Chemometrics applied to the analytical study of the conformational equilibria of two guanine- and cytosine-rich sequences located near the promoter region of the *N-myc* oncogene

Sanae Benabou[1], Joaquim Jaumot, Raimundo Gargallo (*sbenabbe7@alumnes.ub.edu*)

[1]*Department of Analytical Chemistry, University of Barcelona, 08028 Barcelona, Spain*

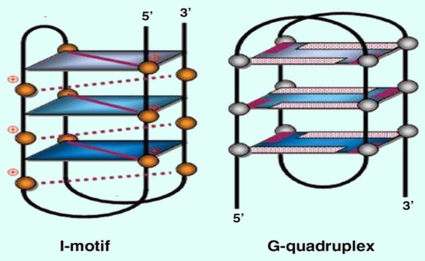
XIII CHEMOMETRICS IN ANALYTICAL CHEMISTRY, 25-29 June 2012, Budapest, Hungary

Introduction

In addition to the double helix, DNA shows other complex configurations such as quadruple structures, which consist on the association of four DNA stretches [1]. The most studied quadruple structures are:

i-motif: This is formed by cytosine-rich sequences in which two cytosines (one of them being protonated) are linked by three hydrogen bonds. This is the only DNA structure showing base pair intercalation.

G-quadruplex: This is formed by guanine-rich sequences. The building block is the G- tetrad, a planar arrangement of four guanines linked by hydrogen bonds.



I-motif G-quadruplex

It has been shown the formation *in vitro* of such structures in DNA sequences corresponding to the end of telomeres and to the promoter regions of several oncogenes, such as *c-kit*, *c-myc* or *bcl-2* [2].

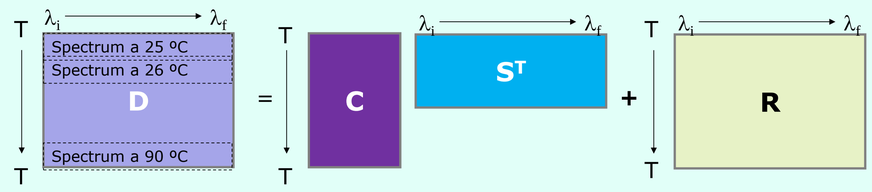
In the present study, we have focused our attention on two **cytosine**- and **guanine**-rich (**n-myc01** and **n-myc02**, respectively) sequences located near the promoter region of the *n-myc* gene. The interest in the study of this gene lies in the fact that it is important in defining the prognosis and treatment of neuroblastic tumors [3].

Methodology

. Spectrophotometric techniques: circular dichroism (CD) and molecular absorption
. Analysis of spectroscopic data: Multivariate Analysis

The whole set of spectra recorded throughout the experiment is arranged in a table or data matrix **D**. Using appropriate methods, it is possible to:

1. Determine the number of species of conformations present throughout the experiment (using SVD),
2. Quantify their relative concentration (distribution diagram, matrix **C**),
3. Recover their circular dichroism and/or molecular absorption spectra (pure spectra, matrix **S**)



In this study we have used two multivariate analysis methods: MCR-ALS (Soft modeling) and EQUISPEC (Hard modeling).

Results

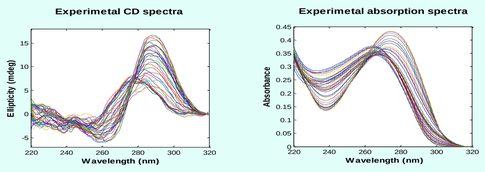
Influence of pH on the solution equilibria of n-myc01 and n-myc02

The experimental CD and molecular absorption spectra recorded throughout an acid-base titration of n-myc01 were analyzed with EQUISPEC.

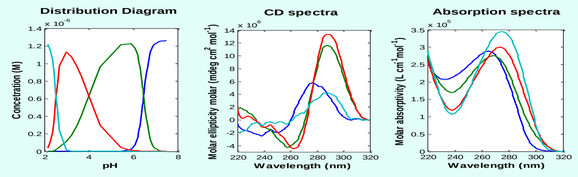
In this case, four acid-base species were needed to fit experimental data. The corresponding CD and molecular absorption spectra were calculated, as well as the distribution diagram.

