Published online January 5, 2006

206–216 *Nucleic Acids Research, 2006, Vol. 34, No. 1*
doi:10.1093/nar/gkj421

Resolution of a structural competition involving dimeric G-quadruplex and its C-rich complementary strand

Joaquim Jaumot, Ramon Eritja[1], Romà Tauler[2] and Raimundo Gargallo*

Department of Analytical Chemistry, University of Barcelona, Diagonal 647, Barcelona, E-08028 Spain, [1]Department of Structural Biology, IBMB-CSIC, Jordi Girona 18-26, Barcelona, E-08034 Spain and [2]Department of Environmental Chemistry, IIQAB-CSIC, Jordi Girona 18-26, Barcelona, E-08034 Spain

Received July 26, 2005; Revised November 28, 2005; Accepted December 14, 2005

ABSTRACT

The resolution of the dimeric intermolecular G-quadruplex/duplex competition of the telomeric DNA sequence 5′-TAG GGT TAG GGT-3′ and of its complementary 5′ ACC CTA ACC CTA-3′ is reported. To achieve this goal, melting experiments of both sequences and of the mixtures of these sequences were monitored by molecular absorption, molecular fluorescence and circular dichroism spectroscopies. Molecular fluorescence measurements were carried out using molecular beacons technology, in which the 5′-TAG GGT TAG GGT-3′ sequence was labelled with a fluorophore and a quencher at the ends of the strand. Mathematical analysis of experimental spectroscopic data was performed by means of multivariate curve resolution, allowing the calculation of concentration profiles and pure spectra of all resolved structures (dimeric antiparallel and parallel G-quadruplexes, Watson–Crick duplex and single strands) present in solution. Our results show that parallel G-quadruplex is more stable than antiparallel G-quadruplex. When the complementary C-rich strand is present, a mixture of both G-quadruplex structures and Watson–Crick duplex is observed, the duplex being the major species. In addition to melting temperatures, equilibrium constants for the parallel/antiparallel G-quadruplex equilibrium and for the G-quadruplex/duplex equilibrium were determined from the concentration profiles.

INTRODUCTION

Telomeres are the physical ends of eukaryotic chromosomes comprising specialized nucleoprotein structures essential for the stability and complete replication of chromosomes (1). Human telomeres consist of DNA tandem repeats of the sequence d(T_2AG_3), with most of them as double-stranded except for the extreme terminal part where the 3′ region of the G-rich strand is single stranded (2). In these overhanging G-rich strands, four-stranded G-quadruplex structures involving planar G-tetrads can be folded. DNA G-quadruplexes have generated considerable attention because they could be molecules of interest for drug design (3). G-quadruplex formation has been shown *in vitro* to inhibit the elongation of telomerase, which is an enzyme essential for the immortalization of tumour cells (4). Recently, a chair-form intramolecular G-quadruplex structure has been proved to suppress the transcriptional activation of the *c-myc* oncogene (5). Variability of the G-quadruplexes structures are related to three main factors: the relative orientation of the strands, the syn/anti glycosidic conformations of guanine bases and the loop connectivities (6,7). Oligonucleotides containing one, two or four G-stretches can form tetrameric, dimeric or monomeric G-quadruplexes, respectively. In dimeric G-quadruplexes, parallel and antiparallel structures have been characterized. The preferred parallel or antiparallel structures depend on the nature of monovalent cations, such as potassium or sodium. In general, the most stable complexes are formed in the presence of potassium (8). Once formed these complexes are extremely stable and dissociate very slowly (much more slowly than duplexes) (9). However, if complementary C-rich strands are present in the solution, competition between duplex and G-quadruplex structures will be present and their formation and predominance will depend on the relative stability of the G-quadruplex and duplexes (10,11). Few studies can be found in the literature analyzing G-quadruplex/duplex competition. Recently, Li *et al*. (11) have shown the preference for the duplex structure at pH 7.0 against the intramolecular G-quadruplex and i-motif structures. Previous works also reported the equilibria between these G-quadruplex structures and the Watson–Crick duplex indicating the influence of the

*To whom correspondence should be addressed. Tel: +34 934034445; Fax: +34 934021233; Email: raimon@apolo.qui.ub.es

oligonucleotide concentration and the cations present in the medium (K^+, Na^+ and Mg^{2+}) (7,10–22).

Melting studies are widely used for the study of equilibria of nucleic acids, including the characterization of the relative stability of a given sequence depending on pH, changes on metal ion nature or concentration or presence of other DNA sequences (23). Melting transitions can be monitored by a variety of techniques including UV molecular absorbance, circular dichroism or calorimetry; although absorbance is the commonly used technique. This is usually achieved by measuring changes in absorbance at 260 nm. In the case of the study of G-quadruplex equilibria, monitoring the melting process at 295 nm has been described as a suitable alternative allowing the interpretation of the process (24). Although this technique is easy to use and the results are qualitatively easy to interpret, it suffers from a number of limitations. Thus, the absorbance changes are not large and in some cases changes cannot be observed. Moreover, higher order nucleic acids structures (like triplexes or G-quadruplexes) have several components in their melting profiles (e.g. more than two states) with overlapped spectra and which do not allow the complete resolution of all transitions involved during a melting experiment. Several approaches have been proposed to overcome these limitations, such as the use of more sensitive techniques and/or the monitoring of melting experiments at more than one wavelength. Among the first approaches, fluorescence based on molecular beacons has been shown to be a useful approach for the monitoring of melting experiments of G-quadruplexes (25–27). In this approach, changes in the fluorescence yield by oligonucleotides containing suitably placed fluorophores and quenchers are measured. The basis of these experiments is that when the oligonucleotide folds into an intermolecular G-quadruplex, the fluorophore and quencher are relatively close and the fluorescence is quenched. When the oligonucleotide melts into the random coil form, the fluorophore and quencher are no longer near and there is a large increase in fluorescence (16,28). On addition of the complementary (C-rich) oligonucleotide, a competition between the G-quadruplex and duplex structures is established in which the duplex will give a high fluorescence signal since fluorophore and quencher are widely separated. In this work, we have used molecular beacons to examine equilibria between two intermolecular G-quadruplex structures and a duplex structure.

Other experimental approaches may involve spectroscopic monitoring of a melting process at more than one wavelength. These methods have been shown to be very useful for the study of DNA sequences whose hyperchromism is larger at wavelengths other than 260 nm, like G-quadruplexes. However, despite this additional information, in some cases, experiments where several overlapped species coexist could not be totally resolved. These difficulties can often be overcome by measuring multiwavelength molecular absorption spectra, and by applying appropriate multivariate data analysis methods to these multiwavelength measurements (29). Among these methods, multivariate curve resolution alternating least squares (MCR-ALS) has been shown to be a powerful tool for the investigation of complex chemical systems, in particular for the investigation of systems where little or no previous knowledge about them exists, such as the dependence upon pH of triplex forming oligonucleotides (30), the dependence upon salt and temperature of the equilibria dumbbell—dimer—random coil (31) or the folding-unfolding equilibria of proteins (31,32).

This paper shows the application of MCR-ALS to resolve the concentration profiles and pure spectra of all species and conformations present during the melting of a mixture of two DNA sequences which show G-quadruplex/duplex competition. These sequences are the two-repeat human telomeric sequence 5′-d(TAG GGT TAG GGT)-3′ and its complementary, 5′-d(ACC CTA ACC CTA)-3′ (Scheme 1). The thermal behaviour of these two sequences and of their mixtures will be monitored by molecular absorption, molecular fluorescence based on molecular beacons and circular dichroism obtaining qualitative structural information as well as thermodynamic parameters.

MATERIALS AND METHODS

Reagents

Solutions were prepared in Ultrapure water (Millipore) with the appropriate pH 7.1 buffer compounds: Na_2HPO_4 (Panreac, a.r.) and KH_2PO_4 (Panreac, a.r.). Salt medium was obtained by addition of KCl (Merck, a.r.) and $MgCl_2$ (Merck, a.r.).

