Hybrid hard-soft-modeling of unfolding processes involving G-quadruplex and i-motif DNA structures

Sanae Benabou, Raimundo Gargallo *(sbenabou_13@ub.edu)*

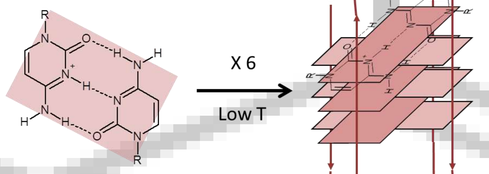
Department of Analytical Chemistry, University of Barcelona, Martí i Franqués 1-11, 08028 Barcelona, Spain (ub.edu/gesq/dna)

Overview

- The stability of complex DNA structures like G-quadruplex or i-motif depends on experimental factors like pH, temperature or ionic strength, among others.
- Biophysical studies, such as melting experiments, provide information about the influence of these factors on the relative stabilities of these structures. In the case of melting experiments, heating induces DNA unfolding.
- Traditionally, melting experiments have been monitored spectrophotometrically in a univariate way, i.e., being the result of the measurement a vector of absorbance values as a function of temperature [1,2].
- Multivariate methods, when applied to these data sets, may be able to discern the existence of intermediates which may be unnoticed when melting experiments are monitored just from a univariate point of view.
- In previous works [3], a mathematical procedure based on hard-modeling was proposed for the determination of ΔH, ΔS, ΔG and melting temperature from spectroscopically-monitored multivariate melting experiments.
- Here, a modification of the method, which is now based on **hybrid hard-soft modeling**, is proposed to deal spectral contributions not related to DNA species

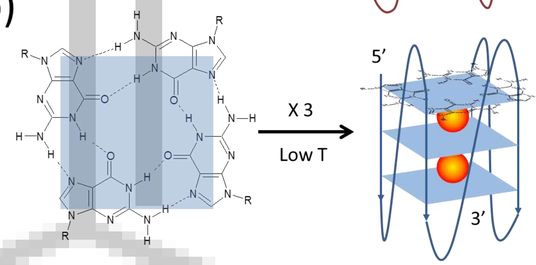
i-motif and G-quadruplex structures

a)



i-motif(a) is formed at acid pH values by the formation and stacking of C^+·C base pairs

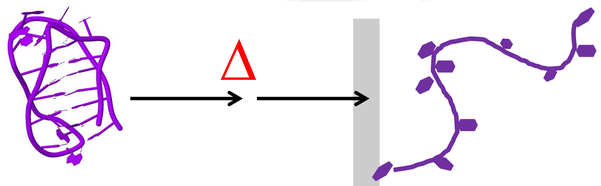
b)



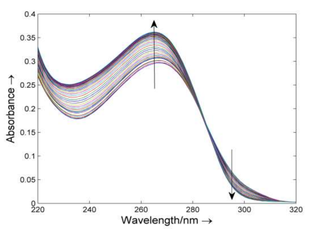
G-quadruplex (b) is formed by the formation and stacking of G-tetrads in presence of cations

Conformational equilibria and meltings

Let us consider the spectroscopically-monitored unfolding of a intramolecular *i*-motif DNA



- Unfolding is characterized by:
  - ΔH^0: energy need to break the attractive interactions in the folded DNA.
  - ΔS^0: its value usually increases when unfolding occurs.
  - ΔG^0: it quantifies the spontaneity of the process at a given temperature.
  - T_m: the value at which half of the initial folded molecules are still in the folded state.



Unfolding is revealed by an increase of the absorbance measured at 260 nm.

For G-quadruplex and *i*-motif structures, hypochromicity is observed at 295 nm.

