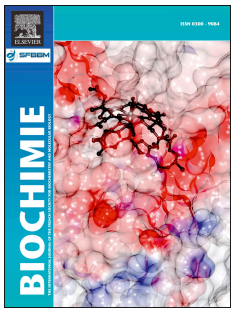
Journal Pre-proof



Stability and context of intercalated motifs (i-motifs) for biological applications

Kelly L. Irving, Jessica J. King, Zoë A.E. Waller, Cameron W. Evans, Nicole M. Smith

PII: S0300-9084(22)00052-9

DOI: https://doi.org/10.1016/j.biochi.2022.03.001

Reference: BIOCHI 6270

To appear in: *Biochimie*

Received Date: 3 February 2021

Revised Date: 21 February 2022

Accepted Date: 3 March 2022

Please cite this article as: K.L. Irving, J.J. King, Zoë.A.E. Waller, C.W. Evans, N.M. Smith, Stability and context of intercalated motifs (i-motifs) for biological applications, *Biochimie* (2022), doi: https://doi.org/10.1016/j.biochi.2022.03.001.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2022 Published by Elsevier B.V.

Journal Pre-proof

Abstract

DNA is naturally dynamic and can self-assemble into alternative secondary structures including the intercalated motif (i-motif), a four-stranded structure formed in cytosine-rich DNA sequences. Until recently, i-motifs were thought to be unstable in physiological cellular environments. Studies demonstrating their existence in the human genome and role in gene regulation are now shining light on their biological relevance. Herein, we review the effects of epigenetic modifications on i-motif structure and stability, and biological factors that affect i-motif formation within cells. Furthermore, we highlight recent progress in targeting i-motifs with structure-specific ligands for biotechnology and therapeutic purposes.

Journal Pre-proof

Journal Pre-proof

Stability and context of intercalated motifs (i-motifs) for biological applications

Kelly L. Irving[1#], Jessica J. King[1#], Zoë A. E. Waller[2], Cameron W. Evans[1]*, and Nicole M. Smith[1]*

[1]School of Molecular Sciences, The University of Western Australia, 35 Stirling Hwy, Crawley WA 6009, Australia.

[2]UCL School of Pharmacy, 29-39 Brunswick Square, London, WC1N 1AX, United Kingdom.

[#]These authors contributed equally to the work.

Keywords (6)

i-Motif, cytosine, methylation, DNA structure, proteins, ligands

Abstract: DNA is naturally dynamic and can self-assemble into alternative secondary structures including the intercalated motif (i-motif), a four-stranded structure formed in cytosine-rich DNA sequences. Until recently, i-motifs were thought to be unstable in physiological cellular environments. Studies demonstrating their existence in the human genome and role in gene regulation are now shining light on their biological relevance. Herein, we review the effects of epigenetic modifications on i-motif structure and stability, and biological factors that affect i-motif formation within cells. Furthermore, we highlight recent progress in targeting i-motifs with structure-specific ligands for biotechnology and therapeutic purposes.

Journal Pre-proof

1. Introduction

Certain guanine- (G) and cytosine- (C) rich sequences within DNA can form four-stranded non-canonical secondary structures known as G-quadruplexes and intercalated (i)-motifs, respectively. G-quadruplexes consist of getpi-stacked planar G-tetrads, with each tetrad containing four guanine bases bound through Hoogsteen hydrogen bonding [1–5]. G-quadruplex structures are known gene regulators and can act as targets for modulating gene expression, providing promising potential therapeutic targets. Although G-quadruplex structures have been more extensively studied, there is increasing interest in the i-motif, a non-canonical secondary structure formed within cytosine-rich DNA sequences, comprising hemi-protonated $C:C^+$ base pairs (Fig. 1a) stacked upon each other to form two intercalated hairpins (Fig. 1b) [6–8].

The i-motif was first reported to form from C-rich DNA under acidic conditions using nuclear magnetic resonance spectroscopy (NMR) by Gehring *et al.* in 1993 [9] and it has been shown that folded i-motifs are relatively stable in comparison to duplex B-DNA [8]. i-Motif structures themselves can vary in different ways: in loop size, the number of stacked $C:C^+$ base pairs that comprise the core of the structure, and in the number of separate nucleic acid strands involved in their formation, including intermolecular (dimeric or tetrameric) or intramolecular (single-stranded) structures [10–12]. i-Motif-forming sequences are prevalent throughout the human genome, but are more frequently found within telomeres [13,14], centromeres [10], and within gene promoter regions [15–19]. Accumulating evidence suggests that these four stranded structures are involved in a variety of biological processes including transcription [20–22], DNA repair [23], oncogene expression [16,17,24], and telomere maintenance [25,26], contributing to interest in these structures as possible therapeutic targets. i-Motif structures have emerged as transcriptional regulators for several oncogenes including *c-MYC* [15,24], *BCL-2* [16,20], *PDGFR-β* [27], *KRAS* [17], *VEGF* [11], *RET* [1], and *Rb* [28]. The link between the presence of i-motif-forming sequences within oncogenes provides a rationale for novel therapies targeting i-motifs for gene regulation in cancer.

It is therefore important to understand the conditions and ligands that promote or disrupt i-motif secondary structure formation, that may in turn be able to alter gene expression patterns for a range of therapeutic applications. DNA secondary structure formation and stability have been linked to cytosine base modifications (*e.g.*, epigenetic marks such as methylation). Furthermore, since i-motifs and G-quadruplexes often have the potential to form on opposing strands at the same genomic locus, the interdependency of these two species is also an important consideration when targeting these structures.

Here, we review cellular aspects that influence the stability and formation of i-motifs, including environmental conditions and epigenetic modifications (Fig. 1d), as well as their relationship with G-quadruplexes. We also discuss current bioinformatic tools used to predict i-motif formation throughout the human genome and review the ligands currently available to modulate i-motif structure.

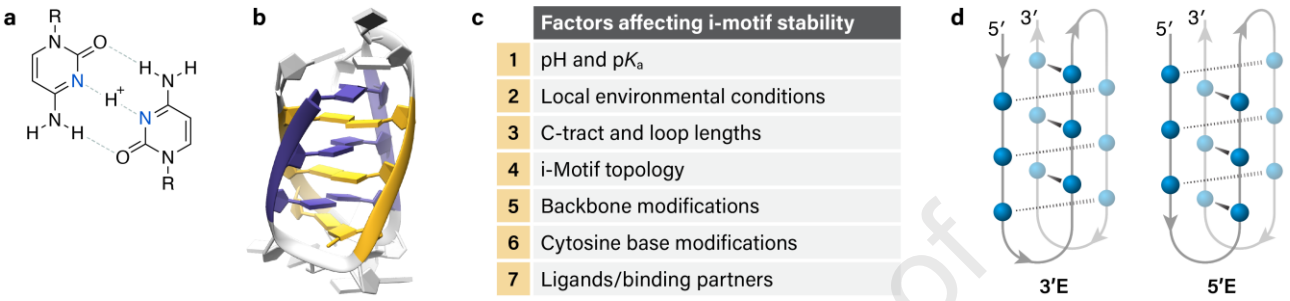


Figure 1. i-Motifs and factors to consider when investigating i-motif stability. (a) Hemi-protonated cytosine–cytosine base pairing shown, highlighting the N3 (blue) protonation of cytosines (centre). **(b)** Intercalated cytosine-cytosine base pairs in a CCCTAA telomeric repeat, NMR structure (PDB: 1ELN) from Phan *et al.* (2000); alternating cytosine–cytosine pairs are shown in dark blue and yellow. **(c)** Factors that impact i-motif formation and stabilisation. **(d)** Comparison of 3′E and 5′E topologies in schematic form.

1.1. Structural and environmental factors affecting i-motif stability

1.1.1. pH

i-Motifs are strongly responsive to pH, even within the cellular environment [29]. It has now been established that i-motifs can form at physiological pH [22,30] and within cells [31–33], overcoming the initial thought that formation could only occur under acidic conditions [6,34]. As pH is such an important defining condition for i-motif formation, one measure of stability is the transitional pH (pH_t). This is defined to be the pH at which the sequence is 50% folded i-motif. Owing to the hemi-protonated base pairing requirement, the stability of an i-motif peaks at the pK_a of the cytosines within the sequence, since C:C$^+$ base pairs require only one cytosine to be protonated. When pH >> pK_a, both cytosines will be deprotonated, and for pH << pK_a, both cytosines will be protonated, each resulting in disruption of the structure [35,36]. Although the pK_a of free cytosine is 4.6, the sequence in which cytosine bases are connected and also the surrounding environment affects the pK_a of individual cytosines within an i-motif. This is the reason why some i-motif-forming sequences are more stable than others. By virtue of the responsiveness of i-motifs to pH, a significant proportion of i-motif research focuses on their use in nanotechnology [37]. The ability of i-motifs to fold and unfold in response to environmental pH changes has enabled the structure to be used for pH sensing [38,29,39], and nanoscale assembly [40,41], which is addressed in Section 5 of this review.

1.1.2. Local environmental conditions

In addition to pH, local environmental conditions such as molecular crowding [42,43], negative superhelicity [34], temperature and the presence of certain cations [44–47] can also alter i-motif stability. These intrinsic and extrinsic factors may override the proposed requirement for acidic conditions and are covered in more detail by Abou Assi *et al.* [48].

1.1.3. C-tract and loop length

i-Motif stability is significantly affected by C-tract length [22,49] as well as loop length and composition [16,49–51]. For sequences of the type $(C_nT_3)C_n$, longer C-tracts are more stable [22], implying that the pK_a of the cytosines increases as the C:C^+ stack height also increases, although for n > 6 there were possibly multiple structures forming and the subsequent increases in melting temperature (T_m) with each additional cytosine within a tract were smaller [49]. Although within this work it was proposed that the two species may be an i-motif and a hairpin structure, it is also possible that once a sequence gets to a certain length there is an equilibrium between a single long i-motif structure and two "beads-on-a-string" i-motifs side by side [49]. In 2011, Hurley and co-workers classified i-motif structures into two distinct classes based on loop length. Class I structures were described to consist of a tetraplex with short intervening loop sizes of 5′-(2:3-4:2)-3′ between the adjacent C-tracts, examples of which are found within the *VEGF*, *RET* and *Rb* oncogene promoter sequences, with experimentally confirmed pH_t values of 5.8–6.4 [1,11,28]. Class II consisted of structures containing longer loop sizes of 5′-(6-8:2-5:6-7)-3′ between the adjacent C-tracts, examples of which may be found within *c-MYC, BCL-2* and *PDGFA* oncogene promoter sequences, all of which display a more physiologically relevant pH_t of 6.6 [15,16,32]. Since this assessment, however, there have been further examples to indicate that i-motif formation and stability is complex. Long loops were initially thought to be inherently more stable, based on the comparison between the published i-motif-forming sequences at the time, but after this several other studies indicated that long loops do not result in more stable i-motifs as a general rule. Brazier and co-workers showed in 2012 that the i-motif structure formed by a HIF-1α promoter sequence shows unexpected stability near neutral pH, despite meeting the requirements to be Class I. Further work in a systematic study of loop length showed that shorter loops are actually more stable in general [50]. Long loops, however, offer the opportunity for additional stabilising interactions (*e.g.*, additional base pairs) that can contribute to overall stability without directly affecting the i-motif [50]. There are now multiple examples of short-looped i-motifs that are stable at neutral pH [22,52]. In comparison to G-quadruplexes, it is clear that sequence requirements for stable i-motif formation are more complicated. Whereas long loops in G-quadruplexes are generally less stable, this is not always the case for i-motifs. As i-motifs are inherently less stable than G-quadruplexes, they are influenced more

by additional stabilising interactions, such as those within loop regions. Further structural details from NMR and crystallography would help in understanding and prediction of loop sequences that confer additional stability.

1.1.4. Topology

Along with loop size, i-motif structures can differ in topology depending on the arrangement of C:C^+ base pairs (Fig. 1d). i-Motifs can be described as having either 3′E intercalation topology, consisting of a terminal C:C^+ base pair at the 3′ end, or a 5′E intercalation topology, consisting of a terminal C:C^+ base pair at the 5′ end [53]. These two different topologies have different stabilities depending on the sequence. The balance of repulsive forces between phosphate groups and attractive forces between sugars determines which conformation is more stable [54]. In the case of human telomeric sequences such as hTeloC, the 3′E conformation is kinetically preferred due to increased sugar–sugar contacts [48] but may slowly refold into a thermodynamically more stable 5′E conformation over time [53,55]. This kinetic partitioning is an important consideration that must be taken into account when considering i-motif stability and response to environmental changes. It is also important to be consistent with handling i-motif-forming sequences, in annealing procedures for example [55], to ensure results are consistent and comparable.

1.1.5. Backbone modifications

One important factor that can dramatically alter i-motif stability is the use of an artificially modified phosphate–sugar backbone. For example, Mergny and Lacroix investigated phosphorothioate and methylphosphonate modifications; phosphorothioate modifications in the backbone were slightly destabilising towards i-motifs, whereas neutral methylphosphonate-modified oligonucleotides were unable to form i-motifs at all [56]. Pérez-Rentero *et al.* reported a negative effect on the stability of i-motifs using threoninol-modified bases [57]. The very close backbone contacts in the narrow i-motif minor grooves (3.1 Å, just over half the distance as that found in B-DNA [58]) may not be able to tolerate significant modification. Damha and González replaced the sugar in DNA by 2-deoxy-2-fluoroarabinose and showed that such a modification significantly stabilizes i-motif formation over a wide pH range, including pH 7 [59]. The modification alters the conformation of the sugar, from C3′-endo to C2′-endo, but does not alter the overall i-motif structure. This combined with the electronegative fluorine atoms, leads to a number of favorable sequential and inter-strand electrostatic interactions. Backbone modifications are often used in therapeutic oligonucleotides to protect against enzymatic degradation *in vivo*, extending the therapeutic effect [60]. If such an approach is to be used for i-motif-forming therapeutic oligonucleotides, careful consideration should be given to the nature of the backbone used.

