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

Journal:	*Organic & Biomolecular Chemistry*
Manuscript ID	OB-ART-05-2018-001198.R1
Article Type:	Paper
Date Submitted by the Author:	02-Jun-2018

Complete List of Authors:	Fleming, Aaron; University of Utah, Chemistry Stewart, Kayla; University of Utah, Department of Chemistry Eyring, Gabriela; University of Utah, Department of Chemistry Ball, Tyler; University of Utah, Department of Chemistry Burrows, Cynthia; University of Utah, Department of Chemistry

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*

Department of Chemistry, 315 S. 1400 East, University of Utah, Salt Lake City, UT 84112-0850, United States

*To whom correspondence should be addressed

E-mail: burrows@chem.utah.edu

†Electronic supplementary information (ESI) available: Additional experimental data for CD titrations, as well as pH_T and T_m values. See DOI:XXX

Abstract

Previously our laboratory identified that poly-2′-deoxycytidine (dC_n) strands of DNA with lengths greater than 12 nucleotides could adopt i-motif folds, while the pH-dependent stabilities follow a 4n-1 repeat pattern with respect to chain length (*J. Am. Chem. Soc.*, **2017**, *139*, 4682–4689). Herein, model i-motif folds in which loop configurations were forced by judiciously mutating dC to non-dC nucleotides allowed a structural model to be proposed to address this phenomenon. The model was developed by systematically studying two i-motifs with either an even or odd number of $d(C•C)^+$ hemiprotonated base pairs in the core. First, a trend in the pH-dependent stability vs. loop nucleotide identity was observed: dC > dT ~ dU >> dA ~ dG. Next, loops comprised of dT nucleotides in the two different core base pair configurations were studied while systematically changing the loop lengths. We found that an i-motif with an even number of base pairs in the core with a single nucleotide in each of the three loops was the

most stable, as well as an i-motif with an odd number of core base pairs having one nucleotide in the two exterior loops and three nucleotides in the central loop. A systematic increase in the central loop from 1–4 nucleotides for an odd number of base pairs in the i-motif core reproduced the 4n-1 repeat pattern observed in the poly-dC_n strands. Additional loop configurations were studied to further support the model. The results are discussed with respect to their biological relevance.

Introduction

Nucleic acid polymers are typically comprised of four canonical nucleotide monomers endowing the strands with the capacity to store chemical information that can be retrieved and replicated with ease. These polymers can be systematically and reproducibly programmed via the primary sequence to adopt a multitude of secondary and tertiary structures. The structural space of nucleic acids include duplexes, hairpins, Holliday junctions, cruciform structures, triple helices, pseudoknots, G-quadruplexes, and i-motifs, as well as higher-order structures found in DNA origami.[1-4] Strands of DNA rich in runs of 2′-deoxycytidine (dC) nucleotides adopt i-motif folds that are held together by $d(C•C)^+$ hemiprotonated base pairs.[3,5] As a result of the requirement for a proton to form the $d(C•C)^+$ base pair, i-motif folds show a strong dependency in their stability on the solution pH, in which they are most stable at low pH and unravel at neutral to high pH. Beyond the primary sequence, the added parameter of pH determining i-motif stability empowers these structures with unique properties that can be modulated by the acidity of the environment.

Strands of DNA capable of i-motif formation are used in pH-dependent biosensors,[6] stimuli-responsive materials based on DNA,[7] and DNA switches.[8] Addition of Ag^+ or Cu^+ ions,[9,10] co-solutes,[11,12] or ligands,[13] as well as strand modifications or alteration of the primary

sequence have been found to tune the stability of i-motifs.[14,15] Recent experiments have found i-motifs can fold in the cellular milieu and in the genomes of human cells.[16,17] Cellular experiments suggest that nature may have harnessed i-motif folds as genomic structural elements to regulate transcription of the *BCL-2*, *HRAS*, and acetyl-CoA carboxylase 1 genes.[13,18,19] Further, folding of i-motif DNA in a gene promoter in response to oxidative modification of the DNA was demonstrated to upregulate transcription.[20] Additional analyses by our laboratory and others have characterized i-motifs in human gene promoters that can fold under near physiological conditions, further supporting the proposal that these secondary structures may have a greater precedence for guiding biological pathways than previously thought.[21,22] In our laboratory, we have contributed by finding that homopolymer runs of dC nucleotides in DNA can adopt i-motif folds when the run is greater than 12 nucleotides.[23] Additionally, our bioinformatic analysis found homopolymer runs of dC in the human genome are favorably biased in gene promoters and 5′- and 3′-untranslated regions.[23] However, the most profound observation we made was that the stability of these poly-dC_n i-motifs followed a 4n-1 repeat frequency for high stability folds with respect to chain length. The greatest stability strands had lengths of 15, 19, 23, or 27 nucleotides (n = 4, 5, 6, or 7) and a strong decrease in stability was observed with chain lengths of 12, 16, 20, or 24 nucleotides (i.e., 4n). The islands of high stability were also found to have the potential of folding under physiological conditions (pH ≥ 7 and 37 °C). Within the group of high stability sequences, dC_{19} was the most stable.[22,23] After our first report on this topic,[23] many questions still remained regarding the structural features that give rise to the 4n-1 repeat pattern observed for the homopolymer i-motifs.

Strands that adopt i-motif folds require at least four runs of dC nucleotides with each run having greater than two nucleotides, and in between each run are one or more nucleotides that occupy loop positions (Fig. 1A).[3,5] The core of the fold houses $d(C•C)^+$ hemiprotonated base pairs (Fig. 1B), in which paired dC nucleotides are in parallel-oriented strands in

intramolecular folds (Fig. 1C); additionally, the dC runs base paired are the first with the third and second with the fourth (Fig. 1A **x** and **x′**). The core has the two paired strands intercalated providing the "i" in the name of the fold. This configuration of strands results in the core having either an even (Fig. 1, **x** = **x′**) or odd number (Fig. 1; **x** $\neq$ **x′**) of base pairs depending on the sequence. An even number of base pairs in the core leads to two possible configurations depending on which base pair is on top of the strand when looking from the 5′ side (Fig. 1C, 5′ *E* or 3′ *E*).[24,25] There exist three loops in the folded structure with the first and third loops existing on the same face of the fold and the second on the opposite (Fig. 1C; **a**, **b**, and **c**). Previous bodies of work have interrogated core base pair configurations with the same loops or maintained the same core base pairs while varying the loop lengths;[12,26-32] however, these studies could not be leveraged to fully address why dC_{19} was the most stable i-motif found in the homopolymer strands studied, and why there was a 4n-1 repeat frequency in the high stability folds. Therefore, we systematically varied the core base pair count (Fig. 1; **x** and **x′**) and loop lengths (Fig. 1; **a**, **b**, and **c**) with a focus centered on a length of 19 nucleotides to address why the repeat pattern occurs from a sequence and structural perspective. Herein, we report unusual observations regarding i-motif folds that will be of interest to the nucleic acid research community.