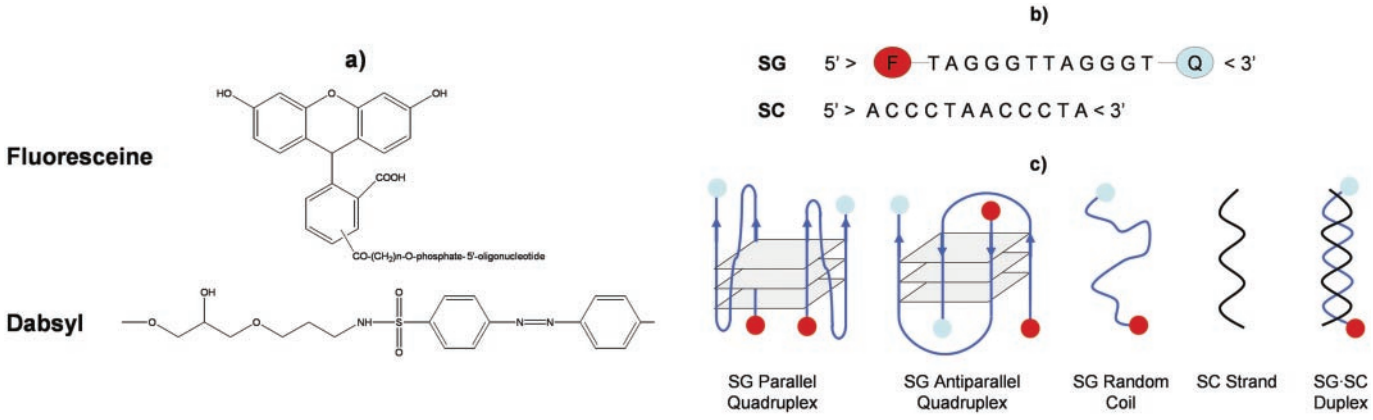
Oligonucleotides

The sequences F-5′-d(TAG GGT TAG GGT)-3′-Q (SG) and 5′-d(ACC CTA ACC CTA)-3′ (SC) (Scheme 1) were synthesized on an Applied Biosystems 392 DNA synthesizer using the 200 nmol scale synthesis cycle. The quencher Q was introduced at the 3′ end using Dabsyl-CPG solid support (Glen Research) and the fluorophore F at the 5′ end using the fluorescein phosphoramidite Fluoroprime (Amersham Biosciences). Standard phosphoramidites were used for the natural bases except for the 2′-deoxyguanosine that was protected with the dimethylformamide (dmf) group. Ammonia deprotection was performed at 55°C for 1 h. The resulting products were desalted by Sephadex G-25 (NAP-10, Amersham Biosciences) and used without further purification. The length and homogeneity of the oligonucleotides was checked by denaturing PAGE and reversed-phase high-performance liquid chromatography (HPLC) using X-Terra® columns. The presence of two diastereoisomers of the glycerol linker used to connect the Dabsyl molecule to the oligonucleotide gave a double band on PAGE and a double peak on HPLC.

DNA concentration was determined by UV absorbance measurements (260 nm) at 90°C considering an extinction coefficient of 157 000 for SG and 113 000 for SC according to the nearest-neighbour approximation as implemented in the web of OligoAnalyzer 3.0 (http://scitools.idtdna.com/scitools/Applications/OligoAnalyzer/). Samples for melting studies were heated at 90°C for 5 min and allowed to renaturalize, cooling slowly until room temperature. Oligonucleotide samples were kept overnight at 4°C until their use.

Procedure

Melting experiments were carried out at pH 7.1 and ionic strength 150 mM provided by buffer compounds (6 mM Na_2HPO_4 and 10 mM KH_2PO_4), 111 mM KCl and 1 mM $MgCl_2$. Melting experiments were carried out step-wise at



Scheme 1. Oligonucleotides studied in this work. (**a**) Chemical structures of fluorescein and dabsyl. (**b**) Sequences of SG and SC oligonucleotides. F and Q stand for fluorophore (fluorescein) and quencher (dabsyl), respectively. (**c**) Schematic presentation of SG-quadruplexes, SC single strand and SG·SC duplex. The structures for SG-quadruplexes have been drawn according to previous NMR studies (19).

3°C increments from 15 to 90°C with a temperature rate of 0.5°C/min and 2 min of stabilization. Each sample was initially stabilized at the starting temperature for 15 min.

Apparatus

UV molecular absorption spectra were recorded on a Perkin-Elmer lambda-19 spectrophotometer between 220 and 340 nm and absorbances were recorded every 1nm. Temperature of the measurements was controlled by a Perkin-Elmer peltier. A Hellma quartz cuvette (path length of 10 mm) was used.

CD spectra were recorded on a Jasco J-720 equipped with a Neslab RET-110 temperature control unit and on a Jasco J-810 equipped with a Julabo F-25/HD temperature control unit. Spectra were recorded between 230 and 320 nm with step resolution of 1 nm, bandwidth of 1nm, sensitivity of 10 mdeg, scan speed of 50 nm/min, response of 8 s and two accumulations for each spectrum. Spectra were recorded using a Hellma quartz cell with pathlength of 10 mm and volume of 700 µl.

Fluorescence emission spectra were recorded on an Aminco Bowman AB-2 fluorimeter, equipped with a cell holder thermostated by a JP Selecta Frigiterm bath. Emission spectra were recorded between 500 and 620 nm and fluorescence intensities were recorded every 1 nm. The excitation wavelength was set to 492 nm, the photomultiplier voltage to 600 V and the excitation and emission bandpass to 4 nm. A 10 mm pathlength quartz cell was used.

Data treatment

Experimental spectra were analyzed using the MCR-ALS procedure described previously (33–35). This procedure is used to estimate the concentration profiles and pure spectra for each spectroscopically active conformation present in analyzed experiments.

MCR-ALS procedure is based on Factor Analysis methods (36) and can be used to analyse spectroscopic data obtained in biochemical or biophysical process monitoring (29,31,32). All MCR-ALS calculations were performed using in-house MATLAB (version 7; The Mathworks Inc., Natick, MA) routines (Codes are freely available at the electronic address www.ub.es/gesq/mcr/mcr.htm).

Spectra recorded in one experiment are collected in a table or matrix $\mathbf{D}$ ($N_r \times N_w$), whose N_r rows being the N_r spectra recorded at successive temperatures and whose N_w columns being the number of wavelengths measured in every spectrum. Mathematically, the goal of MCR-ALS is the calculation of concentration profiles (matrix $\mathbf{C}$) and of pure spectra (matrix $\mathbf{S}^T$, the superscript T means the transpose of matrix $\mathbf{S}$) of species or structures at equilibrium. This is done by applying the multiwavelength extension of Lambert–Beer's law (in matrix form),

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

where $\mathbf{E}$ is the matrix of residual spectral signal not explained by the model that, ideally, should be close to experimental error. Graphically this equation is shown in Scheme 2.

Equation 1 is solved iteratively by an ALS algorithm which calculates concentration $\mathbf{C}$ and pure spectra $\mathbf{S}^T$ matrices optimally fitting experimental data matrix $\mathbf{D}$ using the proposed number of N_s conformations. This N_s is estimated by rank analysis or singular value decomposition (SVD). ALS iterative optimization requires initial estimates of $\mathbf{S}^T$, which can be obtained from pure variable detection methods (37). During the ALS optimization, several constraints were applied including non-negativity for concentration profiles $\mathbf{C}$ and for UV absorbance or fluorescence spectra profiles $\mathbf{S}^T$ (not applied in case of CD spectra profiles), unimodality for concentration profiles $\mathbf{C}$ and closure also for concentration profiles $\mathbf{C}$. See previous works for a more detailed explanation of the ALS iterative optimization procedure.

Concentration profiles $\mathbf{C}$ and pure spectra $\mathbf{S}^T$ resolved for each conformation in the analysis of individual data matrices may differ from the true ones because of possible unresolved underlying factor analysis ambiguities (35). Some of these ambiguity problems can be more easily solved by means of the simultaneous MCR-ALS analysis of several data matrices obtained at different experimental conditions.

When a chemical system is monitored using two or more spectroscopies, a row-wise augmented data matrix can be built up from the individual data matrices corresponding to each spectroscopy, $\mathbf{D}_{CD}$, $\mathbf{D}_F$ and $\mathbf{D}_{abs}$, respectively. The dimensions of the new row-wise augmented matrix will be $(N_r) \times (N_w^{abs} + N_w^F + N_w^{CD})$, and the new linear model used for MCR-ALS analysis is as follows.

