Bioorganic & Medicinal Chemistry 22 (2014) 4407–4418

Contents lists available at ScienceDirect

Bioorganic & Medicinal Chemistry

journal homepage: www.elsevier.com/locate/bmc

i-Motif DNA: Structure, stability and targeting with ligands

Henry A. Day, Pavlos Pavlou, Zoë A. E. Waller *

School of Pharmacy, University of East Anglia, Norwich Research Park, Norwich, Norfolk NR4 7TJ, UK

ARTICLE INFO

Article history:
Received 3 March 2014
Revised 9 May 2014
Accepted 22 May 2014
Available online 4 June 2014

Keywords:
i-Motif
Cytosine
DNA secondary structure
G-quadruplex
DNA binding ligands

ABSTRACT

i-Motifs are four-stranded DNA secondary structures which can form in sequences rich in cytosine. Stabilised by acidic conditions, they are comprised of two parallel-stranded DNA duplexes held together in an antiparallel orientation by intercalated, cytosine–cytosine$^+$ base pairs. By virtue of their pH dependent folding, i-motif forming DNA sequences have been used extensively as pH switches for applications in nanotechnology. Initially, i-motifs were thought to be unstable at physiological pH, which precluded substantial biological investigation. However, recent advances have shown that this is not always the case and that i-motif stability is highly dependent on factors such as sequence and environmental conditions. In this review, we discuss some of the different i-motif structures investigated to date and the factors which affect their topology, stability and dynamics. Ligands which can interact with these structures are necessary to aid investigations into the potential biological functions of i-motif DNA and herein we review the existing i-motif ligands and give our perspective on the associated challenges with targeting this structure.

1. Introduction

Since Watson and Crick first proposed that DNA exists as a B-form double helix[1] there has been extensive study into DNA morphology. It is now well established that DNA can adopt many different secondary structures such as alternative duplexes (including A and Z forms), triplexes, three and four way junctions and quadruplexes.[2] The best studied of these has been the G-quadruplex, a quadruple helical structure formed in sequences rich in guanine. Such sequences which have the potential to form these structures are enriched in telomeric and gene promoter regions within the genome and G-quadruplexes have recently been shown to exist in human cells.[3–5] Ligands which bind G-quadruplex have been used to inhibit telomerase and control the expression of oncogenes.[6] In genomic DNA, wherever there are guanine rich sequences, there are always complementary sequences rich in cytosine. Such sequences can also form quadruplex structures known as i-motifs but much less is known about the prevalence of these structures in vivo and their potential as targets for chemical intervention of cell biology.

In 1993 Gehring, Leroy and Guéron disclosed that DNA sequences containing stretches of cytosines can form intercalated, quadruple-helical structures under acidic conditions.[7] The tetrameric structure consisted of two parallel duplexes combined in an antiparallel fashion by forming intercalated hemiprotonated cytosine–cytosine base pairs (Fig. 1). Due to the unusual nature of the configuration they termed this new type of structure an intercalated (i) motif. The requirement of one of the cytosines in the base pairs to become protonated in order for the structure to form in vitro led to a wealth of applications in DNA nanotechnology. By altering the pH, these sequences can reversibly fold and unfold resulting in potential applications in nanotechnology, for example designing nanomachines[8,9] and applications such as assembly of gold nanoparticles,[10,11] switches for logic operations[12–14] and as sensors to map pH changes in living cells.[15,16] Despite this however, there are still limited published investigations into the biological function of i-motif DNA and relatively few examples of i-motif binding ligands which could be used as probes in such investigations. Herein we review the types of i-motif structures that have been identified and factors which affect their stability. We also review existing i-motif ligands and discuss the challenges associated with designing compounds to target i-motif DNA.