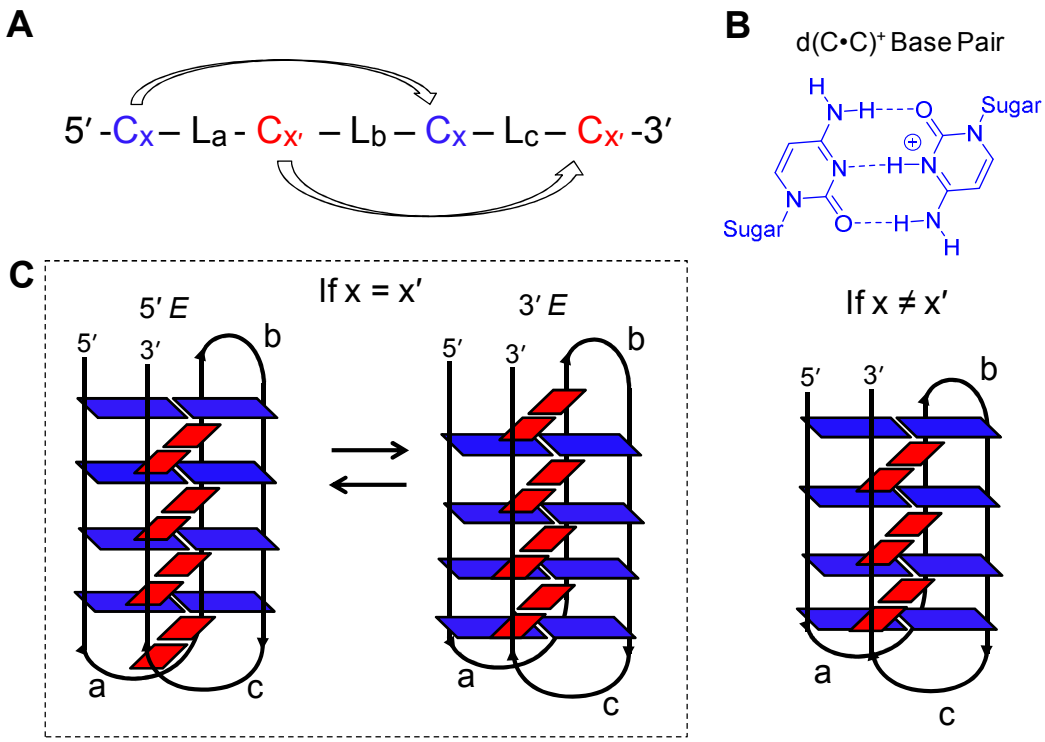


Figure 1. (A) General sequence for an i-motif forming DNA sequence to identify the variables **x**, **x′**, **a**, **b**, and **c**. (B) Structure of a hemiprotonated d(C•C)$^+$ base pair. (C) Scheme to depict the structure of an i-motif fold with either an even (**x** = **x′** = 4) or odd (**x** = 4, **x′** = 3) number of core base pairs. The models represent those that were studied to understand folding of the homopolymer i-motif dC_{19}.

Results and Discussion

Evaluation of i-motif loop nucleotide identity vs. the pH_T and T_m values. When poly-dC_n strands of lengths n = 12–30 nucleotides were inspected for i-motif formation, a 4n-1 repeat frequency in the high stability folds was observed in the pH_T values between the folded and unfolded states (Fig. 2A); a similar pattern in the thermal stability profile was observed in the denaturing T_m values measured at pH 6.0 and 7.0.[23] As an example, one cycle of the pattern begins with low pH_T and T_m values for dC_{16} (pH_T = 6.7; T_m at pH 7.0 < 30 °C). Upon addition of the next nucleotide, the dC_{17} strand showed an increase in both pH_T and T_m at pH 7.0 (6.9 and 40 °C); for dC_{18} the pH_T slightly increased to 7.1 and the T_m at pH 7.0 to 44 °C; dC_{19} had the

highest pH_T = 7.4 and a T_m at pH 7.0 = 41 °C; and finally the cycle was repeated going to dC_{20} that had decreased stability with a pH_T = 6.5 and a T_m at pH 7.0 < 30 °C. As the next step in understanding this cycle, our objective was to synthesize and analyze i-motif forming sequences with loop regions locked in by judiciously converting dC nucleotides to non-dC nucleotides in order to recreate the 4n-1 repeat pattern previously observed.[23] These studies provided a possible structural understanding of the i-motif base pair and loop patterns that give rise to the 4n-1 pattern observed as shown in Figure 2B. The data supporting the proposed model are described next.

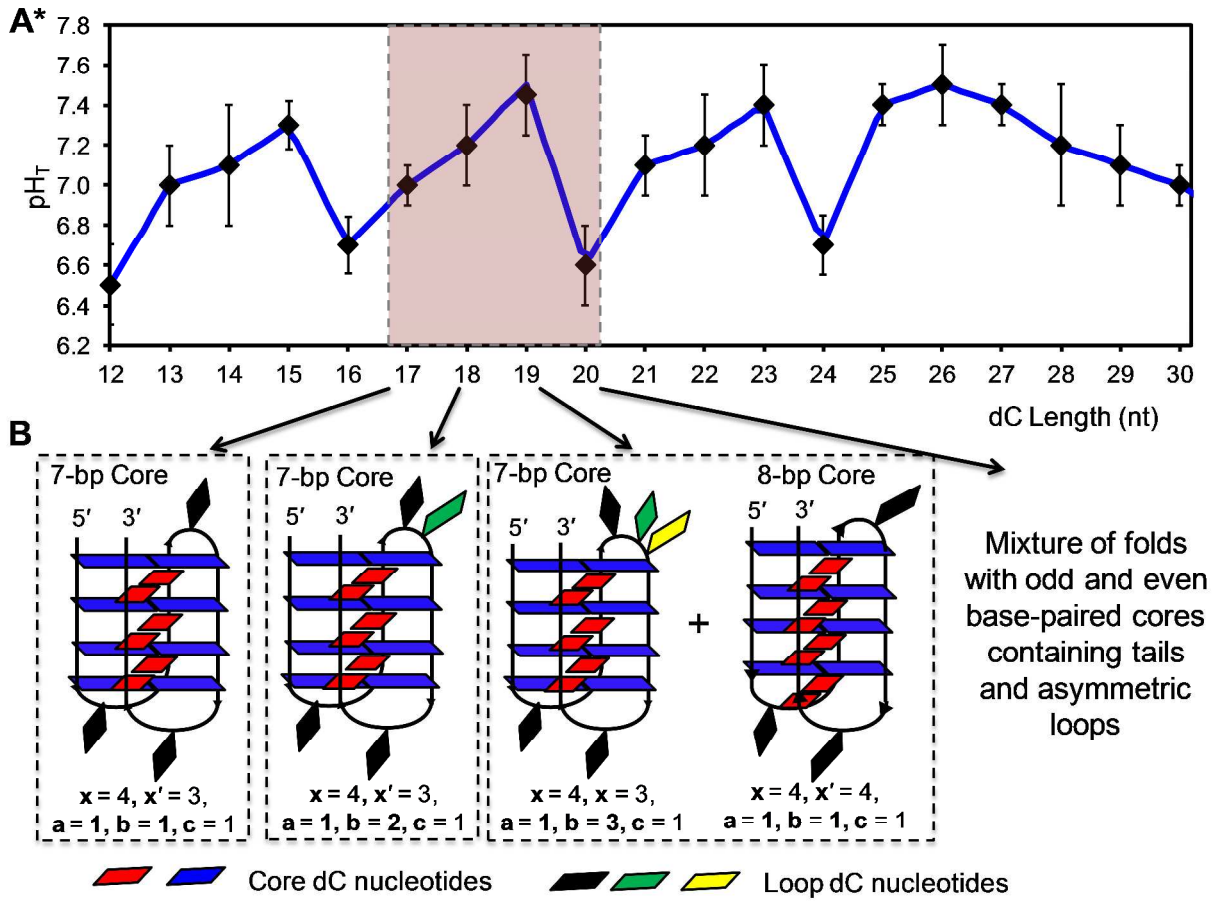


Figure 2. Plot of poly-dC_n i-motif stability versus polymer length (n) and a proposed structural model to support one repeat of the pattern observed. (A) Plot of pH_T values vs. poly-dC_n chain length. (B) Proposed model supporting the structural evolution of poly-dC_n i-motifs yielding a 4n-1 repeat frequency in the stability as the chain length increases. *We previously reported this plot.[23]

