

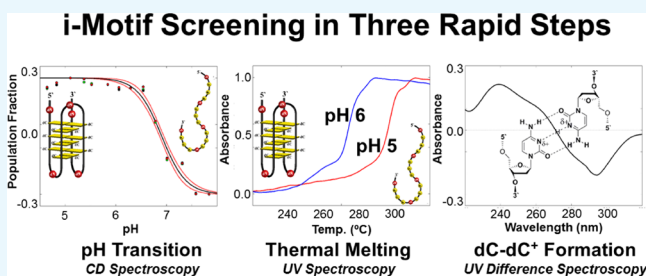
# Rapid Screen of Potential i-Motif Forming Sequences in DNA Repair Gene Promoters

R. Aaron Rogers, Aaron M. Fleming,<sup>1</sup> and Cynthia J. Burrows<sup>\*,2</sup>

Department of Chemistry, University of Utah, 315 South 1400 East, Salt Lake City, Utah 84112, United States

## Supporting Information

**ABSTRACT:** We have studied the in vitro stability of 25 potential i-motif-forming DNA sequences found within the promoter regions of 18 different human DNA repair genes. Three widely available methods of characterization were used to rapidly assess i-motif folding and stability and comprise a simple screen for preliminary identification of physiologically relevant i-motif forming sequences. Four highly pH-stable candidate sequences were identified exhibiting pH transitions (pH at which 50% of the oligodeoxynucleotides in solution are folded) at or above pH 6.6, thermal melting temperatures above 37 °C and isothermal UV difference spectra characteristic of 2'-deoxycytidine imino-nitrogen protonation. These newly identified i-motif forming sequences could represent novel targets for understanding and modulating human DNA repair gene expression.



## INTRODUCTION

The i-motif (iM) is a noncanonical, tetraplex-type secondary structure formed by cytosine-rich DNA sequences, in which at least four polycytidine tracks are proximal to each other; i.e., less than ~10 nucleotides apart. The core of the iM tetraplex is composed of hemiprotonated 2'-deoxycytidine base pairs (dC–dC<sup>+</sup>) between parallel-oriented segments of a single DNA strand (Figure 1).<sup>1–5</sup> Low pH, i.e., pH 4–6, drives iM formation, and the pH value at which iM folding occurs is sequence-dependent. The transitional pH (pH<sub>T</sub>), defined as the pH at which 50% of a solution of a given iM-forming sequence is folded, is the key physical parameter by which an iM's physiological probability is measured.<sup>6</sup> Identifying iM-forming DNA sequences in the human genome that fold at or near neutral pH is, therefore, of high importance for deducing potential iM functions in vivo.

In a genomic sense, potential iM-forming sequences can be found as the complements of potential guanine-quadruplex (G4) forming sequences. Over 700 000 potential G4-forming sequences have been identified within the human genome using a bioinformatics approach and a high-throughput, G4-specific chromatin immunoprecipitation sequencing assay that identify folded G4 sequences within the human genome.<sup>7,8</sup> Interestingly, a propensity for potential G4-forming sequences was noted within promoter regions of genes.<sup>8,9</sup> On the basis of this finding and a variety of experimental data, it has been suggested that certain G4- and iM-forming sequences serve as transcriptional regulatory elements.<sup>10</sup> Though the G4 has received extensive attention in recent years, iMs have remained less studied, primarily due to the pH-dependent nature of their folding transition.

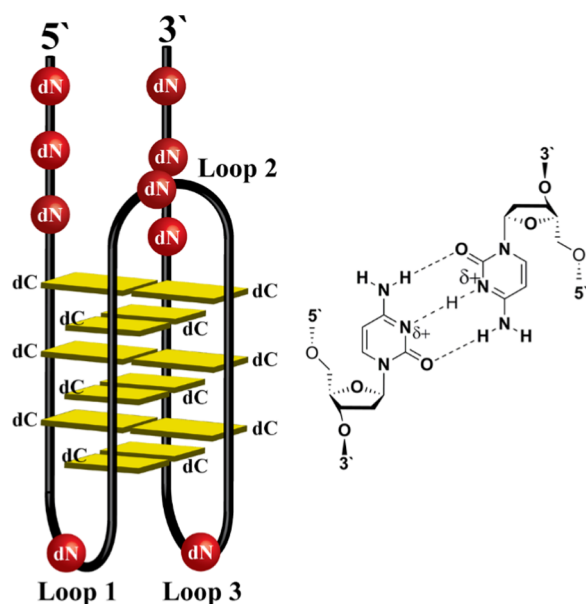
Recently, characterization of endogenous iM-forming oligodeoxynucleotide sequences within the human genome has revealed a significant propensity for many endogenous sequences to form at or near neutral pH.<sup>11–14</sup> Further, in the case of the *BCL2* oncogene promoter iM, a potential transcriptional regulatory role has been identified and manipulated in vivo as a druggable target in efforts to modulate gene expression.<sup>15–19</sup> This has led to the idea that iM structures in the human genome can provide novel drug targets for the regulation of gene expression.<sup>10,16,20</sup> *BCL2* is currently the most-studied iM-forming sequence, both in vitro and in vivo and serves as a promising example of the possibilities for iM research and serves as a primary standard for other iM-forming sequences to be measured by. For these reasons, the search for iM structures that are stable in vitro under physiologically relevant conditions remains a high priority for advancement toward understanding the biological functions of the DNA iM.