2. i-Motif structure

The initial discovery by Gehring et al. was of a four stranded intermolecular i-motif, composed of the sequence d(TCCCCC).[7] Despite the fact that hemiprotonated cytosine base pairs had been identified in 1962[19] and formation of hairpins stabilised by

* Corresponding author. Tel.: +44 (0)1603 591972.
E-mail address: z.waller@uea.ac.uk (Z.A.E. Waller).

http://dx.doi.org/10.1016/j.bmc.2014.05.047
0968-0896/

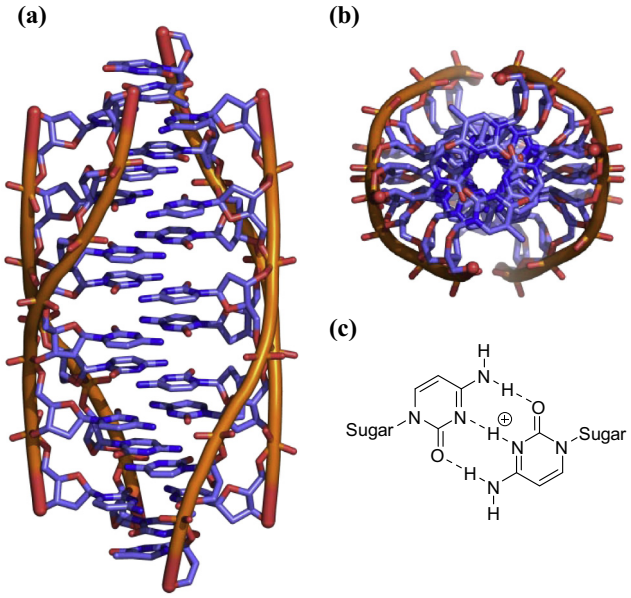

Figure 1. (a) Structure of the d(TC_5) intermolecular i-motif (PDB ID: 225D) Identified by Gehring et al.[7] (b) Top view of d(TC_5). (c) A hemiprotonated cytosine–cytosine$^+$ base pair.

cytosine-cytosine$^+$ base pairs had already been proposed,[20] the identification of an intercalated tetramer came as a complete surprise. With the composite four strands identical by NMR, it was only through nuclear Overhauser effects that the true structure could be characterised. Tetrameric stoichiometry for a range of other short polycytosine sequences was confirmed using gel filtration chromatography. The sequences varied the length of cytosine stretches, ranging from 3 to 12 bases, with and without thymine residues at various points in the sequence.[21] These i-motifs can intercalate in different ways to form two different tetramers: one in which the outmost C–C$^+$ pair is at the 3′ end and the other at the 5′ end. Further studies by Kanaori et al. showed that the tetramers of d(C_4T) and d(C_3T) were both able to slowly interconvert between these two topologies, which they named R-and S-forms.[22] A third less stable topology has also been identified (the T-form), where the cytosine–cytosine$^+$ base pairs at each end are not intercalated.[23] The two main topologies have since been denoted 3′E (R-form) and 5′E (S-form, Fig. 2).[24,25]