Using our previous work as a guide to synthesize the most stable model of dC_{19},[23] four nucleotides of dC were placed in each of the four runs required for i-motif folding (**x** = **x′** = 4) while a single nucleotide was selected for each of the three loops (Fig. 3A). We altered the nucleotide identity of the loops between dA, dG, dT, or dU and compared the results to the homopolymer dC_{19}. We have chosen the nomenclature dC-$4_1 4_1 4_1 4$ for this model system, in which the regular-sized numbers represent the quantity of dC nucleotides in each of the four runs from the 5′ to 3′ side, and the subscripted numbers are the counts of nucleotides in each of the loops (Fig. 3A). The pH_T values were determined via a CD spectroscopy titration study that monitored changes in the folded state from pH 5.00 to 8.00 in 0.25 pH increments at 20 °C (Fig. 3B and ESI Fig. 1 for more example spectra). Our previous report on this topic suggests that the folds inspected under the present conditions were intramolecular and not intermolecular in nature.[23] The folded i-motif at low pH has a λ_{max} at ~286 nm and a λ_{min} at ~260 nm that both decay as the pH is increased to achieve a spectrum for the single-stranded state with a λ_{max} at ~275 nm (Fig. 3B). These spectra for the folded and unfolded states are consistent with the literature.[33] By following the pH-dependent decrease in the 286 nm signal ($[\Theta]_{286nm}$), a titration curve was observed and mathematically fit to determine the midpoint in the transition (i.e., pH_T; Fig. 3C). The studies inspecting how pH_T changed as a function of the nucleotide identity in the loops found a trend in the stability: dC > dT ~ dU >> dA ~ dG (Fig. 3D). These studies identified the dC homopolymer was the most stable i-motif fold (pH_T = 7.4), followed by i-motifs with the pyrimidines dT (pH_T = 7.2) or dU (pH_T = 7.0) in the loops, and the least stable were those with the purines dA (pH_T = 6.2) or dG (pH_T = 6.3) in the loops. Previous studies have alluded to a difference in stability for i-motif folds with changing sequence content suggesting that purine-containing loops destabilize the folds;[24,29,32] however, systematic inspection of this feature in all three loops was not previously conducted. Additionally, we note that the dT- and dU-containing

loops were nearly equivalent in stability suggesting that the added hydrophobic methyl group on dT did not dramatically alter the stability of the folds.

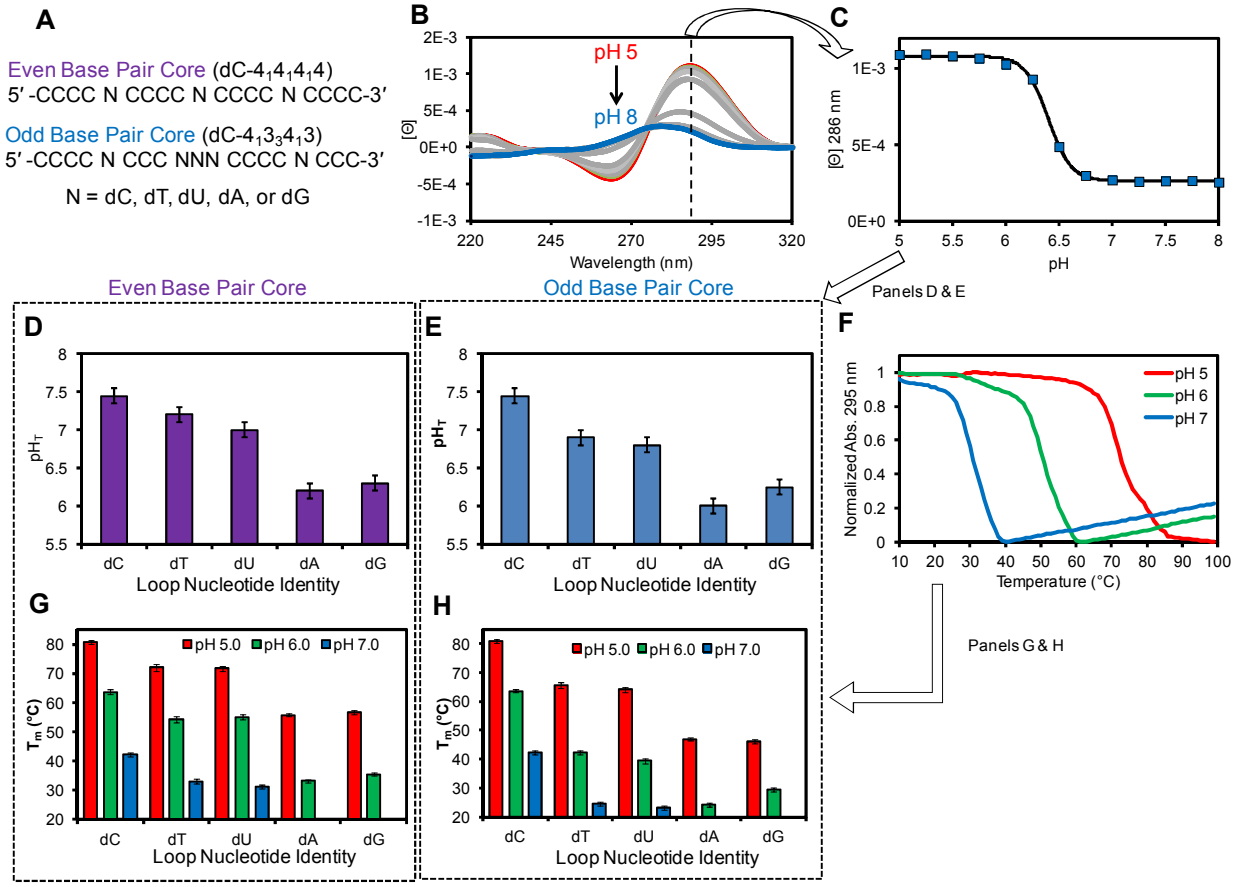


Figure 3. Analysis of model i-motif strands with variations in the loop nucleotide identity to determine the best model for i-motifs of poly-dC$_n$. (A) Sequences studied, (B) example i-motif titration study for dC-4$_1$4$_1$4$_1$4 with dG nucleotides in the loops, (C) CD titration curve at [Θ] = 286 nm for the example study, (D) pH$_T$ values for model i-motifs with an even number of base pairs in the core with each loop nucleotide variant studied, and (E) pH$_T$ values for model i-motifs with an odd number of base pairs in the core with each loop nucleotide studied. (F) Examples of

denaturing T_m curves for dC-$4_1 4_1 4_1 4$ with dT nucleotides in the loop, and T_m values at pH 5.0, 6.0, and 7.0 for the model studies with an (E) even or (F) odd number of base pairs in the i-motif core. The data for the dC_{19} (left most bars of each graph) i-motif in panels D, E, G, and H were previously reported by our laboratory.[23]

Next, the pH_T values were measured for an i-motif sequence with an odd number of base pairs in the core (**x** + **x′** = 7) having a length of 19 nucleotides while varying the loop nucleotide identity. In these studies, the first and third dC runs had four nucleotides (**x** = 4) and the second and fourth runs had three nucleotides (**x′** = 3) while the three loops had lengths of **a** = 1, **b** = 3, and **c** = 1 nucleotides (i.e., dC-$4_1 3_3 4_1 3$; Fig. 3A). These values were selected to maintain the symmetry of the loops by ensuring the first and third loops that are on the same face of the i-motif had the same length; symmetry in the two loops was previously identified in the literature to be a stable configuration.[26] When the loop nucleotide identity was systematically varied, the CD titration study found a similar trend in the pH_T values as determined above in which dC > dT ~ dU >> dA ~ dG (Fig. 3E). Again, i-motif folds containing an odd number of base pairs in the core with pyrimidines in the loops had greater pH_T values than i-motifs with purines in the loops.