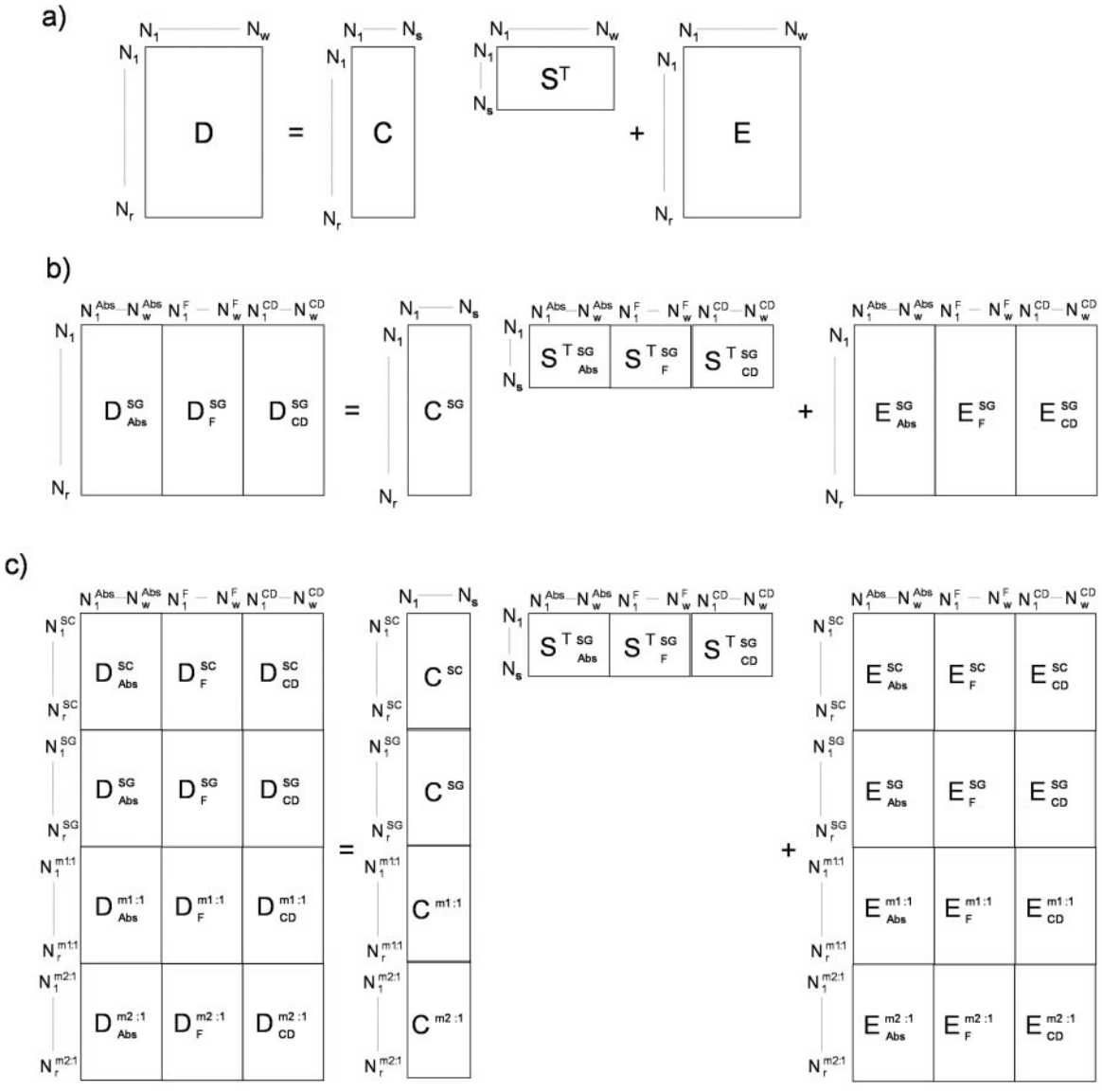
$$[\mathbf{D}_{abs}\mathbf{D}_F\mathbf{D}_{CD}] = [\mathbf{C}]\cdot[\mathbf{S}_{abs}\mathbf{S}_F\mathbf{S}_{CD}]^T + [\mathbf{E}_{abs}\mathbf{E}_F\mathbf{E}_{CD}]. \quad \mathbf{2}$$

Solving Equation 2 for $\mathbf{C}$ and $[\mathbf{S}_{abs}\mathbf{S}_F\mathbf{S}]$ improves identification of the species present in the system especially if their spectra are not well defined in at least one of the spectroscopies.

In some other cases, one of the finally resolved components corresponds in fact to a mixture of two or more species. Simultaneous analysis of different independent experiments in these circumstances is usually very useful and powerful. For instance the model used for MCR-ALS simultaneous analysis of two different melting experiments monitored both by CD spectroscopy, giving experimental data matrices $\mathbf{D}_{CD}^1$ and $\mathbf{D}_{CD}^2$, respectively, is described by Equation 3.

$$\begin{bmatrix}\mathbf{D}_{CD}^1\\ \mathbf{D}_{CD}^2\end{bmatrix} = \begin{bmatrix}\mathbf{C}^1\\ \mathbf{C}^2\end{bmatrix}\cdot\mathbf{S}_{CD}^T + \begin{bmatrix}\mathbf{E}_{CD}^1\\ \mathbf{E}_{CD}^2\end{bmatrix}. \quad \mathbf{3}$$

Finally, a mixed data analysis approach can be proposed for data acquired using different spectroscopies applied



Scheme 2. Data matrix arrangement and MCR-ALS analysis. (**a**) Analysis of a data matrix built by grouping the spectra recorded during a single melting experiment monitored by a single spectroscopy (molecular absorption or molecular fluorescence or CD). (**b**) Simultaneous analysis of three data matrices corresponding to the monitoring of a single melting experiment by three spectroscopies (molecular absorption and molecular fluorescence and CD). (**c**) Simultaneous analysis of 12 data matrices corresponding to the monitoring of four melting experiments (SC, SG and mixtures 1:1 and 2:1) by three spectroscopies (molecular absorption and molecular fluorescence and CD).

simultaneously to multiple experiments. The bilinear model for n experiments studied by several spectroscopies is described by Equation 4.

$$\begin{bmatrix} \mathbf{D}^1_{abs} & \mathbf{D}^1_{F} & \mathbf{D}^1_{CD} \\ \vdots & \vdots & \vdots \\ \mathbf{D}^n_{abs} & \mathbf{D}^n_{F} & \mathbf{D}^n_{CD} \end{bmatrix} = \begin{bmatrix} \mathbf{C}^1 \\ \vdots \\ \mathbf{C}^n \end{bmatrix} \cdot [\mathbf{S}_{abs}\mathbf{S}_{F}\mathbf{S}_{CD}]^{T} + \begin{bmatrix} \mathbf{E}^1_{abs} & \mathbf{E}^1_{F} & \mathbf{E}^1_{CD} \\ \vdots & \vdots & \vdots \\ \mathbf{E}^n_{abs} & \mathbf{E}^n_{F} & \mathbf{E}^n_{CD} \end{bmatrix}. \quad \mathbf{4}$$

This kind of simultaneous analysis is extremely powerful one and allows improvement of resolution of very complex data structures, assembling the profits of both approaches described previously. MCR-ALS analysis of row- and/or column-wise augmented data matrices has been shown to give more reliable solutions, eventually removing totally rotational ambiguities and rank deficiency problems. For a more thorough discussion of these topics see previous references (34,35,38).

Determination of thermodynamic constants

From the resolved concentration profiles ($\mathbf{C}$), information about the thermal behaviour of the species may be obtained. Hence, T_m values can be estimated using MCR-ALS resolved concentration profiles from the crossing point corresponding to the ordered and disordered conformations. Uncertainty related to T_m values determined by this method was estimated to be around ±1°C. Also, a value for the constant corresponding to the conformational equilibrium can be derived as follows.

$$\text{antiparallel G-quadruplex} \underset{\leftarrow}{\longrightarrow} \text{parallel G-quadruplex}$$
$$K_{eq,1} = \frac{[\text{parallel}]}{[\text{antiparallel}]} \quad \mathbf{5}$$

From the calculated equilibrium constants, a value for ΔH^0 and ΔS^0 can be then obtained according to Van't Hoff analysis (39).

Also, an apparent constant for the overall equilibrium can be estimated as follows:

$$\text{SG-quadruplex} + \text{SCstrand} \underset{\leftarrow}{\longrightarrow} \text{duplex}$$
$$K_{eq,2} = \frac{[\text{duplex}]}{[\text{SG-quadruplex}][\text{SCstrand}]} \quad \mathbf{6}$$

where [SG-quadruplex] stands for the overall concentration of SG in antiparallel and parallel structures, [SC strand] stands for the overall concentration of SC in ordered and disordered conformations, and [duplex] stands for duplex species concentration.