1.1.6. Base modifications

Cytosine bases can undergo a number of modifications *in vivo* as well as the plethora of synthetic modifications possible for sequences used in nanotechnology applications. Modification of cytosine bases involved in i-motif formation affects the stability of the resulting structure. The relationship between stability and the number and arrangement of these modifications is complex, and is covered in detail in the following section.

2. Epigenetic modifications of cytosine

Stability of i-motif structures is not only influenced by the abovementioned conditions but can also be affected by DNA base modifications. Cytosine base modifications including methylation may be exploited to increase the stability of i-motif structures in a variety of pH environments [26], by altering the pH_t and/or T_m. Epigenetic modifications such as methylation are of particular interest because there is a natural tendency for non-canonical structures to form within guanine- and cytosine-rich regulatory regions within the genome, which are common substrates for epigenetic modification.

Within mammalian genomes, DNA methylation occurs predominantly *via* the modification of cytosine residues. Methylation is a reversible event that occurs *via* the addition of a methyl group (CH_3) to C5 of a cytosine residue, converting cytosine to 5-methylcytosine (5mC, Fig. 2) [61]. This process is carried out by DNA methyltransferase (DNMT) enzymes using a methyl group donated by *S*-adenosylmethionine (SAM). The initial steps in DNA demethylation are carried out by ten-eleven translocation enzymes (TET), converting 5mC in a stepwise manner to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC) [62,63] (Fig. 2). These cytosine base modifications are of particular interest when investigating i-motif formation and stability due to their different structures and sizes compared to unmodified cytosine. In the following sections, we will discuss such modifications and their effect on i-motif formation separately.

Journal Pre-proof

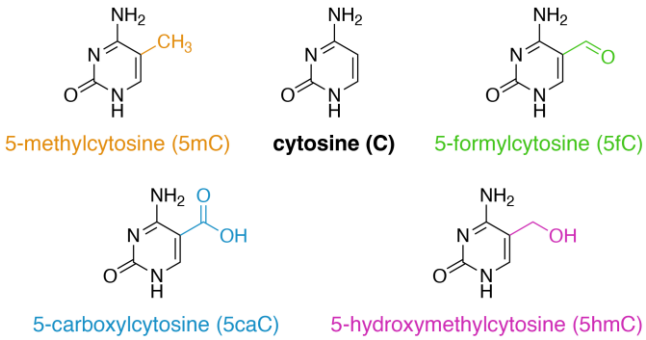


Figure 2. Types of epigenetic modifications to cytosine. Cytosine (C) is methylated to 5-methylcytosine (5mC) by DNA methyltransferases (DNMT) and then may be further oxidised to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC).

2.1. 5-Methylcytosine

DNA methylation in mammalian cells occurs mostly at CpG dinucleotides, which cluster together in CpG islands and are highly prevalent within gene promoter regions [64]. Generally, CpG islands are unmethylated within somatic cells, but may accumulate methylation within abnormal cell types where this hypermethylation state is commonly linked with a decrease in transcriptional activity [64–66]. Regions within CpG islands have a high propensity to fold into non-canonical structures due to the high GC content; in general, it appears that 5mC favours the transition from B-DNA to non-canonical structures by increasing the melting temperature and magnitude of the free energy change upon secondary structure formation (ΔG) within ssDNA [67,68].

It has been shown that 5mC modifications stabilise the *c-MYC* promoter i-motif-forming sequence (Fig. 3a). Bhavsar-Jog *et al.* found that 5mC modifications shifted the pK_a of a synthetic c-Myc i-motif-forming oligonucleotide towards physiological pH (pK_a = 6.3 ± 0.1) compared to that of the unmethylated counterpart (pK_a = 6.1 ± 0.1) while the melting temperatures remained the same within error at 43.7 ± 0.3 °C and 42.5 ± 0.9 °C, respectively [42]. Since hypermethylation of gene promoters is a hallmark of cancer, the effect of 5mC on stability of i-motifs could have implications when studying transcriptional changes in tumours [69].

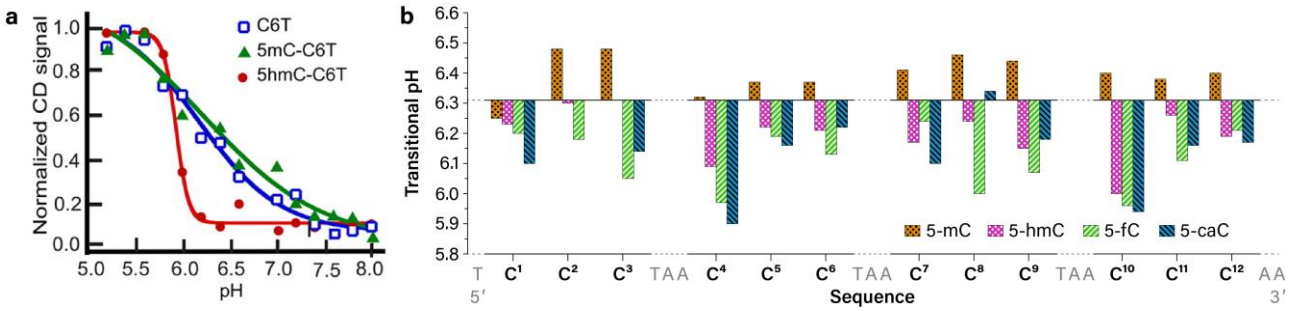


Figure 3. Impact of 5mC and 5hmC modifications on pH-dependent i-motif formation. (a) pH melting curve demonstrating a reduction in transitional pH and significantly increased cooperativity for the 5hmC-modified c-Myc i-motif compared to the unmodified and 5mC modified sequence. **(b)** transitional pH of modified hTelo imotif-forming oligonucleotides compared to unmodified sequence (black line). Subfigure (a) from Bhavsar-Jog *et al.* (2014) and (b) adapted using data from Wright *et al.* (2020).

Because the methyl group of 5mC decreases polarity and increases local rigidity compared to unmethylated cytosine, 5mC-modified i-motifs are slightly more difficult to thermally unfold [42,70]. The number of 5mC modifications within an i-motif-forming sequence has also been shown to be important, with hypermethylation of i-motif-forming sequences destabilising certain i-motif structures. Xu *et al.* demonstrated that 5mC modification of either single or paired cytosine bases stabilised the human telomeric (hTelo) i-motif-forming sequence, resulting in an increase in thermal stability compared to the unmethylated sequence. Hypermethylation (triple 5mC-modified oligonucleotides), on the other hand, destabilised the structure and disrupted i-motif formation, which was attributed to steric hindrance [70].

It has been demonstrated that the position of the 5mC modification is also critical for the stability of the hTelo i-motif structure. Wright *et al.* showed that a singular 5mC modification at ten of the twelve cytosine positions in the hTelo i-motif sequence resulted in an increase in pH_t, while 5mC modification at C1 resulted in a decrease in pH_t and C4 resulted in no change [71] (Fig. 3b). The same group also demonstrated that for the majority of 5mC-modified hTelo sequences, the T_m was increased compared to the unmodified sequence. They further compared 5mC patterns of i-motif-forming sequences within a human breast cancer cell line (MCF7) and a non-tumorigenic mammary epithelial cell line (MCF10A). Of the 44 different i-motif sequences analysed, 27% contained at least one 5mC, and of these, 83% were differentially methylated in MCF7 *vs* MCF10A, highlighting the differences in 5mC methylation between cancer and non-cancerous cell types [71].

Školáková et al. have recently reported that the position of 5mC modifications influences i-motif stability in the *Arabidopsis thaliana* telomeric sequence $(CCCTAAA)_n$ [72]. This work provides insight into the complexity on the impact of 5mC modifications on i-motif formation and stability.

This work identified evolutionally conserved 5mC patterns, including some that did not appear to affect the stability of the i-motif. The 5mC positional effect was best highlighted within asymmetric 5mC modification patterns ($5mC^+$:C base pairs involved in intercalation).

There is emerging evidence that the relationship between G-quadruplex/i-motif structures and 5mC may be bi-directional. Not only can 5mC modifications impact the formation and stability of both G-quadruplex [73,74] and i-motif structures [42,71,70], but recently it has been determined that G-quadruplex structures can bind and inhibit the activity of DNMT1, inhibiting DNA methylation [75]. It is theorised that G-quadruplexes help to regulate gene expression by protecting certain DNA sequences from methylation [75].

2.2. 5-Hydroxymethylcytosine

5mC modified bases can be oxidised to 5hmC by the TET family of oxygenases. Originally thought to exist as an intermediate within the step-wise demethylation pathway, accumulating evidence has demonstrated that 5hmC is a stable epigenetic regulator which is implicated in a range of biological processes and diseases [76–79]. Whereas 5mC tends to inhibit transcription, 5hmC modifications enhance transcriptional activity [78,80,81].

5hmC modification within the c-Myc i-motif-forming sequence is associated with a destabilising effect on the c-Myc i-motif. Bhavsar-Jog *et al.* demonstrated that 5hmC modification within a c-Myc i-motif-forming oligonucleotide slightly lowered the T_m of the i-motif (40.5 ± 0.6 °C) compared to the unmodified oligonucleotide (42.5 ± 0.9 °C) [42]. 5hmC modification makes protonation at the N3 position less favourable, possibly due to the larger molecular size, which imparts steric hindrance and increased hydrophilicity compared to 5mC or an unmodified cytosine. The increased hydrophilicity of 5hmC makes it more favourable for water molecules to interact with ssDNA [42,70], which in turn decreases the pK_a [42] (Fig. 3a). Interestingly, the 5hmC modification substantially increases the cooperativity but it is not currently clear why this should be the case for the c-Myc i-motif, as it is not observed for other sequences such as hTeloC, MSMO1, and PCLB2 [71]. We suggest that the minimal structure (short C-tracts and loops) of the c-Myc i-motif (C6T) may amplify the destabilising effect of 5hmC in this case, but study of more examples of the effects in other i-motif-forming sequences are required.

Unlike 5mC modifications, in which stabilisation is tightly linked to the location of the modification, Xu *et al.* show that 5hmC modifications at different positions throughout the hTelo i-motif sequence resulted in similar levels of destabilisation [70]. In contrast, Wright *et al.* demonstrated that a single

modification of the hTelo i-motif sequence at cytosine position C4 or C10 had the largest change in i-motif stability, regardless of modification type (5mC, 5hmC, 5fC or 5caC) (Fig. 3b). Thus, not only do epigenetic modifications impact i-motif stability, but the position of these modifications play an integral role [64]. Specifically, positions C4 and C10 are the cytosines involved in the top $C:C^+$ base pairing in the i-motif stack, so it is understandable how these will be most sensitive to modification.

While multiple studies have analysed the effects of 5mC and 5hmC *in vitro* using synthetic oligonucleotides [42,70,71], it is important to consider biologically relevant epigenetic modification patterns in the human genome. The prevalence of i-motif-forming sequences 1 kb upstream and downstream of global transcription start sites (TSS) coincide with CpG islands [42,82,83]. Interrogation of 5hmC using TET-assisted bisulfite sequencing, compared with putative i-motif structures using the Quadfinder computational prediction tool, revealed a weak negative correlation of 5hmC density with potential i-motif formation [83]. Furthermore, only a small fraction (15%) of putative i-motif regions contained more than one 5hmC modification. Further analysis revealed that within human embryonic stem cells, sequences that possess three or more 5hmC and colocalise with i-motif-forming sequences are predominantly involved in ligand binding and enzyme regulatory activities [83].

2.3. 5-Formylcytosine and 5-carboxylcytosine

When compared to 5mC and 5hmC, less work has been done on the impact of 5fC and 5caC modifications, which are avenues worthy of further investigation. 5fC and 5caC modifications are larger, more bulky modifications compared to 5mC, and may exhibit greater steric effects. 5fC is implicated in the active demethylation pathway mediated by thymine DNA glycosylase, which can excise both 5fC and 5caC [84]. Given the growing evidence that 5fC and 5caC modifications have biological relevance and are not simply intermediates, research into these base modifications is important [79,85,86].

Previous studies determined that 5fC modifications affect the double helix by leading to a distinct conformation characterised by helical under-winding, impacting local supercoiling and chromatin packaging [87]. Considering the association of i-motif formation and negative supercoiling, under-winding as caused by 5fC modifications may result in the inability of these structures to be stabilised but will need further experimental validation. Previous research has demonstrated that hydrogen bonding at cytosine N3 is weakened by both 5fC and 5caC modifications, reducing base pairing ability [88]; this might be expected to also result in a decrease in i-motif stability and formation at CpG sites containing 5fC and 5caC.

Wright *et al.* determined that 5fC modification at positions C4 or C10 of the hTelo i-motif sequence decreased the melting temperature of the i-motif by 8.3 and 10.3 °C respectively when compared to the unmodified i-motif which displayed a melting temperature of 39.7 °C, reflecting a decrease in stability [71]. In contrast, 5caC modification at position C2 resulted in an increase in thermal stability of the hTelo i-motif oligonucleotide with an increase in melting temperature by 1.5 °C [71]. 5fC and 5caC modifications alter electron density of the cytosines, as these functional groups are electron-withdrawing. When part of i-motif structures, 5fC and 5caC modified cytosines are not able to stabilize the positive charge in the hemi-protonated base pairs as effectively as an unmodified cytosine or 5mC nucleotide [71].

More recently, Ashwood *et al.* have shown that although 5caC modifications tend to destabilise i-motifs, 5caC modifications significantly reduce the free energy of hybridisation transition states [89]. 5caC modification decreases the pK_a of N3, which destabilises the non-canonical C:C$^+$ pairing essential for i-motif formation. In contrast, for canonical C:G base-pairing interactions, 5caC modification increases the stability of duplex DNA. 5caC and other substituents that modify the properties of N3 may therefore provide a useful strategy for tuning the kinetics and mechanism of i-motif folding.