Recent developments employing state-of-the-art in-cell NMR techniques indicate a high potential for iM-folding activity in vivo,<sup>18,21</sup> and new rules have also been established to guide the search for physiologically active iM structures based on perturbations of polycytidine sequences.<sup>13,22</sup> Identification of iM-forming sequences, such as poly-dC<sub>19</sub> and a poly-dC-like iM-forming sequence within the promoter of *RAD17*, has also been coupled with detailed investigations of structural dynamics and folding rate-dependent pH<sub>T</sub> behavior.<sup>13,22,23</sup> As our knowledge of iM-forming DNA progresses, so too does

Received: July 5, 2018

Accepted: August 8, 2018

Published: August 21, 2018



**Figure 1.** Generic model of a DNA i-motif. Each i-motif (left) is composed of a “core” of hemiprotonated dC–dC<sup>+</sup> base pairs (right) between parallel-oriented strands of the tetraplex (yellow). There are three loops and two tails which may be composed of any 2′-deoxynucleotides (dN). Topology is also defined as either 5′E (shown) or 3′E, which is determined by the position of the dC–dC<sup>+</sup> base pair closest to the 5′ or 3′ end of the sequence, respectively.

our ability to efficiently search for and characterize candidate sequences that have the potential to serve as regulatory elements within the human genome.

Here, we provide a concise summary of iM-folding behavior for 25 potential iM-forming sequences located in 18 DNA

repair gene promoters, including the previously studied RAD17 promoter iM.<sup>23</sup> Each iM-forming sequence was identified as the complement to a potential G4-forming sequence found using the QuadParser bioinformatics tool with a modified set of search parameters, as described previously by Fleming et al.<sup>24</sup> These sequences were originally chosen by Fleming et al. due to their presence in the promoter regions of DNA repair genes involved in stress response due to damage induced by reactive oxygen species, which is enhanced in G4-forming sequences. Some of those potential G4-forming sequences were discovered to have a lower probability of folding, so we wished to characterize the complementary sequences to determine whether the dC-rich sequences exhibited different stabilities under physiologically relevant conditions *in vitro*. To accomplish this, we employed three spectroscopic methods of characterization to quantify their pH-dependent stability, thermal stability, and cytosine protonation state under physiological salt and buffer conditions to screen for potential physiologically relevant structures. Each sequence studied varies in polycytidine tract length, polycytidine tract number, and potential loop lengths showing that pH-dependent stability can arise from a wide variety of sequence contexts, confirming some trends previously noted in model studies.

## RESULTS AND DISCUSSION

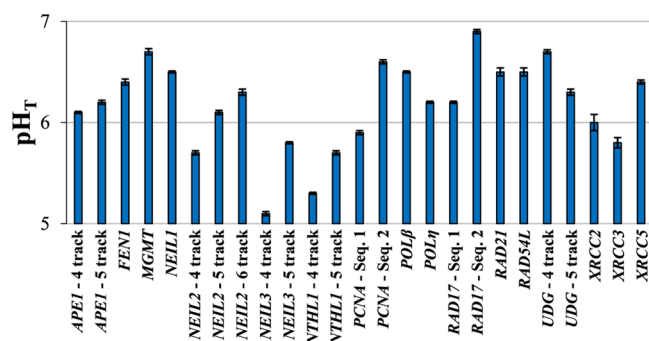
**Oligodeoxynucleotide Sequences.** Table 1 shows the potential iM-forming human gene promoter sequences examined in this study presented with the gene name from which each was extracted along with their polycytidine track lengths and intermediate, nonpolycytidine track lengths, denoted “loop lengths.” Some gene promoters contained two distinct iM-forming sequences (denoted by an appended “sequence #” in Table 1) or multiple polycytidine tracts