From the resolved pure spectra ($\mathbf{S}^T$), qualitative and information can be obtained for each one of the species or conformations.

RESULTS

G-quadruplex/duplex competition

The natural two-repeat human telomeric sequence F-d(TAG GGT TAG GGT)-Q (SG) carrying a fluorescein molecule at the 5′ end and the quencher Dabsyl at the 3′ end and its complementary sequence 5′-d(ACC CTA ACC CTA)-3′ (SC) were prepared using standard protocols.

Competition between the formation of the SG G-quadruplex structures and the SG·SC Watson–Crick duplex was studied at two SG:SC ratios. In the first melting experiment, the denaturation of a 1:1 SG:SC mixture (both SG and SC 1.2 μM as strand concentration) was monitored by means of UV molecular absorption, molecular fluorescence and CD spectroscopies (Figure 1). Molecular absorption spectra show a considerable hyperchromism at 260 nm (∼25%), characteristic of melting processes involving duplex disruption (Figure 1a). Fluorescence spectra (Figure 1b) showed the characteristic maximum of fluorescein dye at 520 nm. An atypical melting profile is obtained upon heating (Figure 1b, inset). Initially, the fluorescence intensity decreases continuously from 14 to 60°C. At temperature values higher than 60°C, the fluorescence intensity increases and reaches a final value below the initial value at 14°C. Finally, CD spectra recorded during the melting process are shown in Figure 1c. At 14°C, it can be distinguished two positive and negative bands located at 266 and 240 nm, respectively, and a shoulder around 290 nm. Upon heating, the intensity of the bands diminishes. At 90°C, two weaker bands, a wide positive centred at 275 nm and a negative around 240 nm, can be observed.

Each of the three datasets (corresponding to the recorded molecular absorption, molecular fluorescence or CD spectra) shown in Figure 1 was individually analysed by MCR-ALS according to Equation 1 and Scheme 2a. In all three cases, the analysis did not provide satisfactory results since resolved concentration profiles and pure spectra could not be related to the expected behaviour of species involved in the denaturation of the mixture of antiparallel and parallel G-quadruplexes, duplex, and single strands. Moreover, the results were highly dependent on the nature of the spectroscopic data analysed. The next step was building-up an augmented data matrix by joining the three datasets and its later analysis according to Equation 2 and Scheme 2b. This augmented data matrix had dimensions 26 rows by 263 columns (86, 91 and 86 channels or wavelengths measured in molecular absorption, molecular fluorescence and CD, respectively). SVD analysis of this augmented data matrix did not allow to determine with absolute certainty whether the number of components (Ns) was two or three. Therefore, MCR-ALS was carried out considering either two or three components (Figure 2). When the system was analysed considering $Ns = 2$ a single transition was observed with a T_m value of 49°C, which could be related to the duplex to random coil transition. This concentration profile gives an overall picture of what is happening in the mixture. However, information about all other species or conformations is not available. Moreover, resolved spectra (Figure 2b–2d), do not provide any good qualitative information about these species or conformations. Hence, as example, resolved CD spectra only reflect the initial and final mixtures. A similar conclusion can be drawn from the inspection of the MCR-ALS results

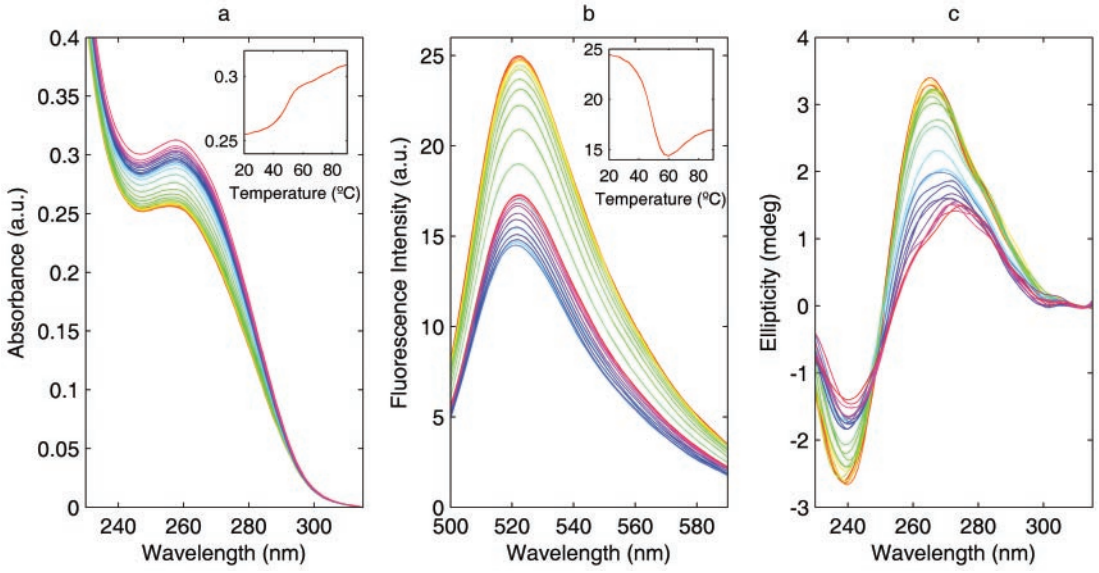


Figure 1. Experimental spectroscopic data recorded during the melting experiment of the 1:1 SG:SC mixture. (**a**) Molecular absorption spectra. Inset: melting profile at 260 nm. (**b**) Molecular fluorescence spectra. Inset: melting profile at 520 nm. (**c**) CD spectra. Experimental conditions: 1.2 μM SG, 1.2 μM SC, pH 7.1, 111 mM K^+ and 1 mM Mg^{2+}.

when $Ns = 3$ (Figure 2e–2h). In this case, the determined T_m values were 47 and 62°C.

Similar experimental trends were observed during the melting of a mixture SG:SC (ratio 2:1, SG: 2.4 μM, SC: 1.2 μM, both as strand concentration). In this case, the hyperchromism found at 260 nm was around 10%. Fluorescence spectra showed similar trends to the 1:1 mixture. Finally, CD spectra at 14°C showed a positive band around 265 nm and a negative band at 240 nm. In this case the shoulder around 290 nm was even less intense than in the case of the 1:1 mixture. The individual and simultaneous analyses of the three spectroscopic datasets provide similar results to those obtained for the 1:1 mixture.

From the results shown until now, it is clear that the analysis of the spectroscopic data recorded during the melting of a SG:SC mixtures did not allow the complete resolution of the system, i.e. the complete resolution of the concentration profiles and pure spectra for all and each one of the species present in the system. In other words, the resolved concentration profiles did not correspond to the expected behaviour of the antiparallel and parallel G-quadruplexes, duplex and single strands present in the mixture. On the other hand, the resolved spectra did not show the spectral characteristics of these structures or conformations. The complete resolution has not been possible even with the simultaneous analysis of the molecular absorption, molecular fluorescence and CD.

Thermal behaviour of SG and SC sequences

It has been shown previously (30,40) that the analytical resolution of complex systems involving the formation of multi-stranded structures monitored by spectroscopic techniques, like the G-quadruplex/duplex competition described in this work, can only be achieved when additional complementary information about the constituent subsystems is provided. This can be done by analysing additional datasets corresponding to complementary experiments. In this case, the thermal behaviour of SG [which can form antiparallel and parallel G-quadruplexes (19)] and SC sequences are provided to help to resolve the melting behaviour of the SG:SC mixtures. With this goal in mind, two new melting experiments were performed, where the thermal behaviour of the SG and of the SC sequences was individually studied.

Molecular absorption, molecular fluorescence and CD experimental spectra recorded during the melting experiment of SG (1.2 μM as strand concentration) are shown in Figure 3. UV molecular absorption spectra show a maximum around 260 nm and an isosbestic point at 285 nm (Figure 3a). The maximum at 260 nm shows a low hyperchromism (~4%) during the melting and any clear sigmoidal curve was observed. This hyperchromism is significantly lower than that observed during the melting of the SG:SC mixtures. Contrarily to the melting recorded at 260 nm, a sigmoid curve is clearly observed with a hyperchromism around 25% when the melting is monitored at 295 nm (T_m of 49°C). This fact clearly reflects the presence of a G-quadruplex structure at low temperatures, according to literature (24). Fluorescence spectra (Figure 3b) show an increase of the intensity and a slightly shift of the maximum at 520 nm. As can be seen in the figure inset, a clear transition can be observed in the melting profile with a T_m of 58°C. Visual inspection of CD spectra (Figure 3c) shows some interesting facts. First, the spectrum recorded at 14°C clearly suggests the presence of a mixture of different structures, because it shows a positive band at 263 nm [characteristic of parallel structures (18)] with a shoulder around 290 nm [characteristic of antiparallel structures (18)] and a negative band at 239 nm. Second, upon heating, the disappearance of the shoulder at 290 nm and the later reduction of the band at 263 nm were observed. Finally, at 90°C only a wide weak positive band (around 265 nm) and a negative band around 240 nm can be seen.

