Studying the solution equilibria of G-quadruplex region upstream of the B-cell lymphoma-2 P1 by means of multivariate data analysis methods

M. del Toro[1], J. Jaumot[1], A. Aviñó[2], R. Eritja[2], R. Gargallo[1]

1 Department of Analytical Chemistry, University of Barcelona, 08028 Barcelona, Spain

2 Institut de Biologia Molecular de Barcelona, CSIC, 08034 Barcelona, Spain

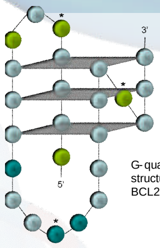
www.ub.es/gesq/dna

INTRODUCTION

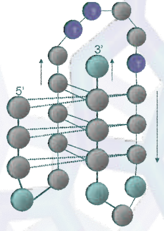
bcl-2 is a protein which is associated with chemotherapy resistant disease, aggressive clinical course, and poor survival in patients with B-cell lymphoproliferative disorders. The human ***bcl-2* gene** which encodes this protein has two promoters, P1 and P2. Promoter P1 is a DNA region rich in guanine and cytosine bases. Recently, it has been shown that guanine-rich DNA regions can form complex structures known as **G-quadruplex**, whereas cytosine-rich regions can form ***i*-motif** structures [1].

Researchers suspect that hundreds of thousands of DNA sequences sprinkled throughout the human genome are potential quadruplex-forming sites [2]. Quadruplex DNA seem to contribute to diverse biological functions, such as the telomere-ends or several promoters, such as *bcl-2*. Directing drugs to these sites might be a way of artificially regulating gene expression and thus providing medicinal benefits such as anticancer activity.

Here, we show the results obtained in the study of the **solution equilibria** of the guanine-rich region (BCL2 sequence) in the promoter region of *bcl-2* by means of **multivariate data analysis** methods [3-5]. The **goal** is the characterization of the acid-base and conformational equilibria of this sequence, as well as the study of its interaction with a porphyrin-based ligand (TMPyP4) and with the cytosine-rich complementary strand (BCL2C).



G-quadruplex structure resolved for BCL2 sequence [1]



Putative *i*-motif structure for BCL2C sequence

DATA ANALYSIS

Spectra recorded in acid-base titrations, melting experiments or mole ratio studies were arranged in a data matrix **D**.

Two methods were applied to analyze **D**: the hard modeling-based **EQUISPEC** program [3] and the soft modeling-based **MCR-ALS** method [4].

The concentration profiles (**C**) and pure spectra ($\mathbf{S}^T$) for all spectroscopically active species present in the system are calculated from the decomposition of **D** according to equation:

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

Decomposition of **D** according to (1) with EQUISPEC requires the fulfillment of a previously proposed simple chemical model. This model is defined by the stoichiometries of all species involved in the considered equilibria, and by approximate values of the equilibrium constants (K_c). In the case of the interaction of a ligand with a DNA sequence, this equilibrium constant can be written as:

$$\mathrm{DNA} + n\,\mathrm{ligand} \rightleftarrows [\mathrm{DNA}:\mathrm{ligand}_n] \qquad K_c = \frac{[\mathrm{DNA}:\mathrm{ligand}_n]}{[\mathrm{DNA}][\mathrm{ligand}]^n} \qquad (2)$$

EQUISPEC assumes that the value of the equilibrium constants do not vary upon advance of the considered reaction.

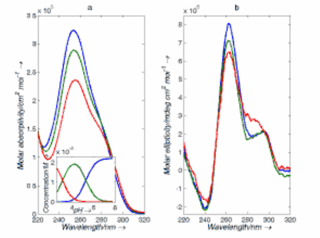
Hard modeling-based programs are especially adequate for the study of chemical equilibria involving monomers or short DNA sequences which do not show secondary effects related to polymeric structures, like polyelectrolyte effects or conformational changes. On the contrary, for large DNA sequences or when analyzing data from melting experiments, it is known that the equilibrium constants vary upon advance of the considered reaction or conformational change. In these cases, analysis of multivariate data is feasible by applying soft modeling-based methods because they do not require the previous proposal or compliance of any species model [5-6].