**Table 1.** Gene Names and Sequences for Potential iM-Forming ODNs Used in the DNA Repair Gene Screen

gene	sequence	C-track lengths	loop lengths
APE1-4 track	5′-TACCCACCCCCACCTGCCCTG	3-5-3-3	1-1-2
APE1-5 track	5′-AACCCCCAGGGCTACCCACCCCCACCTGCCCTG	5-3-5-3-3	7-1-1-2
FEN1	5′-GTCCCCACTCCACCCACACCAGGTCCCCGCAGGCCCTGTCTCCCTC	4-3-4-4-3	6-9-5-4
MGMT	5′-CCGCCCCAGCTCCGCCCCCGCGCGCCCCGGCCCCGCCCCCGC	4-5-4-4-5	7-5-2-1
NEIL1	5′-CGCCCTCCCTGCGCCCCCTCCCCCAC	4-3-6-6	1-4-1
NEIL2-4 track	5′-GGCCCGGGCCCGCCCTCCCTT	3-3-3-3	4-1-1
NEIL2-5 track	5′-GGCCCGGGCCCGCCCTCCCTTCCTGTCCCCCTC	3-3-3-3-4	4-1-1-7
NEIL2-6 track	5′-GGCCCGGGCCCGCCCTCCCTTCCTGTCCCCCTCCCCGA	3-3-3-3-4-3	4-1-1-7-1
NEIL3-4 track	5′-GGCCCGGCCAGGCCCGGCCAA	4-3-4-3	1-3-1
NEIL3-5 track	5′-GGCCCGGCCAGGCCCGGCCAAACAGCACCCCTA	4-3-4-3-3	1-3-1-8
NTHL1-4 track	5′-GTCCCGGGCCCTCACCCGCGCCAC	3-3-3-3	3-3-3
NTHL1-5 track	5′-GTCCCGGGCCCTCACCCGCGCCACTGCAACCCGA	3-3-3-3-3	3-3-3-7
PCNA: sequence 1	5′-TTCCCTAGCCCGACCCGAGAGCTCCCTCTCCCGG	3-4-3-3-3	3-2-7-3
PCNA: sequence 2	5′-CGCCCGGCCCGCCCGGTCCGCTTCCCTCCCTG	4-4-5-3-3	1-1-4-5
POLβ	5′-CGCCCTCTAGCCCCGCCCGCCCCGCCCCAG	4-4-4-4-3	5-1-1-1
POLη	5′-GTCCCGACACCTCTCCAGCCCCAG	3-3-3-4	4-3-2
RAD17: sequence 1	5′-CGCCCCAGCCTGCCAGCCAGTCCCTCCCGG	5-4-3-3-3	6-2-3-1
RAD17: sequence 2	5′-CCACCCCCCCCCGCCCCCCCGGA	9-9	1
RAD21	5′-TTCCCAACCCCTCCCCCGACCTTTTCCCTCCCCCG	4-5-5-3-4-4	1-1-2-4-1
RADS4L	5′-GGCCCGGCCCTCCCCCGCCACCCCCGCCCGGCCCGCCCCCTC	4-4-5-5-5-4-4	1-1-4-1-1-1-1
UDG-4 track	5′-TTCCAGCCCCCTCCCCCGACCCAC	3-5-6-4	2-1-2
UDG-5 track	5′-CACCCCTAAGGGGAGGAACCTTTCTTCCAGCCCCCTCCCCGACCCAC	4-3-5-5-4	21-2-1-2
XRCC2	5′-CGCCACCGGCGGCTTGTTCCTACTCCCTCACTCCCAACCCGG	3-3-3-3-3	15-4-5-2
XRCC3	5′-GACCCGCCCGCCCGCCCGCCCGCCCGC	3-4-4-3-4	1-4-2-2
XRCC5	5′-TACCCACCCATCCCATCCCTCTTCTCCCTC	3-3-3-3-3	1-2-2-6

(denoted by an appended “# tract” in Table 1). Of note are the 5′ and 3′ flanking sequences of RAD17: sequence 2, which are three nucleotides in length, rather than the general two nucleotide flanking sequences in all other oligodeoxynucleotides. This was a deliberate choice to provide consistency with a previously published in-depth investigation of that sequence by our laboratory.<sup>23</sup>

**Sequence Features and Observed  $pH_T$  Values.** pH-dependent circular dichroism (CD) spectra were recorded for each oligodeoxynucleotide sequence in Table 1 in triplicate, and a normalized  $pH_T$  curve was built and evaluated as described in Materials and Methods (Figures S1–S25). Figure 2 shows the  $pH_T$  values ( $\pm$ standard error) for each



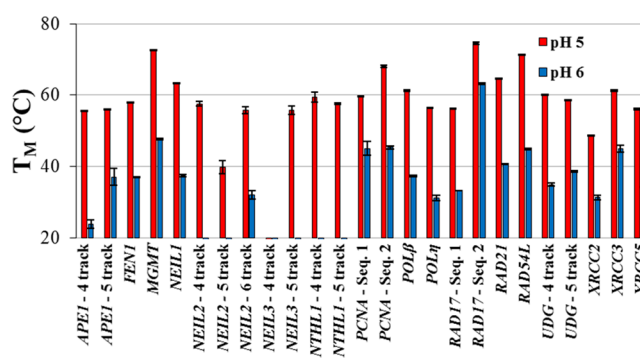
**Figure 2.**  $pH_T$  values for all oligodeoxynucleotide sequences studied under physiological salt and buffer concentrations. Error bars represent the standard error of the calculated  $pH_T$  determined by propagation of the errors in fit parameters  $b_1$  and  $b_2$ , taking into account covariance between the two values. All values and standard errors are presented numerically in Table S1 as well as the CD  $\lambda_{max}$  value at low pH. For ease of comparison, the figure is provided immediately above Figure 3 in a combined format in Figure S26, Supporting Information.

oligodeoxynucleotide studied. Four of the characterized iM-forming sequences have  $pH_T$  values (Table S1) that are greater than 6.6, the  $pH_T$  of the BCL2 iM, which we consider a primary standard of comparison.<sup>11</sup> Stabilities for the top four followed the order RAD17: sequence 2 ( $pH_T$  6.9) > MGMT ( $pH_T$  6.7)  $\approx$  UDG-4 track ( $pH_T$  6.7) > PCNA: sequence 2 ( $pH_T$  6.6).