The thermal stability for each of the strands was then measured at pH 5.0, 6.0, and 7.0 to determine the pH dependency by following the UV absorbance at 260 and 295 nm as a function of temperature between 5 and 100 °C while making measurements at 1 °C increments (Fig. 3F). The results between the two wavelengths analyzed were very similar (ESI Fig. 2); therefore, only data derived from the 295 nm reading are reported. First, as expected for the dC-$4_1 4_1 4_1 4$ and dC-$4_1 3_3 4_1 3$ systems with loops comprised of dC, dT, dU, dA, or dG nucleotides, the thermal stability was highest at pH 5.0, and the T_m values decreased as the pH was increased. The trends in loop nucleotide identity for the T_m values at the three pH conditions studied were similar to those identified in the pH_T values (Fig. 3D and 3E). Specifically, sequences with pyrimidines in the loops were more stable, and those with purines in the loops

were least stable following the trend dC > dT ~ dU >> dA ~ dG (Fig. 3G and 3H). On the basis of the pH_T and T_m values for these sequences, we chose dT nucleotides judiciously to lock in loop lengths for the next set of model studies because they had the least overall impact on the pH_T and T_m values compared to the dC-homopolymer strands. The increased stability with dT loops compared to the other nucleotides will also provide a greater range to monitor the impact of systematic inspection i-motif loop lengths and variation of the core base pair count on the stability of the folds.

Replacement of dC nucleotides with dT to reinforce loop configurations can also impact loop-to-loop interactions. Specifically, loop dC nucleotides are sites of protonation and deprotonation during the pH titration experiments that can have a secondary impact on the stability that is not recreated with non-titratable dT nucleotides. Further, dT nucleotides in the loops can form dT:dT wobble base pairs[34] that dC base pairs cannot and this may impact the structure differently. The ability to utilize available phosphoramidites for the synthesis of i-motif models in which the loop configurations are known while maintaining the unique acid-base properties of dC nucleotides is not available; thus, we point out that the i-motif models with dT nucleotides studied next will not perfectly reproduce the poly-dC_n results. This claim is evident in the pH_T and T_m values reported in Figure 3 and the data described next.

The loop nucleotide identity trend in i-motif stability (dC > dT ~ dU >> dA ~ dG; Fig. 3) may result from a few competing factors. The most stable poly-dC_n strands possess loops that can be titrated and form stabilizing loop-to-loop base pairs that cannot occur with the other nucleotides, as stated above. On the other hand (see above), the dT or dU nucleotides can participate in loop-to-loop interactions that can stabilize the structure. When either purine was placed in the loops sharp decreases in the stabilities were observed (Fig. 3). The data suggest that purines do not participate in stabilizing loop-to-loop interactions, and this feature alone may explain the observation. Additionally, purines are more hydrophobic than pyrimidines[35] resulting

in the decreased stability observed. Likely, a combination of loop-to-loop interactions and solvation effects results in the trend in stability vs. loop nucleotide identity observed in the present studies.

Evaluation of i-motif loop lengths vs. pH_T and T_m values. A systematic study of the impact that loop length has on i-motif stability with a constant number of base pairs in the core was conducted to determine whether a 4n-1 frequency in the stability emerged. First, a core of eight base pairs (**x** = **x′** = 4) was studied while incrementally increasing the central loop length **b** from 1–4 dT nucleotides and maintaining single dT nucleotides in the external loops **a** and **c** (i.e., dC-$4_1 4_n 4_1 4$ where n = 1–4; Fig. 4A). Inspection of the pH_T values measured by the CD titration study found that as the central loop length increased, the pH_T values decreased from 7.1 to 6.6 (Fig. 4B). These results show a clear trend for i-motifs with an even number of base pairs in the core that as the central loop length increased in length, the pH_T values decreased. A similar trend was observed in the denaturing T_m values for this series of sequences at pH 5.0, 6.0, and 7.0 (Fig. 4C). Next, a comparison between these model studies with dT loops and the homopolymers of dC_n with the same overall length did not find a similar trend between the data sets (Fig. 4B). This model study does not address why a 4n-1 repeat pattern was observed in poly-dC_n as the chain length increased. Though, we note that the 19-mer model with an even number of base pairs in the core (**x** = **x′** = 4) and single nucleotide loops (**a** = **b** = **c** = 1) was the most stable, similar to the homopolymer dC_{19}.