All EQUISPEC and MCR-ALS calculations were performed using MATLAB routines (version 6, The Mathworks Inc, Natick, MA).

INFLUENCE OF pH ON THE STABILITY OF THE BCL-2 G-QUADRUPLEX

The results of the analysis with **Equispec** of experimental data recorded along an **acid-base** titration of BCL2 are shown below. Two pH transition midpoints were determined at 5.2 ± 0.2, and 3.1 ± 0.2 (inset). The first value has been related to the protonation of cytosines, whereas the second one has been related to the protonation of adenine bases.

The resolved CD (b) spectrum for the neutral species show positive maxima at 260 and 290 nm. The 260 nm band has been assigned to the parallel strand quartets, and the 290 nm band to the external loop residues. The resolved CD spectra show that the G-quadruplex structure is well maintained in the pH range studied.

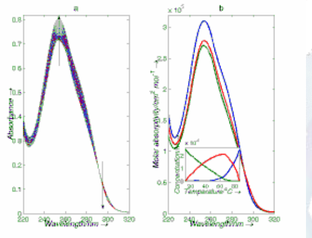


Melting experiments of BCL2 at several pH values in the range 4 - 7 were also carried out and data analyzed with **MCR-ALS**.

At pH 7.1, the determined T_m value for the disruption of the G-quadruplex structure formed by BCL2 was 76 ± 1 °C (see figure).

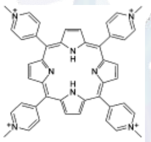
At more acid pH values, T_m increases, indicating the formation of additional bonds at the loops.

T_m is independent of the concentration, indicating the formation of a monomeric structure.



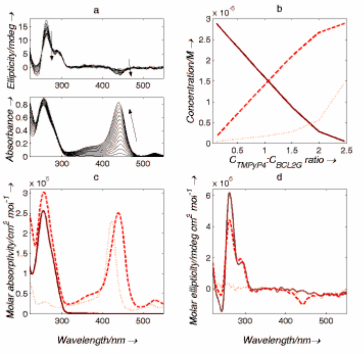
INTERACTION WITH THE PORPHYRIN LIGAND TMPyP4

The interactions of BCL2 with the **porphyrin TMPy4** have been studied.



Simultaneous analysis of CD and molecular absorption data (a) with EQUISPEC showed that only an interaction complex is formed between TMPyP4 and BCL2. The calculated stoichiometry and the value for the equilibrium constant for this complex were **1:2 (BCL2:TMPyP4)** and $K_c = 5.0\ (\pm 2.3) \cdot 10^{13}\ M^{-2}$ (b)

The red shift of the absorption band from 422 nm (free TMPyP4) to 444 nm (complex) (c), the appearance of a weak negative induced CD band around 450 nm and the maintenance of the CD bands at 263 and 240 nm (d) suggest an **end-stacking mechanism**, where the G-quadruplex structure is retained.

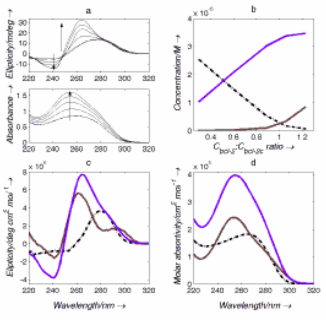


Continuous line: BCL2; dotted line: TMPyP4; dashed line: BCL2:TMPyP4 complex

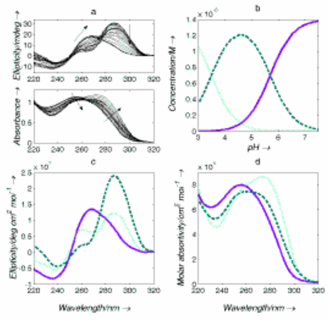
Melting experiments of the ligand:BCL2 mixtures provide information about the relative affinity of the ligand for the G-quadruplex and unfolded conformations of BCL2. The calculated T_m value for the melting of a mixture of BCL2 and TMPyP4 (after data analysis with MCR-ALS) was 81 (± 1) oC, i.e., the presence of **TMPyP4 clearly stabilizes** the G-quadruplex structure.