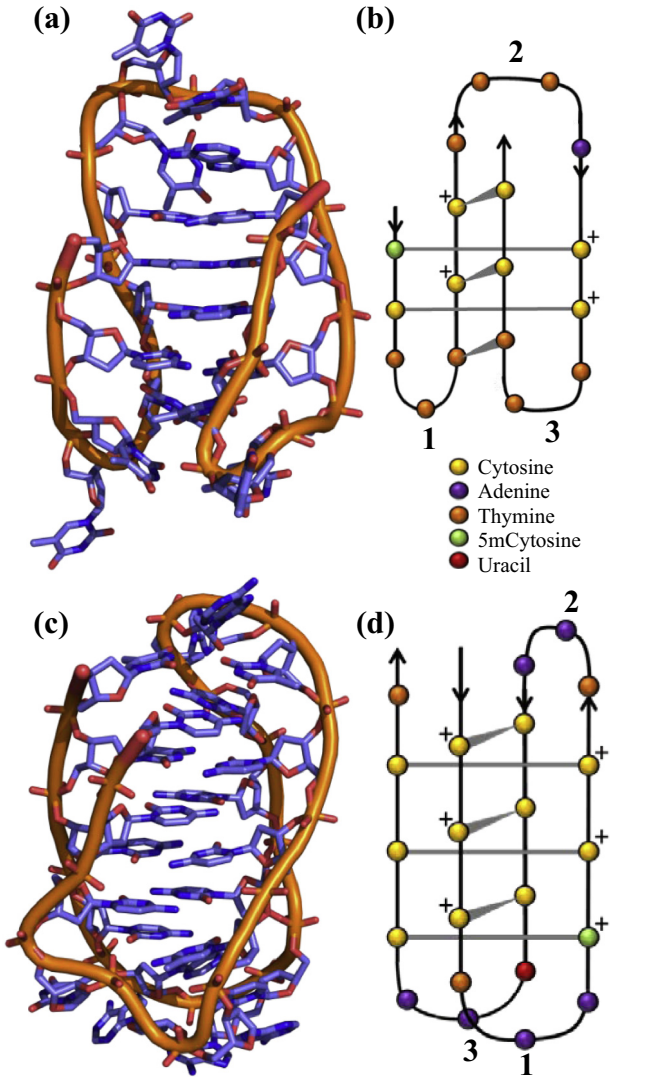

Figure 2. (a) NMR Structure of the 3′E i-motif sequence d(5mCCT_3CCT_3ACCT_3CC)[17] PDB ID: 1A83 (b). A schematic representation of the 3′E structure. (c) NMR Structure of the 5′E i-motif sequence d(CCCT$A_2$5mCCCTA_2CCCUA_2CCCT)[18] PDB ID: 1EL2. (d) A schematic representation of the 5′E structure. These illustrate the different intercalation topologies and the different loops numbered 1, 2 and 3. Arrows depict DNA strand direction from 5′ to 3′.

2.1. Topology

Sequences such as d(TAACCC)$_n$ from the human telomeric repeat region are able to form extra interactions between the non-cytosine bases, further adding to the stability of the i-motif structure. In this case, the TAA section is able to fold back upon itself into a loop and form an A–T base pair which stacks on top of the main cytosine core,[26] similar to the arrangement observed in the crystal structure of d(ACCCT).[27] The telomeric sequence is able to form a hydrogen bonded network between the two adenines and one of the thymines at one end of the i-motif structure, and another adenine–thymine base pair at the other. Each of these additional base pairs further increases the stability of the structure. Other studies have shown that thymine-thymine base pairs may also intercalate into an i-motif structure, for example in the sequence d(5mCCTCTCTCC);[28] however, more than one T–T base pair can have a destabilising effect.[29] Crystal structures of two different 5′E and 3′E i-motif tetramers of d(AACCCC) from the *Tetrahymena* telomere both show novel adenine clusters forming at the 5′ ends of the strands, perpendicular to the cytosine base pairs. These features add extra stability to the structure as demonstrated by the unusually high pH (7.5) at which the crystals were able to form.[30] Longer sequences with C-tracts separated by a section of other bases are also able to form i-motifs in a dimer arrangement. In these structures, each separate strand folds back on itself to form a hairpin which then intercalate to form the i-motif.[31,32] These i-motif dimers can vary significantly depending on the number and type of bases present in the loop structures.[33] Again, it is also possible for the loop regions to add to the stability of the i-motif by forming further interactions. For instance, the sequence d(TCCCGTTTCCA) forms an i-motif by dimerisation of two hairpins; extra stability is gained from an unusual T-G-G-T tetrad which forms between the two loops, enabling i-motif formation to occur at up to pH 6.7.[33] It has also been demonstrated that a minimal i-motif structure with just two C–C$^+$ base pairs can be formed from d(TCGTTTCGT) even at neutral pH by the formation of similar G:T:G:T tetrads at the top and bottom of the structure.[34]

Potentially biologically relevant i-motifs can form from natural sequences containing four tracts of cytosines separated by

stretches of other bases via folding into an intramolecular i-motif. One of the most studied of these is the human telomeric sequence d(CCCTAACCCTAACCCTAACCCT) the NMR structure of which was solved by Phan, Guéron and Leroy (Fig. 2c).[18] This i-motif contains six C–C$^+$ base pairs intercalated in a 5′E topology. Using NMR, they showed that the TAA bases in the loop regions form stabilising interactions but that these alternate between different conformations depending on whether the adenosine 18 nucleotide is in the *syn* or *anti* configuration. The initial structure determination of the intramolecular human telomeric i-motif indicated a 5′E configuration, later confirmed by Schwalbe and co-workers to be the thermodynamically favoured conformation, due to a stabilising T-T pair between loops one and three. Using time-resolved NMR spectroscopy and pH jumps, it was shown that the 3′E configuration could also be formed and was the kinetically favoured topology.[35]

The importance of the loop interactions in the human telomeric intramolecular i-motif have been studied by systematic substitution of the loop bases by tetra(ethylene glycol) units.[36] Thermal stability was found to decrease progressively with increasing substitution from a melting temperature of 64.9 °C (no bases substituted) to 45.4 °C (all 9 loop bases substituted). Overall there was a decrease in melting temperature of ~2 °C per substitution. It also has been shown that substitutions in loop 2 generally had less of an effect compared to substitutions in loops 1 and 3, further supporting the importance of stabilising interactions within the loop regions.