The  $pH_T$  data set shows an apparent trend of increasing  $pH_T$  values with increasing numbers of polycytidine tracks for homologous sequences within a single gene promoter. An exception is the lower  $pH_T$  of the UDG-5 track sequence compared with the UDG-4 track sequence. This is likely due to the longer separation between the first and second polycytidine tracks (21 nucleotides) in the UDG-5 track sequence. Such a long loop is unprecedented, and is likely not involved in the iM structure itself based on the results of previously reported loop length investigations. Model studies of polycytidine oligodeoxynucleotides separated by polythymidine tracks or poly(ethylene glycol) spacers, which serve as forced loops, show a decrease in stability with longer loop length.<sup>14,22,25–28</sup> Further, the CD spectra of UDG-5 track show the loss of an experimentally determined ellipticity minimum ca. 265 nm. We speculate that this indicates the 5′ polycytidine track, and potential 21 nucleotide loop are likely uninvolved in the iM structure, though a detailed structural investigation would be necessary to confirm this.

The identity of the loop nucleotide bases also appears to conform to previously determined investigations of loop nucleotide identity relationships to structural stability. For the first and third loops in model iM-forming sequences, a general preference for T–T or G–T base pairs has been noted, and sequences with higher purine content in the loops are generally disfavored.<sup>26,29,30</sup> Overall, lower purine content in the bases composing the loops correlates with a higher  $pH_T$  in our oligodeoxynucleotide set. All of the NEIL3 and NTHL1 sequences as well as PCNA: sequence 1 and NEIL2-4 track, possess nucleobase loop compositions that are generally higher in purine content, a feature that has been observed to disfavor iM formation due to the bulk of purine bases compared to their pyrimidine counterparts.<sup>29</sup>

**pH-Dependent  $T_M$  Analysis.** All potential iM-forming sequences were subjected to  $T_M$  analysis at pHs 5 and 6. The results are summarized in Figure 3. Representative  $T_M$  curves are provided in Figures S1–S25 in the Supporting Information.



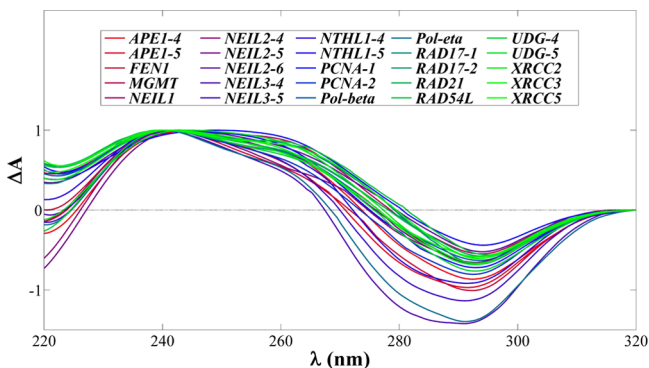
**Figure 3.** Thermal melting temperatures for all oligodeoxynucleotide sequences studied at pH 5 (red) and pH 6 (blue) with the standard error of the mean for each triplicate experiment. Some oligodeoxynucleotide sequences had no discernible thermal melting temperature above 20 °C. All values are tabulated in Table S1 along with their standard errors. For ease of comparison, Figure 2 is provided immediately above the figure in a combined format in Figure S26, Supporting Information.

At pH 5, all but the NEIL3-4 track sequence had a discernible  $T_M$ . This was an unsurprising result considering that this sequence exhibited the lowest  $pH_T$  out of all of the potential iM-forming sequences studied. For all other sequences, the  $T_M$  values at pH 5 were greater than  $\sim$ 40 °C. Interrogation at pH 6, however, showed a significant number of sequences to be thermally unstable to the point that no  $T_M$  could be determined. The majority of these sequences exhibited low  $pH_T$  values and consequently are unlikely to have a significant folded population at pH 6.

For the four highest pH-stability sequences,  $T_M$  values at pH 6 showed the order RAD17: sequence 2 (60.4 °C) > MGMT (47.7 °C) > PCNA: sequence 2 (45.4 °C) > UDG-4 track (35.0 °C). Interestingly, not all of the highest pH-stable sequences gave the highest  $T_M$  values. RAD54L, PCNA: sequence 1, UDG-5 track, XRCC3, RAD21, POLB, APE1-5 track, FEN1, and NEIL1 all had higher pH 6  $T_M$  values than UDG-4 track. Such an observation is likely due to discrepancies in the sharpness of the  $pH_T$  curves, where sequences with higher  $T_M$  values and lower  $pH_T$  values relative to UDG-4 track indicate a more gradual unfolding transition.

**Isothermal UV Difference Spectra (IDS).** Each oligodeoxynucleotide studied was further characterized by taking

the difference between high pH (unfolded state) and low pH (folded state) UV spectra, collected concurrently with pH-dependent CD spectra. The results, shown in Figure 4, are



**Figure 4.** Compilation of the normalized IDS for each oligodeoxynucleotide studied. All spectra are qualitative only and indicative of cytosine protonation. Maxima and minima are tabulated in Table S1.

equivalent to those obtained by thermal UV difference spectroscopy, established as a method to assess the formation of protonated cytosines.<sup>1,31</sup> All IDS are indicative of the formation of protonated cytosine residues and further support the probability of iM folding.