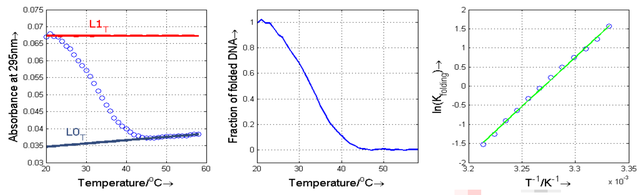
The classical univariate approach

- The univariate approach considers the existence of a **two-state process**.
- Thermodynamic data are calculated from the absorbance at 295 nm following this procedure:
  - First, appropriate baselines are drawn (LO_T and $L1_T$).
  - Second, the relative amount of ordered and unordered DNA is calculated for each one of measured absorbance values:

    Fraction of folded DNA $_T$ = (LO_T – A_T) / (LO_T – $L1_T$)
  - Third, the appropriate expression for the equilibrium constant is applied [1,2]. In our example, the equilibrium for an intramolecular unfolding is straightforward:

    $K_{folding}$ = [fraction of folded DNA] / [fraction of unfolded DNA]
  - Fourth, fitting the van't Hoff equation provides quite reliable values for thermodynamic data:

    $-R \cdot T \cdot \ln K_{folding} = \Delta H^0 - T\Delta S^0$

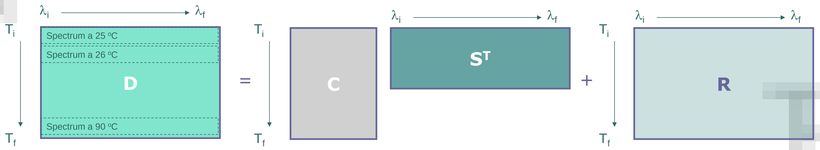


ΔH^0 (kcal/mol)= -52 ± 2 (95%); ΔS^0 (cal/K·mol)= -169 ± 6 (95%)
T_m (ºC)= 32.6

The multivariate approach

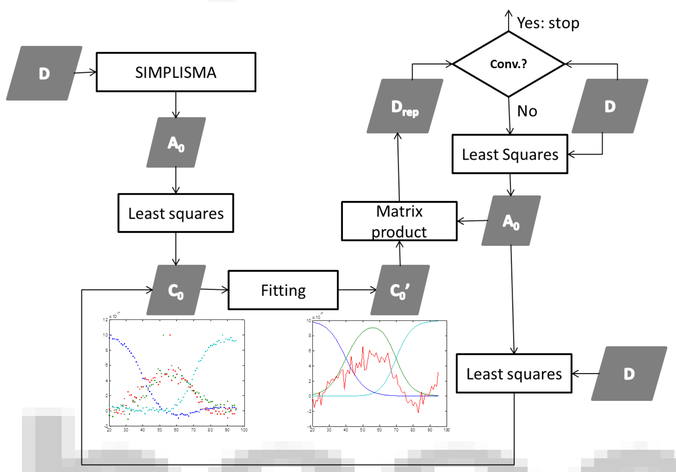
Using appropriate multivariate methods, it may be possible:

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 already done by means of soft-modeling methods, such as Multivariate Curve Resolution [4].
- The main advantage is that the previous proposal of a physico-chemical model is not needed.
- The main drawback are the difficulties found when trying to explain the nature of the considered components and the presence of mathematical ambiguities and rank deficiency.

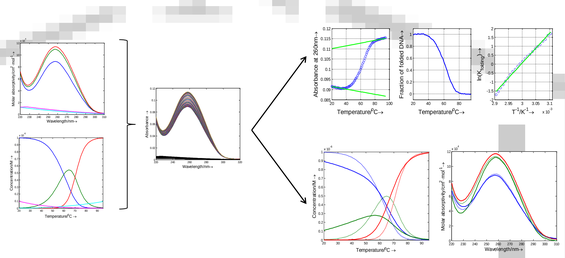
The hybrid hard-soft modelling approach



- The program makes use of the initial estimates of pure spectra obtained with SIMPLISMA [5], as well as of initial estimates for ΔH^0 and ΔS^0.
- Writen in Matlab®, it is based on the use of *lsqcurvefit.m* routine [6] and on hybrid modelling [7].
- Matrix **C** is calculated using the previously developed equations [1,2].

Validation with simulated data

- The simulated distribution diagram includes, in addition to three conformational profiles, the contribution of two interferences. The contribution of the interferences is more pronounced at low and high temperatures, a situation that is often found in the thermal study of biomolecules.
- The simulated spectra for the interferences also try to mimic the experimental contributions of base line drifts, being the molar absorptivity higher at lower wavelengths.