Aside from the telomeric i-motif, other genomic intramolecular i-motif forming sequences have been found in the promoter regions of several oncogenes. The composition of these sequences vary significantly and this is reflected in their respective stability. In 2010, Brooks, Kendrick and Hurley categorised these intramolecular i-motifs into two 'classes', based on the length of their loops.[37] 'class I' i-motifs have shorter loops, whereas 'class II' i-motifs have longer loops. In general, 'class II' i-motifs are reasoned to be the more stable due to extra stabilising interactions within the longer loop regions. Beyond identifying stable i-motif structures with various additional base interactions in unmodified oligonucleotide sequences, other investigations have focused on modifying the structure to try and determine which factors are most important for i-motif stability. Such modifications include those to the phosphate backbone, the sugar and the bases.

2.2. Effect of the phosphate backbone on stability

i-Motif structure consists of two wide grooves and two very narrow grooves. Consequently, there is repulsion between adjacent negatively charged phosphate backbones that define the minor groove; this needs to be balanced for i-motifs to be stable. Molecular dynamics simulations have been used to investigate the effect of phosphate repulsion on the stability of the 5′E and 3′E topologies of the tetrameric i-motif formed from the sequence d(CCCC).[38] These indicated that van der Waals forces and/or CH···O hydrogen bonding between the sugars are responsible for stabilising the narrow grooves of the structure. The simulations suggested that the 3′E topology is more stable due to it having two more favourable sugar interactions. As we have already discussed, the 5′E topology is the thermodynamically stable conformation of the intramolecular human telomeric i-motif,due to an additional stabilising T-T base pair. The free energy of the CH···O hydrogen bonds in the i-motif structure have been calculated to be quite small (2.6 kJ mol^{-1} per bond).[39] Therefore, the most stable conformation for any given i-motif will be a balance between these sugar interactions and any potential interactions within the loop sequences. As yet, there has been no systematic investigation into the dynamics between these features, therefore further work is required to fully determine the contribution of each of these effects.

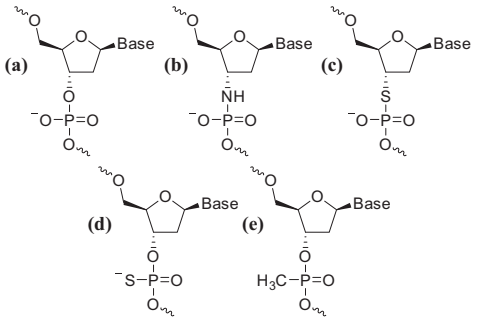

Figure 3. Different phosphate modifications: (a) phosphodiester, (b) phosphoramidate, (c) phosphorothiolate, (d) phosphorothioate, (e) methylphosphonate.

Several different studies have examined alternatives to the phosphate backbone. Oligodeoxycytidine phosphorothioates (Fig. 3d) are able to form both intermolecular and intramolecular i-motifs.[40] Whilst conserving the 16 base nucleic acid sequence Mergny and Lacroix also compared phosphorothioate, the natural phosphodiester, methylphosphonate (Fig. 3a and e) and peptide linkages.[41] In UV melting experiments, they found only the phosphodiester and phosphorothioate oligodeoxynucleotides formed i-motifs. This suggests that the introduction of the bulkier methyl group, although no longer charged, has a destabilising effect. A comparison of phosphodiester and phosphorothioates in longer sequences (18, 20 and 29 bases) also showed i-motif formation, even at neutral pH, and the phosphorothioate was found to be only marginally less stable (by 1–2 °C) in each case. However, phosphoramidate linkages (Fig. 3b) for the sequence d(TCCTCCTTTTCCTCCT) showed no i-motif formation, even at pH 5.6.[42] The chirality of phosphorothioate groups has also been shown to significantly affect stability, for instance the R-form (melting temperature, T_m = 31 °C) is much more stable than the S (T_m = 20 °C).[43] Whilst other phosphate modifications discussed only reduced the stability of the i-motif, substitution for 3′-S-phosphorothiolate linkages (Fig. 3c) resulted in a more stable i-motif structure due to favourable formation of the C3′-*endo* conformation of the sugar which is present in the i-motif.[44]

