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[2], A. Aviñó[3], R. Eritja[4], R. Gargallo[5]

[1] Department of Analytical Chemistry, University of Barcelona, 08028 Barcelona, Spain, miqueldeltoro@gmail.com

[2] Department of Analytical Chemistry, University of Barcelona, 08028 Barcelona, Spain, joaquim@apolo.qui.ub.es

[3] Institut de Biologia Molecular de Barcelona, CSIC, 08034 Barcelona, Spain, aaagma@ibmb.csic.es

[4] Institut de Biologia Molecular de Barcelona, CSIC, 08034 Barcelona, Spain, recgma@cid.csic.es

[5] Department of Analytical Chemistry, University of Barcelona, 08028 Barcelona, Spain, raimon_gargallo@ub.edu

Keywords: MCR-ALS, hard modelling, soft modeling, DNA, solution equilibria.

1 Introduction

Bcl-2 is an apoptosis regulating protein, overexpression of 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 oncoprotein has two promoters, designated P1 and P2, which control its transcriptional initiation. 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 (Figure 1) [1].

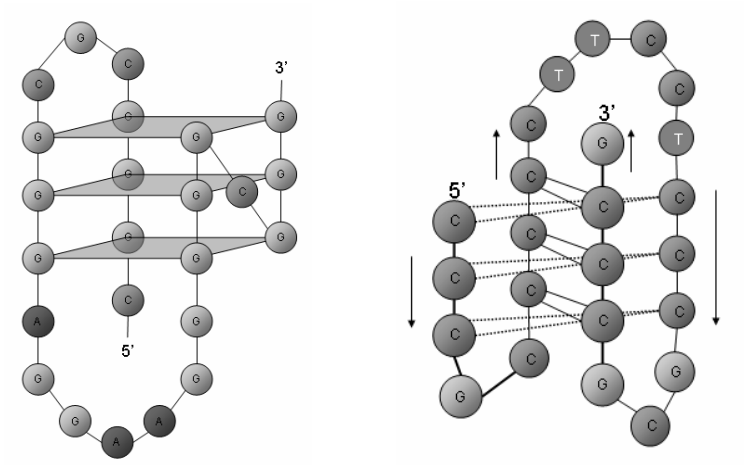

Figure 1 – G-quadruplex and i-motif structures formed by the guanine-rich and cytosine-rich regions in *bcl-2* P1 promoter. G, C, A and T stands for guanine, cytosine, adenine and tymine, respectively.

Researchers suspect that hundreds of thousands of DNA sequences sprinkled throughout the human genome are potential quadruplex-forming sites [2]. Quadruplex DNA displays a diversity of structures and that seem to contribute to diverse biological functions, such as the telomere-ends or several promoters, such as bcl-2 or c-myc. 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 in the promoter region of *bcl-2* by means of multivariate data analysis methods. The goal is the characterization of

the acid-base and conformational equilibria of this sequence, as well as the study of its interaction with several porphyrin-based ligands and with the C-rich complementary strand. To achieve the proposed objectives, spectroscopic techniques as well as Surface Plasmon Resonance, have been used. Spectroscopic data recorded along acid-base, melting and mole-ratio experiments have been analyzed by means of appropriate multivariate data analysis methods [3-5].

2 Material and methods

2.1 Experimental

24-bases long oligonucleotides with sequences 5'-CGG GCG CGG GAG GAA GGG GGC GGG-3' (BCL2) and 5'CCC GCC CCC TTC CTC CCG CGC CCG-3' (BCL2c) were prepared using standard 2-cyanoethyl phosphoramidites (Cruachem Ltd.). DNA strand concentration was determined by absorbance measurements using the nearest-neighbour method. TmPyP4 was purchased from Porphyrin Systems Gbr (Lübeck, Germany) and used without further purification.

