dramatically upon incorporation into the dodecamer. Moreover, the observed stability was still enough to carry out the protonation studies satisfactorily without considering the degradation of the dodecamer. Finally, the non-reversibility observed for the dodecamer unfolding experiment at pH 3.1 probably showed a temperature dependence of depyrimidination process of the K nucleoside.

## **ACKNOWLEDGEMENTS**

This study was supported by the Ministerio de Educación y Cultura (grant MCYT BQ42000-0788) and the Generalitat de Catalunya (grant 2001SGR00056). M.V. also thanks the University of Barcelona for a Ph.D grant. Maqbool A. Siddiqui is thanked for his help.

## REFERENCES

1. Barchi, J.J., Musser, S. and Marquez, V.E. (1992) The decomposition of 1-( $\beta$ -D-ribofuranosyl)-1,2-dihidropyrimidin-2-one (zebularine) in alkali: mechanism and products. *J. Org. Chem.*, **57**, 536-541
2. Barchi, J.J., Haces, A., Marquez, V.E. and McCormack, J.J. (1992) Inhibition of cytidine deaminase by derivatives of 1-( $\beta$ -D-ribofuranosyl)-1,2-dihidropyrimidin-2-one (zebularine). *Nucleosides & Nucleotides*, **11**, 1781-1793
3. Driscoll, J.S., Marquez, V.E., Plowman, J., Liu, P.S., Kelley, J.A. and Barchi, Jr., J.J. (1991) Antitumor properties of 1(1H)-pyrimidinone riboside (zebularine) and its fluorinated analogues. *J. Med. Chem.*, **34**, 3280-3284.
4. Adams, C.J., Murray, J.B., Arnold, J.R.P. and Stockley, P.G. (1994) Incorporation of a fluorescent nucleotide into oligoribonucleotides. *Tetrahedron Lett.*, **35**, 1597-1600.
5. Iocono, J.A., Gildea, B. and McLaughlin, L.W. (1990) Mild acid hydrolysis of 2-pyrimidinone-containing DNA fragments generates apurinic / apyrimidinic sites. *Tetrahedron Lett.*, **31**, 145-178
6. Berthet, N., Constant, J.F., Demeunynck, M., Michon, P. and Lhomme, J. (1997) Search for DNA repair inhibitors: selective binding of nucleic base – acridine conjugates to a DNA duplex containing an abasic site. *J. Med. Chem.*, **40**, 3346-3352
7. Baik, M.H., Friesner, R.A. and Lippard, S. (2002) Theoretical study on the stability of N-glycosyl bonds: Why does N7-platinations not promote depurination, *J. Am. Chem. Soc.*, **124**, 4495-4503
8. Basu, A.K., O'Hara, S.M., Valladier, P., Stone, K., Mols, O. and Marnett, L.J. (1988) Identification of adducts with malondialdehyde and structurally related aldehydes, *Chem. Res. in Toxicol.*, **1**, 53-59
9. Nair, V., Turner, G.A. and Offerman, R.J. (1984) Novel adducts from the modifications of nucleic acid bases by malondialdehyde. *J. Am. Chem. Soc.*, **106**, 3370-3371
10. Brown, T., Brown, D.J.S. in Eckstein, F. (ed.). (1991) Oligonucleotides and analogues, a practical approach, IRL Press, Oxford.
11. Gildea, B. and McLaughlin, L.W. (1989) The synthesis of 2-pyrimidinone nucleosides and their incorporation into oligodeoxynucleotides, *Nucleic Acids Res.*, **17**, 2261-2281
12. Zhou, Y. and Ts'o, P.O. (1996) Solid-phase synthesis of oligo-2-pyrimidinone-2'-deoxyribonucleotides and oligo-2-pyrimidinone-2'-deoxyriboside methylphosphonates. *Nucleic Acids Res.*, **24**, 2652-2659
13. Gargallo, R., Tauler, R. and Izquierdo-Ridorsa, A. (1997) Acid-base and Copper(II) Complexation Equilibria of Poly(inosinic)-Poly(Cytidylic) acid. *Biopol.*, **42**, 271 –283
14. Vives, M., Gargallo, R. and Tauler, R. (2001) Three-way multivariate curve resolution applied to speciation of acid-base and thermal unfolding transitions of an alternating polynucleotide. *Biopol.*, **59**, 477 - 488