2.4. Summary

The prevalence of sequences that have the potential to fold into i-motif structures within CpG-rich regions of the genome – the substrate for 5mC, 5hmC, 5fC and 5caC modification – has led to increasing research into the relationship between epigenetic modifications and DNA secondary structures. i-Motifs and DNA methylation are independently well-established mechanisms that regulate gene expression; further research into the effects of methylation on stability may identify new opportunities that use epigenetic modifications to regulate i-motif formation. Overall, 5mC modifications tend to stabilise i-motif structures, except in the case of hypermethylation, where steric hindrance may become problematic. In contrast, the other forms of cytosine methylation (including 5hmC, 5fC and 5caC) tend to destabilise i-motifs. However, the effects of epigenetic modifications on i-motif stability depend also on sequence, and the position of such modifications within the folded structure. To date, relatively few i-motif structures have been thoroughly investigated for the effects of epigenetic modification on stability. It would also be interesting to map and compare methylation states and i-motif formation sites on a wider scale across the genome to gain insight into the complex relationship between methylation and i-motif stability.

While the majority of methylated cytosines within the human genome are found at CpG dinucleotides, it is important to acknowledge that non-CpG methylation (CpH, where H = A, T or C) can also occur 90–93]. The propensity for i-motifs to form in repetitive C-rich DNA stretches suggests that CpC methylation may also be of interest when investigating i-motif stability. CpC methylation has been observed in particular cell types including neurons and glia, and utilises the alternative enzymes DNMT3A and DNMT3B, which have been associated with learning and memory [90,91].

Epigenetic modifications and their occurrence in a genomic context need to be considered when predicting structures, as well as the effect that aberrant methylation profiles may have on i-motif stability. The presence of 5hmC and 5mC modifications is thought to modulate the binding and recognition abilities of transcription factors and other biologically relevant proteins. Sprujit *et al.* determined that 5hmC and subsequent TET-mediated modifications (5fC and 5caC) were recognised by different proteins within a mouse embryonic stem cell line [94]. For G-quadruplexes, it has been shown that 5hmC modifications within the *VEGF* promoter alters that sequence's ability to be recognised by nucleolin, which facilitates transcription [18]. Thus, epigenetic differences, whether associated with disease or ageing, could have a considerable effect on the stability of non-canonical secondary nucleic acid structures, with downstream changes in transcription and other cellular processes.

3. Prediction of i-motif-forming sequences

To fully understand the biological function of i-motif structures, it is important to be able to examine their abundance and distribution within the human genome. Whilst several computational prediction algorithms have been developed to predict sites at which G-quadruplex structures may form, including Quadparser [94], G4Hunter [95], and QGRS Mapper [96], tools for predicting the formation of i-motifs are less extensive.

Early algorithms for predicting i-motifs were heavily based on G-quadruplex algorithms, owing to the complementary nature of the two tetraplex structures. Since then, Quadparser has been adapted to search for the prevalence of potential i-motif sequences, taking into account the requirement of the longer tract length required for i-motif formation than for a G-quadruplex [22]. Sequences with at least five cytosines per tract have been demonstrated to fold into an i-motif at room temperature and neutral pH [22]; these findings resulted in a postulated folding rule with the predictive sequence being $C_5(N_{1\text{-}19}C_5)_3$ where 'N' represents any nucleotide including C, with intervening loop lengths of 1-19 nucleotides. Using this algorithm, a total of 5,125 sequences with i-motif-forming potential were identified in the human genome, with 12.4% of these putative i-motif-forming sequences being

located within promoter regions suggesting that i-motifs are not randomly distributed throughout the human genome [9]. However, it should be noted that this model is limited by constraints placed on tract length and folding rules. The aim of this study was to identify sequences that would fold at neutral pH, so does not detect some well-established i-motif sequences, such as hTelo, which do not conform to this pattern and are not stable under neutral conditions. Using these simple algorithms it has been shown that regulatory regions, and particularly gene promoters, are heavily enriched in i-motifs [92,94,97,98].

Bhavsar-Jog *et al.* also highlighted the high abundance of putative i-motif-forming sequences within 100 bp of promoters of 15,760 reference genes from the human GRCh37.p10 primary assembly [83]. i-Motif-forming sequences were found to be enriched within genes involved in skeletal system development, DNA binding, DNA templated transcription, and catalytic activity, and depleted in genes involved with immune responses and G-protein coupled receptor activity [22,83].

i-Motif structures are not limited to the human genome; their complementary nature to G-quadruplex structures is supportive evidence that these structures are present within chimpanzee, rat and mouse genomes [99]. While limited studies have probed for i-motif structures individually, Niu *et al.* used Quadparser to probe the *Bombyx mori* (silkworm) genome [91]. The results demonstrated 3,104 putative i-motif sequences, with 210 being within promoter regions [100]. This implies that i-motif structures are frequent within other genomes. Of note, the same group showed that the i-motif within the *BmPOUM2* promoter bound BmILF to activate transcription [100].

G4-iM Grinder is the latest advancement in the search for sequences with physiological i-motif-forming potential [19]. This R package uses G-quadruplex searching algorithms and allows for putative i-motif sequences to be predicted based on GC content, tetrad, bulge, and loop sizes; large negative scores indicate high potential for i-motif formation, and large positive scores indicate high potential for G-quadruplex formation. Initial tests with G4-iM Grinder involved predictions for 95 known i-motif sequences. Strong correlation was found between Grinder test scores and known *in vitro* i-motif formation from past biophysical studies, revealing a mean score of -55.8 ± 13.6 when searching for i-motif sequences, similar to G-quadruplex formation probability 52.9 ± 13.1 [19].

Interestingly, G4-iM Grinder reveals that chromosome 19 had the highest i-motif density (number of putative i-motifs per 100,000 nucleotides) of all human chromosomes, followed by chromosomes 22 and 17, while chromosomes 13 and Y had the lowest densities [19] (Fig. 4). In this study, G4-iM Grinder was also employed to analyse genomes of species that cause pathologies in humans, including a range of viruses, fungi, bacteria, and parasites. These results demonstrated that putative

i-motif sites were notably dense in Rubella and Epstein-Barr viruses and also in *Leishmania* parasites, whereas other parasites and fungi had lower densities of predicted non-canonical structures within their genomes compared to the human average [19].

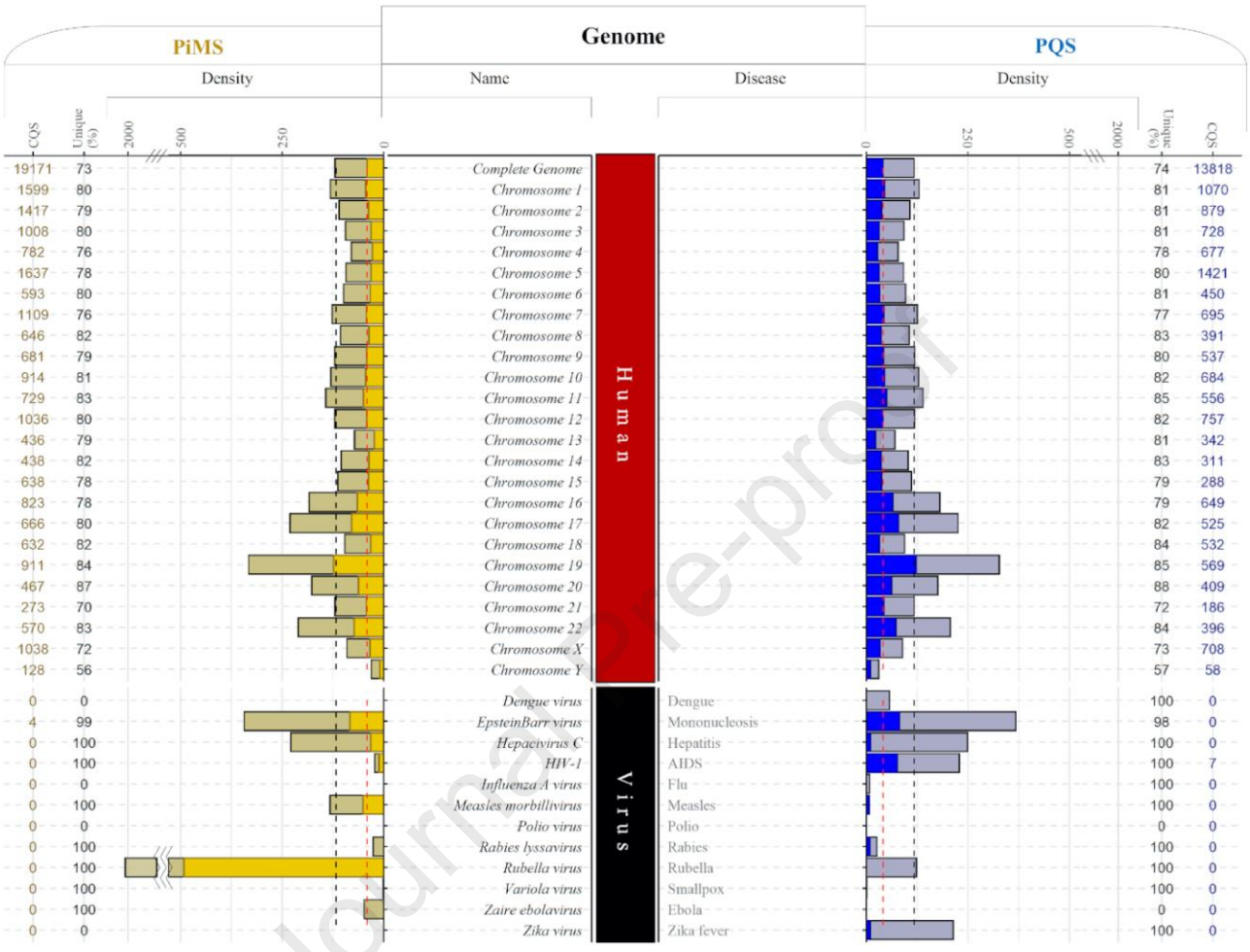


Figure 4. G4–iM Grinder results (method 2A) in the human genome and disease-causing virus genomes. Putative i-motif sequences (PiMS) are displayed on left side, while PQS (putative quadruplex-forming sequences) are displayed on the right. Density is per 100,000 nucleotides; unique (%) is the percentage of unique sequences; CQS is number of results with confirmed-to-form quadruplex sequences). Genomic densities are shown as coloured bars (yellow for PiMS that score -40 and less; blue for PQS that score at least 40). Grey bars are the unfiltered genomic densities. Black dotted line shows the human average density, and red dotted line shows the human average density that scores at least 40 for PQS or -40 for PiMS. Figure adapted from Belmonte-Reche *et al.* (2020).

Both Quadparser and G4-iM Grinder are essential computational tools to enable the understanding of where i-motif structures have the propensity to form genome-wide. Both tools have identified i-motif sequences within regulatory regions of the human genome, further supporting their role in gene expression. Prediction algorithms have identified i-motif structures within regulatory regions, but experimental evidence (*e.g.*, chromatin immunoprecipitation) is essential to accurately map i-motif

formation throughout the genome. Current tools are limited somewhat by restricted search patterns, which can be overcome by machine-learning implementations (*e.g.,* Quadron, G4NN) [101,102]. It is important to note, however, that current prediction tools do not take into account environmental parameters (pH, ionic strength) or methylation, nor any competitive secondary structure formation (*i.e.,* G-quadruplex and i-motif) which may occur at the same genomic locus on opposing strands, which is the subject of the following section. Until more is known about the precise types of loop sequences and additional interactions that stabilise (and destabilise) i-motif structures, the ability to predict i-motif formation from sequence will inevitably be limited. Identification of types of stabilising loop interactions in i-motif-forming sequences is an important step towards accurate i-motif predictions.

4. Relationship between i-motifs and G-quadruplexes

G-quadruplex and i-motif sequences may often be found on complementary strands at a given locus. While i-motif formation occurs in strands complementary to G-quadruplex forming sequences, there is increasing evidence suggesting that these tetraplex structures do not form concomitantly, but rather the formation of one structure precludes formation of the other. The development of structure-specific scFv antibodies for G-quadruplexes (BG4, 2013) [103] and i-motifs (iMab, 2018) [31] have allowed the visualisation and quantification of G-quadruplexes and i-motifs in cells. G-quadruplexes recognised by BG4 were greatest in number during DNA replication (S phase of the cell cycle) [103] whereas i-motifs were most prominent during high levels of transcription (late G1 phase) [31]. Taken together, these studies suggest that the two structures may form independently of each other.

The first genomic loci supporting the notion that G-quadruplex and i-motif formation on complementary DNA strands may be mutually exclusive was the insulin-linked polymorphic region (ILPR) [104]. Since then, studies combining the use of molecular tweezers and mechanical/chemical foot-printing have demonstrated numerous genomic loci that have mutually exclusive G-quadruplex and i-motif structures, many of which owe this exclusivity to steric hindrance [104,105].

In 2016, Cui *et al.* predicted the folding probabilities of G-quadruplex and i-motifs in dsDNA using force ramp-up assays [105]. i-Motif and G-quadruplex structures in duplex B-DNA fully unfold under mechanical force up to 60 pN [104]. The authors used force-ramp assays under conditions that encouraged both G-quadruplex and i-motif formation and a population analysis technique to assess the simultaneous folding and unfolding events of tetraplexes. This investigation into the formation of G-quadruplex and i-motif utilized a technique which calculated the probability of tetraplex unfolding within a variety of buffers that promote formation of i-motif only (100 mM Li^+, pH 5.5), G-

quadruplex only (100 mM K^+, pH 7.4) or both (100 mM K^+, pH 5.5) [105]. A general pattern was observed in complementary strands of G-quadruplex- and i-motif-forming sequences including hTelo, *hTERT*, and *hINS*; these sequences displayed tetraplex formation in only one of the complementary strands, supporting the mutually exclusive hypothesis. In 2020, King *et al.* reported that G-quadruplex and i-motif formation are interdependent in human cells, where ligand-induced stabilisation of G-quadruplex structures resulted in a concomitant decrease in i-motif formation and ligand-induced stabilisation of i-motif structures resulted in decreased G-quadruplex formation [106]. When the G-quadruplex and i-motif-forming sequences in the hTelo dsDNA were offset by a certain distance, simultaneous formation of G-quadruplex and i-motif was evident in the complementary strands, confirming that steric hindrance is a major factor which is a consequence of the proximity of the two tetraplexes (Fig. 5) [105]. This effect is more likely within the genome in longer or tandem repeated sequences, as both structures can potentially form adjacent to each other.

