STUDY OF THE CONFORMATIONAL EQUILIBRIA OF G-QUADRUPLEX-FORMING DNA SEQUENCES BY MEANS OF SPECTROSCOPIC AND CHEMOMETRIC TECHNIQUES

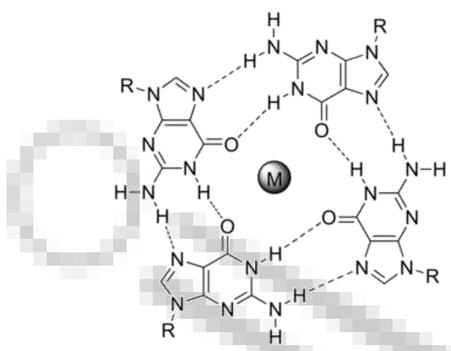
Alba Ferrer, Sanae Benabou, Raimundo Gargallo

*Department of Chemical Engineering and Analytical Chemistry, University of Barcelona, Diagonal 647, 08028 Barcelona, Spain (**www.ub.es/gesq/dna**)*

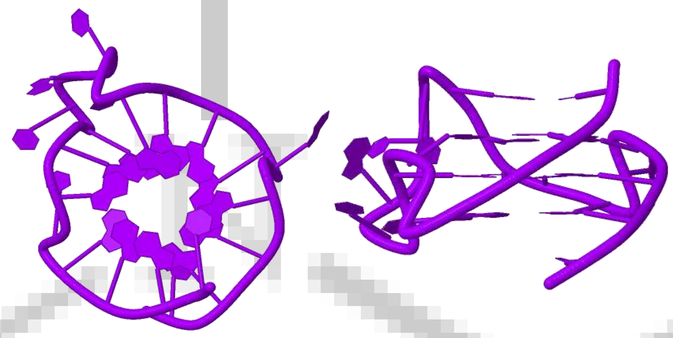
Overview

- Apart from the well-known double-stranded helix firstly proposed by Watson and Crick, DNA may adopt other complex structures, such as *i*-motif or G-quadruplex.
- The *i*-motif structures, which are formed by cytosine-rich sequences, have not been observed *in vivo*. However, their conformational equilibria have been used in the development of sensors and nanomachines [1].
- The G-quadruplex structures, which are formed by guanine-rich sequences, have been already observed *in vivo*. They seem to be related with several cancer diseases and aging [2,3].
- The stability of these structures depends on factors like pH, temperature, ionic strength, and the presence of ligands, such as proteins or drugs.
- The knowledge about the stability of these structures is obtained from biophysical experiments. Traditionally, these experiments have been monitored measuring an instrumental signal at one single channel. The univariate approach has several drawbacks like the difficulty when modeling systems where intermediates are present.
- Here, several examples of the application of multivariate approaches to the study of complex DNA structures are given.

G-quadruplex structures



The figure shows the structure of a G-tetrad (planar ensemble of four guanine bases) stabilized by hydrogen bonding and a monovalent cation resided in the central channel

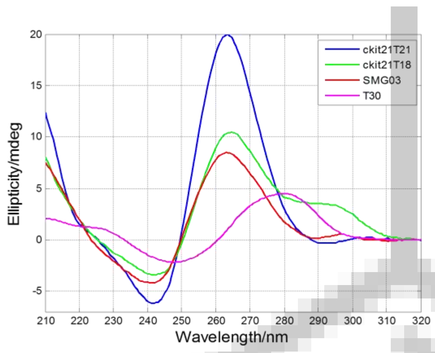


Top and lateral views of the 3D model for a parallel G-quadruplex from the Protein Data Bank (reference 2KYP). The top view shows the central channel where monovalent cations accommodate.
The lateral view shows the stacking of several G-tetrads

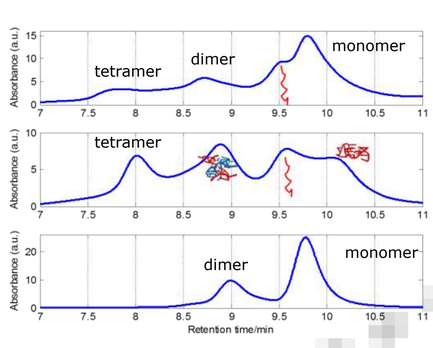
DNA sequences studied in this work

DNA sequence	Base sequence (5′ → 3′)	Comment
ckitG1T18	C GGG C GGG CGCGA GGG **AT** GGG	Mutation of a G-rich sequence in *c-kit* gene
ckitG1T21	C GGG C GGG CGCGA GGG **A** GGG **T**	Mutation of a G-rich sequence in *c-kit* gene
SMG03	AA GGG CGA GG CA GG ACA GGG A	G-rich sequence in *SMCARCA4* gene
T30	TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT	Control sequence not forming G-quadruplex

