Electronic Supplementary Information

Unraveling the 4n-1 rule for DNA i-motif stability: base pairs vs. loop lengths

Aaron M. Fleming, Kayla M. Stewart, Gabriela M. Eyring, Tyler E. Ball, and Cynthia J. Burrows*

Dept. of Chemistry, University of Utah, Salt Lake City, UT 84112-0850

*To whom correspondence should be addressed

E-mail: <u>burrows@chem.utah.edu</u>

Item	Page
Figure S1. Example titration data for i-motif models with variable loop nucleotide identiti	ies. S1
Figure S2 . Example denaturing T_m curves recorded at 260 and 295 nm for an i-motif.	S2
Figure S3 . The T_m values for model i-motifs with a 5` tail increasing in length.	S 5
Figure S4 . The T_m values for model i-motifs with variable exterior loop lengths.	S6
Figure S5. Data for additional loop-length dependent studies.	S8
Figure S6. Data to reproduce the 4n-1 pattern from dC_{13} - dC_{16} with i-motif models.	S10
Figure S7. Structures of i-motifs illustrating the different grooves.	S11









dC-4141414 with dA loops



8



Figure S2. Example denaturing T_m curves recorded at 260 and 295 nm for an i-motif.

The example T_m curves at 260 and 295 nm were collected for the sequence 5'-CCCC A CCCC A CCCC A CCCC-3'.

Figure S3. The T_m values for model i-motifs with a 5` tail increasing in length.

A T_m values for the even base-paired core i-motif with a 5` tail in the sequence 5`-<u>Y</u>CCCCTCCCCTCCCCC where Y = 0–3.



B T_m values for the odd base-paired core i-motif with a 5` tail in the sequence 5`-CCCTCCCTCCCTCCC \underline{Y} where Y = 0-3.



The T_m values for odd base-paired core i-motifs at pH 7 with a 5` tail were below 20 °C, and therefore, these values could not be reliably measured.

Figure S4. The T_m values for model i-motifs with variable exterior loop lengths.







T_m vs. Asymmetric Exterior Loops for the Odd Core i-motif

Figure S5. Data for additional loop-length dependent studies.





*The data for the poly-dCn strands were previously reported by our laboratory (*J. Am. Chem. Soc.*, **2017**, *139*, 4682–4689).





*The data for the poly-dCn strands were previously reported by our laboratory (*J. Am. Chem. Soc.*, **2017**, *139*, 4682–4689).

Figure S6. Data to reproduce the 4n-1 pattern from dC_{13} - dC_{16} with i-motif models. Even core model sequence 5`-CCC T CCC T_n CCC T CCC-3` where n = 1–4



Odd core model sequence 5`-CCC T CC Tn CCC T CC-3` where n = 1-4



*The data for the poly-dCn strands were previously reported by our laboratory (*J. Am. Chem. Soc.*, **2017**, *139*, 4682–4689).

These data demonstrate that an i-motif core comprised of an odd number of base pairs and a single nucleotide in the exterior loops with a systematic increase in the central loop length reproduce the 4n-1 repeat pattern the best for dC_{13} – dC_{16} .



Figure S7. Structures of i-motifs illustrating the different grooves.

PDB 1ELN- Phan, A.T., Gueron, M., Leroy, J.L. J.Mol.Biol., 2000, 299, 123-144.

The structure was derived from the sequence 5`-d(CCCTAACCCTAACCCTAACCCT) that was solved by NMR. The structure provided above and the one on the following page illustrats that i-motif folds have four grooves, two that are wide and two that are narrow; further, the central loop and the two outer loops span grooves with different depths. Comparison of the grooves between the structure on this page and the one on the following page shows that the sequence impacts the narrow vs. wide grooves. These structures aid in understanding why the outer two loops are most likely they same length while the central loop can be a different length to span different groove depths; however, in the present case it appears as though the best structures have minimized the loop lengths while maximizing the number of base pairs in the core. Without deeper structural analysis into the dC_n strands, we do not know the detailed structural reason behind the biophysical studies presented in the present report.



PDB 1A83 Han, X.; Leroy, J. L.; Gueron, M. J. Mol. Biol. 1998, 278, 949-965.

The structure provided was derived from the sequence 5`-d(5mCCTTTCCTTTACCTTTCC) that was solved by NMR.