Conversely, some studies have demonstrated promoter regions that are able to form G-quadruplex and i-motif structures simultaneously [1]. In 2009, Sun and Hurley, using the knowledge of transcriptionally induced negative superhelicity, investigated i-motif and G-quadruplex formation in the *c-MYC* promoter [107]. Using enzymatic and chemical footprinting they showed that i-motif and G-quadruplex formation was encouraged by negative superhelicity in physiological conditions, and demonstrated that in the *c-MYC* promoter the i-motif and G-quadruplex structures form at the same time in opposite strands with slight displacements [107]. However, later research revealed that i-motif and G-quadruplex formation in the *c-MYC* promoter is mutually exclusive using dual-beam optical tweezers and population-based analysis under conditions that should allow both i-motif and G-quadruplex formation (50 mM MES pH 5.5, 100 mM KCl) [24]. Furthermore, Abou Assi *et al.* used fluorinated sugar backbone modifications to slow down the kinetics of nucleic acid folding. They found that while i-motif and G4 structures may be mutually exclusive in some cases (including gene promoters), there was evidence that i-motifs and G4 structures may co-exist transiently as intermediates in telomeric sequences [108].

The discovery of pseudo-symmetrical G-quadruplex and i-motif structures in the *RET* oncogene promoter region also revealed a stable low energy model in which the *RET* G-quadruplex and i-motif were formed concomitantly [1]. Wolski *et al.* examined the likelihood of simultaneous G-quadruplex and i-motif existence in human telomeric DNA using NAB (nucleic acid builder) language from AmberTools16 package [14]. In neutral and acidic pH conditions the i-motif was stabilised by the presence of the complementary G4, however when the i-motif was triggered into folding, spontaneous formation of the G-quadruplex was not observed. At neutral pH the simultaneous formation of the G-quadruplex and i-motif was preserved, indicating that in cellular environments enzymatic machinery

would be required for their unfolding [14]. They concluded that in *ex vivo* cases both the i-motif and G-quadruplex can form concomitantly without steric hindrance.

As described above, there is evidence supporting both the concomitant and mutually exclusive formation of G-quadruplex and i-motif structures in duplex DNA. However, examples of simultaneous formation require an offset as shown in Fig. 5, or are observed only transiently. The reasons for mutual exclusivity have not yet been fully elucidated. Whether steric hindrance prevents both structures from coexisting remains to be confirmed within cells, where other environmental factors also play a role in regulating G-quadruplex and i-motif formation. Overall, the relative abundance of G-quadruplexes and i-motifs at different cell cycle stages and the effects observed by the addition of stabilising ligands indicates that G-quadruplex and i-motif structures likely do not coexist simultaneously on opposing strands in a cellular context.

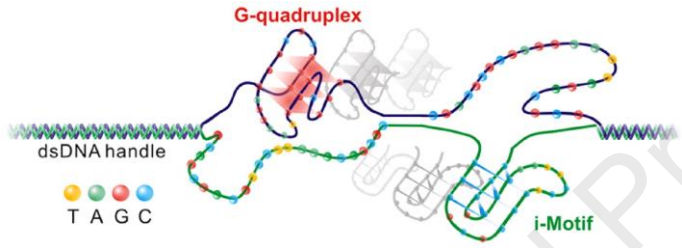


Figure 5. Offset arrangements of i-motif and G-quadruplex sequences in complementary DNA strands enables formation of both structures. Schematic of G-quadruplex and i-motif structures with a spatial offset, allowing both structures to fold. Figure adapted with permission from Cui, Y., Kong, D., Ghimire, 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* 55, 2291–2299 (2016). Copyright 2016 American Chemical Society.

5. Modulating i-motif formation and therapeutic applications

Alternative secondary DNA structure formation provides an additional layer in regulating gene expression. i-Motifs can act as binding sites for proteins, including transcription factors, to facilitate activation of gene expression. The unique structure and biological roles of i-motifs make them attractive potential targets for small molecule binding to modulate transcriptional activity. A recent review by Brown and Kendrick provides a detailed overview of i-motif structures as molecular targets in therapeutic applications [109]. Furthermore, the i-motif's dynamic nature and sensitivity to pH has been used for important DNA nanotechnology purposes.

5.1. Natural i-motif binding proteins

i-Motifs in promoter regions recruit particular nuclear proteins and transcription factors to regulate gene expression [21,24,110]. Poly-C binding proteins (PCBP) interact with C-rich sequences and are involved in regulating transcription and gene expression [110]. Heterogenous nuclear riboprotein K (hnRNP K), a member of the PCBP family, recognises and binds to the i-motif in the *c-MYC* promoter nuclease hypersensitive element (NHE) III_1 to activate *c-MYC* transcription [24]. However, Sutherland *et al.* demonstrated that in order to form a stable complex, a fifth tandem repeat of consensus sequence CCCTCCCC (CT element) must undergo additional torsional stress, provided by specificity protein 1 (SP1) [24]. SP1 binding induces negative superhelicity [111], allowing the i-motif structure to form and provide two binding sites for hnRNP K [24]. In a later study, Kang *et al.* identified ninety-five nuclear proteins from HeLa nuclear extract that bound to either the wild-type *BCL-2* i-motif-forming sequence, a mutant, or both [21]. Of particular interest was the heterogenous nuclear ribonucleoprotein LL (hnRNP LL), which also belongs to the PCBP family. hnRNP LL is related to the hnRNP L protein, a mRNA splicing factor which binds and stabilises *BCL2* mRNA [112]. In humans, hnRNP LL protein has four RNA recognition motifs, two of which bind to the lateral loops *via* sequence recognition of the *BCL2* i-motif, causing the structure to unfold into a stable single-stranded conformation, activating transcription [21].

The nuclear protein hnRNP A1 is overexpressed in many cancers and is involved in mRNA biogenesis. Chromatin immunoprecipitation revealed that hnRNP A1 binds to the guanine/cytosine-rich (GC) element in the *KRAS* and *HRAS* promoters, which contain high sequence and functional homology [113]. Furthermore, hnRNP A1 knockout in T24 bladder cancer cells by shRNA caused a reduction in *HRAS* transcript levels (44 ± 5% of the control), indicating the hnRNP A1 is crucial in activating gene expression [113]. FRET titrations and melting experiments revealed that hnRNP A1 binding causes unfolding of the *HRAS* i-motif and activation of gene expression [113]. This study, along with others involving *c-MYC* [24,114], *BCL-2* [21], *KRAS* [115,116], and *c-KIT* [117], suggest that oncogene expression may be controlled by a G-quadruplex/i-motif molecular switch which is regulated by protein binding.

POU (Pit-Oct-Unc) domain transcription factors have key roles in development across multiple species [118–120]. In insects, such as *B. mori*, they are involved in several processes such as cell proliferation and differentiation in developing wings [121]. The transcription factor BmPOUM2 has a GC-rich promoter region between -237 and +1 nucleotides (relative to the transcription start site), which can adopt i-motif and/or G-quadruplex conformations [100]. Niu *et al.* tested the transcriptional effect of the i-motif using luciferase expression activity assays and found that the i-

motif structure enhanced transcription of BmPOUM2, this study identified the nuclear transcription factor BmILF binds to the i-motif structure with high specificity in *B. mori*, and that the formation of the i-motif structure regulates BmPOUM2 transcriptional activity (Fig. 6) [100]. More recently, i-motif structures were observed *in vivo* in *B. mori* testis by immunofluorescence with the BmILF transcription factor antibody [122]. There was an observable increase in foci under acidic pH and a decrease when cells were treated with an i-motif inhibitor, tetra(*N*-methyl-4-pyridyl)porphyrin (TMPyP4), indicating that i-motif formation influences transcription [122], although the use of TMPyP4 is always complex as it is a promiscuous ligand and can interact with and stabilise/destabilise a number of DNA and RNA secondary structures [123–126].

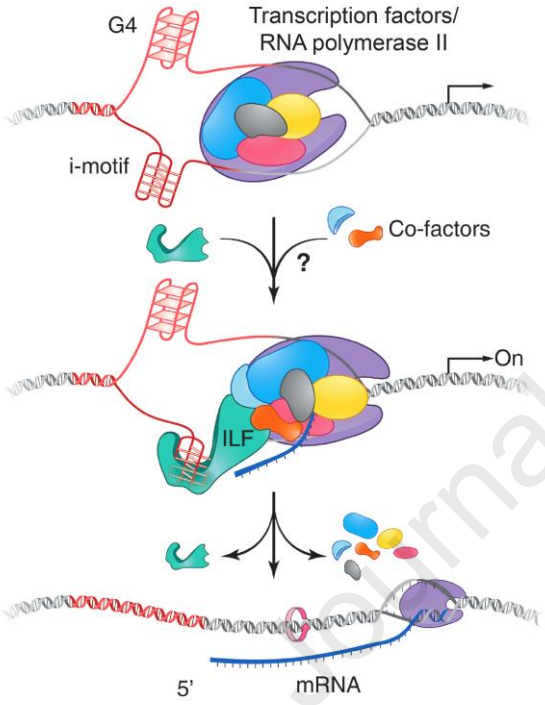


Figure 6. Schematic of epigenetic regulatory mechanism model of BmPOUM2. Transcription factors bind to the core promoter, separating the dsDNA strands, allowing the GC-rich region to fold into G-quadruplex and i-motif structures. The i-motif recruits BmILF, which in turn may recruit other co-factors, for transcription. After transcription the transcription factors dissemble and the G-quadruplex and i-motifs can unfold to form dsDNA. Figure adapted from Niu *et al.* (2018).

With i-motifs enriched in regulatory regions, one of the key challenges is to identify i-motif-protein interactions in detail to characterize an understanding of how i-motif structures influence gene expression and protein function. There are no doubt further native i-motif binding/interacting proteins that remain to be explored, and this constitutes a significant area that could be expanded in the field.

5.2. Natural i-motif ligands

Flavonoids have been identified as a family of natural ligands that bind to i-motifs. A recent study using pH and PEG-200 to mimic cellular crowding conditions that favour *BCL-2* i-motif formation and stability, screened the binding of natural flavonoids to the *BCL-2* i-motif *via* electrospray ionisation mass spectrometry (ESI-MS) to identify potential *BCL-2* i-motif binding ligands [127]. Three flavonoids with phenolic hydroxyl groups were identified with approximate binding constants of 10^4 mol^{-1} L at 1:1 stoichiometry [127]. The plant flavonol, fisetin, binds to the *VEGF*, *MYC*, and hTelo i-motifs resulting in a transition to a hairpin structure, allowing replication to occur [128]. When the *VEGF* i-motif is formed it inhibits DNA polymerase activity, therefore hindering replication and contributing to the progression of diseases, including cancer [128]. The specific binding of flavonoids to i-motif structures offers a novel mechanism to regulate gene expression in a therapeutic context.

5.3. Synthetic i-motif ligands

i-Motif structure and potential binding sites must be known prior to designing i-motif-selective ligands. There are four main modes of interaction for i-motif ligands: intercalation, end stacking, external groove-binding, and binding to the loop regions [37]. The loop regions of i-motifs are not tightly packed with intercalated cytosine pairs, and they form the unique features that differentiate between different i-motifs, thus they are more desirable targets for ligand binding [129]. i-Motif ligands are limited in comparison to the number of G-quadruplex ligands already identified, but a number of ligand families have been reported, including carbon nanotubes [130], mitoxantrone [131], thiazole orange derivatives [132], acridones [133], benzothiophenes [27], and transition metal complexes [129], among others.

In two companion papers, Kang *et al.* and Kendrick *et al.* reported two i-motif binding compounds (IMC-48, an i-motif stabiliser, and IMC-76, an i-motif destabiliser) and showed that by targeting i-motif-forming sequences, the formation and disassociation of i-motifs can significantly influence expression of *BCL-2* [20,21]. Overexpression of anti-apoptotic genes enables cells to evade programmable death, a hallmark of cancer [134]. The authors screened 1,990 compounds from the NCI Diversity Set I library before identifying IMC-48 (NSC 138948), a cholestane derivative, which binds and stabilises the i-motif in the *BCL-2* gene promoter. A pregnanol derivative, IMC-76 (NSC 59276), was found to bind to the hairpin conformation formed in the C-rich sequence of the *BCL-2* promoter region and thus destabilises i-motif formation. The authors used IMC-76 to stabilise the *BCL-2* DNA hairpin, and IMC-48 to bind and stabilise the i-motif central loop to demonstrate for the

first time that targeting i-motif structures with small molecules can modulate transcription [20]. The authors determined that these two molecules can be used to manipulate the equilibrium between i-motif and hairpin structure formation and discovered that there were contrasting implications for *BCL-2* expression. IMC-48 stabilisation of the *BCL-2* i-motif led to an increase in both *BCL-2* mRNA and protein expression, while the i-motif-destabilising agent IMC-76 resulted in reduced *BCL-2* mRNA levels [20,21]. They determined that the lateral loops of the *BCL-2* i-motif were essential for IMC-48 binding affinity and both ligands were specific for the *BCL-2* i-motif-forming sequence, compared to the loop regions from the *c-MYC* i-motif.