Absorbance spectra were recorded on an Agilent HP8453 diode array spectrophotometer. Temperature was controlled by a 89090A Agilent peltier device. Circular dichroism (CD) spectra were recorded on a Jasco J-810 spectropolarimeter equipped with Julabo F-25/HD temperature control unit. Fluorescence measurements were done on an Aminco-Bowman AB2 spectrofluorimeter. pH measurements were performed with an Orion SA 720 pH/ISE meter and micro combination pH electrode (Thermo).

2.2 Spectroscopic data analysis

Spectra recorded in acid-base titrations, melting experiments or mole ratio studies were arranged in a table or data matrix **D**. Two multivariate data analysis methods were applied to analyze matrix **D**: the hard modeling-based EQUISPEC program [3] and the soft modeling-based MCR-ALS method [4]. Both approaches calculate concentration profiles and pure spectra for all spectroscopically active species present in the system from the decomposition of the experimental data matrix **D** according to equation:

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

Where **C** and $\mathbf{S}^{\mathbf{T}}$ are, respectively, the data matrices of the concentration profiles and of the pure spectra for each one of the spectroscopically active species or conformations present in the experiment. **E** is the data matrix of residuals not explained by the proposed species or conformations in **C** and $\mathbf{S}^{\mathbf{T}}$.

Decomposition of data matrix **D** according to equation 1 with EQUISPEC requires the fulfilment 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:

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

EQUISPEC assumes that the value of the equilibrium constants do not vary upon advance of the considered reaction.Hard modeling-based programs, like EQUISPEC, are especially adequate for the study of chemical equilibria involving monomers (or porphyrins) 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, application of hard modeling-based methods is rather difficult or even impossible since it is difficult to propose a simple species model describing the spectral behavior observed. In these cases, the analysis of multivariate data is possible by applying soft modeling-based methods because they do not require the previous proposal or compliance of any species model. Multivariate Curve Resolution using Alternating

Least Squares (MCR-ALS) is a soft modeling-based method that has been already widely applied in the study of acid-base and conformational transitions of DNAs and proteins [5-6].

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

3 Results and discussion

First, we have characterized the acid-base properties of BCL2 in the pH range 3 - 8. The analysis of experimental data recorded along an acid-base titration of BCL2 with Equispec allowed the calculation of the distribution diagram and of the pure spectra for each one of the species considered. The model which provided the best results was the one with pH transition midpoint values equal to 5.2 ± 0.2, and 3.1 ± 0.2 (Figure 2). 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 spectrum for the neutral species show positive maxima near 260 nm and near 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, probably because protonation of cytosines and adenines takes place at the loops (Figure 1).

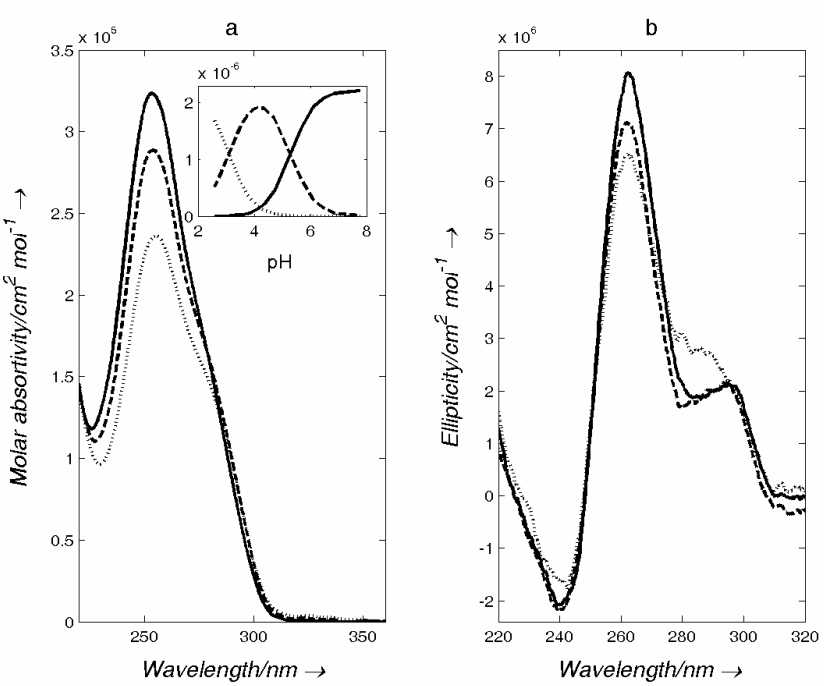

Figure 2. Calculated molecular absorption and CD spectra for the three acid-base species considered in the acid-base titration of BCL2. Inset: calculated distribution diagram. These results were obtained after simultaneous analysis of molecular absoprtion and CD data with EQUISPEC.