Blue line: a neutral structure with a simple helix.
Green line:"*i-motif*" with some deprotonated bases.
Red line: "*i-motif" with some protenated bases*.
Cyan line:protonated form of "*i-motif*".

The ***i*-motif structure** is formed at pH slightly lower than 7 and it reaches its maximal stability around the pKa of free cytosine.



Acid-base titration of n-myc01 1.3 μM, 25 ºC



Results of EQUISPEC analysis

A similar study was carried out for the complementary guanine-rich sequence (**n-myc02**).

In this case, only three acid-base species were needed to fit experimental data:

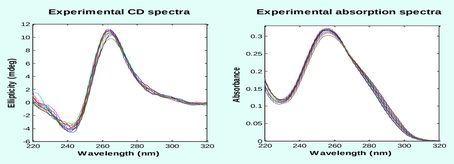
Blue line: "G-quadruplex" with A and C bases deprotonated.

Green line: "G-quadruplex" with protonated C bases and deprotonated A bases.

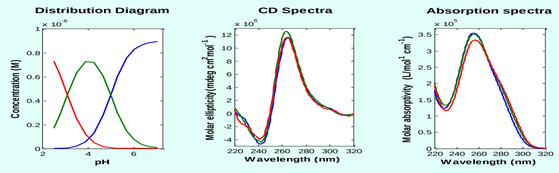
Red line: "G-quadruplex" with A and C bases protonated.

The G-quadruplex structure is well maintained within the considered pH range.

The shape of the pure CD spectra allows the identification of a **parallel** G-quadruplex



acid-base titration of n-myc02 0,9 μM, 25 ºC

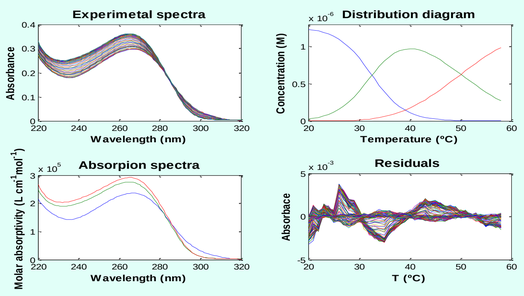


Results of EQUISPEC analysis

Influence of temperature on the stability of *i*-motif formed by n-myc01

Melting experiment of cytosine-rich sequence

The analysis of one of the molecular absorption-monitored melting experiment *of i-motif* is shown here.

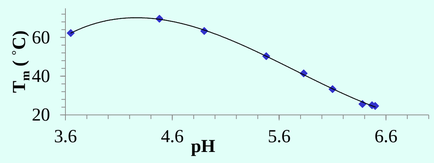


$C_{n\text{-}myc01}$ = 1,26 μM, pH = 6.1, 150 mM of KCl.
— "*i*-motif";
— partially unfolded species;
— fully unfolded species

The value of T_m is 33 ± 1 ºC.

Dependence of T_m with pH

In the pH range 6.8 to 4.0, T_m values were almost a linear function of pH



The value of T_m is maximal around the pK_a of free cytosine (4.5 approx.).

Molecularity of structure

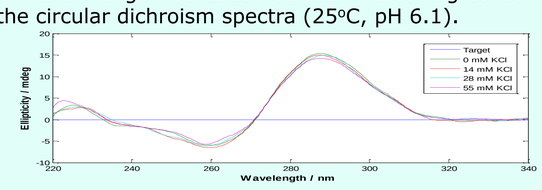
The variation of the melting temperature (T_m) upon the concentration of oligonucleotide allows the assessment of the molecularity.

C_{DNA}(M)	Tm (ºC)
5,0E-07	33,0
1,3E-06	33,0
5,0E-06	32,0

C_{DNA} ↑ ⇨ constant T_m ⇨ Intramolecular folding

Effect of ionic strength

No significant change was observed with increasing concentration of KCl in the circular dichroism spectra (25ºC, pH 6.1).