Transition metal complexes, such as polypyridyl ruthenium complexes, have high DNA binding affinity and luminescence properties, which make them valuable as therapeutic agents [135–137]. While these complexes have been thoroughly studied for medicinal purposes, in the field of i-motif research they have only just recently received attention. A recent study reported the first interactions between ruthenium complexes and i-motifs with various loop lengths [129]. Using synchrotron radiation circular dichroism (SRCD), UV melting, and luminescence spectroscopy the authors found that ruthenium complexes bind and stabilise i-motifs with long loop sequences, binding to loops 1 or 3 [129]. However, while luminescence spectroscopy indicated that ruthenium emission increased with increasing loop length for shorter loops, ruthenium binding did not stabilise the i-motif structure with shorter loops [129]. This work highlights the potential for transition metal drugs targeting i-motifs with long loop lengths. Since then, Spence *et al.* demonstrated the potential of $[Ru(bqp)_2]^{2+}$ isomers as i-motif binding probes [138]. Using emission lifetime measurements, they show that the cis- isomer of the complex could “light up” and detect the presence of i-motif, even when other DNA structures were present. Moreover, separated cis enantiomers revealed that the Λ-cis analogue prefers the i-motif whereas the Δ-cis prefers double stranded DNA.

In recent years it has become clear that when designing i-motif or G-quadruplex ligands one must also evaluate interactions with their counterparts. Many G-quadruplex ligands that were thought to be specific have recently been showed to also interact with i-motif structures [139,140], with initial testing neglecting i-motif binding as i-motif structures were thought to be biologically irrelevant at the time of publication. Randazzo and co-workers demonstrated that seven commonly used G-quadruplex-binding compounds (berberine, BRACO-19, mitoxantrone, Phen-DC3, pyridostatin, RHPS4, and TMPyP4) act as multi-targeted directed agents by also interacting with the i-motif from the human telomere at acidic pH [140]. Further work by Abdelhamid and Waller showed this also to be true for the i-motif sequences from the promoter regions of the ATXN2L and DAP genes, which are stable at neutral pH [141]. While these compounds stabilise G-quadruplexes, their interaction with i-motif structures primarily leads to i-motif destabilisation [141]. Given that previously regarded

specific G-quadruplex ligands also interact with i-motifs, in order to gain a better understanding of their mechanisms of action and biological consequences, there is a necessity to consider both i-motif and G-quadruplex formation when using such ligands. It also opens up the question about what properties are required for using ligands in biological experiments. For example, a lot of excellent work has been performed using berberine, BRACO-19, mitoxantrone, Phen-DC3, pyridostatin, RHPS4, and TMPyP4 as ligands. There are no perfect ligands, but we must consider the tools we have available and also their limitations when using them in biological experiments.

Only a limited number of i-motif ligands have been tested in cellular systems, indicative of the difficulty in developing selective i-motif ligands. Dash's group recently demonstrated targeted *in situ* cycloaddition using i-motif-linked gold coated magnetic nanoparticles as templates to produce i-motif-specific ligands [133]. Thiolated *c-MYC* and *BCL-2* i-motif sequences were immobilised onto gold-coated magnetic nanoparticles to efficiently isolate and identify triazole ligands that are selective for i-motifs, using the complementary guanine-rich and dsDNA sequences to prepare control templates [133]. Two triazole compounds (**3be** and **3bm**) were identified with high specificity for *c-MYC* and *BCL-2* i-motifs respectively by FRET-melting, fluorescence titrations and competitive binding [133]. Cellular assays demonstrated that ligand binding modulated *c-MYC* and *BCL-2* transcription (Fig. 7). Binding of **3be** to the *c-MYC* i-motif significantly reduced expression without affecting *BCL-2* expression. Likewise, **3bm** upregulated *BCL-2* expression *via* i-motif stabilisation, indicating that a target-guided synthesis approach generates highly potent, target-specific compounds and is a promising strategy for developing novel drug candidates [133].

Despite the recent increase in i-motif binding ligands, there remains only a limited number of compounds that specifically bind to i-motifs in comparison to G-quadruplex ligands. It is understandable that the number of i-motif targeting ligands is limited. Study of i-motifs has previously been restricted considering the pH at which these structures are stable. However, there are now many identified i-motifs that form naturally at neutral pH, as well as conditions that can be used *in vitro* for i-motifs that are less stable. It is also easy to compare i-motif-interacting compounds with G-quadruplex ligands, yet the fields are at significantly different stages with respect to the potency of ligands. The limited number of generic and specific i-motif ligands means that this is an area in need of significant growth. Further understanding into the pharmacophore for i-motif binding and better design of ligands is required to develop this field. These efforts would be supported by increasing the number of i-motif crystal and NMR solution structures, to allow for more *in silico* screening and computer-based drug design strategies. Furthermore, determining the binding mechanism of such ligands will be key in the development of potential therapeutic drugs in this area.

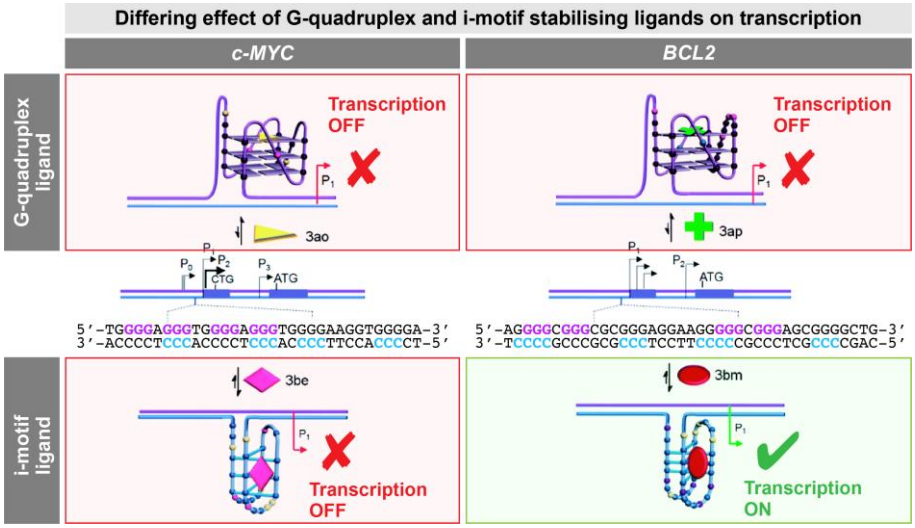


Figure 7. Proposed model for targeting the i-motif structures in the *c-MYC* and *BCL-2* gene promoters to regulate transcription. Figure adapted from Saha P, Panda D, Müller D, *et al.* (2020) *In situ* formation of transcriptional modulators using non-canonical DNA i-motifs, copyright © 2020 Royal Society of Chemistry. Used by permission.

5.4. Biotechnology and therapeutic applications

One of the earliest uses of i-motif structures in biotechnology is the i-motif switch (I-switch), first developed by Krishnan's group to detect pH changes in living cells [38], and later in multicellular organisms [29]. Gold nanoparticles (AuNPs) were later utilized to enhance I-switch monitoring of pH changes [142]. Along with monitoring pH, i-motifs have also been used to control DNA nanostructure assemblies, drug release platforms, and biosensors in the nanotechnology field [37,143–147].

Kendrick *et al.* simultaneously targeted the *c-MYC* G-quadruplex and the *BCL-2* i-motif promoter structures to modulate gene expression and subsequent protein expression which contribute to the development of cancerous cells [148]. This novel dual targeting demonstrated promising therapeutic results with the downregulation of oncogene expression and enhanced sensitivity to cyclophosphamide, a chemotherapy drug [148].

Platelet-derived growth factor receptor β (PDGFR-β) expression is essential to maintaining homeostasis in adults, with defects or high levels of PDGFR-β signalling contributing to numerous diseases including cancer [149,150]. Through fluorescence resonance energy transfer (FRET) screening a benzothiophene-2-carboxamide (NSC309874) compound was identified to bind selectively to the PDGFR-β i-motif. This compound targets the i-motif resulting in downregulation of gene expression, contrary to the i-motif roles observed in other promoter regions such as *c-MYC*

and *BCL2* [27]. It is hypothesised that the binding of benzothiophene-2-carboxamide destabilises hnRNP K binding, preventing PDGFR-β transcription [27].

The area of i-motif based therapeutics is in its infancy and there is scope for significant development in this area both through targeting the i-motif-forming sequences in the genome and using i-motif based nanotechnologies for the delivery of drugs. The study of modified backbones and bases will naturally influence this area and such a toolkit could provide users with exquisite capabilities of fine-tuning i-motif stability if required.

6. Future challenges and conclusion

Although work on i-motifs is increasing, there remain significant knowledge gaps. One of the major challenges in enhancing the potential of i-motifs as therapeutic targets is to identify which i-motifs form in living cells and develop an understanding of factors that regulate their formation. These questions could be addressed by developing genome-wide maps of i-motif formation sites with methods such as chromatin immunoprecipitation sequencing (ChIP-seq). The development of imaging probes including ligands with validated specificity for i-motifs would enable live cell imaging to study the dynamics of these structures within the cellular context. Both approaches have been used to map and visualise G-quadruplexes. Furthermore, uncovering the relationship between i-motifs and G-quadruplexes is crucial to understanding their folding mechanisms and for the development of therapeutic targeting strategies due to their overrepresentation in multiple diseases. Given the occurrence of natural i-motif binding proteins, ChIP-seq methods to identify i-motif/protein binding partners genome-wide could bring forth the identification of endogenous proteins that regulate i-motif formation. Co-crystal or nuclear magnetic resonance structures of protein/i-motif and ligand/i-motif complexes are essential to identify the structural determinants that attract specific proteins/ligands to specific i-motifs.

In recent years the extent and complexity of research into the function of i-motif structures has increased significantly. The i-motif's sensitivity to pH changes has allowed them to be widely used in nanotechnology applications, however, further investigation is required to confirm the biological roles of i-motifs *in vivo*. As we have discussed, many components in cells can affect the formation of i-motif structures, and by extension act as regulators of gene expression. Understanding the relationship between cytosine base modifications and i-motif structures will allow for an alternative level of programmability over these structures. In turn this will give rise to further research into aberrant modifications and subsequent impact on these biologically relevant structures within abnormal cell types. While there is much more to discover about the biological roles of i-motifs, the

current research provides a basis for i-motifs not only as potential therapeutic targets, but also for their use as pH and environment-responsive nanodevices in drug development.

7. References

[1] K. Guo, A. Pourpak, K. Beetz-Rogers, V. Gokhale, D. Sun, L.H. Hurley, Formation of pseudosymmetrical G-quadruplex and i-motif structures in the proximal promoter region of the RET oncogene, J. Am. Chem. Soc. 129 (2007) 10220–10228. https://doi.org/10.1021/ja072185g.

[2] V.S. Chambers, G. Marsico, J.M. Boutell, M. Di Antonio, G.P. Smith, S. Balasubramanian, High-throughput sequencing of DNA G-quadruplex structures in the human genome, Nat. Biotechnol. 33 (2015) 877.

[3] M. Adrian, B. Heddi, A.T. Phan, NMR spectroscopy of G-quadruplexes, Methods. 57 (2012) 11–24. https://doi.org/10.1016/j.ymeth.2012.05.003.

[4] S. Burge, G.N. Parkinson, P. Hazel, A.K. Todd, S. Neidle, Quadruplex DNA: sequence, topology and structure, Nucleic Acids Res. 34 (2006) 5402–5415. https://doi.org/10.1093/nar/gkl655.

[5] G.N. Parkinson, M.P.H. Lee, S. Neidle, Crystal structure of parallel quadruplexes from human telomeric DNA, Nature. 417 (2002) 876–880. https://doi.org/10.1038/nature755.

[6] H.A. Day, P. Pavlou, Z.A.E. Waller, i-Motif DNA: structure, stability and targeting with ligands, Bioorg. Med. Chem. 22 (2014) 4407–4418. https://doi.org/10.1016/j.bmc.2014.05.047.

[7] T.A. Brooks, S. Kendrick, L. Hurley, Making sense of G-quadruplex and i-motif functions in oncogene promoters, FEBS J. 277 (2010) 3459–3469.

[8] M. Guéron, J.-L. Leroy, The i-motif in nucleic acids, Curr. Opin. Struct. Biol. 10 (2000) 326–331.

[9] K. Gehring, J.-L. Leroy, M. Guéron, A tetrameric DNA structure with protonated cytosine-cytosine base pairs, Nature. 363 (1993) 561.

[10] M. Garavís, N. Escaja, V. Gabelica, A. Villasante, C. González, Centromeric Alpha-Satellite DNA Adopts Dimeric i-Motif Structures Capped by AT Hoogsteen Base Pairs, Chem. Eur. J. 21 (2015) 9816–9824.

[11] K. Guo, V. Gokhale, L.H. Hurley, D. Sun, Intramolecularly folded G-quadruplex and i-motif structures in the proximal promoter of the vascular endothelial growth factor gene, Nucleic Acids Res. 36 (2008) 4598–4608. https://doi.org/10.1093/nar/gkn380.

[12] P. Catasti, X. Chen, L.L. Deaven, R.K. Moyzis, E.M. Bradbury, G. Gupta, Cytosine-rich strands of the insulin minisatellite adopt hairpins with intercalated Cytosine+· Cytosine pairs, J. Mol. Biol. 272 (1997) 369–382.

[13] A.T. Phan, M. Guéron, J.-L. Leroy, The solution structure and internal motions of a fragment of the cytidine-rich strand of the human telomere, J. Mol. Biol. 299 (2000) 123–144.

[14] P. Wolski, K. Nieszporek, T. Panczyk, G-Quadruplex and I-Motif Structures within the Telomeric DNA Duplex. A Molecular Dynamics Analysis of Protonation States as Factors Affecting Their Stability, J. Phys. Chem. B. 123 (2018) 468–479. https://doi.org/10.1021/acs.jpcb.8b11547.

[15] J. Dai, E. Hatzakis, L.H. Hurley, D. Yang, I-motif structures formed in the human c-MYC promoter are highly dynamic–insights into sequence redundancy and I-motif stability, PLoS One. 5 (2010) e11647.

[16] S. Kendrick, Y. Akiyama, S.M. Hecht, L.H. Hurley, The i-Motif in the bcl-2 P1 Promoter Forms an Unexpectedly Stable Structure with a Unique 8:5:7 Loop Folding Pattern, 131 (2009) 17667–17676. https://doi.org/10.1021/ja9076292.