Molecular absorption and CD monitored melting experiments of BCL2 at several pH 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 ^oC. The melting temperature is independent of the concentration, indicating the formation of a monomeric structure.

The interactions of BCL2 with the porphyrin TmPy4 have been studied by mole-ratio, Surface Plasmon Resonance and melting experiments. Figure 3 shows the results obtained after the simultaneous analysis of molecular absorption and CD data recorded along a mole-ratio experiment. The analysis of the whole data set 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 (± 2.3) • 10^{13} M^{-2}. The resolved pure molecular absorption and CD spectra give some idea about the nature of the complex. The red shift of the absorption band from 422 nm (free TmPyP4) to 444 nm (complex), the appearance of a weak negative induced CD band around 450 nm and the maintenance of the CD bands at 263 and 240 nm suggest an end-stacking mechanism, where the G-quadruplex structure is retained.

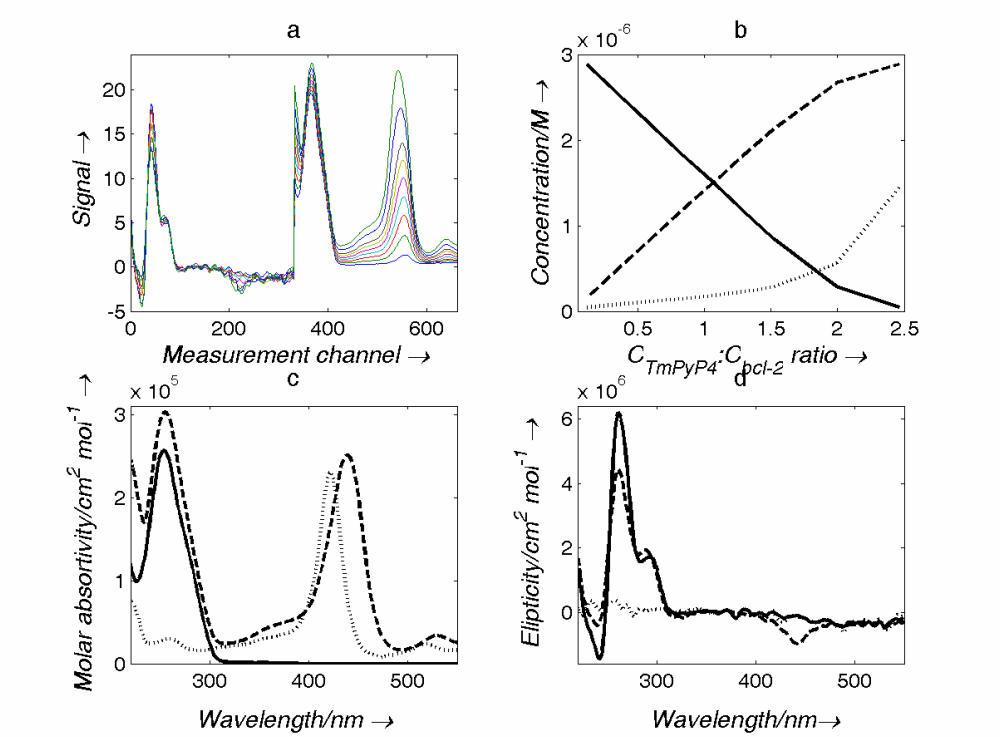

Figure 3. Interaction of BCL2 with TmPyP4. (a) Experimental data (absorbance values have been multiplied by 25); (b) Calculated distribution diagram. (c) Resolved pure molecular absorption spectra. (d) Resolved pure CD spectra. Solid line: BCL2 G-quadruplex, Dotted line: TmPyP4, Dashed line: interaction 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.