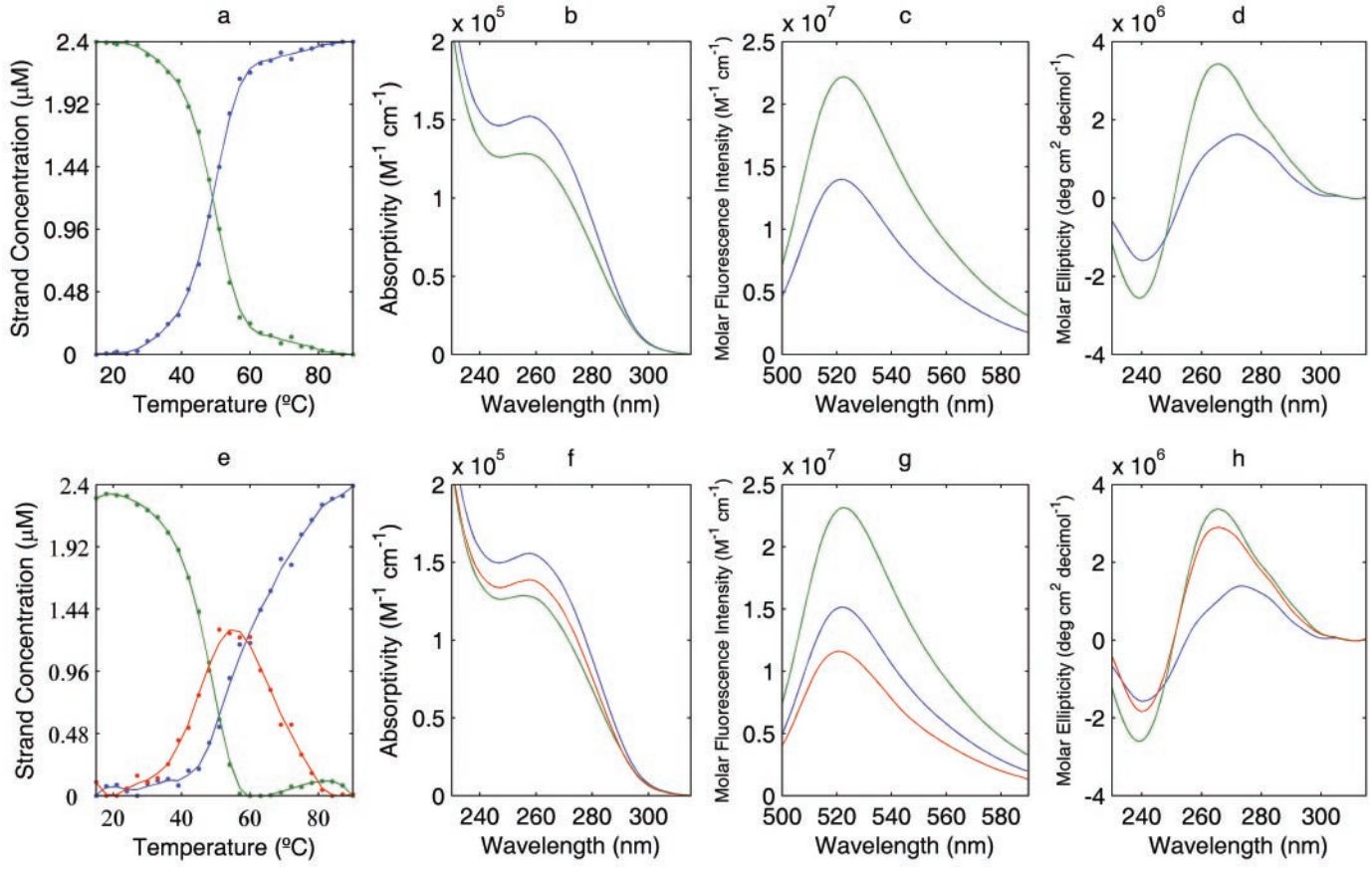


Figure 2. MCR-ALS analysis of the spectroscopic data recorded during the melting experiment of the 1:1 SG:SC mixture according to Equation 2. (a–d) Refer to the resolved concentration profile and pure spectra when only two components were considered. (**a**) Resolved concentration profile. (**b**) Resolved molecular absorption spectra. (**c**) Resolved molecular fluorescence spectra. (**d**) Resolved CD spectra. (e–h) Refer to the resolved concentration profiles and pure spectra when three components were considered. (**e**) Resolved concentration profiles. (**f**) Resolved molecular absorption spectra. (**g**) Resolved molecular fluorescence spectra. (**h**) Resolved CD spectra.

The second additional dataset corresponded to the melting of the SC sequence. It is expected that spectroscopic data recorded during this experiment only reflect the loss of the partially stacked single strand upon heating. A small hyperchromism effect (~5%) was observed in the UV molecular absorption (spectra not shown). CD spectra recorded at 15°C gave a positive band at 275 nm and a negative band at 249 nm. Upon heating, the amplitude of both bands decreased.

Analysis of the whole dataset

Once the complete set of spectroscopic data for the thermal denaturation of the sequences SG, SC and of their mixtures was available, a simultaneous analysis was carried out. Hence, 12 data matrices, corresponding to the four independent melting experiments monitored by three spectroscopies, were grouped into an augmented data matrix, according to Equation 4 and Scheme 2c. The dimensions of this augmented data matrix are 104 rows (26 spectra from each experiment) by 263 columns. The first step in MCR-ALS analysis was the determination of the number of components. In this case, SVD analysis of the augmented data matrix indicated the presence of six species ($Ns = 6$). Using this number of components, initial estimates of the spectra for the ALS optimization were determined by means of a pure variable detection method.

Figure 4 shows resolved concentration profiles corresponding to the thermal denaturation of sequences SC, SG and of their mixtures at 1:1 and 2:1 ratios. Figure 5 shows resolved pure molecular absorption, molecular fluorescence and CD spectra for the six proposed structures or conformations. Residual noise (matrices E in Scheme 2c) not explained by the proposed model in Figures 4 and 5 is given as Supplementary Data. According to the shape and absolute values of these residuals the existence of any other spectroscopically active species or conformation was dismissed.

As expected, only two conformations are present during the melting of the SC sequence (Figure 4a). These conformations can easily be related to the initial partially stacked strand, and to the final disordered strand. The formation of i-motif structures was not considered because these structures are formed at more acid pH values (10,11). The resolved molecular absorption, molecular fluorescence and CD spectra for each conformation are in agreement with the experimental spectral trends.

Three structures or conformations were resolved in the thermal denaturation of SG sequence (Figure 4b). At the