[17] C.E. Kaiser, N.A. Van Ert, P. Agrawal, R. Chawla, D. Yang, L.H. Hurley, Insight into the complexity of the i-motif and G-quadruplex DNA structures formed in the KRAS promoter and subsequent drug-induced gene repression, J. Am. Chem. Soc. 139 (2017) 8522–8536.

[18] R.K. Morgan, M.M. Molnar, H. Batra, B. Summerford, R.M. Wadkins, T.A. Brooks, Effects of 5-Hydroxymethylcytosine Epigenetic Modification on the Stability and Molecular Recognition of VEGF i-Motif and G-Quadruplex Structures, J. Nucleic Acids. 2018 (2018).

[19] E. Belmonte-Reche, J.C. Morales, G4-iM Grinder: when size and frequency matter. G-Quadruplex, i-Motif and higher order structure search and analysis tool, NAR Genomics Bioinforma. 2 (2020). https://doi.org/10.1093/nargab/lqz005.

[20] S. Kendrick, H.-J. Kang, M.P. Alam, M.M. Madathil, P. Agrawal, V. Gokhale, D. Yang, S.M. Hecht, L.H. Hurley, 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. 136 (2014) 4161–4171. https://doi.org/10.1021/ja410934b.

[21] H.J. Kang, S. Kendrick, S.M. Hecht, L.H. Hurley, 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. 136 (2014) 4172–85. https://doi.org/10.1021/ja4109352.

[22] E.P. Wright, J.L. Huppert, Z.A.E. Waller, Identification of multiple genomic DNA sequences which form i-motif structures at neutral pH, Nucleic Acids Res. 45 (2017) 2951–2959. https://doi.org/10.1093/nar/gkx090.

[23] R.A. Rogers, A.M. Fleming, C.J. Burrows, Rapid Screen of Potential i-Motif Forming Sequences in DNA Repair Gene Promoters, ACS Omega. 3 (2018) 9630–9635. https://doi.org/10.1021/acsomega.8b01551.

[24] C. Sutherland, Y. Cui, H. Mao, L.H. Hurley, A Mechanosensor Mechanism Controls the G-Quadruplex/i-Motif Molecular Switch in the MYC Promoter NHE III1, J Am Chem Soc. 138 (2016) 14138–14151. https://doi.org/10.1021/jacs.6b09196.

[25] Y. Chen, K. Qu, C. Zhao, L. Wu, J. Ren, J. Wang, X. Qu, Insights into the biomedical effects of carboxylated single-wall carbon nanotubes on telomerase and telomeres, Nat. Commun. 3 (2012) 1074. https://doi.org/10.1038/ncomms2091.

[26] A. Aviñó, M. Dellafiore, R. Gargallo, C. González, A.M. Iribarren, J. Montserrat, R. Eritja, Stabilization of Telomeric I-Motif Structures by (2′ S)-2′-Deoxy-2′-C-Methylcytidine Residues, ChemBioChem. 18 (2017) 1123–1128.

[27] R.V. Brown, T. Wang, V.R. Chappeta, G. Wu, B. Onel, R. Chawla, H. Quijada, S.M. Camp, E.T. Chiang, Q.R. Lassiter, C. Lee, S. Phanse, M.A. Turnidge, P. Zhao, J.G.N. Garcia, V. Gokhale, D. Yang, L.H. Hurley, The Consequences of Overlapping G-Quadruplexes and i-Motifs in the Platelet-Derived Growth Factor Receptor β Core Promoter Nuclease Hypersensitive Element Can Explain the Unexpected Effects of Mutations and Provide Opportunities for Selective Targeting of Both Structures by Small Molecules To Downregulate Gene Expression, J. Am. Chem. Soc. 139 (2017) 7456–7475. https://doi.org/10.1021/jacs.6b10028.

[28] Y. Xu, Formation of the G-quadruplex and i-motif structures in retinoblastoma susceptibility genes (Rb), Nucleic Acids Res. 34 (2006) 949–954. https://doi.org/10.1093/nar/gkj485.

[29] S. Surana, J.M. Bhat, S.P. Koushika, Y. Krishnan, An autonomous DNA nanomachine maps spatiotemporal pH changes in a multicellular living organism, Nat. Commun. 2 (2011) 340. https://doi.org/10.1038/ncomms1340.

[30] J. Zhou, C. Wei, G. Jia, X. Wang, Z. Feng, C. Li, Formation of i-motif structure at neutral and slightly alkaline pH, Mol. Biosyst. 6 (2010) 580–586.

[31] M. Zeraati, D.B. Langley, P. Schofield, A.L. Moye, R. Rouet, W.E. Hughes, T.M. Bryan, M.E. Dinger, D. Christ, I-motif DNA structures are formed in the nuclei of human cells, Nat. Chem. 10 (2018) 631–637. https://doi.org/10.1038/s41557-018-0046-3.

[32] S. Dzatko, M. Krafcikova, R. Hänsel-Hertsch, T. Fessl, R. Fiala, T. Loja, D. Krafcik, J. Mergny, S. Foldynova-Trantirkova, L. Trantirek, Evaluation of the Stability of DNA i-Motifs in the Nuclei of Living Mammalian Cells, Angew. Chem. Int. Ed. 57 (2018) 2165–2169.

[33] W. Tang, K. Niu, G. Yu, Y. Jin, X. Zhang, Y. Peng, S. Chen, H. Deng, S. Li, J. Wang, Q. Song, Q. Feng, In vivo visualization of the i-motif DNA secondary structure in the Bombyx mori testis, Epigenetics Chromatin. 13 (2020) 12. https://doi.org/10.1186/s13072-020-00334-y.

[34] D. Sun, L.H. Hurley, The Importance of Negative Superhelicity in Inducing the Formation of G-Quadruplex and i-Motif Structures in the c-Myc Promoter: Implications for Drug Targeting and Control of Gene Expression, J. Med. Chem. 52 (2009) 2863–2874. https://doi.org/10.1021/jm900055s.

[35] B.G. Kim, T.V. Chalikian, Thermodynamic linkage analysis of pH-induced folding and unfolding transitions of i-motifs, Biophys. Chem. 216 (2016) 19–22.

[36] J.-L. Mergny, L. Lacroix, X. Han, J.-L. Leroy, C. Helene, Intramolecular folding of pyrimidine oligodeoxynucleotides into an i-DNA motif, J. Am. Chem. Soc. 117 (1995) 8887–8898.

[37] M. Debnath, K. Fatma, J. Dash, Chemical Regulation of DNA i-Motifs for Nanobiotechnology and Therapeutics, Angew. Chem. Int. Ed. 58 (2019) 2942–2957. https://doi.org/10.1002/anie.201813288.

[38] S. Modi, S. M G, D. Goswami, G.D. Gupta, S. Mayor, Y. Krishnan, A DNA nanomachine that maps spatial and temporal pH changes inside living cells, Nat. Nanotechnol. 4 (2009) 325–330. https://doi.org/10.1038/nnano.2009.83.

[39] D. Liu, S. Balasubramanian, A Proton-Fuelled DNA Nanomachine, Angew. Chem. Int. Ed. 42 (2003) 5734–5736. https://doi.org/10.1002/anie.200352402.

[40] W. Wang, H. Liu, D. Liu, Y. Xu, Y. Yang, D. Zhou, Use of the interparticle i-motif for the controlled assembly of gold nanoparticles, Langmuir. 23 (2007) 11956–11959.

[41] J. Sharma, R. Chhabra, H. Yan, Y. Liu, pH-driven conformational switch of "i-motif" DNA for the reversible assembly of gold nanoparticles, Chem. Commun. (2007) 477–479.

[42] Y.P. Bhavsar-Jog, E. Van Dornshuld, T.A. Brooks, G.S. Tschumper, R.M. Wadkins, Epigenetic modification, dehydration, and molecular crowding effects on the thermodynamics of i-motif structure formation from C-rich DNA, Biochemistry. 53 (2014) 1586–1594.

[43] A. Rajendran, S. Nakano, N. Sugimoto, Molecular crowding of the cosolutes induces an intramolecular i-motif structure of triplet repeat DNA oligomers at neutral pH, Chem. Commun. 46 (2010) 1299. https://doi.org/10.1039/b922050j.

[44] H.A. Day, C. Huguin, Z.A.E. Waller, Silver cations fold i-motif at neutral pH, Chem. Commun. Camb. Engl. 49 (2013) 7696–7698. https://doi.org/10.1039/c3cc43495h.

[45] S.E. Kim, I.-B. Lee, C. Hyeon, S.-C. Hong, Destabilization of i-motif by submolar concentrations of a monovalent cation, J. Phys. Chem. B. 118 (2014) 4753–4760.

[46] M.A. Abdelhamid, L. Fábián, C.J. MacDonald, M.R. Cheesman, A.J. Gates, Z.A.E. Waller, Redox-dependent control of i-Motif DNA structure using copper cations, Nucleic Acids Res. 46 (2018) 5886–5893. https://doi.org/10.1093/nar/gky390.

[47] H.A. Day, E.P. Wright, C.J. MacDonald, A.J. Gates, Z.A.E. Waller, Reversible DNA i-motif to hairpin switching induced by copper(II) cations, Chem. Commun. 51 (2015) 14099–14102. https://doi.org/10.1039/C5CC05111H.

[48] H. Abou Assi, M. Garavís, C. González, M.J. Damha, i-Motif DNA: structural features and significance to cell biology, Nucleic Acids Res. 46 (2018) 8038–8056.

[49] P. Školáková, D. Renčiuk, J. Palacký, D. Krafčík, Z. Dvořáková, I. Kejnovská, K. Bednářová, M. Vorlíčková, Systematic investigation of sequence requirements for DNA i-motif formation, Nucleic Acids Res. 47 (2019) 2177–2189.

[50] S.P. Gurung, C. Schwarz, J.P. Hall, C.J. Cardin, J.A. Brazier, The importance of loop length on the stability of i-motif structures, Chem. Commun. 51 (2015) 5630–5632.

[51] S.M. Reilly, R.K. Morgan, T.A. Brooks, R.M. Wadkins, Effect of interior loop length on the thermal stability and pKa of i-motif DNA, Biochemistry. 54 (2015) 1364–1370. https://doi.org/10.1021/bi5014722.

[52] J.A. Brazier, A. Shah, G.D. Brown, I-motif formation in gene promoters: unusually stable formation in sequences complementary to known G-quadruplexes, Chem. Commun. 48 (2012) 10739–10741.

[53] A.L. Lieblein, J. Buck, K. Schlepckow, B. Fürtig, H. Schwalbe, Time-Resolved NMR Spectroscopic Studies of DNA i-Motif Folding Reveal Kinetic Partitioning, Angew. Chem. Int. Ed. 51 (2012) 250–253. https://doi.org/10.1002/anie.201104938.

[54] T.E. Malliavin, J. Gau, K. Snoussi, J.-L. Leroy, Stability of the I-motif Structure Is Related to the Interactions between Phosphodiester Backbones, Biophys. J. 84 (2003) 3838–3847. https://doi.org/10.1016/S0006-3495(03)75111-X.

[55] M.A.S. Abdelhamid, Z.A.E. Waller, Tricky Topology: Persistence of Folded Human Telomeric i-Motif DNA at Ambient Temperature and Neutral pH, Front. Chem. 8 (2020) 40. https://doi.org/10.3389/fchem.2020.00040.

[56] J. Mergny, L. Lacroix, Kinetics and thermodynamics of i-DNA formation: phosphodiester versus modified oligodeoxynucleotides, Nucleic Acids Res. 26 (1998) 4797–4803. https://doi.org/10.1093/nar/26.21.4797.

[57] S. Pérez-Rentero, R. Gargallo, C. González, R. Eritja, Modulation of the stability of i-motif structures using an acyclic threoninol cytidine derivative, RSC Adv. 5 (2015) 63278–63281. https://doi.org/10.1039/C5RA10096H.

[58] H. Abou Assi, M. Garavís, C. González, M.J. Damha, i-Motif DNA: structural features and significance to cell biology, Nucleic Acids Res. 46 (2018) 8038–8056. https://doi.org/10.1093/nar/gky735.

[59] H. Abou Assi, R.W. Harkness, N. Martin-Pintado, C.J. Wilds, R. Campos-Olivas, A.K. Mittermaier, C. González, M.J. Damha, Stabilization of i-motif structures by 2′-β-fluorination of DNA, Nucleic Acids Res. 44 (2016) 4998–5009. https://doi.org/10.1093/nar/gkw402.

[60] T.C. Roberts, R. Langer, M.J.A. Wood, Advances in oligonucleotide drug delivery, Nat. Rev. Drug Discov. 19 (2020) 673–694. https://doi.org/10.1038/s41573-020-0075-7.

[61] D.H. Lim, E.R. Maher, DNA methylation: a form of epigenetic control of gene expression, Obstet. Gynaecol. 12 (2010) 37–42.

[62] K.P. Koh, A. Yabuuchi, S. Rao, Y. Huang, K. Cunniff, J. Nardone, A. Laiho, M. Tahiliani, C.A. Sommer, G. Mostoslavsky, R. Lahesmaa, S.H. Orkin, S.J. Rodig, G.Q. Daley, A. Rao, Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells, Cell Stem Cell. 8 (2011) 200–213. https://doi.org/10.1016/j.stem.2011.01.008.

[63] Y.-F. He, B.-Z. Li, Z. Li, P. Liu, Y. Wang, Q. Tang, J. Ding, Y. Jia, Z. Chen, L. Li, Y. Sun, X. Li, Q. Dai, C.-X. Song, K. Zhang, C. He, G.-L. Xu, Tet-Mediated Formation of 5-Carboxylcytosine and Its Excision by TDG in Mammalian DNA, Science. 333 (2011) 1303–1307. https://doi.org/10.1126/science.1210944.