G refer to a guanine base involved in the formation of G-tetrads. Both ckitG1T18 and ckitG1T21 G-quadruplex structures contain four G-tetrads



CD spectra measured at 20ºC and pH 7
T30: linear, partially stacked structure. **ckit21T21** and **SMG03** are typical of a parallel G-quadruplex structure.
ckit21T18 shows hybrid (parallel + antiparallel) structure

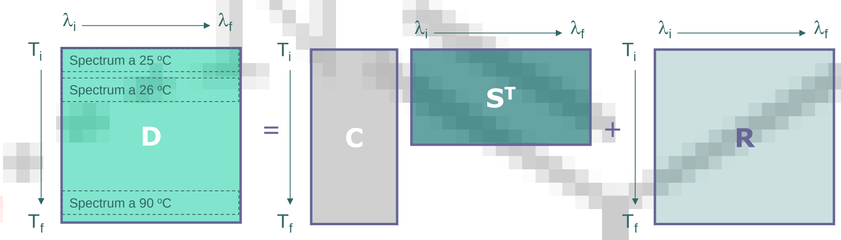


Size-Exclusion Chromatography of **ckit21T18** (top), **ckit21T21** (middle) and **SMG03** (bottom). The chromatograms show that these sequences may form a rather complex mixture of monomer and multimer structures, depending on the sample pretreatment.

The multivariate approach

Using appropriate multivariate methods, it may be possible:

1. To determine the number of species of conformations present throughout the experiment,
2. To quantify their relative concentration (distribution diagram, matrix **C**),
3. To recover their pure spectra (matrix **S**)

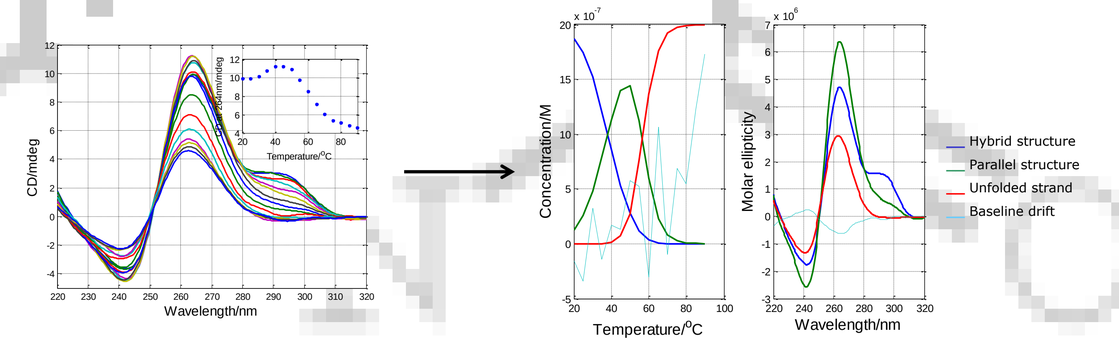


- The analysis of spectroscopic data measured along melting experiments has been done by means of an hybrid hard-soft-modeling method [4].
- The analysis of spectroscopic data measured along mole-ratio experiments has been done by means of hard-modeling methods, such as Equispec [5]

Thermally-induced G-quadruplex unfolding

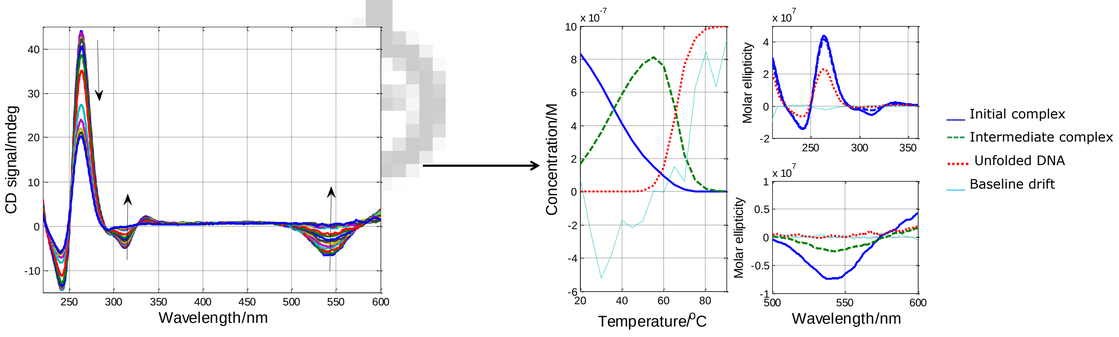
Upon heating, the three studied sequences showed different behavior:

- **ckit21T21** showed a simple unfolding process, and only two components (G-quadruplex and unfolded strand) were observed.
- **ckit21T18** showed a complex unfolding process, and multivariate analysis based on hybrid modeling [4] was needed to solve all components (see below)
- **SMG03** showed an intermediate behavior

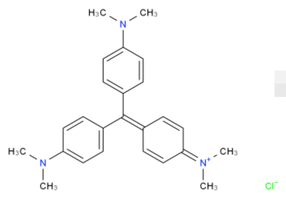


T_m values provide qualitative information on the stability of DNA folded structures.
The hybrid approach not only provides thermodynamic parameters related to the folding/unfolding equilibria, but also allows modeling of baseline drifts.
Baseline drifts at high temperatures are usually due to evaporation phenomena.

