

# i-motif structures in long cytosine-rich sequences found upstream of the promoter region of the SMARCA4 gene

R. Gargallo<sup>1</sup>, S. Benabou<sup>1</sup>, A. Aviñó<sup>2</sup>, S. Lyonnais<sup>3</sup>, C. González<sup>4</sup>, R. Eritja<sup>2</sup>, A. De Juan<sup>1</sup>

<sup>1</sup> Department of Chemical Engineering and Analytical Chemistry, University of Barcelona, Barcelona, Spain (ub.edu/gesq/dna)

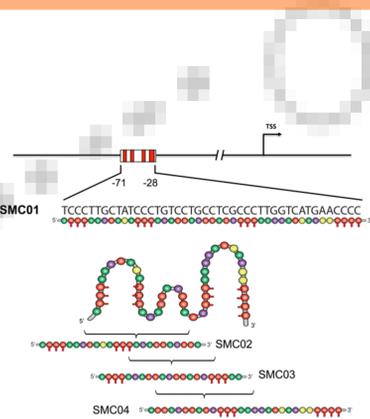
<sup>2</sup> Institute for Advanced Chemistry of Catalonia (IQAC-CSIC), Networking Center on Bioengineering, Biomaterials and Nanomedicine, Barcelona, Spain

<sup>3</sup> Institute of Molecular Biology of Barcelona (IBMB-CSIC), Barcelona, Spain

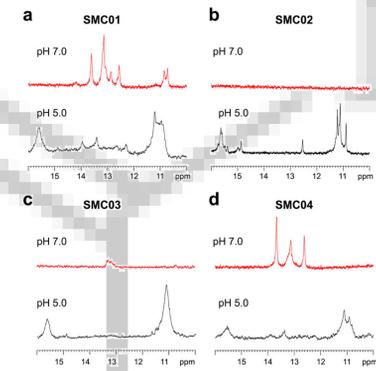
<sup>4</sup> Institute of Physical Chemistry "Rocasolano" (CSIC), Madrid, Spain

## 1. Overview

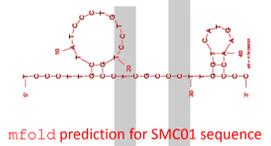
- Cytosine-rich oligonucleotides are capable of forming complex structures known as i-motif with increasingly studied biological properties [1,2].
- The study of sequences prone to form i-motifs located near the promoter region of genes may be difficult because these sequences not only contain repeats of cytosine tracts of disparate length but also these may be separated by loops of varied nature and length [3].
- In this work, the formation of an intramolecular i-motif structures by a long sequence located upstream of the promoter region of the SMARCA4 gene has been demonstrated. NMR, CD, Gel Electrophoresis, SEC, and multivariate analysis have been used. Not only the wild sequence (5'-TC<sub>3</sub>T<sub>2</sub>GCTATC<sub>3</sub>TGTC<sub>2</sub>TGC<sub>2</sub>TCGC<sub>3</sub>T<sub>2</sub>G<sub>2</sub>TCATGA<sub>2</sub>C<sub>4</sub>-3') has been studied but also several other truncated and mutated sequences.
- Despite the apparent complex sequence, the results showed that the wild sequence may form a relatively stable and homogeneous unimolecular i-motif structure, both in terms of pH or temperature.
- The model ligand TMPyP4 destabilizes the structure, whereas the presence of 20% (w/v) PEG200 stabilized it slightly.



## 2. NMR data

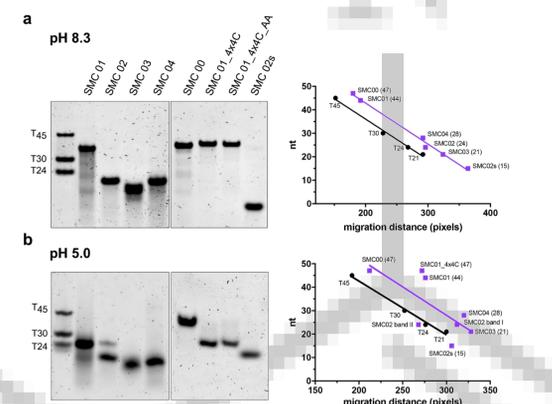


<sup>1</sup>H NMR spectra of SMC01 (a), SMC02 (b), SMC03 (c), and SMC04 (d) sequences at pH 7.0 and 5.0. The experimental conditions were 0.3 mM DNA, 100 mM KCl, 20 mM phosphate or acetate buffer, 5 °C.