[64] V.N. Babenko, I.V. Chadaeva, Y.L. Orlov, Genomic landscape of CpG rich elements in human, BMC Evol. Biol. 17 (2017) 19.

[65] P.A. Jones, S.B. Baylin, The fundamental role of epigenetic events in cancer, Nat. Rev. Genet. 3 (2002) 415. https://doi.org/10.1038/nrg816.

[66] J.F. Costello, M.C. Frühwald, D.J. Smiraglia, L.J. Rush, G.P. Robertson, X. Gao, F.A. Wright, J.D. Feramisco, P. Peltomäki, J.C. Lang, Aberrant CpG-island methylation has non-random and tumour-type–specific patterns, Nat. Genet. 24 (2000) 132.

[67] N.A. Temiz, D.E. Donohue, A. Bacolla, B.T. Luke, J.R. Collins, The role of methylation in the intrinsic dynamics of B-and Z-DNA, PLoS One. 7 (2012) e35558.

[68] E. Isaakova, A. Varizhuk, G.E. Pozmogova, CpG Methylation in G-Quadruplex and I-Motif DNA Structures, Significances Bioeng. Biosci. 1 (2018) 55–61. https://doi.org/10.31031/SBB.2018.01.000514.

[69] J.A. Kretzmann, K.L. Irving, N.M. Smith, C.W. Evans, Modulating gene expression in breast cancer via DNA secondary structure and the CRISPR toolbox, NAR Cancer. 3 (2021) zcab048. https://doi.org/10.1093/narcan/zcab048.

[70] B. Xu, G. Devi, F. Shao, Regulation of telomeric i-motif stability by 5-methylcytosine and 5-hydroxymethylcytosine modification, Org. Biomol. Chem. 13 (2015) 5646–5651. https://doi.org/10.1039/C4OB02646B.

[71] E.P. Wright, M.A.S. Abdelhamid, M.O. Ehiabor, M.C. Grigg, K. Irving, N.M. Smith, Z.A.E. Waller, Epigenetic modification of cytosines fine tunes the stability of i-motif DNA, Nucleic Acids Res. 48 (2020) 55–62. https://doi.org/10.1093/nar/gkz1082.

[72] P. Školáková, Z. Badri, S. Foldynová-Trantírková, J. Ryneš, J. Šponer, M. Fojtová, J. Fajkus, R. Marek, M. Vorlíčková, J.-L. Mergny, L. Trantírek, Composite 5-methylations of cytosines modulate i-motif stability in a sequence-specific manner: Implications for DNA nanotechnology and epigenetic regulation of plant telomeric DNA, Biochim. Biophys. Acta BBA - Gen. Subj. 1864 (2020) 129651. https://doi.org/10.1016/j.bbagen.2020.129651.

[73] J. Lin, J. Hou, H. Xiang, Y. Yan, Y. Gu, J. Tan, D. Li, L. Gu, T. Ou, Z. Huang, Stabilization of G-quadruplex DNA by C-5-methyl-cytosine in bcl-2 promoter: implications for epigenetic regulation, Biochem. Biophys. Res. Commun. 433 (2013) 368–373.

[74] A.J. Stevens, S. Stuffrein-Roberts, S.L. Cree, A. Gibb, A.L. Miller, K. Doudney, A. Aitchison, M.R. Eccles, P.R. Joyce, V.V. Filichev, G-quadruplex structures and CpG methylation cause drop-out of the maternal allele in polymerase chain reaction amplification of the imprinted MEST gene promoter, PloS One. 9 (2014) e113955.

[75] S.-Q. Mao, A.T. Ghanbarian, J. Spiegel, S.M. Cuesta, D. Beraldi, M. Di Antonio, G. Marsico, R. Hänsel-Hertsch, D. Tannahill, S. Balasubramanian, DNA G-quadruplex structures mold the DNA methylome, Nat. Struct. Mol. Biol. 25 (2018) 951–957.

[76] W. Sun, L. Zang, Q. Shu, X. Li, From development to diseases: The role of 5hmC in brain, Genomics. 104 (2014) 347–351. https://doi.org/10.1016/j.ygeno.2014.08.021.

[77] J.A. Hackett, R. Sengupta, J.J. Zylicz, K. Murakami, C. Lee, T.A. Down, M.A. Surani, Germline DNA Demethylation Dynamics and Imprint Erasure Through 5-Hydroxymethylcytosine, Science. 339 (2013) 448–452. https://doi.org/10.1126/science.1229277.

[78] K.P. Koh, A. Yabuuchi, S. Rao, Y. Huang, K. Cunniff, J. Nardone, A. Laiho, M. Tahiliani, C.A. Sommer, G. Mostoslavsky, Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells, Cell Stem Cell. 8 (2011) 200–213.

[79] M. Iurlaro, G. Ficz, D. Oxley, E.-A. Raiber, M. Bachman, M.J. Booth, S. Andrews, S. Balasubramanian, W. Reik, A screen for hydroxymethylcytosine and formylcytosine binding proteins suggests functions in transcription and chromatin regulation, Genome Biol. 14 (2013) R119. https://doi.org/10.1186/gb-2013-14-10-r119.

[80] S.J. Clark, J. Melki, DNA methylation and gene silencing in cancer: which is the guilty party?, Oncogene. 21 (2002) 5380. https://doi.org/10.1038/sj.onc.1205598.

[81] F. Matarese, E. Carrillo-de Santa Pau, H.G. Stunnenberg, 5-Hydroxymethylcytosine: a new kid on the epigenetic block?, Mol. Syst. Biol. 7 (2011) 562. https://doi.org/10.1038/msb.2011.95.

[82] Y. Zhao, Z. Du, N. Li, Extensive selection for the enrichment of G4 DNA motifs in transcriptional regulatory regions of warm blooded animals, FEBS Lett. 581 (2007) 1951–1956.

[83] Y.P. Bhavsar-Jog, E. Van Dornshuld, T.A. Brooks, G.S. Tschumper, R.M. Wadkins, Co-Localization of DNA i-Motif-Forming Sequences and 5-Hydroxymethyl-cytosines in Human Embryonic Stem Cells, Molecules. 24 (2019) 3619. https://doi.org/10.3390/molecules24193619.

[84] S. Ito, L. Shen, Q. Dai, S.C. Wu, L.B. Collins, J.A. Swenberg, C. He, Y. Zhang, Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine, Science. 333 (2011) 1300–1303.

[85] M. Bachman, S. Uribe-Lewis, X. Yang, H.E. Burgess, M. Iurlaro, W. Reik, A. Murrell, S. Balasubramanian, 5-Formylcytosine can be a stable DNA modification in mammals, Nat. Chem. Biol. 11 (2015) 555–557. https://doi.org/10.1038/nchembio.1848.

[86] J.-B. Yan, C.-C. Lai, J.-W. Jhu, B. Gongol, T.L. Marin, S.-C. Lin, H.-Y. Chiu, C.-J. Yen, L.-Y. Wang, I.-C. Peng, Insulin and Metformin Control Cell Proliferation by Regulating TDG-Mediated DNA Demethylation in Liver and Breast Cancer Cells, Mol. Ther. - Oncolytics. 18 (2020) 282–294. https://doi.org/10.1016/j.omto.2020.06.010.

[87] E.-A. Raiber, P. Murat, D.Y. Chirgadze, D. Beraldi, B.F. Luisi, S. Balasubramanian, 5-Formylcytosine alters the structure of the DNA double helix, Nat. Struct. Mol. Biol. 22 (2015) 44.

[88] Q. Dai, P.J. Sanstead, C.S. Peng, D. Han, C. He, A. Tokmakoff, Weakened N3 Hydrogen Bonding by 5-Formylcytosine and 5-Carboxylcytosine Reduces Their Base-Pairing Stability, ACS Chem. Biol. 11 (2016) 470–477. https://doi.org/10.1021/acschembio.5b00762.

[89] B. Ashwood, P.J. Sanstead, Q. Dai, C. He, A. Tokmakoff, 5-Carboxylcytosine and Cytosine Protonation Distinctly Alter the Stability and Dehybridization Dynamics of the DNA Duplex, J. Phys. Chem. B. 124 (2020) 627–640. https://doi.org/10.1021/acs.jpcb.9b11510.

[90] J.U. Guo, Y. Su, J.H. Shin, J. Shin, H. Li, B. Xie, C. Zhong, S. Hu, T. Le, G. Fan, H. Zhu, Q. Chang, Y. Gao, G. Ming, H. Song, Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain, Nat. Neurosci. 17 (2014) 215–222. https://doi.org/10.1038/nn.3607.

[91] J.-H. Lee, S.-J. Park, K. Nakai, Differential landscape of non-CpG methylation in embryonic stem cells and neurons caused by DNMT3s, Sci. Rep. 7 (2017). https://doi.org/10.1038/s41598-017-11800-1.

[92] W. Xiao, X. Liu, X. Niu, C. Li, Y. Guo, J. Tan, W. Xiong, L. Fan, Y. Li, The frequency of CpG and non-CpG methylation of Notch3 gene promoter determines its expression levels in breast cancer cells, Exp. Cell Res. 386 (2020) 111743. https://doi.org/10.1016/j.yexcr.2019.111743.

[93] A.J. Price, L. Collado-Torres, N.A. Ivanov, W. Xia, E.E. Burke, J.H. Shin, R. Tao, L. Ma, Y. Jia, T.M. Hyde, J.E. Kleinman, D.R. Weinberger, A.E. Jaffe, Divergent neuronal DNA methylation patterns across human cortical development reveal critical periods and a unique role of CpH methylation, Genome Biol. 20 (2019) 196. https://doi.org/10.1186/s13059-019-1805-1.

[94] J.L. Huppert, Prevalence of quadruplexes in the human genome, 33 (2005) 2908–2916. https://doi.org/10.1093/nar/gki609.

[95] A. Bedrat, L. Lacroix, J.-L. Mergny, Re-evaluation of G-quadruplex propensity with G4Hunter, Nucleic Acids Res. 44 (2016) 1746–1759.

[96] O. Kikin, L. D'Antonio, P.S. Bagga, QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences, Nucleic Acids Res. 34 (2006) W676–W682.

[97] A.M. Fleming, Y. Ding, R.A. Rogers, J. Zhu, J. Zhu, A.D. Burton, C.B. Carlisle, C.J. Burrows, 4n-1 Is a "Sweet Spot" in DNA i-Motif Folding of 2'-Deoxycytidine Homopolymers, J. Am. Chem. Soc. 139 (2017) 4682–4689. https://doi.org/10.1021/jacs.6b10117.

[98] A.M. Fleming, K.M. Stewart, G.M. Eyring, T.E. Ball, C.J. Burrows, Unraveling the 4 n− 1 rule for DNA i-motif stability: base pairs vs. loop lengths, Org. Biomol. Chem. 16 (2018) 4537–4546.

[99] A. Verma, K. Halder, R. Halder, V.K. Yadav, P. Rawal, R.K. Thakur, F. Mohd, A. Sharma, S. Chowdhury, Genome-wide computational and expression analyses reveal G-quadruplex DNA motifs as conserved cis-regulatory elements in human and related species, J. Med. Chem. 51 (2008) 5641–5649.

[100] K. Niu, X. Zhang, H. Deng, F. Wu, Y. Ren, H. Xiang, S. Zheng, L. Liu, L. Huang, B. Zeng, S. Li, Q. Xia, Q. Song, S.R. Palli, Q. Feng, BmILF and i-motif structure are involved in transcriptional regulation of BmPOUM2 in Bombyx mori, Nucleic Acids Res. 46 (2018) 1710–1723. https://doi.org/10.1093/nar/gkx1207.

[101] A.B. Sahakyan, V.S. Chambers, G. Marsico, T. Santner, M. Di Antonio, S. Balasubramanian, Machine learning model for sequence-driven DNA G-quadruplex formation, Sci. Rep. 7 (2017) 14535. https://doi.org/10.1038/s41598-017-14017-4.

[102] J.-M. Garant, J.-P. Perreault, M.S. Scott, Motif independent identification of potential RNA G-quadruplexes by G4RNA screener, Bioinformatics. 33 (2017) 3532–3537. https://doi.org/10.1093/bioinformatics/btx498.

[103] G. Biffi, D. Tannahill, J. McCafferty, S. Balasubramanian, Quantitative visualization of DNA G-quadruplex structures in human cells, Nat. Chem. 5 (2013) 182–186. https://doi.org/10.1038/nchem.1548.

[104] S. Dhakal, Z. Yu, R. Konik, Y. Cui, D. Koirala, H. Mao, G-Quadruplex and i-Motif Are Mutually Exclusive in ILPR Double-Stranded DNA, Biophys. J. 102 (2012) 2575–2584. https://doi.org/10.1016/j.bpj.2012.04.024.

[105] Y. Cui, D. Kong, C. Ghimire, C. Xu, H. Mao, Mutually Exclusive Formation of G-Quadruplex and i-Motif Is a General Phenomenon Governed by Steric Hindrance in Duplex DNA, Biochemistry. 55 (2016) 2291–2299. https://doi.org/10.1021/acs.biochem.6b00016.

[106] J.J. King, K.L. Irving, C.W. Evans, R.V. Chikhale, R. Becker, C.J. Morris, C.D. Peña Martinez, P. Schofield, D. Christ, L.H. Hurley, Z.A.E. Waller, K.S. Iyer, N.M. Smith, DNA G-Quadruplex and i-Motif Structure Formation Is Interdependent in Human Cells, J. Am. Chem. Soc. 142 (2020) 20600–20604. https://doi.org/10.1021/jacs.0c11708.

[107] D. Sun, L.H. Hurley, The Importance of Negative Superhelicity in Inducing the Formation of G-Quadruplex and i-Motif Structures in the c-Myc Promoter: Implications for Drug Targeting and Control of Gene Expression, J. Med. Chem. 52 (2009) 2863–2874. https://doi.org/10.1021/jm900055s.