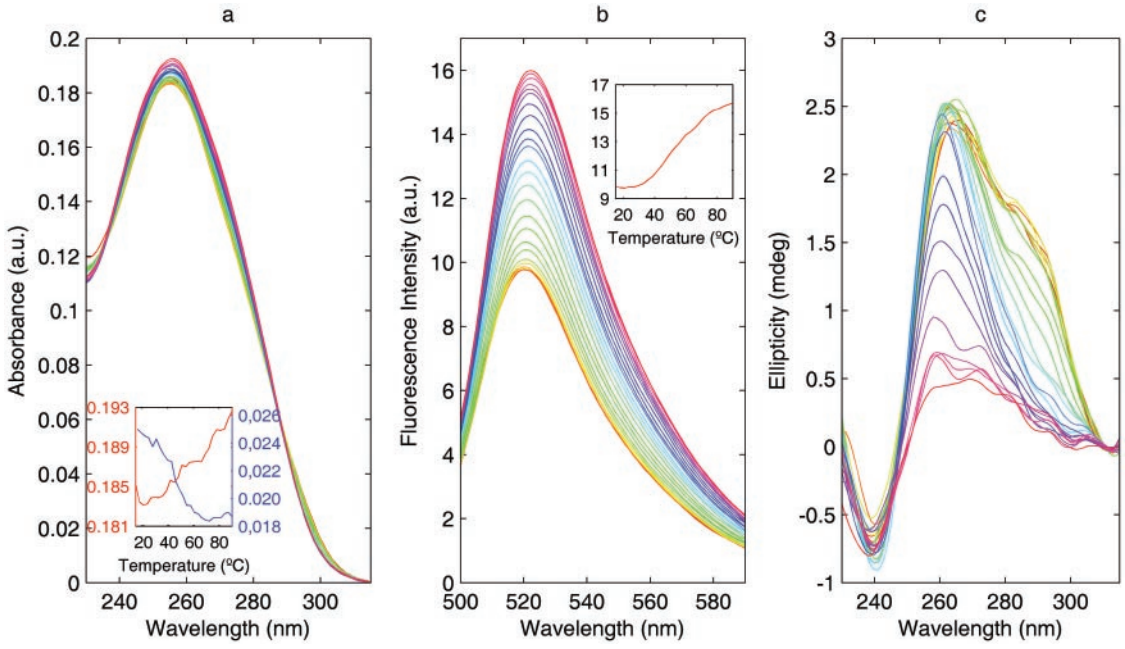


Figure 3. Experimental spectroscopic data recorded during the melting experiment of the SG sequence. (**a**) Molecular absorption spectra. Inset: melting profiles at 260 nm (red line) and 295 nm (blue line). (**b**) Molecular fluorescence spectra. Inset: melting profile at 520 nm. (**c**) CD spectra. Experimental conditions: 1.2 μM SG, pH 7.1, 111 mM K^+ and 1 mM Mg^{2+}.

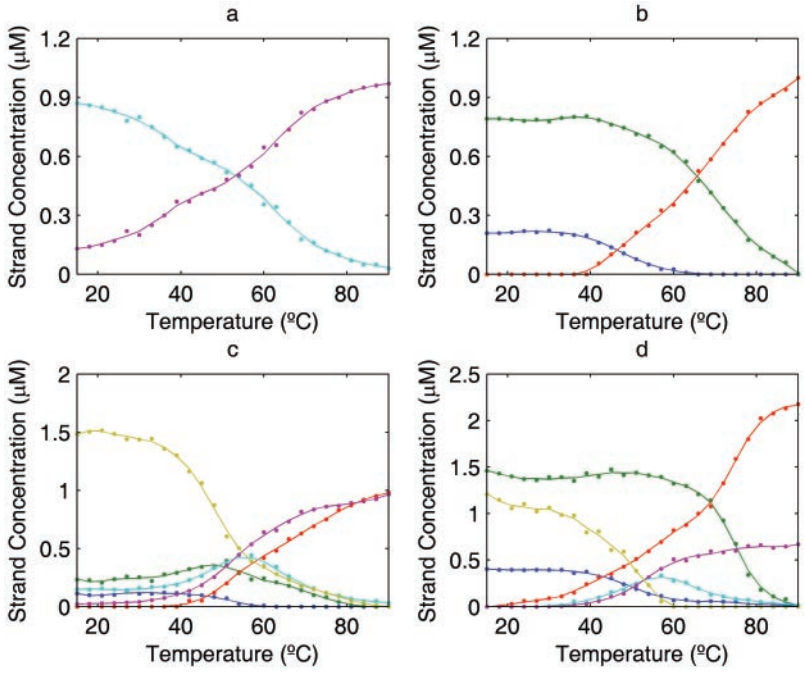


Figure 4. MCR-ALS analysis of the spectroscopic data recorded during the four melting experiments according to Equation 4. (**a**) Resolved concentration profiles for the melting of the SC sequence. (**b**) Resolved concentration profiles for the melting of the SG sequence. (**c**) Resolved concentration profiles for the mixture 1:1 SG:SC. (**d**) Resolved concentration profiles for the mixture 2:1 SG:SC. Cyan, ordered SC strand; magenta, disordered SC strand; green, parallel SG G-quadruplex; blue, antiparallel SG G-quadruplex; red, disordered SG strand; yellow, SG·SC duplex.

beginning of the experiment two structures were present in solution, which evolve to a disordered conformation upon heating. Interpretation of these concentration profiles is possible due to the complementary information provided by the resolved corresponding spectra (Figure 5). Hence, the major structure at low temperatures was assigned to a dimeric parallel G-quadruplex since the corresponding resolved CD spectrum shows a positive band at 262 nm and a negative band at 240 nm, which is considered characteristic of a parallel G-quadruplex (18,41,42). On the other hand, the minor structure present at low temperatures was assigned to a dimeric antiparallel G-quadruplex (19) (Scheme 1). Again,

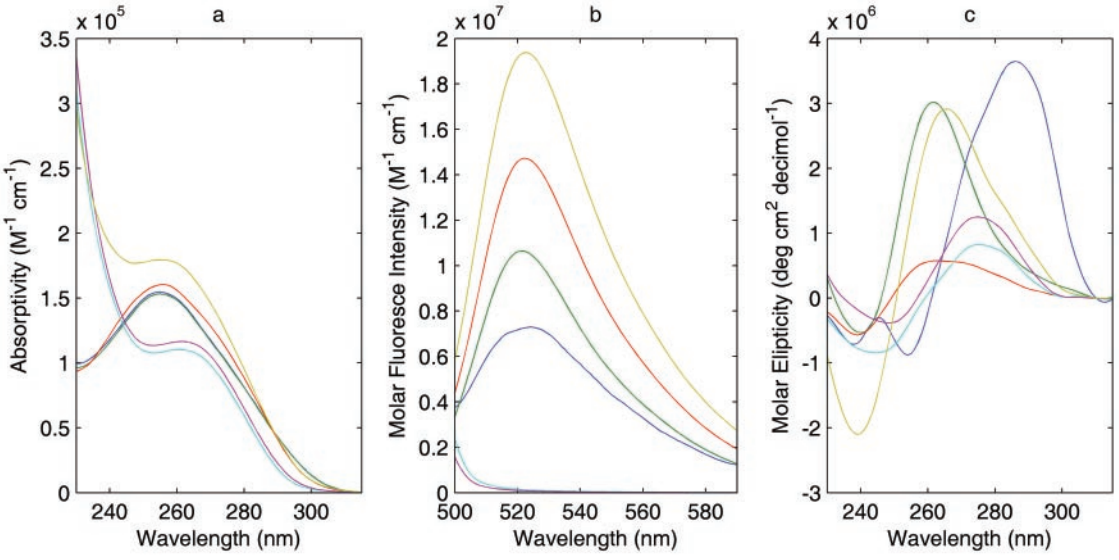


Figure 5. MCR-ALS analysis of the spectroscopic data recorded during the four melting experiments according to Equation 4. (**a**) Resolved pure molecular absorption spectra. (**b**) Resolved molecular fluorescence spectra. (**c**) Resolved CD spectra. Colours as in Figure 4.

the position of the resolved bands in the corresponding CD spectrum (positive band at 290 nm, negative band at 258 nm and weak positive band at 242 nm) agrees with the description of the spectrum of the antiparallel G-quadruplex found in literature (42). Finally, the conformation present at high temperatures was assigned to disordered SG as the corresponding CD spectrum shows wide weak bands characteristic of non-structured DNA. From the resolved concentration profiles, melting temperatures (T_ms) for the parallel and antiparallel G-quadruplex can be determined, these temperatures being 46 and 65°C for the antiparallel and parallel structures, respectively. Also, a value for the apparent equilibrium constant corresponding to the transition between the antiparallel and parallel structures was obtained as $K_{eq,1} = 3.7 \pm 0.2$ ($n = 7$), which corresponds to a $\Delta G^o = -3.1 \pm 0.1$ kJ/mol (25°C). The calculated values of ΔH^0 and ΔS^0 for the overall transition from SG G-quadruplex (antiparallel and parallel structures) to SG-disordered coil according to traditional Van't Hoff analysis are 97 ± 7 kJ/mol and 286 ± 25 J/K·mol, respectively.

Figure 5 also includes the resolved molecular absorption and molecular fluorescence spectra for each of these three SG structures or conformations. Hence, while the resolved molecular absorption spectra for the antiparallel and parallel structures are very similar, molecular fluorescence spectra are slightly different, the resolved spectrum for the parallel structure being more intense.