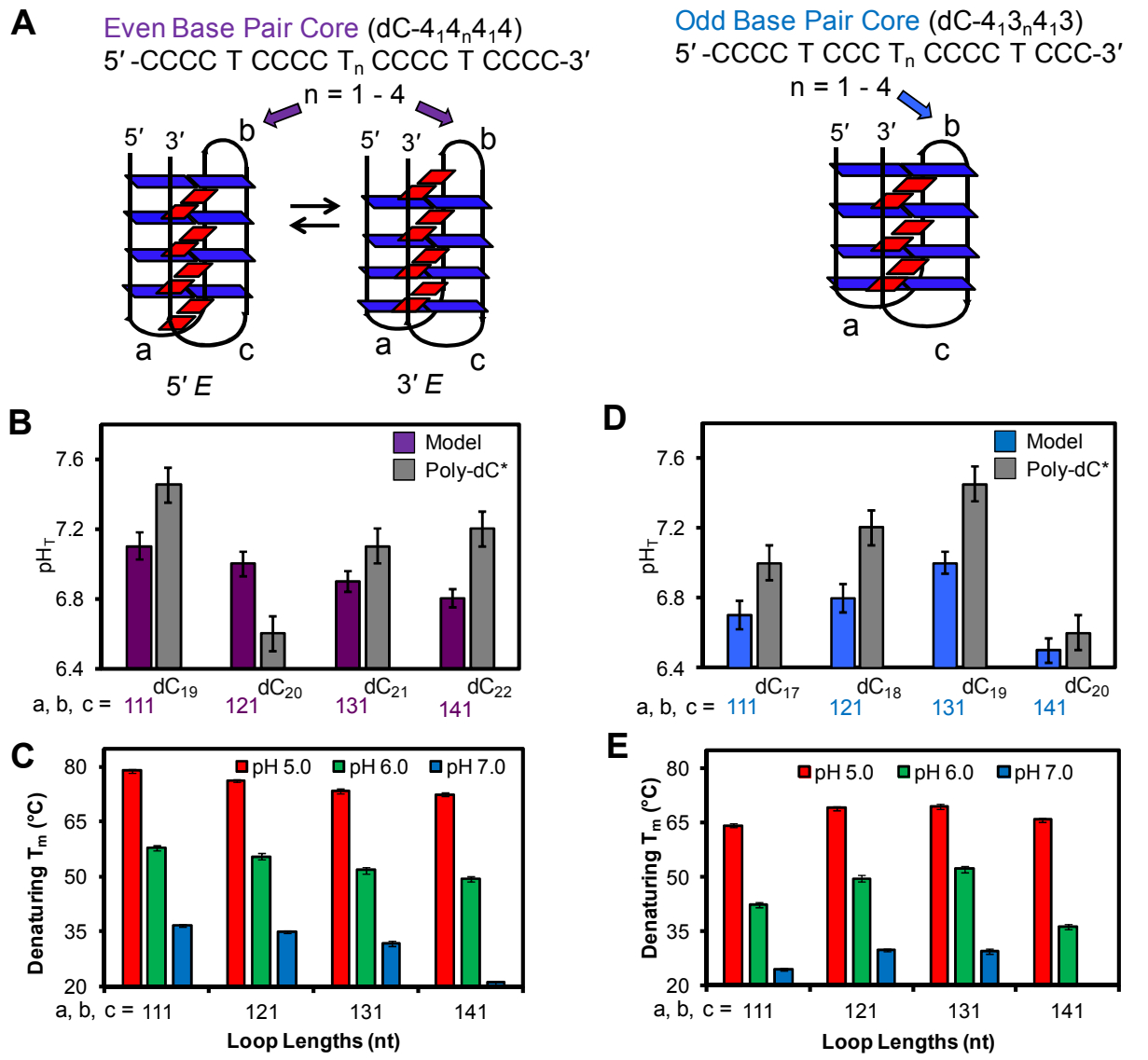


Figure 4. Comparison of pH_T and T_m values from the model i-motif sequences varied in the length of the central loop versus poly-dC_n. (A) Sequences studied and their cartoon models. (B) pH_T and (C) T_m values for comparisons with models containing an even number of $d(C•C)^+$ base pairs in the core. (D) pH_T and (E) T_m values for comparisons with models containing an odd number of $d(C•C)^+$ base pairs in the core. *These data were previously reported by our laboratory.[23]

The next model studies were conducted with seven base pairs in the core (**x** = 4 and **x′** = 3) and the loops were locked with dT nucleotides by first maintaining one nucleotide in the first (**a**) and third (**c**) loops while systematically increasing the central loop length from 1–4 dT

nucleotides (i.e., dC-$4_1 3_n 4_1 3$ where n = 1–4; Fig. 4A). The results of these models found as the central loop length increased from one to three nucleotides, the pH_T value increased from 6.7 to 6.9 to reach a maximum with three dT nucleotides in the central loop (Fig. 4D). Studies with four nucleotides in the central loop led to a significant decrease in the pH_T value to 6.5 relative to the maximally stable model sequence (n = 3; Fig. 4D). More interestingly, when the models with an odd number of core base pairs and a systematic increase in the central loop length were compared to the poly-dC_n,[23] a similar trend between the two data sets emerged (Fig. 4D). Specifically, in the model with the central loop length of one nucleotide (length = 17 nucleotides) a lower pH_T was observed as well as for poly-dC_{17}. Addition of a second nucleotide to the central loop increased the pH_T for the model that tracked with the poly-dC_{18} strand. Next, the maximum stability was found for the model with three nucleotides in the central loop and poly-dC_{19}. Finally, addition of the fourth nucleotide to the model sequence resulted in a drastic decrease in the stability that was also found for poly-dC_{20}. Additionally, the T_m values for the models at pH 6.0 and 7.0 also revealed a pattern in stability that followed a trend with the poly-dC_n strands (Fig. 4E). In summary, the model i-motif with 19 nucleotides containing an odd base pair count in the core (7 base pairs; **x** = 4 and **x′** = 3) and having loop lengths of **a** = 1, **b** = 3, and **c** = 1 was the most stable, which aligned with the stability of poly-dC_{19}.

Before conclusions were drawn regarding how the i-motif structure evolves as each nucleotide is added to the poly-dC_n strand to follow a 4n-1 repeat pattern,[23] additional model studies were conducted. First, a set of experiments was conducted with a tail increasing in length from 0–3 nucleotides long on the 5′ end of the i-motif for both core compositions. This study addresses whether the 4n-1 repeat pattern results from a growing tail off of the core (Fig. 5, **y** = tail length). As the tail length was increased in length with dT nucleotides, the pH_T (Fig. 5A and 5B) and T_m (ESI Fig. 3) values for the i-motifs decreased in both even and odd base-paired cores. The pH_T values for the tail length dependent models were then compared to poly-

dC_n strands of the same length, and a trend that tracked between the two data sets was not observed (Fig. 5A and 5B). On the basis of these tail length-dependent model studies, we conclude that as the poly-dC_n chain grows, the added nucleotides must not occupy tail positions because these data do not reproduce the 4n-1 repeat pattern. The observation that that the stability of DNA secondary structures decrease with increasing tail length has been documented for duplex[36] and G-quadruplex forming sequences.[37]