15. Dyson,R., Kaderli,S., Lawrence,G.A., Maeder,M. and Zuberbühler,A.D. (1997) Second order global analysis: the evaluation of series of spectrophotometric titrations for improved determination of equilibrium constants. *Anal. Chim. Acta*, **353**, 381-393
16. Tauler,R., Smilde,A. and Kowalski,B.R. (1995) Selectivity, local rank, three-way data analysis and ambiguity in multivariate curve resolution. *J. Chemom.*, **9**, 31-58
17. Jaumot, J., Escaja,N., Gargallo,R., González,C., Pedroso,E. and Tauler,R. (2002) Multivariate curve resolution: a powerful tool for the analysis of conformational transitions in nucleic acids. *Nucleic Acids Res.*, **30**, e92
18. Gargallo,R., Vives,M., Tauler,R. and Eritja.R. (2001) Protonation studies and multivariate curve resolution on oligodeoxynucleotides carrying the mutagenic base 2-aminopurine. *Biophys. J.*, **81**, 2886-2896
19. Vives,M., Gargallo,R. and Tauler, R. (1999) Study of the intercalation equilibrium between the polynucleotide poly(adenylic)-poly(uridylic) acid and the ethidium bromide dye by means of multivariate curve resolution and multivariate extension of the continuous variation and mole ratio methods. *Anal. Chem.*, **71**, 4328-4337
20. Izquierdo-Ridorsa,A., Saurina,J., Hernández-Cassou,S. and Tauler, R. (1997) Second order Multivariate Curve Resolution applied to rank deficiency data obtained from acid-base spectrophotometric titrations of mixtures of nucleic acids. *Chemom. Intel. Lab. Syst.*, **38**, 183-196
21. Saurina,J., Hernández-Casou,S., Tauler,R. and Izquierdo-Ridorsa,A. (1998) Multivariate Resolution of rank-deficient spectrophotometric data from first-order kinetic decomposition reactions. *J. Chemom.*, **12**, 183-203
22. Amrhein,M., Sirmivaman,B, Bonvin,D. and Schumacher, M. (1996) On the rank deficiency and rank augmentation of the spectral measurement matrix, *Chemom. Intel. Lab. Syst.*, **33**, 17-33
23. Matlab® routines available at <http://www.ub.es/gesq/mcr/mcr.htm>
24. Fox,J.J., Van Praag,D., Wempen,I., Doerr,I.L., Cheong,L., Knoll,J.E., Eidinoff,M.L., Bendich,A.B. and George, B. (1959) Thiation of nucleosides. II. Synthesis of 5-methyl 2'-deoxycytidine and related pyrimidine nucleosides. *J. Am. Chem. Soc.*, **81**, 178-87
25. Izatt,R.M., Christensend,J.J. and Rytting,J.H. (1971) Sites and thermodynamic quantities associated with proton and metal ions. *Chem. Rev.*, **71**, 439-480
26. Gargallo,R., Tauler,R. and Izquierdo-Ridorsa,A. (1996) Influence of selectivity and polyelectrolyte effects on the performance of soft-modeling and hard-modeling approaches applied to the study of acid-base equilibria of polyelectrolytes by spectrometric titrations. *Anal. Chim. Acta*, **331**, 195 – 205
27. Davis,T.M., McFail-Isom,L., Keane,E. and Williams,L.D. (1998) Meltings of a DNA hairpin without hyperchromism. *Biochem.*, **37**, 6795-6798