Figure 4c shows the resolved concentration profiles for the 1:1 mixture. In this case, six species or conformations are present during the melting of this mixture. At low temperatures, the SG·SC duplex is the major structure in equilibrium. SG parallel G-quadruplex, SG antiparallel G-quadruplex and ordered SC (both in equilibrium concentration practically negligible) can also be detected. Upon heating, the duplex melts with a T_m of 48°C yielding a mixture of structures related both to SG (parallel G-quadruplex and random coil) and to SC (partially stacked and disordered). Also, a value for the apparent constant corresponding to the formation of SG·SC duplex from SG G-quadruplex (from both parallel and antiparallel structures) and SC (from both partially stacked and disordered conformations) was obtained, as being $K_{eq,2} = 7.4$ (at 25°C), which corresponds to $\Delta G^o = -42$ kJ/mol (at 25°C).

Finally, Figure 4d shows the resolved concentration profiles for the 2:1 mixture. The same six species resolved in the previous experiment are also observed but, in this case, the distribution of these species is slightly different. Now, the major species at low temperatures are the SG parallel G-quadruplex and the SG·SC duplex. The SG antiparallel G-quadruplex could also be observed but again its contribution is minor. All three structures or conformations disappeared upon heating, with T_m of 48, 50 and 72°C for the SG antiparallel G-quadruplex, SG·SC duplex and SG parallel G-quadruplex, respectively. The resolved spectra for the SG·SC duplex (Figure 5) are in agreement with the experimental trends. For instance, the intensity of the molecular fluorescence spectrum is high, reflecting the distance between the fluorophore and the quencher in the double-stranded structure. CD spectrum also reflects quite well the expected shape for a B-DNA.

These results were partially confirmed by PAGE in native conditions (Supplementary Data). Unfortunately, quantification of the different bands was not possible because the G-quadruplex structures run as a broad smear but it was possible to observe that the duplex structure was the most abundant structure in the equimolar SG:SC mixture.

DISCUSSION

The results presented in this paper describe quantitatively the equilibria between two different G-quadruplex structures and the competition between them and the Watson–Crick duplex when the complementary strand is in solution. Despite the high complexity of the system and of the experimental (non-selective) spectral data, application of an appropriate experimental methodology and later analysis by MCR-ALS allowed the total resolution and characterization of the system.

Concentration profiles for all detected structures and conformations and their corresponding spectra for each of the three spectroscopic techniques used in this work were obtained. Limitations of the applied methodology in the detection and quantification of all possible different parallel and antiparallel structures present in the system are based on the quality of the experimental data analysed. Thus, it should be stressed that the applied methodology could only resolve another parallel and/or antiparallel structure in case: (i) its pure spectra are different to those depicted in Figure 5a–c, and/or (ii) its melting profile is different and linearly independent to those depicted in Figure 4b. According to the information provided by the spectroscopic techniques used in this work, any other structure could be detected.

The equilibrium between the two different G-quadruplexes formed was studied first. The presence of both dimeric parallel G-quadruplex and dimeric antiparallel G-quadruplex structures in potassium had already been suggested by other authors. Parkinson *et al*. (43) described the parallel structure and Phan and Patel (19) confirmed that two different structures could coexist in solution. In this work, the predominance of the parallel G-quadruplex at low temperatures was shown. Resolved concentration profiles of these two G-quadruplex structures and of the random coil conformation allowed the calculation of the thermodynamic parameters for these equilibria. The calculated values are in good agreement with those obtained by other authors, like Darby *et al*. (28) (a value of $\Delta H^0 = 135$ kJ/mol) for the melting of an intermolecular G-quadruplex formed by the sequence 5′-GGGTTAGGG-3′ in 50 mM K^+, pH 7.4.

On the other hand, our results seem to confirm that if the complementary strand is added, the predominant structure at physiological pH and temperature is the Watson–Crick duplex, which was already pointed out by other authors (10,11,16). Moreover, our results confirm that it is possible to form quadruplex structures, even in the presence of the complementary C-rich strand. The fraction of duplex found by Li *et al*. (11) for an intramolecular quadruplex/duplex competition was ~60% in 100 mM Na^+. In our case, for a dimeric intermolecular G-quadruplex, this fraction is ~70% for the 1:1 SG:SC mixture, and 36% for the 2:1 mixture, at 150 mM K^+. This coexistence of the G-quadruplex and duplex structures might have a relevant biological significance. Probably, not only the major concentration of the duplex is significant but also the relatively small quantities of the G-quadruplex structure. Resolved concentration profiles for the mixtures also show that a fraction of the G-quadruplex melts just after the duplex. This fact had already been described by Phan *et al*. (10) for different sequences at this pH value. Those authors show that G-quadruplex melts around 63°C, while a 22 bp duplex melts around 61°C. On the other hand, we calculated thermodynamic parameters from the resolved concentration profiles for the 1:1 mixture. The calculated value of ΔG^o for the formation of duplex from any structure of SG and SC was found to be of the same order of magnitude as the value reported by Li *et al*. (11) for the formation of a duplex from an intramolecular G-quadruplex and its complementary strand (-40 ± 6 kJ/mol at pH 7.0, 100 mM Na^+, 25°C).

Resolved spectra (Figure 5) also provided interesting information. CD spectra for the antiparallel and the parallel G-quadruplex structures were resolved despite the fact that these structures are always mixed in the experimental conditions we used. As expected, the CD spectrum for the antiparallel structure shows a clear band around 290 nm while the parallel structure does not show a positive band until 260 nm. As the SG strand included a fluorophore and a quencher at the two ends, it is also interesting to analyse the resolved molar fluorescence emission spectra in order to extract more structural information. As expected, the duplex structure showed a large intensity of fluorescence due to the distance between the fluorophore and the quencher when the SG and SC strands adopted the Watson–Crick duplex structure. Similar behaviour is observed for the SG random coil structure. In this case, when the two strands forming the intermolecular G-quadruplex or the duplex structures dissociate, the distance between the fluorophore and the quencher is rather large. Comparison of the relative intensities between the resolved fluorescence emission spectra of the antiparallel and parallel G-quadruplexes showed larger fluorescence intensity for the last one. This fact seems to agree with the structures proposed in previous works (19,43). In our case, considering the structure for the intermolecular parallel G-quadruplex proposed by Parkinson *et al*., the two fluorophores and the two quenchers would be on opposite sides of the quadruplex and, therefore, small quenching effects could be possible. On the other hand, the intermolecular antiparallel structure [considering the structure proposed by Phan *et al*. (19)] would have a fluorophore and a quencher in each side of the quadruplex. This would mean that the fluorophore and the quencher would remain at a relatively small distance and much less intensity of fluorescence would be detected because of fluorescence quenching.

Overall, CD spectroscopy is the technique, among those used in this work, which provides the highest information content about the system because CD spectra are more sensitive to structural and conformational changes. This makes CD data very useful to finally solve the contribution of each species or structure in the mixture. In this case, molecular absorption does not help very much because the pure spectra of all species and structures (Figure 5a) show small spectral changes (mostly only in intensity) and a strong overlap. On the other hand, fluorescence based on molecular beacons is a very useful spectroscopic technique for labelled oligonucleotides. It has the additional advantage of being more selective than absorption. However, in the case of multiple fluorescent species, the spectra may be strongly overlapped (Figure 5b).

Finally, it must be stressed that the methodology applied in this work for the characterization of intermolecular G-quadruplex/duplex competition can be also applied for the study of similar equilibria involving intramolecular G-quadruplex. In this sense, we have also studied the unfolding/folding equilibria of the 5′-GGT TGG TGT GGT TGG-3′ (15-base-long thrombin aptamer), which is known to form a monomeric chair G-quadruplex structure (Supplementary Data). The individual and simultaneous analysis of molecular absorption, molecular fluorescence and CD revealed a simpler system than those described here, where only a broad transition from the initial chair structure to a random coil was observed. Because of this, we have focused our study on the equilibria involving the two-repeat human telomeric sequence, which shows a greater structural variability.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

Thanks are due to Dr Shantanu Chowdhury (IGIB, Delhi, India) for his helpful comments. J.J. acknowledges a PhD scholarship from the Spanish Ministerio de Educación y Ciencia (MEC). This research was supported by the grants BQU2003-0191 and BFU2004-02048/BMC. Funding to pay the Open Access publication charges for this article was provided by Spanish Ministerio de Educacion y Ciencia.

Conflict of interest statement. None declared.