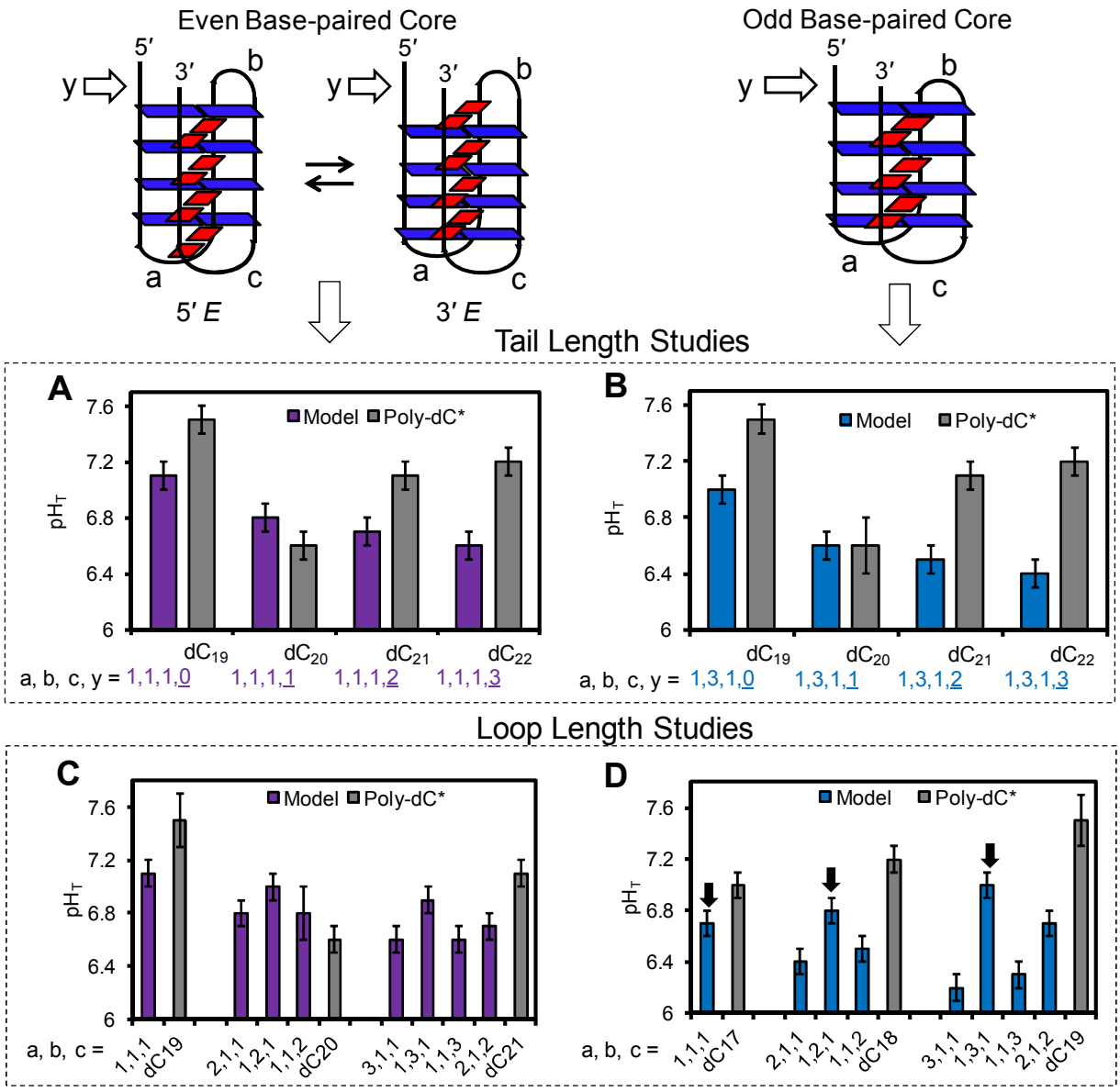


Figure 5. Evaluation of i-motif models with tails or changes in the exterior loop lengths. Models with the addition of a grow tail to either an even (A) or odd (B) base-paired i-motif core were studied and compared to poly-dC_n strands of the same length. Models with expansion of the exterior loops for either an even (C) or odd (D) base-paired i-motif core were studied and compared to poly-dC_n strands of the same length. In panels C and D, the data were also compared to models with expanded central loops that were initially shown in Figure 4; these comparisons further allow visualization of the trends in which the models pointed out with arrows fit the poly-dC_n pattern in stability. *These data were previously reported by our laboratory.[23]

In the final set of model experiments, we examined both even and odd base-paired i-motif cores, systematically filled the exterior loops with one dT nucleotide at a time, and measured the pH_T values and pH-dependent T_m values. The model data were compared to poly-dC_n data of the same overall length (Fig. 5). The comparison initiated with **a** = 1, **b** = 1, and **c** = 1 for each core type (these data were initially shown in Fig. 4). In the even base-paired core model, increasing any of the three loop lengths from one (compared to dC_{19}) to two (compared to dC_{20}) nucleotides decreased the pH_T and T_m values (Fig. 5C and ESI Fig. 4). A noteworthy observation was that having uneven loop lengths of **a** $\neq$ **c** (i.e., asymmetric model) was more destabilizing than the case in which **a** = **c** = 1 and **b** = 2 (Fig. 5C). In the odd base-paired core model that was initially compared to dC_{17} and then compared to dC_{18}, when a nucleotide was added, the addition of a dT nucleotide to either loops **a** or **c** resulted in a decrease in stability when compared to all single nucleotide loops (**a** = **b** = **c** = 1; Fig. 5D). In contrast, as stated above (Fig. 4D and 4E), adding the nucleotide to the central loop increased the pH_T and T_m values (Fig. 5D). Also, in the odd base-paired core i-motif model, having asymmetry in the **a** and **c** loops was more destabilizing when compared to the case in which the second nucleotide was added to the central loop (Fig. 5D and ESI Fig. 4). These comparisons add further confirmation to the best model for the structural evolution of the poly-dC_n strand to have an odd base-paired core and fill the nucleotides into the central loop (Fig. 2B and 4D).

The loops were then expanded by one more dT nucleotide. The even base-paired core model filled with the next dT nucleotide was compared to dC_{21}. Increasing the asymmetry in loops **a** and **c** to be a combination of one and three nucleotides long was even more destabilizing than observed with the one- and two-nucleotide combination (Fig. 5C). Additionally, adding the dT nucleotide to the central loop **b** was also destabilizing but less so than addition to either exterior loop. Another combination of loop lengths that achieves 21 nucleotides long is **a** = 2, **b** = 1, and **c** = 2, which maintains symmetry in the **a** and **c** loops. This

model was also destabilizing, and more importantly, none of these models followed the trend set by the poly-dC_n strands (Fig. 5C). Lastly, models with an odd base-paired core i-motif were studied with an additional dT nucleotide added (Fig. 5D). This model was significantly destabilized by having asymmetric **a** and **c** loops with either combination of one and three nucleotides. In contrast, having the third dT nucleotide in the central loop with single nucleotide exterior loops was very stabilizing, as previously stated (Fig. 4D and 5D). Next, we found that the **a** = 2, **b** = 1, and **c** = 2 model was additionally destabilizing. The strands that followed the trend set by poly-dC_n are shown with arrows in Figure 5D that were also first discussed in Figure 4D. Some additional models were explored, but they all failed to reproduce the trend established by the poly-dC_n strands (ESI Fig. 5). These data provide even more support for the best model to address the evolution of the i-motif structures in the poly-dC_n strands being one which has an odd base-paired core, and the nucleotides incrementally fill the central loop while maintaining single nucleotide exterior loops.

Proposed Model for the 4n-1 Repeat Pattern Observed for Poly-dC_n i-motifs. Examination of these model studies led us to the following conclusion regarding how the i-motif structure evolves in the poly-dC_n strands to give rise to a 4n-1 repeat frequency in the high stability folds (Fig. 2B). Starting with poly-dC_{17}, the structure is best explained by an i-motif with seven base pairs in the core (**x** = 4 and **x′** = 3) and single nucleotides in each of the three loops (**a** = **b** = **c** = 1). Upon addition of the 18^{th} nucleotide, it is filled into the central loop (**b** = 2) of an i-motif with a seven base-paired core (**x** = 4 and **x′** = 3) and symmetry is maintained with a single nucleotide in the first and third loops (**a** = **c** = 1). Next, the most stable homopolymer dC_{19} can be explained by either an i-motif with seven base pairs in the core (**x** = 4 and **x′** = 3) and loop lengths of **a** = 1, **b** = 3, and **c** =1 or an eight base-paired core (**x** = **x′** = 4) with one nucleotide in each of the three loops (**a** = **b** = **c** = 1). Both structures have high stability, and having two nearly equivalent structures for this sequence may explain the overall higher stability for dC_{19}

(Fig. 3). Lastly, for dC_{20}, a seven base-paired core (**x** = 4 and **x′** = 3) with loops of **a** = 1, **b** = 4, and **c** =1 or a structure with a single nucleotide tail (**y** = 1) and loops of **a** = 1, **b** = 3, and **c** = 1 results in a structure that has decreased stability, fitting with the trend observed with poly-dC_n strands. Both decreased stability models fit the trend. Additional structural options exist for an i-motif of 20 nucleotides with possible asymmetric exterior loops (Fig. 2B and 5C, ESI Fig. 5); however, determination of the actual structure is not possible with the present data, and therefore, we selected the structures that best fit the trend (Fig. 2B). Finally to confirm that the core base pair count and loop lengths hold for another 4n-1 cycle, we conducted similar studies with model i-motifs focused on dC_{13}-dC_{16} (ESI Fig. 6). These final experiments provided a similar behavior with two fewer base pairs in the core, and we conclude that the proposed model holds at other cycles of the 4n-1 repeat pattern (Fig. 2B).

Implications. The data regarding i-motif stability as a function of loop nucleotide identity lead to the conclusion that these folds show greater resilience to unfolding as the pH is increased with pyrimidines in the loops compared to purines (Fig. 3). This concept was hinted at in previous reports,[12,27,29] but was not systematically explored throughout all three loops as reported here. More interestingly, the poly-dC_n i-motifs were most resilient to increased pH. The observation that poly-dC_n i-motifs are more stable than i-motifs with non-dC loops suggests that homopolymer dC_n microsatellites (n > 12; Fig. 2A) in the genome have the greatest potential for folding in vivo as was recently demonstrated in the genomes of human cells.[17] These insights can aid future studies inspecting genomes for potential i-motif folding sequences with possible biological function; although, this assumption neglects any stabilizing interactions that may occur between heterogeneous loop sequences that, for instance, have been documented for the *BCL-2* and *n-MYC* promoter i-motifs to increase the stability of this sequence.[38,39] On an additional note, modulation of the pH dependency for i-motif folds (i.e., pH_T value) via the loop identity may have utility for biosensors developed around these sequences.