## FIGURE CAPTIONS

**Scheme 1.** Different arrangements of matrices used in data analysis. (a) Analysis of a single matrix containing molecular absorption, CD or fluorescence spectral data recorded along an acid-base, thermal or kinetic experiment. Simultaneous analysis of a set of correlated data matrices by (b) row-wise matrix augmentation of a molecular absorption and CD (or fluorescence) acid-base study, ( $[\mathbf{D}_{\text{abs}}, \mathbf{D}_{\text{CD}}]_{(n)}$ ) (c) row- and column-wise matrix augmentation of a molecular absorption and CD thermal study together with the pure spectra resolved by EQUISPEC in the previous protonation studies, ( $[[\mathbf{S}_{\text{abs}}^{\text{T}}, \mathbf{S}_{\text{CD}}^{\text{T}}]^{\text{acid-base}}, [\mathbf{D}_{\text{abs}}, \mathbf{D}_{\text{CD}}]^{\text{pH } 1}, \dots; [\mathbf{D}_{\text{abs}}, \mathbf{D}_{\text{CD}}]^{\text{pH } n}]$ ). (d) column-wise matrix augmentation of a molecular absorption kinetic study together with the pure spectra resolved by EQUISPEC in previous protonation studies, ( $[\mathbf{S}_{\text{abs}}^{\text{T}}]^{\text{acid-base}}, \mathbf{D}_{\text{abs}}^{\text{kinetic } 1}, \dots; \mathbf{D}_{\text{abs}}^{\text{kinetic } n}]$ ).

**D:** data matrix; **C:** concentration profiles matrix, **S<sup>T</sup>** pure spectra matrix; **E:** residual matrix; **N:** active spectroscopic species, conformations or degradation products.

**Figure 1.** (a) Reversed-phase HPLC analysis from the 5'CGTACGKGTACG3' synthesis ended with the last 5'-DMT protector group, with a 0 – 80 % B linear gradient in 20 min. (b) Reversed-phase HPLC analysis after extraction of the 5'-DMT protector group from the collected peak eluted in (a) at 12.18 min, with a 0 – 50 % B linear gradient in 20 min. (c) Reversed-phase HPLC analysis after extraction of the 5'-DMT protector group from the collected peak eluted in (a) at 11.62 min, with a 0 – 50 % B linear gradient in 20 min. Analysis carried out with a polystyrene column Hamilton PRP-1, 250 x 10 mm; flow rate: 3ml/min; solvent A was 0.1M triethylammonium acetate/acetonitrile (95:5) and solvent B was 0.1M triethylammonium acetate/acetonitrile (3:7).

**Figure 2.** Protonation studies of K nucleoside. (a) Experimental molecular absorption spectra recorded along the pH range 2.1 – 7.4,  $\mathbf{D}_{\text{abs}}$ . (b) Experimental fluorescence spectra recorded in the same pH range,  $\mathbf{D}_{\text{fl}}$ . (c) Pure molecular absorption spectra for the acid-base species calculated by EQUISPEC,  $\mathbf{S}_{\text{abs}}^{\text{T}}$ . (d) Pure fluorescence spectra,  $\mathbf{S}_{\text{fl}}^{\text{T}}$ . (e) Concentration profile, **C**. Dashed line (1'): protonated base. Continuous line (2'): deprotonated base. The arrows indicate the pH values where K nucleoside kinetics experiments were carried out.

**Figure 3.** Protonation studies of the 5'CGTACGKGTACG3' dodecamer. (a) Experimental molecular absorption spectra recorded along the pH range 1.9 – 10.4,  $\mathbf{D}_{\text{abs}}$ . (b) Experimental CD spectra recorded along the same pH range,  $\mathbf{D}_{\text{CD}}$ . (c) Experimental fluorescence spectra recorded along the pH