**High-Stability Sequences.** All sequences met the experimental criteria presented here to be deemed probable iM structures. The candidate sequences best suited for further investigation, however, are the four aforementioned high-stability sequences; *MGMT*, *PCNA*: sequence 2, *RAD17*: sequence 2, and *UDG-4* track. Of those four, *MGMT* and *RAD17*: sequence 2 pose the greatest potential for further investigation.

The promoter iM of *MGMT* presents a unique potential for gene regulation pertinent to a number of different disease states. *MGMT* is a gene encoding a protein responsible for the direct, stoichiometric repair of *O*<sup>6</sup>-methyl-2'-deoxyguanosine, a primary mutagenic lesion in DNA.<sup>32</sup> Its promoter region has been described as a prognostic factor in patient response to alkylating agent chemotherapy, determined by the methylation status of that region of the gene.<sup>33–39</sup> The potential iM-forming sequence studied here contains eight CpG dinucleotides and is located 75 bases upstream of the transcription start site. 5-Methyl-2'-deoxycytidine, the methylated DNA epigenetic marker in CpG dinucleotides, has been shown to slightly enhance the thermal stability of the human telomere iM and increase the *pK*<sub>a</sub> of the imino nitrogen of the nucleobase.<sup>40,41</sup> Regardless of whether methylation affects the iM-forming region within the *MGMT* promoter, further investigation is clearly justified to determine its potential for biological activity, which we have not yet determined.

*RAD17*: sequence 2 presents a unique sequence context atypical of most iM-forming oligodeoxynucleotides in that it is composed of two nine-nucleotide polycytidine tracks separated by a single guanine nucleotide. This is extremely similar to the recently discovered poly-C<sub>19</sub> oligodeoxynucleotide, which is the most stable polycytidine sequence length currently known.<sup>13</sup> The *pH*<sub>T</sub> of *RAD17*: sequence 2 is nearest to neutral pH out of all of the sequences studied here and is only slightly lower than that of poly-C<sub>19</sub>. The *T*<sub>M</sub> values of *RAD17*: sequence 2 are also slightly lower than that of poly-C<sub>19</sub>. This is likely due to the presence of the single guanine nucleotide in

the middle of the polycytidine region, which would likely force a loop to form and reduce structural dynamics. Given the relatively high *pH*<sub>T</sub> and *T*<sub>M</sub> values of these sequences, further investigations into the stability, folding behavior, and structure are currently ongoing.

## CONCLUSIONS

Here, we have provided a survey of 25 oligodeoxynucleotides extracted from the promoter regions of 18 DNA repair genes and assessed their folding behavior and potential for further investigation using standard spectroscopic methods. Among the oligodeoxynucleotides studied, four were identified as potential high-stability iM structures relative to the most biologically investigated iM to date, that in the *BCL2* oncogene promoter.<sup>11,15,16</sup> These results are not an exhaustive characterization of these potential iMs, but rather are presented as a method of rapidly assessing the potential for iM folding under physiologically relevant conditions. Among various endogenous DNA repair gene promoters, the iM could play a regulatory role important to a variety of DNA damage-related disease states. Further, our survey shows a variety of sequence contexts that can give rise to high-stability sequences, which could be useful for designing intelligent biosensors and pH-responsive devices.

## MATERIALS AND METHODS

### Oligodeoxynucleotides Synthesis and Purification.

All oligodeoxynucleotides used herein were synthesized using standard phosphoramidite procedures by the DNA synthesis core facility at the University of Utah. All samples were then treated with 200 mM piperidine for 30 min at 90 °C prior to purification by high-performance liquid chromatography using a high pH eluent to prevent iM folding on column. Samples were then dialyzed 3 × 12 h against ddH<sub>2</sub>O, lyophilized, resuspended in ddH<sub>2</sub>O, quantified, and then frozen as concentrated stock solutions until needed for experiments. Molar absorptivity values for all oligodeoxynucleotides studied were determined using the nearest-neighbor method.

**Circular Dichroism (CD) Spectroscopy.** CD spectra and *pH*<sub>T</sub> values were recorded and determined, respectively, using a set of Britton–Robinson buffers (20 mM HOAc, 20 mM H<sub>3</sub>PO<sub>4</sub>, 20 mM H<sub>3</sub>BO<sub>3</sub>, 140 mM KCl, 12 mM NaCl, pH adjusted with NaOH) ranging from pH 4.5 to 8 in increments of 0.25 pH units. It should be noted that Na<sup>+</sup> concentrations ranging from 0 to 100 mM can destabilize iM structures by up to 0.4 pH units, and our use of NaOH as a pH adjustment does place our samples in this range.<sup>1</sup> For that reason, internal comparison of our data set is validated for the work presented here, and cross-examination with stabilities presented in other studies may not be straightforward. Concentrated stock oligodeoxynucleotides were spiked into triplicate samples of 500 μL of each buffer to give a concentration of ~5 μM, and the pH values were recorded. Each sample was then loaded into a 2 mm cuvette, and the CD spectrum was recorded. CD and ultraviolet–visible (UV) spectra were recorded concurrently from 200 to 320 nm at 21 °C for each scan. The ellipticities at all wavelengths were converted to molar ellipticity,  $[\theta]$  by the equation  $[\theta] = \theta \times (10 \times C \times l)^{-1}$ , where *C* is the concentration of the sample based on the absorbance at 260 nm in the UV spectrum and *l* is the path length of the cuvette. The *pH*<sub>T</sub> values were found by nonlinear regression of normalized ellipticities at the experimentally

determined wavelength maxima ca. 287 nm (low pH) versus pH using the following sigmoid model equation.