Melting of a 1:5 mixture (**ckitG1T21:CV**) at pH 7 monitored with CD spectroscopy
Data were also analyzed with the hybrid multivariate approach [4]

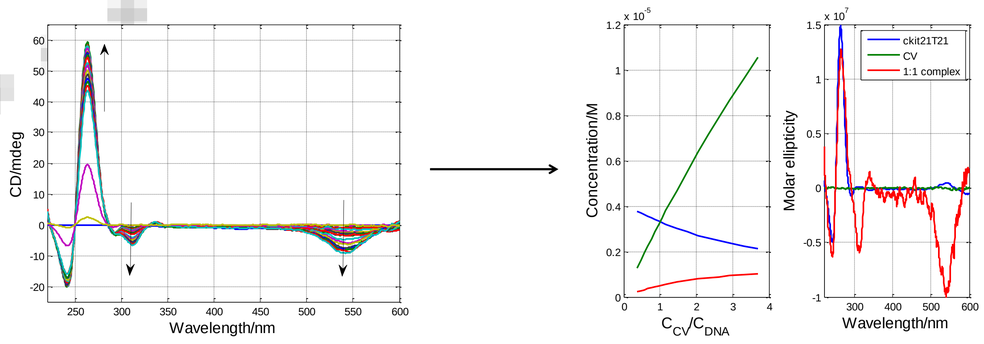


Interaction with a ligand



- Crystal Violet (CV) is a fluorescent triphenylmethane dye with antifungal and antimitotic properties.
- It has been shown to bind selectively to the G-quadruplex over duplex or single stranded DNA.
- Based on the interaction of CV with DNAs, many sensors have been developed to detect analytes like K^+, Cu^{2+} or Pb^{2+} ions, ATP, and activity of proteins.

The figure below shows the titration of **ckitG1T21** with increasing volumes of CV stock solution at pH 7 monitored with CD spectroscopy
Data were analyzed with the Equispec program, based on hard-modeling [5]



Upon binding, two induced CD bands appeared at 313 and 545nm., where CV absorbs.
The structure of G-quadruplex remains unaltered , which may indicate electrostatics-based interaction

Next table summarizes the results obtained:

System	T_m (ºC)	ΔH^0 (kcal·mol^{-1})	ΔS^0 (cal·K^{-1}·mol^{-1})	$\Delta G_{37ºC}$ (kcal·mol^{-1})	DNA:CV stoichiometry	Logarithm of the overall binding constant
ckit21T21	61±2	-36±2	-109±6	-2.6		
ckit21T21 + CV	36±3 66±2	-18±1 -65±3	-58±4 -191±13	0.0 -5.5	1:1	4.7±0.2
ckit21T18	38±2 57±2	27±3 -54±8	88±6 -162±18	-0.3 -2.5		
ckit21T18 + CV	46±2 56±2	-6±1 -37±2	-20±2 -112±6	-0.1 -2.0	1:1	4.9±0.2
SMG03	35±3	-27±2	-83±5	0.8		
SMG03 + CV	55±2	-16±1	-50±1	-0.5	1:1	5.3±0.2

References

[1] Benabou, S.; Aviñó, A.; Eritja, R.; González, C.; Gargallo, R. *RSC Advances* **2014**, 4, 26956-26980
[2] Neidle, S. *The FEBS Journal,* **2010**, 277, 1118-1125
[3] Biffi, G.; Antonio, M.Di; Tannahill, D.; Balasubramanian. S. *Nat. Chem.* **2014**, 6, 75
[4] Gargallo, R. *Anal. Biochem.* **2014**, 466, 4-15
[5] Dyson, R.; Kaderli, S.; Lawrence, G.A.; Maeder, M.; Zuberbühler, A.D. *Anal. Chim. Acta* **1997**, 353, 381-393

Acknowledgments

Thanks are due to Dra. Anna Aviñó and Dr. Ramon Eritja. Funding from the Spanish government and recognition from the Generalitat de Catalunya are acknowledged.

The background shows the CH⁺·C base pair, the building block of the i-motif DNA.