range 2.2 – 7.7 ,  $\mathbf{D}_{fl}$ . (d) Pure molecular absorption spectra for the acid-base species calculated by EQUISPEC,  $\mathbf{S}_{abs}^T$ . (e) Pure CD spectra,  $\mathbf{S}_{CD}^T$ . (f) Pure fluorescence spectra,  $\mathbf{S}_{fl}^T$ . (g) Concentration profile,  $\mathbf{C}$ . Symbol ('star') indicates the pH values where melting experiments were carried out. The arrow indicates the pH value where dodecamer kinetics experiment was carried out. (1) dodecamer with all the bases protonated; (2) dodecamer with protonated cytosine and adenine and deprotonated 2-pyrimidinone, thymine and guanine; (3) dodecamer with neutral bases; (4) dodecamer with neutral cytosine, adenine and 2-pyrimidinone but with anionic thymine and guanine.

**Figure 4.** MCR-ALS results of the analysis of melting experiments of 5'CGTACGKGTACG3' dodecamer (a) Pure molecular absorption spectra  $\mathbf{S}_{abs}^T$  and (b) pure CD spectra  $\mathbf{S}_{CD}^T$  for the species calculated by simultaneous MCR-ALS analysis of melting at pH 3.9, melting at pH 6.7 and annealing at pH 6.7. (c) Pure normalized fluorescence spectra,  $\mathbf{S}_{fl}^T$ , for the melting experiment at pH 6.7 (d) Concentration profiles for the melting at pH 3.9. (e) Concentration profile for the melting at pH 6.7. (f) Concentration profile for the annealing at pH 6.7. (2) and (3) as in Figure 3; (5) random coil conformation from the dodecamer with protonated adenine and cytosine and deprotonated thymine, guanine and 2-pyrimidinone; (6) random coil conformation obtained from the dodecamer with neutral nucleobases. Preparation of solutions: 8.2 mM acetic acid, 1.93 mM sodium acetate, 150 mM NaCl, pH 3.9; 10 mM  $\text{KH}_2\text{PO}_4$ , 10 mM  $\text{Na}_2\text{HPO}_4$ , 149 mM NaCl: pH 6.7.

**Figure 5.** Kinetics experiments of K nucleoside. (a) Pure molecular absorption spectra  $\mathbf{S}_{abs}^T$ . (b) Concentration profile  $\mathbf{C}$  at pH 2.1. (c) Concentration profile  $\mathbf{C}$  at pH 3.1, (d) Concentration profile  $\mathbf{C}$  at pH 4.1. (1') protonated K nucleoside; (2') deprotonated K nucleoside; (3') mixture of ribose and protonated nucleobase after glycosyl bond cleavage; (4') mixture of ribose and deprotonated K nucleobase after glycosyl bond cleavage. Preparation of solutions: 17.7 mM HCl, 147.8 mM NaCl, pH 2.1; 4.7 mM HCl, 149.9 mM NaCl, pH 3.1; 8.2 mM acetic acid, 1.93 mM sodium acetate, 150 mM NaCl pH 4.1.

**Figure 6.** Kinetics experiment of K nucleoside at pH 11.9 (18.4 mM NaOH, 135.1 mM NaCl) (a) Pure spectra resolved by MCR-ALS,  $\mathbf{S}_{abs}^T$ . (b) Concentration profiles resolved by MCR-ALS,  $\mathbf{C}$ . Inset: amplification of the interval 0 - 2h. Continuous line: K nucleoside. Dashed line: intermediate from the nucleophilic attack by hydroxyl ions over the 2-pyrimidinone positions C5=C6. Dotted line: breakage of the 2-pyrimidinone ring.