$$[x] = \frac{-1}{1 + e^{((-b_1 \times \text{pH}) + b_2)}} + 1, \text{pH}_T = \frac{b_2}{b_1}$$

[x] is the folded state population at a given pH, and  $b_{\#}$  is parameter to be fit via nonlinear regression in Matlab. The 95% confidence intervals ( $\Delta_{95\%}$ ) were calculated in addition to the regression curve itself to provide an accurate representation of the accuracy of the data set and regression analysis.

**Thermal Melting Temperature Analysis ( $T_M$ ).** Using the same Britton–Robinson buffer as noted for the CD spectroscopy screen, UV spectroscopy was used to monitor the heat-induced denaturation of all sequences studied in triplicate. For each potential iM-forming sequence in the DNA repair gene screen, pHs 5 and 6 buffers were used. Stock concentration oligodeoxynucleotides were spiked into the solution to give a concentration that would result in  $\sim 0.20$ – $0.25$  absorbance units at 262 nm ( $\sim 1$ – $3.5 \mu\text{M}$  for all ODNs) to inhibit intermolecular structure formation. Each solution was equilibrated at 20 °C (DNA repair gene sequences for the screen) for 5 min and then heated at 1 °C  $\text{min}^{-1}$ , paused for 30 s after a 1 °C increase in temperature, and the UV absorbance recorded at 262 nm. This was continued to a final temperature of 100 °C. The derivative method was then used to analyze the resultant thermally induced denaturation curves at 262 nm.

**Isothermal Difference Spectroscopy.** During CD spectroscopic experiments, the UV spectrum was recorded concurrently with each CD spectrum. The UV spectra for the lowest and highest pH scans were subtracted from one another according to protocols established in the literature for thermal difference spectra and isothermal difference spectra (IDS).<sup>31</sup> The results of this approach are identical to those obtained via the thermal method described by Mergny and co-workers and are characteristic of the protonation state of 2'-deoxycytidine residues.<sup>1</sup> The only differentiating factor between a thermal and isothermal difference spectrum is that the unfolded random coil state is obtained thermally in a thermal difference spectrum and is obtained via pH-induced unfolding in the isothermal difference spectrum.

## ■ ASSOCIATED CONTENT

### ■ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acsomega.8b01551.

For each oligodeoxynucleotide studied; pH-dependent CD spectra, thermal denaturation profiles and  $\text{pH}_T$  profiles, and all presented data in tabulated form (PDF)

## ■ AUTHOR INFORMATION

### Corresponding Author

\*E-mail: burrows@chem.utah.edu (C.J.B.).

### ORCID

Aaron M. Fleming: 0000-0002-2000-0310

Cynthia J. Burrows: 0000-0001-7253-8529

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

This work was supported by a National Cancer Institute grant (R01 CA090689). The oligonucleotides were provided by the DNA/peptide core facility at the University of Utah that is supported in part by the NCI Cancer Center Support grant (P30 CA042014).