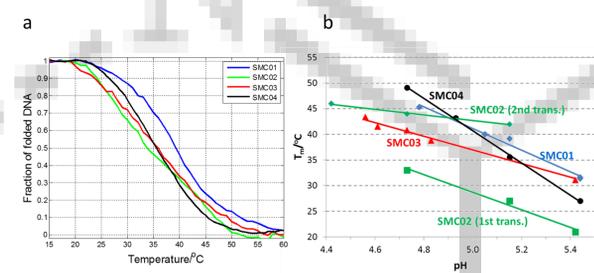
- NMR signals around 15.5 ppm indicate the existence of C-C<sup>+</sup> base pairs at pH 5.0.
- At pH 7.0, signals between 12 and 14 ppm indicate Watson-Crick base pairs, in agreement with mfold [4] predictions.

## 3. Molecularly



Name	Sequence (5'→3')
SMC00	TCAACTTGCTATCAACTGCTCCTGCTCGCAACTTGGTCATGAACAAC
SMC01	TCCC TTGCTATCCC TGTCCTGCTCGCC TGGTCATGAACCCC
SMC01_4x4C	TCCC TTGCTATCCC TGTCCTGCTCGCC TGGTCATGAACCCC
SMC01_4x4C_AA	TCCC TTGCTATCCC TGTCCTGCTCGCC TGGTCATGAACCCC
SMC02	TCCC TTGCTATCCC TGTCCTGCTCGCC TGGTCATGAACCCC
SMC02s	TCCC TTGCTATCCC T
SMC03	TCCC TGTCCTGCTCGCC TT
SMC04	TCCC TTGCTATCCC TGTCCTGCTCGCC TT
TT	TTCCC TTCCC TTCCC TTCCC TT

## 4. Thermal stability

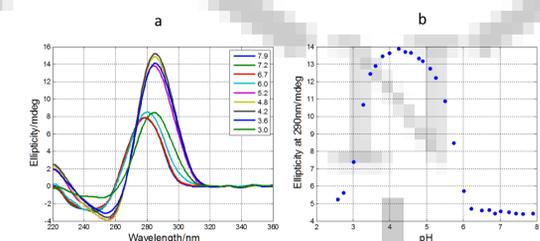


(a) Fraction of folded DNA calculated from the absorbance trace at 295 nm. (b) Plot of determined  $T_m$  values vs. pH. The experimental conditions were 2 mM DNA, 150 mM KCl, 20 mM acetate buffer, pH 5.2.

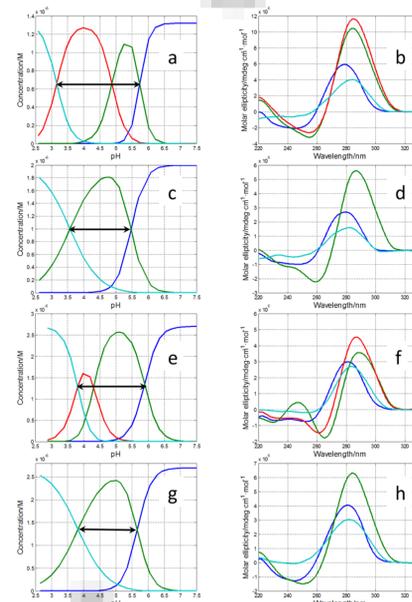
Sequence	pH 4.7				pH 5.2			
	$T_m$ (°C)	$\Delta H^0$ (kcal·mol <sup>-1</sup> )	$\Delta S^0$ (cal·K <sup>-1</sup> ·mol <sup>-1</sup> )	$\Delta G^0_{370c}$ (kcal·mol <sup>-1</sup> )	$T_m$ (°C)	$\Delta H^0$ (kcal·mol <sup>-1</sup> )	$\Delta S^0$ (cal·K <sup>-1</sup> ·mol <sup>-1</sup> )	$\Delta G^0_{370c}$ (kcal·mol <sup>-1</sup> )
SMC01	45	-40.5	-127.1	-1.0	38	-39.0	-124.9	-0.3
SMC02	33 / 44	n.d.	n.d.	n.d.	27 / 42	n.d.	n.d.	n.d.
SMC03	41	-27.9	-88.9	-0.3	36	-30.7	-99.5	0.1
SMC04	49	-26.1	-80.7	-1.0	36	-32.6	-105.7	0.1