# Scheme 1

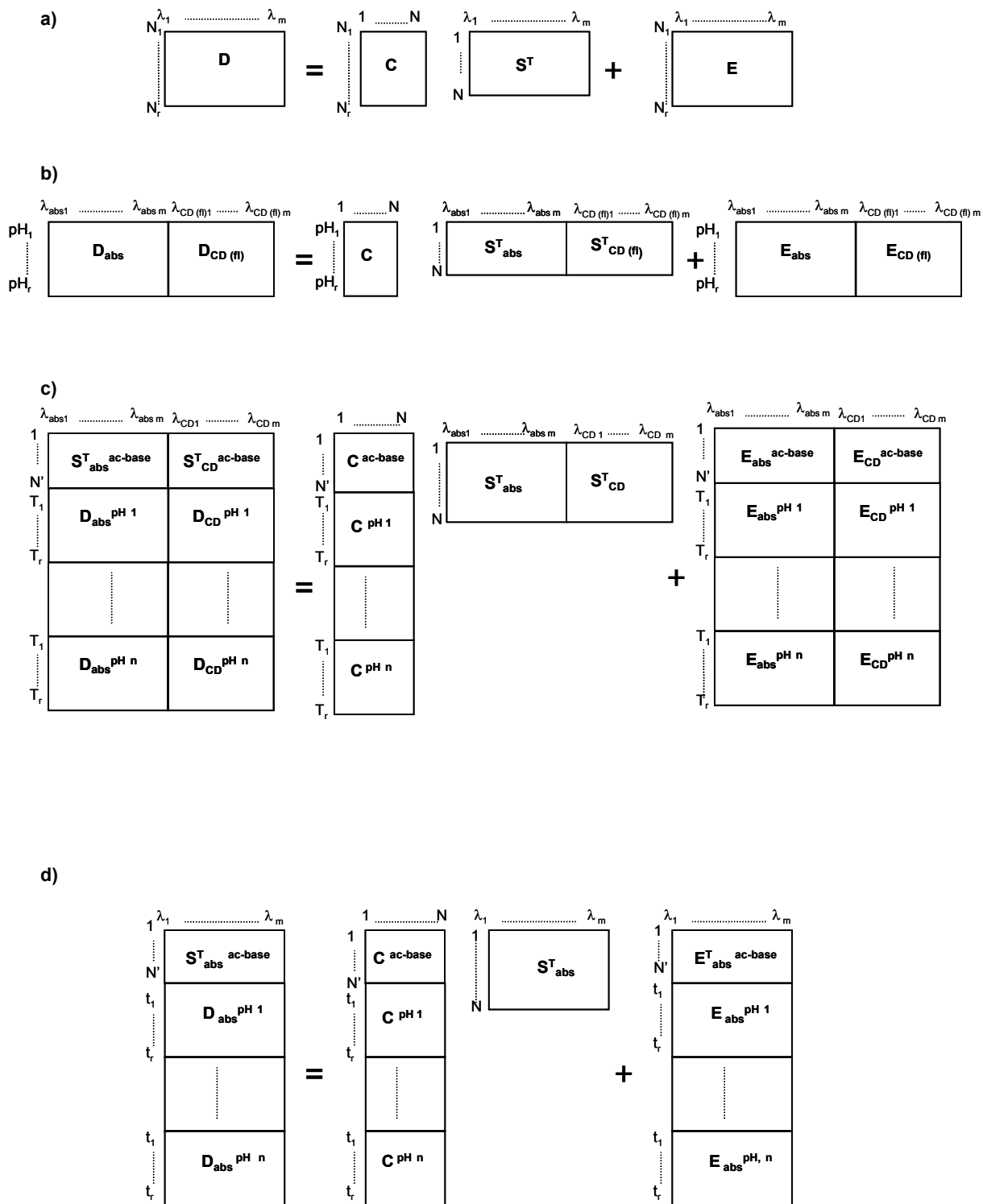


Figure 1

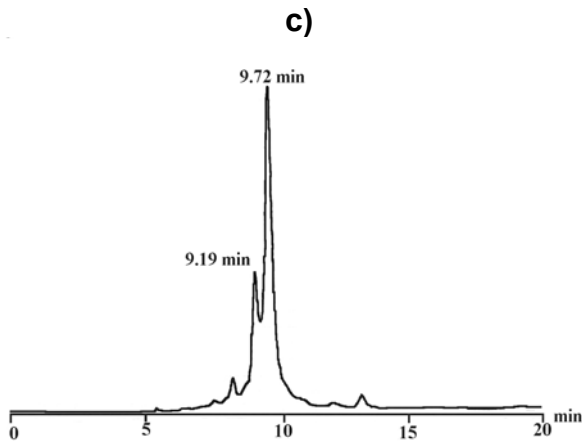
