Hard-modeling of multivariate spectroscopic data recorded throughout melting experiments of DNA structures

Raimundo Gargallo *(raimon_gargallo@ub.edu)*

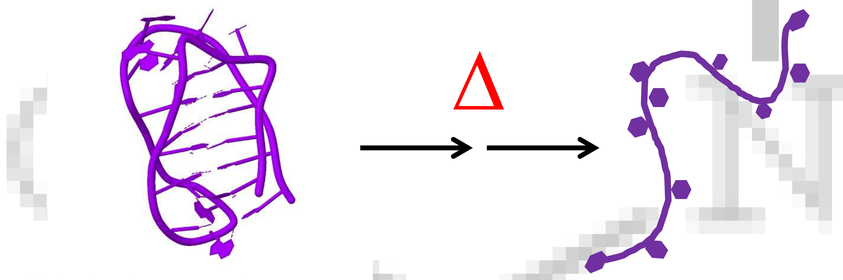
Department of Analytical Chemistry, University of Barcelona, Diagonal 647, 08028 Barcelona, Spain (www.ub.es/gesq/dna)

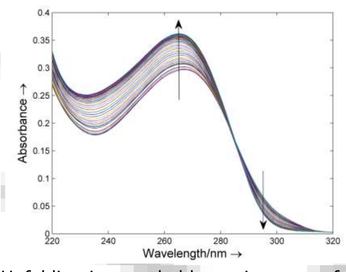
Overview

- The stability of DNA secondary structures depends on factors like pH, temperature, ionic strength... and on the presence of ligands, such as proteins or drugs.
- The knowledge about the stability of these structures is obtained from biophysical experiments such as meltings. In these, a controlled displacement of a conformational equilibrium occurs by means of a change of temperature.
- Traditionally, melting experiments have been monitored spectrophotometrically measuring the absorbance at 260 nm, being the result of the measurement a vector of absorbance values as a function of temperature. Appropriate univariate methods have been developed to analyze such data [1,2].
- The univariate approach has several drawbacks like the difficulty when modelling systems where intermediates are present.
- Here, a hard-modeling-based procedure is proposed for the determination of thermodynamic data (ΔH^0, ΔS^0, ΔG^0 and the melting temperature, T_m) from spectroscopically-monitored multivariate melting experiments.

Conformational equilibria and meltings

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





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.

- 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.

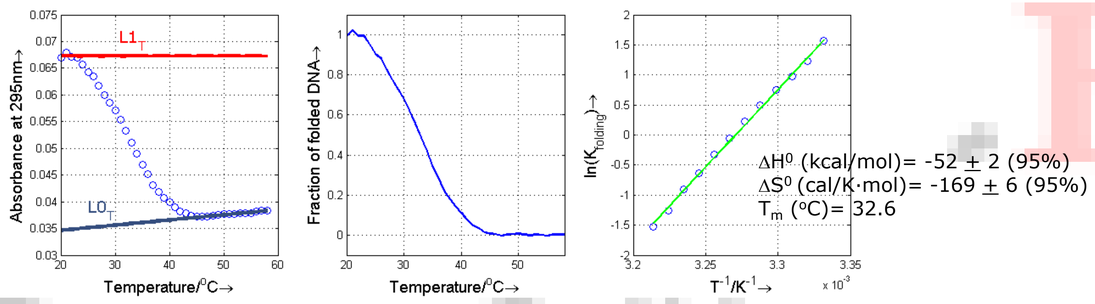
The classical univariate approach

- 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 (**$L0_T$** 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 $_T$ = (**$L0_T$** – A_T) / (**$L0_T$** – **$L1_T$**)
  - 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:

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

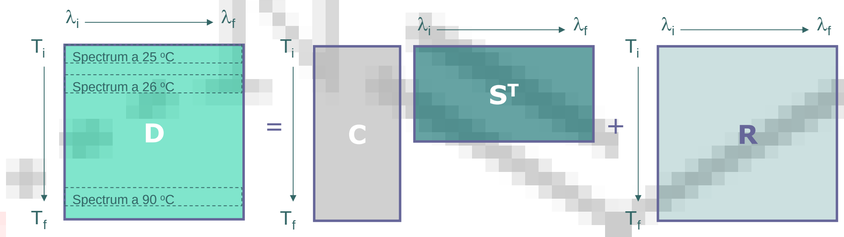


ΔH^0 (kcal/mol)= -52 ± 2 (95%)
ΔS^0 (cal/K·mol)= -169 ± 6 (95%)
T_m (ºC)= 32.6

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**)

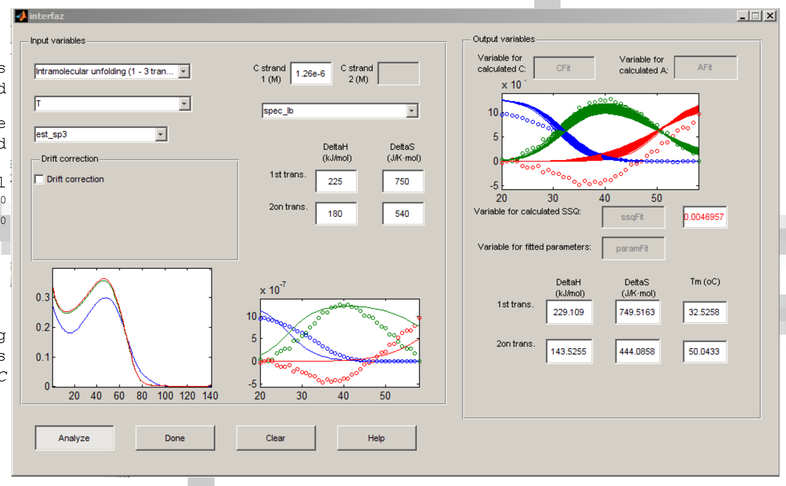


- The analysis of spectroscopic data measured along melting experiments has been already done by means of soft-modeling methods, such as Multivariate Curve Resolution [3,4].
- The main advantage is that the previous proposal of a physico-chemical model is not needed.
- The main drawback are the difficulties found when trying to explain the nature of the considered components and the presence of mathematical ambiguities and rank deficiency.

The hard-modelling-based 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].
- Matrix **C** is calculated using the previously developed equations [1,2].

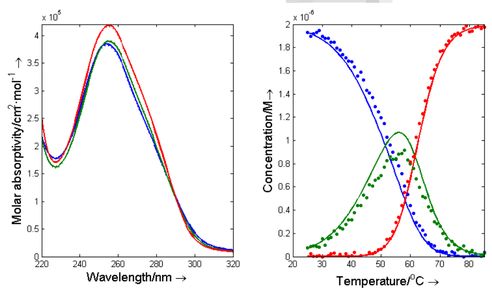
1. The model is selected
2. The variables are selected
3. Initial estimates for ΔH^0 and ΔS^0
4. Graphics showing the initial estimates of **S** and *C*



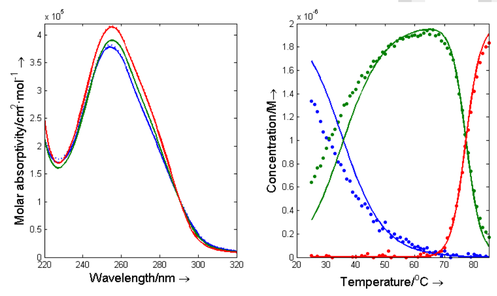
5. Graphical output of matrix **C**
6. Calculated fit
7. Output values for ΔH^0, ΔS^0 and T_m

Validation with simulated data

- Until now, it has been developed and validated for modelling intramolecular and duplex unfolding.
- The influence of noise and the concentration overlap has been studied using simulated data.
- In general, the method recovers satisfactorily the thermodynamic values used to build up the simulated data.