2.3. Effect of the sugar; the RNA i-motif

Initial reports suggested that oligoribonucleotides of cytidines could not form an i-motif structure above pH 6.[45] This is surprising, as in normal DNA i-motif, the deoxyribose sugars adopt the C3′-*endo* conformation, which is more commonly found in RNA (Fig. 4a). For example, a comparison of the stability of a deoxyribonucleotide sequence d(CCCTCCCTTTTCCCTCCC) and the corresponding ribonucleotide sequence r(CCCUCCCUUUUCCCUCCC) found that the DNA formed an i-motif with a melting temperature of 54 °C whilst the RNA was far less stable (T_m = 25 °C). Melting experiments on the analogousuracil substituted DNA sequence gave a much higher melting temperature (T_m = 56 °C). This suggests that repulsive interactions between the 2′ OH groups of the ribose sugars in the minor groove are the likely cause of RNA i-motif instability (Fig. 4c). Steric interactions in the sugar unsurprisingly causes destabilisation of the structure; substitutions of deoxyribose with arabinose (Fig. 4d) in which the 2′ OH group points towards the major groove, were found not to disrupt i-motif formation whereas substitutions of deoxyribose with 2′-O-methylated ribose (Fig. 4e) completely perturb formation of the structure.[46,47] Mixed RNA/DNA i-motifs have been found to be less stable than their DNA counterparts, with a decrease in melting

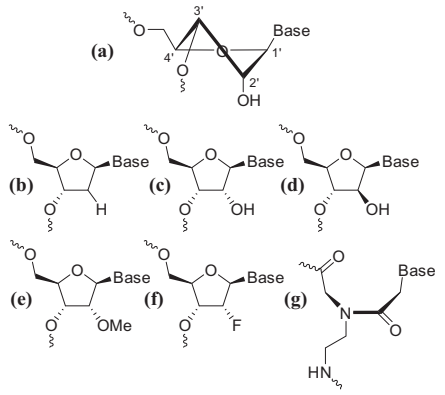

Figure 4. (a) The C3′-*endo* sugar configuration and different types of sugar modifications, (b) deoxyribose, (c) ribose, (d) arabinose, (e) 2′-methoxyribose, (f) 2′-fluororibose, (g) peptide nucleic acid.

temperature of ~6.5 °C for the substitution of just one RNA strand.[46] The formation of DNA_2-RNA_2 hybrid i-motifs have been characterised at pH 4.5. Two DNA strands were found to form one of the minor grooves whilst the two RNA strands defined the other; these were found to eventually dissociate and reform into pure DNA i-motif and single stranded RNA.[48] The introduction of a single 2′-fluorine substitution (Fig. 4f) was also found to form a more stable i-motif structure, with a change in melting temperature (ΔT_m) of +4 °C. This supports the hypothesis that sugar-sugar interactions are central to i-motif stability.[49] Reduction in the charge repulsion of the backbones can be alleviated by using neutral analogues, for example peptide nucleic acids (PNA, Fig. 4g). However, PNA i-motifs are less stable compared to the analogous DNA sequences.[50–54] Interestingly however, hybrid PNA-DNA i-motifs are more stable than those composed of either pure DNA or PNA.[53]

2.4. Effect of base modifications

Modification of the bases within i-motifs could potentially affect their stability.[42,55] When cytosine is replaced with 5-methylcytosine, the i-motif still forms without any decrease in stability.[42] Furthermore, replacement of thymine with 5-propynyl uracil, resulted in a more stable i-motif structure (ΔT_m = +5–9 °C). A recent study by Wadkins and co-workers indicated that modification of cytosines has different effects, depending on the experimental crowding conditions. For example, in dilute aqueous solutions, methylation of cytosine increased the pK_a and T_m of i-motifs, whereas modification to hydroxymethyl cytosine lowered the pK_a and T_m.[56] The authors described that this data indicates that modification could be useful for fine-tuning the pH- or temperature-dependent folding/unfolding of i-motifs. Given the rise in interest in the function of epigenetic modification of DNA, this is an area which requires further investigation.