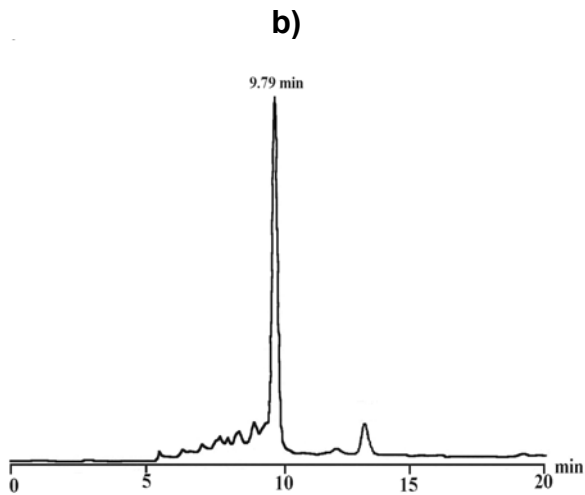
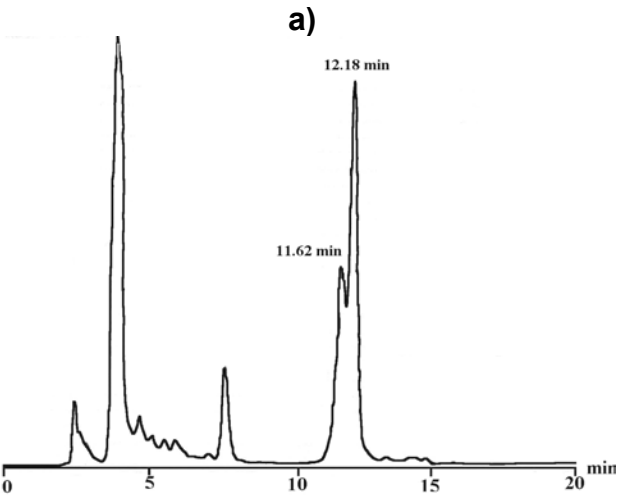
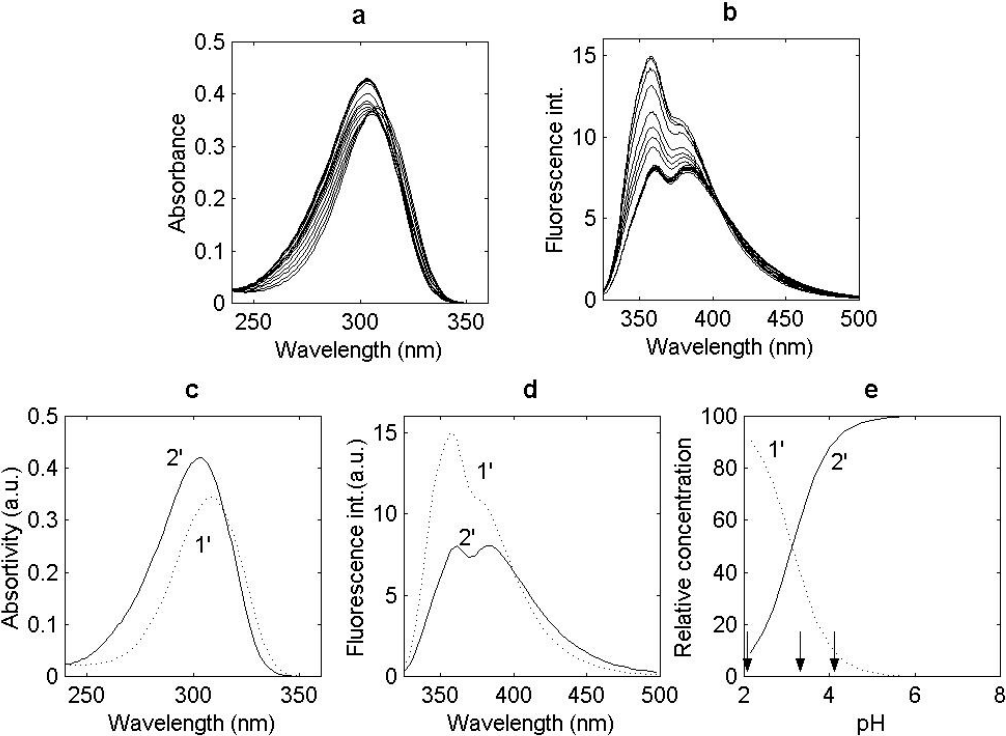