These data provide a direct comparison of model i-motif forming sequences with either an even or odd number of base pairs in the core while systematically changing the loop lengths. Interestingly, the trends observed between the two base-paired cores show differences in the preferred loop lengths leading to the most stable folds. When the i-motif had an even number of d(C•C)$^+$ base pairs (**x** = **x′** = 4), the most stable fold observed had a single nucleotide in each of the three loops (**a** = **b** = **c** = 1); in contrast, when there was an odd number of d(C•C)$^+$ base pairs in the core (**x** = 4 and **x′** = 3), the most stable fold had loop lengths of **a** = 1, **b** = 3, and **c** = 1 (Fig. 3). The reason for this difference is not immediately clear; although, the difference between the two different base-paired core i-motif models is the preferred length of the central loop.

In an i-motif fold there exist four grooves (ESI Fig. 7),[40,41] in which the two grooves opposite one another have nearly identical features; additionally, loops **a** and **c** span opposite yet similar grooves while loop **b** spans the other groove that might have different geometric features (Fig. 1C). The present data suggest that when there are an even number of d(C•C)$^+$ base pairs in the core, the groove widths are all similar resulting in the preferred fold to have three loops of the same length. To determine whether the preferred loop lengths were a single nucleotide or another combination of the same length, we conducted additional model studies. A study of the sequence dC-4$_n$4$_n$4$_n$4 (n = 1–3) identified single nucleotide loops as the most stable, and as the loops all increased in length together, the i-motif was destabilized (ESI Fig. 5). Thus, this final evaluation further confirms the conclusion that i-motifs with an even number of base pairs in the core are most stable with a single nucleotide in each of the three loops; this model is a valid representation of the most stable homopolymer, dC$_{19}$.

On the other hand, i-motif folds with an odd number of d(C•C)$^+$ base pairs in the core favor loop lengths of **a** = 1, **b** = 3, and **c** = 1 and likely have grooves with different widths. The grooves that loops **a** and **c** span are likely more narrow than the groove that loop **b** spans

because the preferred **b** loop length was longer. This led us to probe the impact on the pH_T value as the exterior loops were systematically expanded while maintaining the preferred central loop length of **b** = 3. This study found that as the exterior loop lengths were increased the pH_T values decreased, again supporting the favorability of the **a** = 1, **b** = 3, and **c** = 1 loop lengths with an odd number of $d(C•C)^+$ base pairs in the i-motif core (ESI Fig. 6). These data further validate the i-motif of 19 nucleotides long with **x** = 4, **x′** = 3, **a** = 1, **b** = 3, and **c** = 1 as the other best model for the homopolymer dC_{19}.

Conclusions

The present report developed a structural model to understand why poly-dC_n strands adopt i-motif folds with the stability of the folds following a 4n-1 repeat pattern in the chain length (Fig. 2A).[23] Our attention was focused on understanding the pattern around the most stable fold dC_{19}. We first determined the best nucleotide to replace dC in a loop of an i-motif with either an even or odd number of $d(C•C)^+$ hemiprotonated base pairs in the core was dT because this nucleotide impacted the stability the least (Fig. 3); in contrast, loops comprised of purine nucleotides caused a significant decrease in the i-motif stability (Fig. 3). Next, we systematically evaluated loop lengths with dT nucleotides in i-motifs with either an even or odd number of core base pairs to identify structures that tracked the trend in stability established by the poly-dC_n strands. This allowed us to propose a structural model to understand why poly-dC_n strands have a 4n-1 repeat pattern in their stability. Further, these studies allowed us to identify that the folded dC_{19} strand is likely most stable because an i-motif forms with an odd base pair core and loops of **a** = 1, **b** = 3, and **c** = 1, or alternatively with an even base pair core and all three loops as a single nucleotide (**a** = **b** = **c** = 1), both of which have similar stabilities (Fig. 2B and 3). These equivalent folds likely provide the additional stability observed for dC_{19}. Further, following i-motif stability while systematically varying the core base pair count and loop lengths allowed us to identify that folds with an odd number of base pairs in the core versus those with

an even number of base pairs favor different loop configurations to yield the most stable fold. Additionally, in both models, having asymmetry in the exterior loops (**a** ≠ **c**) was always destabilizing, as well as having a tail appended to the core of the fold (Fig. 5 and ESI Fig. 5). The i-motif community can leverage these data to better understand how i-motifs' pH_T values and pH-dependent thermal stabilities will change as a function of loop nucleotide identity and loop length. Lastly, as a result of the recent observation that i-motifs can fold in human genomic DNA inside cells,[17] the search for folded sequences is pressing, and these data can guide the discovery of new possible i-motif sequences with high pH-dependent stability.

Experimental Section

Oligodeoxynucleotide preparation. The i-motif forming strands were synthesized via standard protocols by the DNA/Peptide core facility at the University of Utah. The synthesized strands were purified using a semi-preparative, anion-exchange HPLC running the following mobile phase system: A = 1:9 MeCN:ddH_2O, B = 20 mM Tris buffer at pH 8, and 1 M NaCl in 1:9 MeCN:ddH_2O. The separation method was initiated at 1% B followed by a linear increase to 100% B over 30 min with a flow rate of 3 mL/min while monitoring the oligodeoxynucleotide elution by their absorbance at 260 nm. The purified oligodeoxynucleotides were dialyzed against ddH_2O for 36 h while changing the water three times. Next, the dialyzed oligodeoxynucleotides were lyophilized to dryness and resuspended in ddH_2O followed by determination of the concentration on the basis of the 260 nm absorbance reading. The primary sequences were used to estimate the extinction coefficients for calculation of the concentrations.

Circular Dichroism (CD) spectroscopy studies. The pH-dependent CD studies were conducted from pH 5.00–8.00 by taking measurements at intervals of 0.25 pH units. The variable pH values were established in 20 mM Britton-Robinson buffer with 140 mM KCl added.

The oligodeoxynucleotides with concentrations of 5 μM were incubated at each pH for 30 min prior to CD analysis. The CD spectra were recorded from 220–320 nm with a scan interval of 0.1 nm. The differential absorbance units measured were then converted to molar ellipticity values ([Θ]) and then plotted. Next, the pH-dependent spectra were stacked for presentation, followed by making a secondary plot of $[\Theta]_{286\ nm}$ vs. pH yielding a titration curve that was fit with the Henderson-Hasselbalch equation to determine the transition midpoint (pH_T). The titrations were conducted in triplicate to obtain experimental error bars.

Thermal melting (T_m) studies. The i-motif strands were prepared at 3 μM concentration in the same buffer used in the CD studies at pH 5.0, 6.0, or 7.0. After incubating the strands for 30 min at 20 °C, they were placed in T_m analysis cuvettes. The T_m experiments were conducted by thermally equilibrating the strands at 5 °C for 10 min followed by heating the samples to 100 °C at a rate of 1 °C/min while thermally equilibrating at each step for 60 s prior to measuring the absorbance. Absorbance readings were taken at 260 and 295 nm. The T_m values were determined by plotting the 260 or 295 nm reading as a function of temperature to obtain the thermal denaturing profile. The transition point was determined using a two-point average method. All measurements were made in triplicate to obtain the experimental error bars.