In vivo, the region upstream of the P1 promoter on the human *bcl-2* gene contains both guanine and cytosine rich strands. In addition to the studies carried out with the guanine-rich region (BCL2), the cytosine-rich sequence (BCL2c) has been already studied recently [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 (Figure 1) versus 24-base pair Watson-Crick duplex. In order to have a quantitative plot of this competition, mole-ratio, acid -base and melting experiments of BCL2:BCL2c mixtures have been carried out.

Here we show the formation of duplex structure at neutral pH and the possible presence of minor concentrations of BCL2 and BCL2 isolated strands. Figure 4 shows the results obtained from the titration of a BCL2c sample with increasing concentrations of BCL2 at pH 7.1. Experimental spectra 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 hairpin at pH 7.13 in the equimolar mixture.

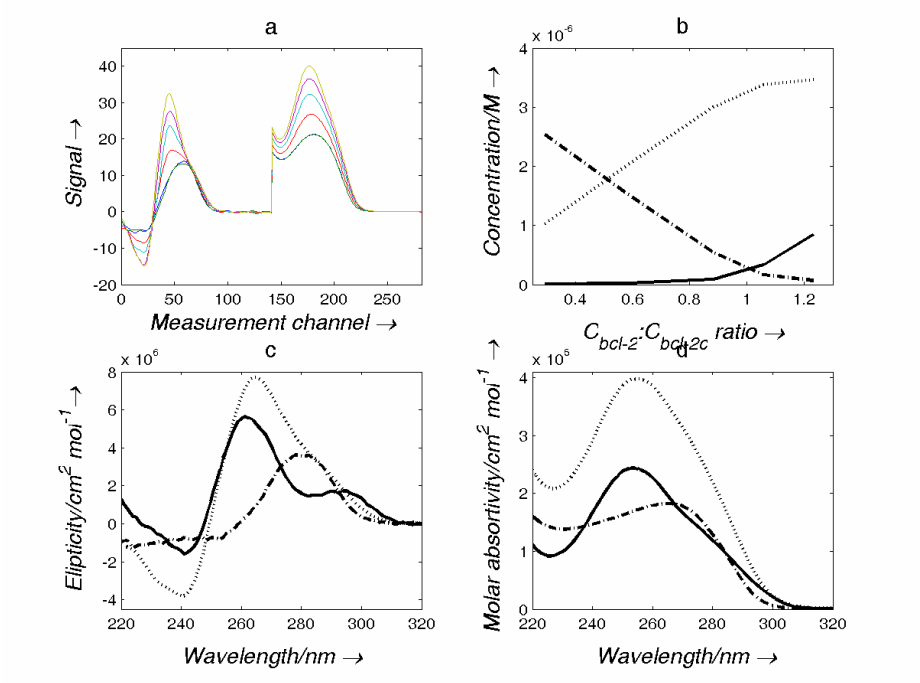

Figure 4. Interaction of BCL2c with BCL2. (a) Experimental data (absorbance values have been multiplied by 25); (b) Calculated distribution diagram. (c) Resolved pure CD spectra. (d) Resolved pure molecular absorption spectra. Dash-dot line: BCL2c hairpin, solie line: BCL2 G-quadruplex, dotted line: Watson-Crick 24 base pair duplex. C_{bcl2c} : 3.5 μM, pH 7.1.

4 Conclusion

Multivariate methods have been shown to be a powerful tool in the analysis of spectroscopic data recorded along DNA conformational changes. Appropriate selection of hard- or soft-modelling methods provide reliable results which can be compared with those obtained from complementary techniques, such as PAGE or SPR. In this work, the conformational and acid-base equilibria and the interaction of TmPyP4 with the guanine-rich region in *bcl-2* gene have been described.

5 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.