On the contrary, the melting temperature at pH 6.1 decreased from 52°C (without added salt) to 33°C (at 150 mM KCl). This behavior can be explained because of the shift of the pK_a of cytosine to higher values in a low-salt buffer.

pH	C_{KCl}(mM)	Tm (ºC)
6,10	0,0	52,5
6,10	55,0	44,2
6,10	150,0	33,3

Effect of temperature on the thermodynamic parameters: Results of analysis according to the Van't Hoff equation

The thermodynamic parameters associated to the unfolding of **n-myc01** have been calculated from the spectra recorded throughout melting experiments and using a multivariate hard-modeling approach.

folded DNA ↔ unfolded DNA **K= [unfolded DNA]/[folded DNA]**

pH	ΔH° (kcal/mol)	ΔS° (cal/K·mol)	T_m (ºC)	ΔG °$_{25ºC}$ (kcal/mol)
3.7	60	179	62	6.6
4.5	78	228	70	10.0
4.9	83	248	63	9.1
5.5	73	225	50	5.9
5.8	57	179	41	3.6
6.1	52	170	33	1,2
6.4	59	197	25	0.3
7.0	Not calc.	Not calc.	< 15	Not calc.

- The *i*-motif structure has a maximal stability around the pK_a of cytosine.
- At pH values around 6.5 an increase of ΔH^o is observed, probably due to the formation of Watson-Crick pairs (G·C and A·T).

A similar procedure has been used to analyze melting data of **n-myc02** sequence.

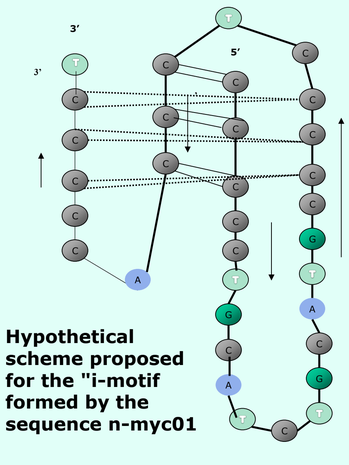
pH	ΔH° (kcal/mol)	ΔS° (cal/K·mol)	T_m (°C)	ΔG°$_{25°C}$ (kcal/mol)
3.9	114	313	89	20.7
4.5	85	236	86	14.4
5.1	73	207	80	11.4
6.1	74	209	79	11.2
7.1	73	209	77	10.9

- The G-quadruplex is very stable throughout the studied pH range.
- The stability increases at low pH values, probably because of the formation of additional base pairs involving protonated cytosine and/or adenine

The estimated uncertainties for the thermodynamic parameters are around 10% due to the small spectral changes observed. The uncertainty associated to the T_m values is 1 ºC.

Conclusions

- n-myc01 and n-myc02 sequences form intramolecular *i*-motif and G-quadruplex structures. According to CD data, the G-quadruplex shows a parallel conformation.
- The *i-motif* structure is stable within the pH range 2.5 - 6.5. Its maximal stability is around 4.5, the pK_a of free cytosine. It is not stable at pH 7 and 25ºC.
- The structure of the G-quadruplex is stable under physiological conditions within a whole pH range. Its stability is higher at pH 3.1
- Hard-modeling has been shown to be useful for the analysis of spectral data involving complex structures of DNA.



Hypothetical scheme proposed for the "i-motif formed by the sequence n-myc01

Bibliography

[1] Bloomfield, V.A.C., D. M.; Tinoco, I. JR., Nucleic Acids. Structures, properties and functions, ed. U.S. Books. 2000, Sausalito, California.

[2] Bucek, P.; Jaumot, J.; Aviñó, A.; Eritja, R.; Gargallo, R. pH-modulated Watson-Crick duplex – quadruplex equilibria of guanine-rich and cytosine-rich DNA sequences 140 bp upstream of the c-kit transcription initiation site. Chemistry - A European Journal, Vol. 15, 12663-12671(2009).

[3] G. Perini, D. Diolaiti, A. Porro, and G.D. Valle. In vivo transcriptional regulation of N-Myc target genes is controlled by E-box methylation. PNAS, Vol. 102, 12117 (2004).

Acknowledgements

We acknowledge funding from both the Spanish government (CTQ2009-11572 and CTQ2010-20541-C03- 01) and Catalan government (2009 SGR 45 and 2009 SGR 238).