Acknowledgments

These studies were supported by a National Science Foundation grant (CHE-1507813). The oligomers were synthesized by the DNA/Peptide core facility at the University of Utah that is supported in part by a NCI Cancer Center Support Grant (P30 CA042014).

Notes

The authors declare no completing financial interest regarding this work.

References

1. J. Choi and T. Majima, Chem. Soc. Rev., 2011, 40, 5893-5909.

2. A. N. Lane, J. B. Chaires, R. D. Gray and J. O. Trent, Nucleic Acids Res., 2008, 36, 5482-5515.

3. H. A. Day, P. Pavlou and Z. A. Waller, Bioorg. Med. Chem., 2014, 22, 4407-4418.

4. A. Rajendran, M. Endo and H. Sugiyama, Angew. Chem. Int. Ed. Engl., 2012, 51, 874-890.

5. K. Gehring, J. L. Leroy and M. Gueron, Nature, 1993, 363, 561-565.

6. A. Dembska, Anal. Chim. Acta, 2016, 930, 1-12.

7. C. H. Lu and I. Willner, Angew. Chem. Int. Ed. Engl., 2015, 54, 12212-12235.

8. F. Wang, X. Liu and I. Willner, Angew. Chem. Int. Ed. Engl., 2015, 54, 1098-1129.

9. H. A. Day, C. Huguin and Z. A. Waller, Chem. Commun. (Camb), 2013, 49, 7696-7698.

10. M. A. Abdelhamid, L. Fabian, C. J. MacDonald, M. R. Cheesman, A. J. Gates and Z. A. Waller, Nucleic Acids Res., 2018, doi: 10.1093/nar/gky1390.

11. S. Takahashi, J. A. Brazier and N. Sugimoto, Proc. Natl. Acad. Sci. U. S. A., 2017, 114, 9605-9610.

12. T. Fujii and N. Sugimoto, Phys. Chem. Chem. Phys., 2015, 17, 16719-16722.

13. S. Kendrick, H. J. Kang, M. P. Alam, M. M. Madathil, P. Agrawal, V. Gokhale, D. Yang, S. M. Hecht and L. H. Hurley, J. Am. Chem. Soc., 2014, 136, 4161-4171.

14. H. Abou Assi, Y. C. Lin, I. Serrano, C. Gonzalez and M. J. Damha, Chemistry, 2018, 24, 471-477.

15. L. Lannes, S. Halder, Y. Krishnan and H. Schwalbe, ChemBioChem, 2015, 16, 1647-1656.

16. S. Dzatko, M. Krafcikova, R. Hansel-Hertsch, T. Fessl, R. Fiala, T. Loja, D. Krafcik, J. L. Mergny, S. Foldynova-Trantirkova and L. Trantirek, Angew. Chem. Int. Ed. Engl., 2018, 57, 2165-2169.

17. M. Zeraati, D. B. Langley, P. Schofield, A. L. Moye, R. Rouet, W. E. Hughes, T. M. Bryan, M. E. Dinger and D. Christ, Nat. Chem., 2018, doi: 10.1038/s41557-41018-40046-41553.

18. G. Miglietta, S. Cogoi, E. B. Pedersen and L. E. Xodo, Sci. Rep., 2015, 5, 18097.

19. M. H. Kaulage, S. Bhattacharya and K. Muniyappa, ChemBioChem, 2018, doi: 10.1002/cbic.201800021.

20. A. M. Fleming, J. Zhu, Y. Ding and C. J. Burrows, ACS Chem. Biol., 2017, 12, 2417-2426.

21. E. P. Wright, J. L. Huppert and Z. A. Waller, Nucleic Acids Res., 2017, 45, 2951-2959.

22. R. A. Rogers, A. M. Fleming and C. J. Burrows, Biophys. J., 2018, 114, 1804-1815.

23. A. M. Fleming, Y. Ding, R. A. Rogers, J. Zhu, A. D. Burton, C. B. Carlisle and C. J. Burrows, J. Am. Chem. Soc., 2017, 139, 4682-4689.

24. A. L. Lieblein, B. Furtig and H. Schwalbe, ChemBioChem, 2013, 14, 1226-1230.

25. A. L. Lieblein, J. Buck, K. Schlepckow, B. Furtig and H. Schwalbe, Angew. Chem. Int. Ed. Engl., 2012, 51, 250-253.

26. J.-L. Mergny, L. Lacroix, X. Han, J.-L. Leroy and C. Helene, J. Am. Chem. Soc., 1995, 117, 8887-8898.

27. M. McKim, A. Buxton, C. Johnson, A. Metz and R. D. Sheardy, J. Phys. Chem. B, 2016, 120, 7652-7661.

28. S. P. Gurung, C. Schwarz, J. P. Hall, C. J. Cardin and J. A. Brazier, Chem. Commun., 2015, 51, 5630-5632.

29. S. Benabou, M. Garavis, S. Lyonnais, R. Eritja, C. Gonzalez and R. Gargallo, Phys. Chem. Chem. Phys., 2016, 18, 7997-8004.

30. I. V. Nesterova and E. E. Nesterov, J. Am. Chem. Soc., 2014, 136, 8843-8846.

31. S. Nonin-Lecomte and J. L. Leroy, J. Mol. Biol., 2001, 309, 491-506.

32. S. Fernandez, R. Eritja, A. Avino, J. Jaumot and R. Gargallo, Int. J. Biol. Macromol., 2011, 49, 729-736.

33. K. Guo, V. Gokhale, L. H. Hurley and D. Sun, Nucleic Acids Res., 2008, 36, 4598-4608.

34. G. He, C. K. Kwok and S. L. Lam, FEBS Lett., 2011, 585, 3953-3958.

35. K. M. Guckian, B. A. Schweitzer, R. X. Ren, C. J. Sheils, D. C. Tahmassebi and E. T. Kool, J. Am. Chem. Soc., 2000, 122, 2213-2222.

36. L. Di Michele, B. M. Mognetti, T. Yanagishima, P. Varilly, Z. Ruff, D. Frenkel and E. Eiser, J. Am. Chem. Soc., 2014, 136, 6538-6541.

37. A. Bugaut and P. Alberti, Biochimie, 2015, 113, 125-133.

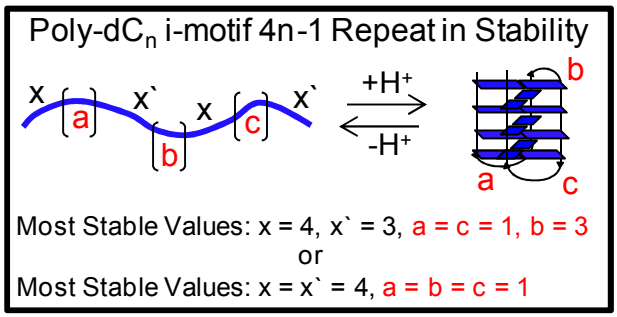
38. S. Kendrick, Y. Akiyama, S. M. Hecht and L. H. Hurley, J. Am. Chem. Soc., 2009, 131, 17667-17676.

39. S. Benabou, R. Ferreira, A. Avino, C. Gonzalez, S. Lyonnais, M. Sola, R. Eritja, J. Jaumot and R. Gargallo, Biochim. Biophys. Acta, 2014, 1840, 41-52.

40. A. T. Phan, M. Gueron and J. L. Leroy, J. Mol. Biol., 2000, 299, 123-144.

41. X. Han, J. L. Leroy and M. Gueron, J. Mol. Biol., 1998, 278, 949-965.

Table of Contents Entry



The 4n-1 repeat pattern for poly-2′-deoxycytidine i-motifs results from ideal loop lengths and core base pairs.