## ■ REFERENCES

- (1) Mergny, J.-L.; Lacroix, L.; Han, X.; Leroy, J.-L.; Helene, C. Intramolecular Folding of Pyrimidine Oligodeoxynucleotides into an I-DNA Motif. *J. Am. Chem. Soc.* **1995**, *117*, 8887–8898.
- (2) Gehring, K.; Leroy, J.-L.; Gueron, M. A Tetrameric DNA Structure with Protonated Cytosine-Cytosine Base Pairs. *Nature* **1993**, *363*, 561–565.
- (3) Han, X.; Leroy, J.-L.; Gueron, M. An Intramolecular I-Motif: The Solution Structure and Base-Pair Opening Kinetics of d-(5mCCT<sub>3</sub>CCT<sub>3</sub>ACCT<sub>3</sub>CC). *J. Mol. Biol.* **1998**, *278*, 949–965.
- (4) Guéron, M.; Leroy, J.-L. The I-Motif in Nucleic Acids. *Curr. Opin. Struct. Biol.* **2000**, *10*, 326–331.
- (5) Leroy, J.-L.; Gueron, M.; Mergny, J.-L.; Helene, C. Intramolecular Folding of a Fragment of the Cytosine-Rich Strand of Telomeric DNA into an i-Motif. *Nucleic Acids Res.* **1994**, *22*, 1600–1606.
- (6) Phan, A. T.; Guéron, M.; Leroy, J.-L. The Solution Structure and Internal Motions of a Fragment of the Cytidine-Rich Strand of the Human Telomere. *J. Mol. Biol.* **2000**, *299*, 123–144.
- (7) Huppert, J. L.; Balasubramanian, S. Prevalence of Quadruplexes in the Human Genome. *Nucleic Acids Res.* **2005**, *33*, 2908–2916.
- (8) Chambers, V. S.; Marsico, G.; Boutell, J. M.; Di Antonio, M.; Smith, G. P.; Balasubramanian, S. High-Throughput Sequencing of DNA G-Quadruplex Structures in the Human Genome. *Nat. Biotechnol.* **2015**, *33*, 877–881.
- (9) Huppert, J. L.; Balasubramanian, S. G-Quadruplexes in Promoters throughout the Human Genome. *Nucleic Acids Res.* **2007**, *35*, 406–413.
- (10) Kendrick, S.; Hurley, L. H. The Role of G-Quadruplex/i-Motif Secondary Structures as Cis-Acting Regulatory Elements. *Pure Appl. Chem.* **2010**, *82*, 1609–1621.
- (11) Kendrick, S.; Akiyama, Y.; Hecht, S. M.; Hurley, L. H. The I-Motif in the BCL2 P1 Promoter Forms an Unexpectedly Stable Structure with a Unique 8:5:7 Loop Folding Pattern. *J. Am. Chem. Soc.* **2009**, *131*, 17667–17676.
- (12) Brazier, J. A.; Shah, A.; Brown, G. D. I-Motif Formation in Gene Promoters: Unusually Stable Formation in Sequences Complementary to Known G-Quadruplexes. *Chem. Commun.* **2012**, *48*, 10739.
- (13) Fleming, A. M.; Ding, Y.; Rogers, R. A.; Zhu, J.; Zhu, J.; Burton, A. D.; Carlisle, C. B.; Burrows, C. J. 4 n – 1 Is a “Sweet Spot” in DNA i-Motif Folding of 2'-Deoxycytidine Homopolymers. *J. Am. Chem. Soc.* **2017**, *139*, 4682–4689.
- (14) Wright, E. P.; Huppert, J. L.; Waller, Z. A. E. Identification of Multiple Genomic DNA Sequences Which Form I-Motif Structures at Neutral PH. *Nucleic Acids Res.* **2017**, *45*, 2951–2959.
- (15) Kang, H.-J.; Kendrick, S.; Hecht, S. M.; Hurley, L. H. The Transcriptional Complex Between the BCL2 i-Motif and HnRNP LL Is a Molecular Switch for Control of Gene Expression That Can Be Modulated by Small Molecules. *J. Am. Chem. Soc.* **2014**, *136*, 4172–4185.
- (16) Kendrick, S.; Kang, H.-J.; Alam, M. P.; Madathil, M. M.; Agrawal, P.; Gokhale, V.; Yang, D.; Hecht, S. M.; Hurley, L. H. The Dynamic Character of the BCL2 Promoter i-Motif Provides a Mechanism for Modulation of Gene Expression by Compounds That Bind Selectively to the Alternative DNA Hairpin Structure. *J. Am. Chem. Soc.* **2014**, *136*, 4161–4171.
- (17) Cui, Y.; Kong, D.; Ghimire, C.; Xu, C.; Mao, H. Mutually Exclusive Formation of G-Quadruplex and i-Motif Is a General Phenomenon Governed by Steric Hindrance in Duplex DNA. *Biochemistry* **2016**, *55*, 2291–2299.