Figure 2



**Figure 3**

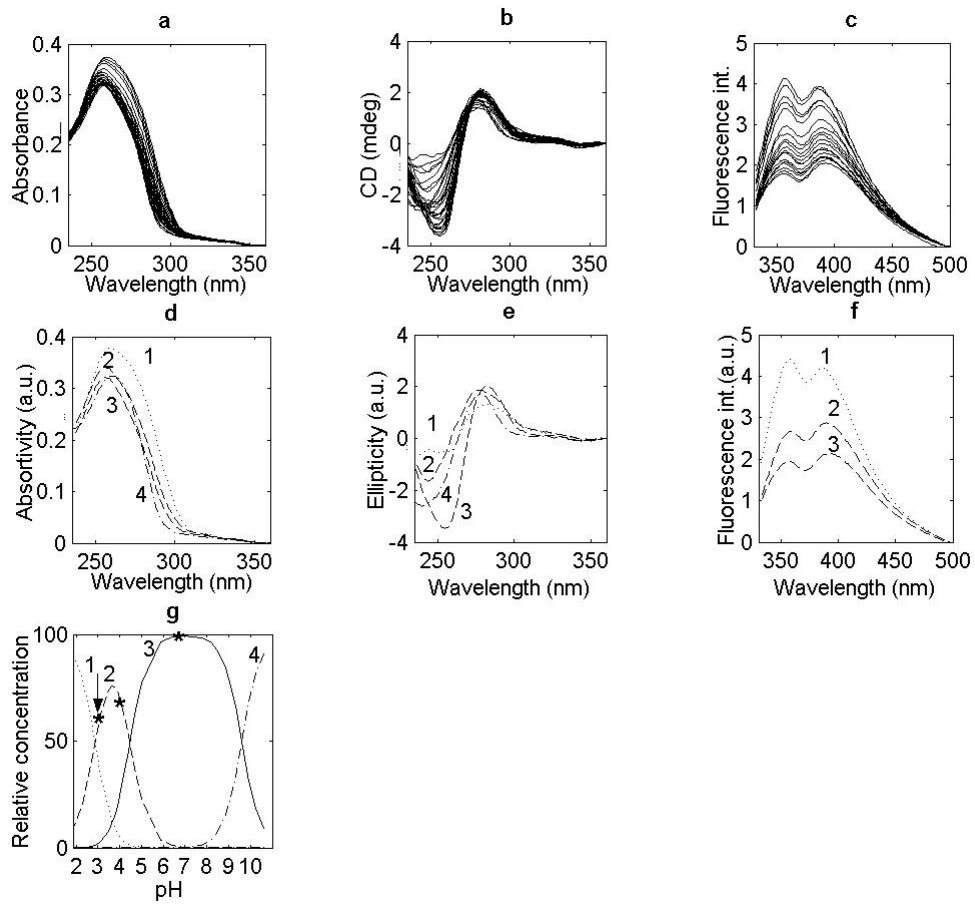


Figure 4