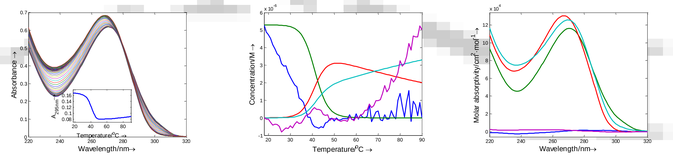


- Univariate analysis did not allow the resolution of the two simulated transitions, even after the successful subtraction of baseline drifts.
- Application of the pure hard-modeling method (dashed lines) considering a three species model did not allow the complete recovery of the thermodynamic parameters used for the simulation.
- Application of the proposed hybrid approach allowed an acceptable recovery of the concentration profiles related to the unfolding process (continuous line).

Application to experimentally measured spectroscopic data

Unfolding of an intramolecular *i*-motif DNA at pH 6.1 (*5′-$C_3T_3C_3T_3C_3T_3C_3T_3$-3′*)

- In this case, univariate analysis allows a good estimation of the thermodynamic parameters related to the main transition but it is difficult to discern what happens at higher temperatures.
- Analysis of the multivariate data by means of the hybrid approach was done considering five components, three of them related to the unfolding process and two of them related to interferences.



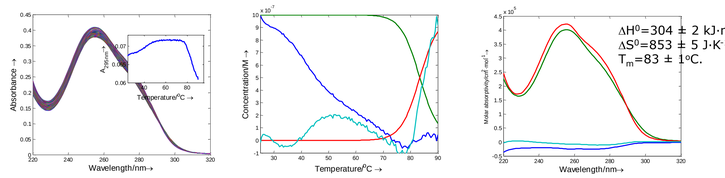
—: i-motif
—: partially stacked strand
—: completely unstacked strand

ΔH^0_1=271 ± 2 kJ·mol^{-1}
ΔS^0_1=857.4 ± 0.2 J·K^{-1}·mol^{-1}
T_{m1}=42 ± 1ºC

ΔH^0_2=24 ± 5 kJ·mol^{-1}
ΔS^0_2=70 ± 1 J·K^{-1}·mol^{-1}
T_{m2}=68 ± 15ºC

Unfolding of an intramolecular G-quadruplex (*5′-T GGG T GGG TGTGT GGG T GGGG-3′*)

- In this case, because of the high stability of the structure in front of temperature the accurate definition of the lower and, especially, upper baselines is difficult, which in turn would eventually lead to imprecise estimations of the thermodynamic parameters.
- Analysis of the multivariate data by means of the hybrid approach was done considering four components, two of them related to the unfolding process and two of them related to interferences.



ΔH^0=304 ± 2 kJ·mol^{-1}
ΔS^0=853 ± 5 J·K^{-1}·mol^{-1}
T_m=83 ± 1ºC.

References

[1] Mergny, J.-L.; Lacroix, L. *Oligonucleotides*, **2003**, 13, 515-537
[2] Breslauer, K.J. *Methods in Enzymology* **1987**, 259, 221-245
[3] Fernandez, S.; Eritja, R.; Aviñó, A.; Jaumot, J. Gargallo, R. *Int. J. Biol. Macromol.* **2011**, 49, 729-736
[4] Kumar, P.; Verma, A.; Maiti, S.; Gargallo, R.; Chowdhury, S. *Biochemistry* **2005**, 44, 16426-16434
[5] Windig, W.; Guilment, J. *Anal. Chem.* **1991**, 63, 1425-1432
[6] Optimization Toolbox. Version 4.0 (R2008a)
[7] de Juan, A.; Maeder, M.; Martínez, M.; Tauler, R. *Anal. Chim. Acta* **2001**, 442, 337–350

Acknowledgments

The assistance provided by Joaquim Jaumot (IQAC-CSIC, Barcelona, Spain) in some of the stages of programming is appreciated.

Funding from Spanish government (CTQ2012-38616-C02-02) and Autonomous Catalan government (2014SGR1106) is acknowledged.