- (18) Zeraati, M.; Langley, D. B.; Schofield, P.; Moye, A. L.; Rouet, R.; Hughes, W. E.; Bryan, T. M.; Dinger, M. E.; Christ, D. I-Motif DNA Structures Are Formed in the Nuclei of Human Cells. *Nat. Chem.* **2018**, *10*, 631–637.
- (19) Sutherland, C.; Cui, Y.; Mao, H.; Hurley, L. H. A Mechanosensor Mechanism Controls the G-Quadruplex/i-Motif Molecular Switch in the MYC Promoter NHE III<sub>1</sub>. *J. Am. Chem. Soc.* **2016**, *138*, 14138–14151.
- (20) Day, H. A.; Pavlou, P.; Waller, Z. A. E. I-Motif DNA: Structure, Stability and Targeting with Ligands. *Bioorg. Med. Chem.* **2014**, *22*, 4407–4418.
- (21) Dzatko, S.; Krafcikova, M.; Hänsel-Hertsch, R.; Fessl, T.; Fiala, R.; Loja, T.; Krafcik, D.; Mergny, J.-L.; Foldynova-Trantirkova, S.; Trantirek, L. Evaluation of the Stability of DNA I-Motifs in the Nuclei of Living Mammalian Cells. *Angew. Chem., Int. Ed.* **2018**, *57*, 2165–2169.
- (22) Fleming, A. M.; Stewart, K. M.; Eyring, G. M.; Ball, T. E.; Burrows, C. J. Unraveling the  $4n - 1$  Rule for DNA i-Motif Stability: Base Pairs vs. Loop Lengths. *Org. Biomol. Chem.* **2018**, *16*, 4537–4546.
- (23) Rogers, R. A.; Fleming, A. M.; Burrows, C. J. Unusual Isothermal Hysteresis in DNA I-Motif PH Transitions: A Study of the RAD17 Promoter Sequence. *Biophys. J.* **2018**, *114*, 1804–1815.
- (24) Fleming, A. M.; Zhu, J.; Ding, Y.; Visser, J. A.; Zhu, J.; Burrows, C. J. Human DNA Repair Genes Possess Potential G-Quadruplex Sequences in Their Promoters and 5'-Untranslated Regions. *Biochemistry* **2018**, 991–1002.
- (25) Yang, Y.; Sun, Y.; Yang, Y.; Xing, Y.; Zhang, T.; Wang, Z.; Yang, Z.; Liu, D. Influence of Tetra(Ethylene Glycol) (EG<sub>4</sub>) Substitution at the Loop Region on the Intramolecular DNA i-Motif. *Macromolecules* **2012**, *45*, 2643–2647.
- (26) Fujii, T.; Sugimoto, N. Loop Nucleotides Impact the Stability of Intrastrand I-Motif Structures at Neutral PH. *Phys. Chem. Chem. Phys.* **2015**, *17*, 16719–16722.
- (27) Gurung, S. P.; Schwarz, C.; Hall, J. P.; Cardin, C. J.; Brazier, J. A. The Importance of Loop Length on the Stability of I-Motif Structures. *Chem. Commun.* **2015**, *51*, 5630–5632.
- (28) Reilly, S. M.; Morgan, R. K.; Brooks, T. A.; Wadkins, R. M. Effect of Interior Loop Length on the Thermal Stability and  $pK_a$  of i-Motif DNA. *Biochemistry* **2015**, *54*, 1364–1370.
- (29) Benabou, S.; Garavis, M.; Lyonnais, S.; Eritja, R.; González, C.; Gargallo, R. Understanding the Effect of the Nature of the Nucleobase in the Loops on the Stability of the I-Motif Structure. *Phys. Chem. Chem. Phys.* **2016**, *18*, 7997–8004.
- (30) Benabou, S.; Aviñó, A.; Eritja, R.; González, C.; Gargallo, R. Fundamental Aspects of the Nucleic Acid I-Motif Structures. *RSC Adv.* **2014**, *4*, 26956.
- (31) Mergny, J.-L.; Li, J.; Lacroix, L.; Amrane, S.; Chaires, J. B. Thermal Difference Spectra: A Specific Signature for Nucleic Acid Structures. *Nucleic Acids Res.* **2005**, *33*, e138.
- (32) Soll, J. M.; Sobol, R. W.; Mosammaparast, N. Regulation of DNA Alkylation Damage Repair: Lessons and Therapeutic Opportunities. *Trends Biochem. Sci.* **2017**, *42*, 206–218.
- (33) Amatu, A.; Sartore-Bianchi, A.; Moutinho, C.; Belotti, A.; Bencardino, K.; Chirico, G.; Cassingena, A.; Rusconi, F.; Esposito, A.; Nichelatti, M.; et al. Promoter CpG Island Hypermethylation of the DNA Repair Enzyme MGMT Predicts Clinical Response to Dacarbazine in a Phase II Study for Metastatic Colorectal Cancer. *Clin. Cancer Res.* **2013**, *19*, 2265–2272.
- (34) Puppa, A. D.; Persano, L.; Masi, G.; Rampazzo, E.; Sinigaglia, A.; Pistollato, F.; Denaro, L.; Barzon, L.; Palù, G.; Basso, G.; et al. MGMT Expression and Promoter Methylation Status May Depend on the Site of Surgical Sample Collection within Glioblastoma: A Possible Pitfall in Stratification of Patients? *J. Neurooncol.* **2012**, *106*, 33–41.
- (35) Toyota, M.; Suzuki, H.; Yamashita, T.; Hirata, K.; Imai, K.; Tokino, T.; Shinomura, Y. Cancer Epigenomics: Implications of DNA Methylation in Personalized Cancer Therapy. *Cancer Sci.* **2009**, *100*, 787–791.
- (36) Thon, N.; Kreth, S.; Kreth, F. W. Personalized Treatment Strategies in Glioblastoma: MGMT Promoter Methylation Status. *Oncotargets Ther.* **2013**, 1363.
- (37) Gerson, S. L. Clinical Relevance of MGMT in the Treatment of Cancer. *J. Clin. Oncol.* **2002**, *20*, 2388–2399.
- (38) Xie, H.; Tubbs, R.; Yang, B. Detection of MGMT Promoter Methylation in Glioblastoma Using Pyrosequencing. *Int. J. Clin. Exp. Pathol.* **2015**, *8*, 636.
- (39) Kitange, G. J.; Carlson, B. L.; Schroeder, M. A.; Grogan, P. T.; Lamont, J. D.; Decker, P. A.; Wu, W.; James, C. D.; Sarkaria, J. N. Induction of MGMT Expression Is Associated with Temozolomide Resistance in Glioblastoma Xenografts. *Neuro-Oncology* **2009**, *11*, 281–291.
- (40) Xu, B.; Devi, G.; Shao, F. Regulation of Telomeric I-Motif Stability by 5-Methylcytosine and 5-Hydroxymethylcytosine Modification. *Org. Biomol. Chem.* **2015**, *13*, 5646–5651.
- (41) Bhavsar-Jog, Y. P.; Van Dornshuld, E.; Brooks, T. A.; Tschumper, G. S.; Wadkins, R. M. Epigenetic Modification, Dehydration, and Molecular Crowding Effects on the Thermodynamics of i-Motif Structure Formation from C-Rich DNA. *Biochemistry* **2014**, *53*, 1586–1594.