2.5. Environmental conditions

Studies into i-motif stability have been driven towards finding a natural sequence which forms a stable i-motif at physiological rather than acidic pH. One study has shown that i-motifs can form from natural sequences at pH 7.5, but only at low temperature (4 °C).[57] Molecular crowding has also been shown to be important in stabilising DNA secondary structures. Simulation of the crowded cellular and nuclear environment can be achieved using high molecular weight polyethylene glycols. Under these conditions, G-quadruplexes and i-motifs are favoured over duplex[58] and single stranded structures.[59] For i-motif, these conditions raise the pK_a of the cytosine N3, enabling protonation and i-motif formation at neutral pH.[56,60] Such studies indicate that the formation of DNA secondary structures may be more favoured under physiological conditions, compared to the simplified experiments using buffers.

The extent which alternative secondary structures, such as G-quadruplex and i-motif, are able to compete with duplex DNA formation is of considerable interest. The presence of alternative secondary structures is highly dependent on the conditions in which the experiments are performed.[56,61–63] Reports suggest that formation of i-motif destabilises the accompanying double stranded structure and a minimum spacing of five base pairs between the i-motif and duplex structures is required to maintain the stability of the adjacent duplex.[64] Something that may help mitigate this is the effect of negative superhelicity. During transcription, duplex DNA must be unwound into its component single strands, this creates negative superhelical stress in the single strand which can be relieved by formation of a secondary structure.[65] Hurley and co-workers demonstrated this effect by placing the G-quadruplex/i-motif forming sequence of the c-MYC oncogene promoter into a supercoiled plasmid to simulate the negative superhelicity generated during transcription. Then, using a combination of enzymatic and chemical footprinting, they observed formation of both G-quadruplex and i-motif structures at neutral pH. This study outlines a mechanism by which negative superhelical stress promotes formation of i-motif DNA structures under physiological conditions.

Cationic conditions are commonly used to influence the formation of different types of DNA secondary structure. For example, Saxena and co-workers have investigated the effects of Na^+, K^+ and Mg^{2+} on the formation of i-motif structure in the c-jun protooncogene.[66] They showed that at pH 5.7 in the presence of 100 mM Na^+, only duplex DNA was observed whereas, at the same pH but in the presence of 100 mM K^+, i-motif and G-quadruplex structures were predominant. Moreover, in the presence of 10 mM Mg^{2+} and 100 mM Na^+, all three structures were observed. However, Mergny et al. found that the pK_a of the cytosine N3 was reduced in low salt conditions, favouring i-motif formation.[55] They indicated that at pH 4.8 (around the pK_a of cytosine) increasing the concentration of NaCl to 100 mM resulted in a destabilisation of the structure; though increasing the salt concentration further to 300 mM did not increase the destabilisation any further. Likewise, they observed no differences in i-motif stability upon addition of 5 mM magnesium, calcium, zinc, lithium or potassium cations in the presence of 100 mM NaCl at pH 6.4. Other examples, including a study on sequences from n-MYC (see Table 1), also show a decrease in stability of i-motif with increasing ionic concentration.[67] It may be then, that the observations by Saxena et al. are due to the stabilisation of G-quadruplex, favouring the dissociation of the double strand.

In our own experiments we have investigated the effect of Ag^+ ions on the folding of i-motif at neutral pH.[68] We showed using a combination of CD, UV difference spectroscopy and Förster Resonance Energy Transfer (FRET) experiments that the C-rich human telomeric sequence folded into a secondary structure at pH 7.4 in the presence of Ag^+ ions. Such structure formation was reversible by chelation of the Ag^+ ions with cysteine. This was later supported by Goncharova's study, using vibrational and electronic CD spectroscopy, which showed that silver cations stabilise cytosine base pairing and enable the formation of an i-motif-like structure at conditions up to pH 10.[69] Aside from this, there needs to be more investigation into potential stabilisation and folding of the i-motif by other types of cations.