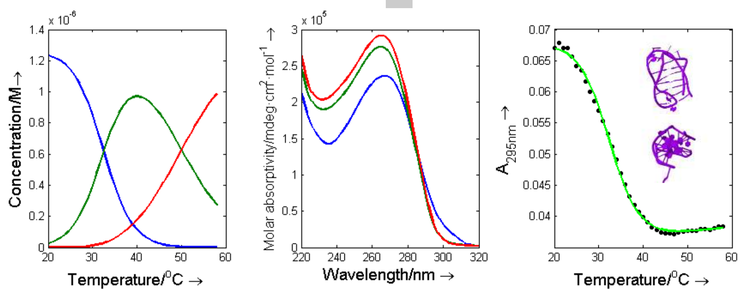
The results may be not enough accurate when strong overlap in both spectral and concentration dimensions are present. In this case, the values for the change in enthalpy and entropy for both transitions show prediction errors around 6%.



The results may also be not enough accurate when rotational ambiguities are present. In this example, the values for the change in enthalpy and entropy for the first transition show prediction errors around 30%.

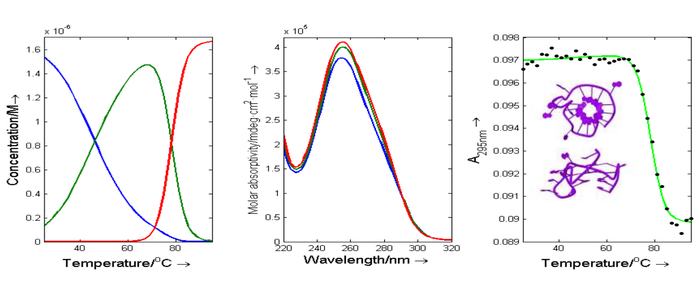
Application to experimentally measured spectroscopic data

Unfolding of an intramolecular *i*-motif DNA at pH 6.1 (*nmyc01*)



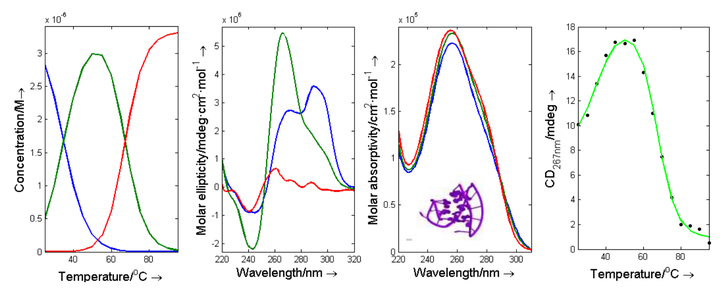
Proposed nature of the transition	ΔH^0 (kcal/mol)	ΔS^0 (cal/K·mol)	T_m (ºC)
Intramolecular *i*-motif → partially unfolded strand	55	180	32
Partially unfolded strand → completely unfolded strand	34	107	50

Unfolding of an intramolecular parallel G-quadruplex (*nmyc02*)



Proposed nature of the transition	ΔH^0 (kcal/mol)	ΔS^0 (cal/K·mol)	T_m (ºC)
G-quadruplex→ partially unfolded G-quadruplex (unstacked loops?)	22	69	46
Partially unfolded G-quadruplex→ completely unfolded strand	88	251	78

Unfolding of an intramolecular mixed G-quadruplex (*ckitG2T18*)



Proposed nature of the transition	ΔH^0 (kcal/mol)	ΔS^0 (cal/K·mol)	T_m (ºC)
Mixed G-quadruplex → parallel G-quadruplex	32	105	35
Parallel G-quadruplex→ completely unfolded strand	45	134	67

References

[1] Mergny, J.-L.; Lacroix, L. *Oligonucleotides*, **2003**, 13, 515-537
[2] Breslauer, K.J. *Methods in Enzymology* **1987**, 259, 221-245
[3] Bucek, P.; Jaumot, J.; Aviñó, A.; Eritja, R.; Gargallo, R. *Chem. Eur. J.* **2009**, 15, 12663- 12671
[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)

Acknowledgments

Thanks are due to Sanae Benabou, Sergio Fernández and Sintayehu Manaye. Funding from the Spanish government (CTQ2009-11572 and CTQ2010-20541-C03-01) and the support from the Generalitat de Catalunya (2009SGR45 and 2009SGR238) are thanked.

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