[108] H. Abou Assi, R. El-Khoury, C. González, M.J. Damha, 2′-Fluoroarabinonucleic acid modification traps G-quadruplex and i-motif structures in human telomeric DNA, Nucleic Acids Res. 45 (2017) 12055–12055. https://doi.org/10.1093/nar/gkx962.

[109] S.L. Brown, S. Kendrick, The i-Motif as a Molecular Target: More Than a Complementary DNA Secondary Structure, Pharmaceuticals. 14 (2021) 96. https://doi.org/10.3390/ph14020096.

[110] L. Lacroix, Identification of two human nuclear proteins that recognise the cytosine-rich strand of human telomeres in vitro, Nucleic Acids Res. 28 (2000) 1564–1575. https://doi.org/10.1093/nar/28.7.1564.

[111] S. Selvam, D. Koirala, Z. Yu, H. Mao, Quantification of Topological Coupling between DNA Superhelicity and G-quadruplex Formation, J. Am. Chem. Soc. 136 (2014) 13967–13970. https://doi.org/10.1021/ja5064394.

[112] M.-H. Lim, D.-H. Lee, S.E. Jung, D.-Y. Youn, C.S. Park, J.-H. Lee, Effect of Modulation of hnRNP L Levels on the Decay of bcl-2 mRNA in MCF-7 Cells, Korean J. Physiol. Pharmacol. 14 (2010) 15. https://doi.org/10.4196/kjpp.2010.14.1.15.

[113] G. Miglietta, S. Cogoi, E.B. Pedersen, L.E. Xodo, GC-elements controlling HRAS transcription form i-motif structures unfolded by heterogeneous ribonucleoprotein particle A1, Sci. Rep. 5 (2016) 18097. https://doi.org/10.1038/srep18097.

[114] A. Siddiqui-Jain, C.L. Grand, D.J. Bearss, L.H. Hurley, Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription, Proc. Natl. Acad. Sci. 99 (2002) 11593–11598. https://doi.org/10.1073/pnas.182256799.

[115] S. Cogoi, L.E. Xodo, G-quadruplex formation within the promoter of the KRAS proto-oncogene and its effect on transcription, Nucleic Acids Res. 34 (2006) 2536–2549. https://doi.org/10.1093/nar/gkl286.

[116] S. Cogoi, M. Paramasivam, B. Spolaore, L.E. Xodo, Structural polymorphism within a regulatory element of the human KRAS promoter: formation of G4-DNA recognized by nuclear proteins, Nucleic Acids Res. 36 (2008) 3765–3780. https://doi.org/10.1093/nar/gkn120.

[117] H. Fernando, A.P. Reszka, J. Huppert, S. Ladame, S. Rankin, A.R. Venkitaraman, S. Neidle, S. Balasubramanian, A Conserved Quadruplex Motif Located in a Transcription Activation Site of the Human c-kit Oncogene, Biochemistry. 45 (2006) 7854–7860. https://doi.org/10.1021/bi0601510.

[118] G. Ruvkun, M. Finney, Regulation of transcription and cell identity by POU domain proteins, Cell. 64 (1991) 475–478. https://doi.org/10.1016/0092-8674(91)90227-P.

[119] W. Herr, R.A. Sturm, R.G. Clerc, L.M. Corcoran, D. Baltimore, P.A. Sharp, H.A. Ingraham, M.G. Rosenfeld, M. Finney, G. Ruvkun, The POU domain: a large conserved region in the mammalian pit-1, oct-1, oct-2, and Caenorhabditis elegans unc-86 gene products., Genes Dev. 2 (1988) 1513–1516. https://doi.org/10.1101/gad.2.12a.1513.

[120] A.K. Ryan, M.G. Rosenfeld, POU domain family values: flexibility, partnerships, and developmental codes., Genes Dev. 11 (1997) 1207–1225. https://doi.org/10.1101/gad.11.10.1207.

[121] K. Certel, A. Hudson, S.B. Carroll, W. Johnson, Restricted patterning of vestigial expression in Drosophila wing imaginal discs requires synergistic activation by both Mad and the Drifter POU domain transcription factor, Development. 127 (2000) 3173–3183.

[122] W. Tang, K. Niu, G. Yu, Y. Jin, X. Zhang, Y. Peng, S. Chen, H. Deng, S. Li, J. Wang, Q. Song, Q. Feng, In vivo visualization of the i-motif DNA secondary structure in the Bombyx mori testis, Epigenetics Chromatin. 13 (2020) 12. https://doi.org/10.1186/s13072-020-00334-y.

[123] S. Haldar, Y. Zhang, Y. Xia, B. Islam, S. Liu, F.L. Gervasio, A.J. Mulholland, Z.A.E. Waller, D. Wei, S. Haider, Mechanistic Insights into the Ligand-Induced Unfolding of an RNA G-Quadruplex, J. Am. Chem. Soc. 144 (2022) 935–950. https://doi.org/10.1021/jacs.1c11248.

[124] P. Weisman-Shomer, E. Cohen, I. Hershco, S. Khateb, O. Wolfovitz-Barchad, L.H. Hurley, M. Fry, The cationic porphyrin TMPyP4 destabilizes the tetraplex form of the fragile X syndrome expanded sequence d(CGG)n, Nucleic Acids Res. 31 (2003) 3963–3970. https://doi.org/10.1093/nar/gkg453.

[125] S.Y. Rha, E. Izbicka, R. Lawrence, K. Davidson, D. Sun, M.P. Moyer, G.D. Roodman, L. Hurley, D. Von Hoff, Effect of telomere and telomerase interactive agents on human tumor and normal cell lines, Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res. 6 (2000) 987–993.

[126] C.L. Grand, H. Han, R.M. Muñoz, S. Weitman, D.D. Von Hoff, L.H. Hurley, D.J. Bearss, The cationic porphyrin TMPyP4 down-regulates c-MYC and human telomerase reverse transcriptase expression and inhibits tumor growth in vivo, Mol. Cancer Ther. 1 (2002) 565–573.

[127] Y. Yang, H. Fu, C. Qian, H. Li, D.D.Y. Chen, Characterization of interaction between Bcl-2 oncogene promoter I-Motif DNA and flavonoids using electrospray ionization mass spectrometry and pressure-assisted capillary electrophoresis frontal analysis, Talanta. 215 (2020) 120885. https://doi.org/10.1016/j.talanta.2020.120885.

[128] S. Takahashi, S. Bhattacharjee, S. Ghosh, N. Sugimoto, S. Bhowmik, Preferential targeting cancer-related i-motif DNAs by the plant flavonol fisetin for theranostics applications, Sci. Rep. 10 (2020) 2504. https://doi.org/10.1038/s41598-020-59343-2.

[129] B.J. Pages, S.P. Gurung, K. McQuaid, J.P. Hall, C.J. Cardin, J.A. Brazier, Stabilization of Long-Looped i-Motif DNA by Polypyridyl Ruthenium Complexes, Front. Chem. 7 (2019) 744. https://doi.org/10.3389/fchem.2019.00744.

[130] X. Li, Y. Peng, J. Ren, X. Qu, Carboxyl-modified single-walled carbon nanotubes selectively induce human telomeric i-motif formation, Proc. Natl. Acad. Sci. 103 (2006) 19658–19663. https://doi.org/10.1073/pnas.0607245103.

[131] E.P. Wright, H.A. Day, A.M. Ibrahim, J. Kumar, L.J.E. Boswell, C. Huguin, C.E.M. Stevenson, K. Pors, Z.A.E. Waller, Mitoxantrone and Analogues Bind and Stabilize i-Motif Forming DNA Sequences, Sci. Rep. 6 (2016) 39456. https://doi.org/10.1038/srep39456.

[132] Q. Sheng, J.C. Neaverson, T. Mahmoud, C.E.M. Stevenson, S.E. Matthews, Z.A.E. Waller, Identification of new DNA i-motif binding ligands through a fluorescent intercalator displacement assay, Org. Biomol. Chem. 15 (2017) 5669–5673. https://doi.org/10.1039/C7OB00710H.

[133] P. Saha, D. Panda, D. Müller, A. Maity, H. Schwalbe, J. Dash, *In situ* formation of transcriptional modulators using non-canonical DNA i-motifs, Chem. Sci. 11 (2020) 2058–2067. https://doi.org/10.1039/D0SC00514B.

[134] D. Hanahan, R.A. Weinberg, Hallmarks of Cancer: The Next Generation, Cell. 144 (2011) 646–674. https://doi.org/10.1016/j.cell.2011.02.013.

[135] M.R. Gill, J.A. Thomas, Ruthenium(ii) polypyridyl complexes and DNA—from structural probes to cellular imaging and therapeutics, Chem. Soc. Rev. 41 (2012) 3179. https://doi.org/10.1039/c2cs15299a.

[136] K. Deo, B. Pages, D. Ang, C. Gordon, J. Aldrich-Wright, Transition Metal Intercalators as Anticancer Agents—Recent Advances, Int. J. Mol. Sci. 17 (2016) 1818. https://doi.org/10.3390/ijms17111818.

[137] F.E. Poynton, S.A. Bright, S. Blasco, D.C. Williams, J.M. Kelly, T. Gunnlaugsson, The development of ruthenium(II) polypyridyl complexes and conjugates for in vitro cellular and in vivo applications, Chem. Soc. Rev. 46 (2017) 7706–7756. https://doi.org/10.1039/c7cs00680b.

[138] P. Spence, J. Fielden, Z.A.E. Waller, Beyond Solvent Exclusion: i-Motif Detecting Capability and an Alternative DNA Light-Switching Mechanism in a Ruthenium(II) Polypyridyl Complex, J. Am. Chem. Soc. 142 (2020) 13856–13866. https://doi.org/10.1021/jacs.0c04789.

[139] M.A.S. Abdelhamid, A.J. Gates, Z.A.E. Waller, Destabilization of i-Motif DNA at Neutral pH by G-Quadruplex Ligands, Biochemistry. 58 (2019) 245–249. https://doi.org/10.1021/acs.biochem.8b00968.

[140] A. Pagano, N. Iaccarino, M.A.S. Abdelhamid, D. Brancaccio, E.U. Garzarella, A. Di Porzio, E. Novellino, Z.A.E. Waller, B. Pagano, J. Amato, A. Randazzo, Common G-Quadruplex Binding Agents Found to Interact With i-Motif-Forming DNA: Unexpected Multi-Target-Directed Compounds, Front Chem. 6 (2018) 281. https://doi.org/10.3389/fchem.2018.00281.

[141] M.A.S. Abdelhamid, Z.A.E. Waller, Tricky Topology: Persistence of Folded Human Telomeric i-Motif DNA at Ambient Temperature and Neutral pH, Front. Chem. 8 (2020) 40. https://doi.org/10.3389/fchem.2020.00040.

[142] Y. Zhao, L. Cao, J. Ouyang, M. Wang, K. Wang, X.-H. Xia, Reversible Plasmonic Probe Sensitive for pH in Micro/Nanospaces Based on i-Motif-Modulated Morpholino-Gold Nanoparticle Assembly, Anal. Chem. 85 (2013) 1053–1057. https://doi.org/10.1021/ac302915a.

[143] S. Benabou, A. Aviñó, R. Eritja, C. González, R. Gargallo, Fundamental aspects of the nucleic acid i-motif structures, RSC Adv. 4 (2014) 26956–26980. https://doi.org/10.1039/C4RA02129K.

[144] D. Miao, Y. Yu, Y. Chen, Y. Liu, G. Su, Facile Construction of i-Motif DNA-Conjugated Gold Nanostars as Near-Infrared and pH Dual-Responsive Targeted Drug Delivery Systems for Combined Cancer Therapy, Mol. Pharm. 17 (2020) 1127–1138. https://doi.org/10.1021/acs.molpharmaceut.9b01159.

[145] P. Shende, I. Kataria, Potential applications of folded and unfolded DNA nanocarriers in medicine, J. Drug Deliv. Sci. Technol. 57 (2020) 101729. https://doi.org/10.1016/j.jddst.2020.101729.

[146] L. Shi, P. Peng, J. Zheng, Q. Wang, Z. Tian, H. Wang, T. Li, I-Motif/miniduplex hybrid structures bind benzothiazole dyes with unprecedented efficiencies: a generic light-up system for label-free DNA nanoassemblies and bioimaging, Nucleic Acids Res. 48 (2020) 1681–1690. https://doi.org/10.1093/nar/gkaa020.

[147] J. Duan, X. Wang, M.E. Kizer, Biotechnological and Therapeutic Applications of Natural Nucleic Acid Structural Motifs, Top. Curr. Chem. 378 (2020) 26. https://doi.org/10.1007/s41061-020-0290-z.

[148] S. Kendrick, A. Muranyi, V. Gokhale, L.H. Hurley, L.M. Rimsza, Simultaneous Drug Targeting of the Promoter MYC G-Quadruplex and BCL2 i-Motif in Diffuse Large B-Cell Lymphoma Delays Tumor Growth, J Med Chem. 60 (2017) 6587–6597. https://doi.org/10.1021/acs.jmedchem.7b00298.

[149] J. Andrae, R. Gallini, C. Betsholtz, Role of platelet-derived growth factors in physiology and medicine, Genes Dev. 22 (2008) 1276–1312. https://doi.org/10.1101/gad.1653708.

[150] K. Pietras, T. Sjöblom, K. Rubin, C.-H. Heldin, A. Östman, PDGF receptors as cancer drug targets, Cancer Cell. 3 (2003) 439–443. https://doi.org/10.1016/S1535-6108(03)00089-8.

Acknowledgements

We thank the National Health and Medical Research Council (NHMRC) of Australia for financial support. N.M.S. is the recipient of an NHMRC Research Fellowship.

Journal Pre-proof

Journal Pre-proof

Highlights

- Cytosine-rich DNA can adopt alternative conformations including the i-motif
- i-Motifs are associated with transcription, repair, and telomere maintenance
- i-Motif stability depends on environment, sequence, modifications, and ligands
- i-Motifs are potential therapeutic targets and nanodevices in drug development

Journal Pre-proof

The Authors declare no conflict of interest.

Journal Pre-proof