INTERACTION WITH THE COMPLEMENTARY CYTOSINE-RICH STRAND

In vivo, the P1 promoter on the human *bcl-2* gene contains both guanine (BCL2) and cytosine (BCL2C) rich strands [6]. In an equimolar mixture of BCL2 and BCL2C at biological conditions of pH and ionic strength, it is expected a competition between quadruplex structures (G-quadruplex and *i*-motif) *versus* 24-base pair Watson-Crick duplex. In order to have a quantitative plot of this competition, mole-ratio and acid-base experiments of BCL2:BCL2C mixtures have been carried out:



Purple: Watson-Crick BCL2·BCL2C duplex; dashdot line: BCL2C; brown: BCL2



Purple: Watson-Crick BCL2·BCL2C duplex; dashed line: mixture of G-quadruplex BCL2 and *i*-motif BCL2C; dotted line: mixture of G-quadruplex BCL2 and random coil BCL2C

Here we show the formation of duplex structure at neutral pH and the possible presence of **minor concentrations of BCL2 and BCL2** isolated strands. Experimental spectra from the mole-ratio experiment were fitted with a simple model which described the duplex formation from the isolated BCL2 and BCL2C strands yielding an equilibrium constant (K_c) equal to 6.3 (± 2.9)•$10^7\ M^{-1}$. The proposed distribution diagram denotes the existence of minor concentrations of BCL2 G-quadruplex and BCL2C at pH 7.13 in the equimolar mixture. This is also observed from the results of the acid-base titration of a mixture: pH can modulate the formation of quadruplex or duplex structures.

CONCLUSIONS

- Chemometrics:
  - Multivariate methods have been shown to be a **powerful tool** in the analysis of spectroscopic data recorded along DNA studies
  - Appropriate selection of hard- or soft-modeling methods provide **reliable results** which can be compared with those obtained from complementary techniques, such as ITC, PAGE or SPR.
- Biophysics:
  - The G-quadruplex structure of BCL2 is well maintained through wide pH (3-8) and temperature (20 – 70 °C) ranges.
  - Addition of the complementary strand BCL2C clearly shifts the equilibrium to the formation of Watson-Crick duplex **only at neutral pH**.
  - TMPyP4 and BCL2 form a 2:1 complex. The **stability** of BCL2 G-quadruplex is **enhanced** upon interaction with the drug.

ACKNOWLEDGMENTS

We gratefully acknowledge grants BFU2007-63287/BMC and CTQ2006-15052-C02-01/BQU from the **Spanish *Ministerio de Educación y Ciencia***.

REFERENCES

[1] J. Dai, T. S. Dexheimer, D. Chen, M. Carver, A. Ambrus, R. A. Jones, D. Yang. An intramolecular G-quadruplex structure with mixed parallel / antiparallel G-strands in the human bcl-2 promoter region in solution. *J. Am. Chem. Soc.* 2006, 128, pp. 1096 - 1098.

[2] S. Borman. Ascent of Quadruplexes. *Chemical & Engineering News* **2007**, 85, pp. 12-17.

[3] R. Dyson, S. Kaderli, G.A. Lawrence, M. Maeder, A.D. Zuberbühler. Second order global analysis: the evaluation of series of spectrophotometric titrations for improved determination of equilibrium constants. *Analytica Chimica Acta* **1997**, 353, pp. 381-393.

[4] R. Tauler, Multivariate curve resolution applied to second order data. *Chemometrics and Intelligent Laboratory Systems* **1995**, 30, pp. 133-146.

[5] J. Jaumot, M. Vives, R. Gargallo. Application of multivariate resolution methods to the study of biochemical and biophysical processes. *Analytical Biochemistry* **2004**, 327, pp. 1-13.

[6] N. Khan, A. Aviñó, R. Tauler, C. Gonzalez, R. Eritja, R. Gargallo. Solution equilibria of the i-motif-forming region upstream of the B-cell lymphoma-2 P1 promoter. *Biochimie* **2007**, 89, pp. 1562-172.