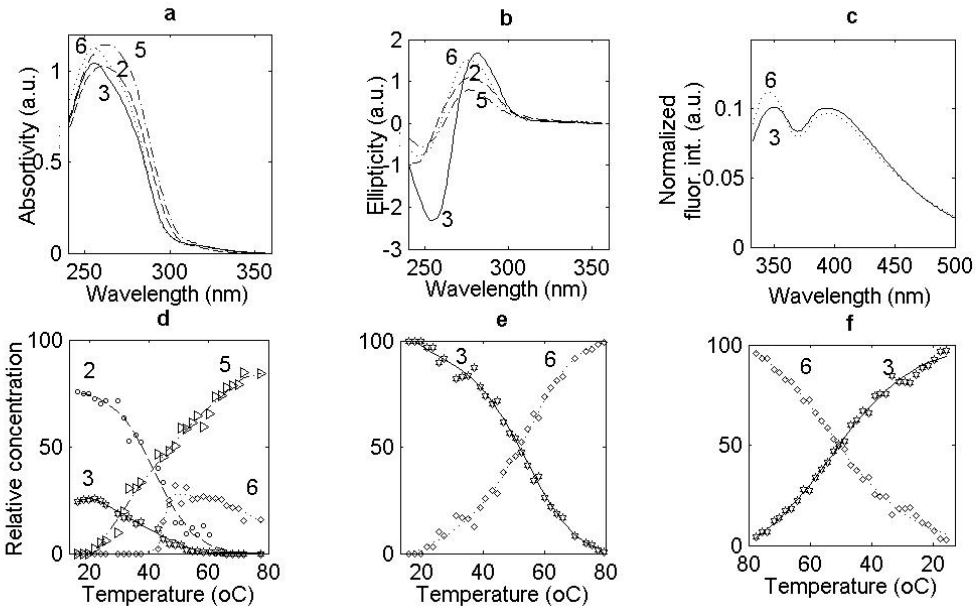




Figure 5

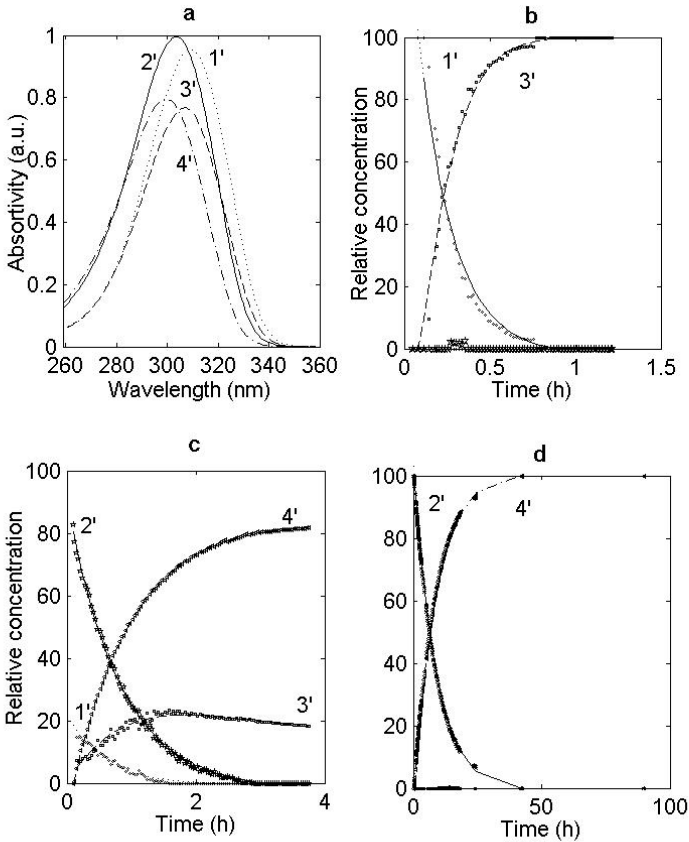


Figure 6