REFERENCES

1. Blackburn,E.H. and Szostak,J.W. (1984) The molecular-structure of centromeres and telomeres. *Annu. Rev. Biochem.*, **53**, 163–194.
2. Blackburn,E.H. (1991) Structure and function of telomeres. *Nature*, **350**, 569–573.
3. Neidle,S. and Read,M.A. (2000) G-quadruplexes as therapeutic targets. *Biopolymers*, **56**, 195–208.
4. Zahler,A.M., Williamson,J.R., Cech,T.R. and Prescott,D.M. (1991) Inhibition of telomerase by G-quartet DNA structures. *Nature*, **350**, 718–720.
5. Siddiqui-Jain,A., Grand,C.L., Bearss,D.J. and Hurley,L.H. (2002) Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proc. Natl Acad. Sci. USA*, **99**, 11593–11598.
6. Simonsson,T. (2001) G-quadruplex DNA structures—variations on a theme. *Biol. Chem.*, **382**, 621–628.
7. Phan,A.T., Modi,Y.S. and Patel,D.J. (2004) Two-repeat Tetrahymena telomeric d(TGGGGTTGGGGT) sequence interconverts between asymmetric dimeric G-quadruplexes in solution. *J. Mol. Biol.*, **338**, 93–102.
8. Shafer,R.H. and Smirnov,I. (2000) Biological aspects of DNA/RNA quadruplexes. *Biopolymers*, **56**, 209–227.
9. Mergny,J.L., De Cian,A., Ghelab,A., Sacca,B. and Lacroix,L. (2005) Kinetics of tetramolecular quadruplexes. *Nucleic Acids Res.*, **33**, 81–94.
10. Phan,A.T. and Mergny,J.L. (2002) Human telomeric DNA: G-quadruplex, i-motif and Watson–Crick double helix. *Nucleic Acids Res.*, **30**, 4618–4625.
11. Li,W., Miyoshi,D., Nakano,S. and Sugimoto,N. (2003) Structural competition involving G-quadruplex DNA and its complement. *Biochemistry*, **42**, 11736–11744.
12. Hardin,C.C., Perry,A.G. and White,K. (2000) Thermodynamic and kinetic characterization of the dissociation and assembly of quadruplex nucleic acids. *Biopolymers*, **56**, 147–194.
13. Smirnov,I. and Shafer,R.H. (2000) Effect of loop sequence and size on DNA aptamer stability. *Biochemistry*, **39**, 1462–1468.
14. Miyoshi,D., Nakao,A., Toda,T. and Sugimoto,N. (2001) Effect of divalent cations on antiparallel G-quartet structure of d(G(4)T(4)C(4)). *FEBS Lett.*, **496**, 128–133.
15. Miyoshi,D., Matsumura,S., Li,W. and Sugimoto,N. (2003) Structural polymorphism of telomeric DNA regulated by pH and divalent cation. *Nucleosides Nucleotides Nucleic Acids*, **22**, 203–221.
16. Risitano,A. and Fox,K.R. (2003) Stability of intramolecular DNA quadruplexes: Comparison with DNA duplexes. *Biochemistry*, **42**, 6507–6513.
17. Miyoshi,D., Nakao,A. and Sugimoto,N. (2003) Structural transition from antiparallel to parallel G-quadruplex of d(G(4)T(4)G(4)) induced by Ca2+. *Nucleic Acids Res.*, **31**, 1156–1163.
18. Dapic,V., Abdomerovic,V., Marrington,R., Peberdy,J., Rodger,A., Trent,J.O. and Bates,P.J. (2003) Biophysical and biological properties of quadruplex oligodeoxyribonucleotides. *Nucleic Acids Res.*, **31**, 2097–2107.
19. Phan,A.T. and Patel,D.J. (2003) Two-repeat human telomeric d(TAGGGTTAGGGT) sequence forms interconverting parallel and antiparallel G-quadruplexes in solution: distinct topologies, thermodynamic properties, and folding/unfolding kinetics. *J. Am. Chem. Soc.*, **125**, 15021–15027.
20. Petraccone,L., Erra,E., Esposito,V., Randazzo,A., Mayol,L., Nasti,L., Barone,G. and Giancola,C. (2004) Stability and structure of telomeric DNA sequences forming quadruplexes containing four G-tetrads with different topological arrangements. *Biochemistry*, **43**, 4877–4884.
21. Sacca,B., Lacroix,L. and Mergny,J.L. (2005) The effect of chemical modifications on the thermal stability of different G-quadruplex-forming oligonucleotides. *Nucleic Acids Res.*, **33**, 1182–1192.
22. Petraccone,L., Erra,E., Esposito,V., Randazzo,A., Galeone,A., Barone,G. and Giancola,C. (2005) Biophysical properties of quadruple helices of modified human telomeric DNA. *Biopolymers*, **77**, 75–85.
23. Mergny,J.L. and Lacroix,L. (2003) Analysis of thermal melting curves. *Oligonucleotides*, **13**, 515–537.
24. Mergny,J.L., Phan,A.T. and Lacroix,L. (1998) Following G-quartet formation by UV-spectroscopy. *FEBS Lett.*, **435**, 74–78.
25. Tyagi,S. and Kramer,F.R. (1996) Molecular beacons: Probes that fluoresce upon hybridization. *Nat. Biotechnol.*, **14**, 303–308.
26. Tan,W.H., Wang,K.M. and Drake,T.J. (2004) Molecular beacons. *Curr. Opin. Chem. Biol.*, **8**, 547–553.
27. Simonsson,T. and Sjoback,R. (1999) DNA tetraplex formation studied with fluorescence resonance energy transfer. *J. Biol. Chem.*, **274**, 17379–17383.
28. Darby,R.A.J., Sollogoub,M., McKeen,C., Brown,L., Risitano,A., Brown,N., Barton,C., Brown,T. and Fox,K.R. (2002) High throughput measurement of duplex, triplex and quadruplex melting curves using molecular beacons and a LightCycler. *Nucleic Acids Res.*, **30**, e39.
29. Jaumot,J., Vives,M. and Gargallo,R. (2004) Application of multivariate resolution methods to the study of biochemical and biophysical processes. *Anal. Biochem.*, **327**, 1–13.
30. Jaumot,J., Avino,A., Eritja,R., Tauler,R. and Gargallo,R. (2003) Resolution of parallel and antiparallel oligonucleotide triple helices formation and melting processes by multivariate curve resolution. *J. Biomol. Struct. Dyn.*, **21**, 267–278.
31. Jaumot,J., Escaja,N., Gargallo,R., Gonzalez,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.
32. Navea,S., de Juan,A. and Tauler,R. (2002) Detection and resolution of intermediate species in protein folding processes using fluorescence and circular dichroism spectroscopies and multivariate curve resolution. *Anal. Chem.*, **74**, 6031–6039.
33. Jaumot,J., Gargallo,R., de Juan,A. and Tauler,R. (2005) A graphical user-friendly interface for MCR-ALS: a new tool for multivariate curve resolution in MATLAB. *Chemometr. Intell. Lab.*, **76**, 101–110.
34. Tauler,R. (1995) Multivariate curve resolution applied to second order data. *Chemometr. Intell. Lab.*, **30**, 133–146.
35. Tauler,R., Smilde,A. and Kowalski,B. (1995) Selectivity, local rank, 3-way data-analysis and ambiguity in multivariate curve resolution. *J. Chemometr.*, **9**, 31–58.
36. Malinowski,E.R. and Howery,D.G. (2002) *Factor Analysis in Chemistry, 3rd edn*. Wiley, New York, NY.
37. Windig,W. and Guilment,J. (1991) Interactive self-modeling mixture analysis. *Anal. Chem.*, **63**, 1425–1432.
38. Tauler,R. (2001) Calculation of maximum and minimum band boundaries of feasible solutions for species profiles obtained by multivariate curve resolution. *J. Chemometr.*, **15**, 627–646.
39. Marky,L.A. and Breslauer,K.J. (1987) Calculating thermodynamic data for transitions of any molecularity from equilibrium melting curves. *Biopolymers*, **26**, 1601–1620.
40. Kumar,P., Kanchan,K., Gargallo,R. and Chowdhury,S. (2005) Application of multivariate curve resolution for the study of folding processes of DNA monitored by fluorescence resonance energy transfer. *Anal. Chim. Acta*, **536**, 135–143.
41. Balagurumoorthy,P. and Brahmachari,S.K. (1994) Structure and stability of human telomeric sequence. *J. Biol. Chem.*, **269**, 21858–21869.
42. In Berova,N., Nakanishi,K. and Woody,R.W. (eds) (2000) *Circular Dichroism. Principles and Applications, 2nd Edn*. Wiley-VCH Inc., NY.
43. Parkinson,G.N., Lee,M.P.H. and Neidle,S. (2002) Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature*, **417**, 876–880.