- Analysis of thermal melting curves confirmed the dimeric structure of SMC02 at higher concentrations.
- In general, thermal stability of these i-motif structure is low

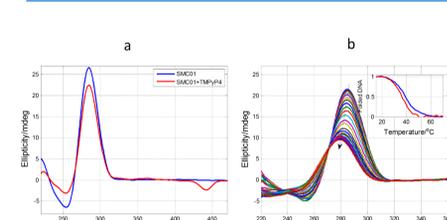
## 5. Acid-base properties



Acid-base titration of SMC01 sequence. (a) Selected experimental CD spectra. (b) Variation with pH of ellipticity values measured at 290nm. The experimental conditions were 2 mM DNA, 150 mM KCl, 25 °C.



## 6. Interaction with TMPyP4

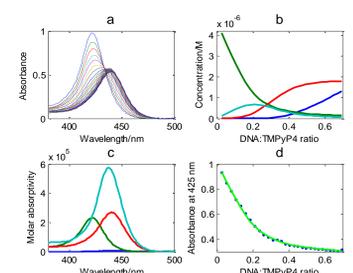


(a) CD spectra of SMC01 sequence in absence and in presence of TMPyP4. (b) CD spectra recorded along the melting of the previous mixture from 15 to 70 °C. Inset shows the fraction of folded DNA in absence and in presence of ligand. The experimental conditions were 2 μM DNA concentration, 4 μM TMPyP4, 150 mM KCl, 20 mM acetate buffer, pH 5.2, 25 °C.

**Determination of the SMC01:TMPyP4 stoichiometry and binding constant.** (a) Experimental molecular absorbance data. (b) Calculated distribution diagram. (c) Calculated pure spectra. (d) Experimental (blue circles) and calculated (green line) absorbance data at 425 nm.

In (b) and (c) free SMC01, free TMPyP4, 1:2 SMC01:TMPyP4 complex, and 1:4 SMC01:TMPyP4 complex. The experiment was carried out at 25 °C, 150 mM KCl, 20 mM acetate buffer, pH 5.2.

Experimental data in (a) were analyzed with a previously described multivariate analysis method that enables the calculation of the binding constants for the proposed model of species, and the corresponding pure spectra [6].



- TMPyP4 binds to the SMC01 i-motif structure with two different stoichiometries (1:2 and 1:4).
- Binding constants are 10<sup>6.9</sup> and 10<sup>6.1</sup> M<sup>-1</sup>.

## References

- Benabou, S.; Aviñó, A.; Eritja, R.; González, C.; Gargallo, R. *RSC Advances* **2014**, *4*, 26956-26980
- Zeraati, M.; Langley, D. B.; Schofield, P.; Moye, A.L.; Rouet, R.; Hughes, W. E.; Bryan, T. M.; Dinger, M. E.; Daniel Christ, C. *Nat. Chem.* **2018**, *10*, 631
- Kendrick, S.; Akiyama, Y.; Hecht, S. M.; Hurley, L. H. *J. Am. Chem. Soc.* **2009**, *131*, 17667-17676
- Zuker, M. *Nucleic Acids Res.* **2003**, *31*, 3406-3415
- Jaumot, J.; Eritja, R.; Tauler, R.; Gargallo, R. *Nucleic Acids Res.* **2006**, *34*, 206-216
- Benabou, S.; Aviñó, Lyonnais, S.; González, C.; Eritja, R.; de Juan, A.; Gargallo, R. *Biochimie* **2017**, *140*, 20-33

## Acknowledgments

Funding from the Spanish government and recognition from the Catalan autonomous government are acknowledged.

**The poster background shows the CH<sup>+</sup>·C base pair, the building block of the i-motif DNA.**