3. Biological significance

Computational studies have indicated that putative G-quadruplex forming sequences can be found in 43% of all gene

Table 1
Example sequences of i-motifs which have been investigated

Name	Sequence 5′ → 3′
hTelo	d(CCCTAACCCTAACCCTAACCCT)
c-MYC	d(CCCCACCTTCCCCACCCTCCCCACCCTCCCC)
bcl-2	d(CAGCCCCGCTCCCGCCCCCTTCCTCCCGCGCCCGCCCCT)
Rb	d(GCCGCCCAAAACCCCCCG)
RET	d(CCGCCCCCCGCCCCGCCCCGCCCCTA)
VEGF-A	d(GACCCCGCCCCCGGCCCGCCCCGG)
c-ki-ras	d(GCTCCCTCCCTCCCTCCTTCCCTCCCTCCC)
c-kit	d(CCCTCCTCCCAGCGCCCACCCT)
PDGF-A	d(CCGCGCCCCTCCCCCGCCCCCGCCCCCGCCCCCCCCCCCCCC)
c-myb	d(TCCTCCTCCTCCTTCTCCTCCTCCTCCGTGTCCTCCTCCTCC)
hTERT	d(CCCCGCCCCGTCCCGACCCCTCCCGGGTCCCCGGCCCAGCCCCCACCGGGCCCTCCCAGCCCCTCCCC)
HIF-1α	d(CGCGCTCCCGCCCCCTCTCCCCTCCCCGCGCGCCCGAGCGCGCCTCCGCCCTTGCCCGCCCCCTG)
c-jun	d(TAACCCCCTCCCCCTCCCCCCTTTAAT)
ILPR	d(TGTCCCCACACCCCTGTCCCCACACCCCTGT)
n-MYC	d(ACCCCCTGCATCTGCATGCCCCCTCCCACCCCCT)

promoters and 69% of oncogene promoters.[70] As yet there has been no systematic analysis of the genome to find sequences capable of forming i-motifs. Due to the complimentary nature of G-quadruplex and i-motif forming sequences, it is likely that there is some overlap in their prevalence; though the stability and the dynamics of i-motif structures are likely to vary much more, depending on the type of sequence. Initial investigations into native i-motif forming sequences have centred around those from the complimentary C-rich strands of known promoter G-quadruplex forming sequences. A range of native i-motif forming sequences have now been identified and characterised (Table 1).

3.1. The telomeric i-motif

The sequence of bases that occur at the ends of eukaryotic DNA are known as the telomeres. They form part of a protective DNA-protein complex and comprise tandem repeat sequences which are rich in cytosine and guanine. These serve to cap the chromosome and protect the genetic material from degradation during cell replication. Telomerase, an enzyme which elongates the telomeres, is normally inactive but it has been shown to be activated in around 85% of cancers, enabling cell immortality.[71,72] Stabilisation of human telomeric G-quadruplex structures is known to inhibit telomerase activity with a range of different ligands.[73] However, it has also now been shown that stabilisation of the human telomeric i-motif by carboxyl-modified single-walled carbon nanotubes (SWNTs) can also inhibit telomerase resulting in telomere uncapping, DNA damage response and apoptosis.[74] The telomeric i-motif was one of the first intramolecular i-motif forming sequences to be identified, forming a 'class I' structure with short 3:3:3 loop topology.[18,75] Under mildly acidic conditions the i-motif structure has been shown to be able to compete with duplex formation.[61] Furthermore, several proteins have been identified in human cells which bind specifically to the C-rich single strand repeat sequence of human telomeric DNA.[76,77]

3.2. The c-MYC promoter i-motif

One of the best-studied promoter i-motif forming sequences is that of the nuclease hypersensitive element (NHE) III_1 region in the c-MYC oncogene.[78] The c-MYC gene codes for a transcription factor with a wide range of functions, in particular regulation of cell cycle progression and cell growth;[79] dysregulation of c-MYC is characteristic in many cancers. The nuclease hypersensitive element III_1 is one of seven NHEs that are part of the promoter; however, NHE III_1 is of particular interest as it not only contains a cytosine rich strand capable of forming an i-motif, but is responsible for the control of around 90% of c-MYC transcription.[80] In vitro, the c-MYC i-motif is only stable up to pH 6.5 and has a transitional pH_T (the pH at which it is 50% folded) of 6.6.[37] Laurence Hurley's group showed that under conditions of negative superhelical stress, similar to that generated during transcription, the i-motif can form at neutral pH. However, it does so with a different and larger loop topology of 6:2:6 ('class II') with only six cytosine base pairs,[65] in comparison to the 5:5:5 loop topology which forms at low pH with eight cytosine base pairs.[78]

There are several proteins involved in the transcription of c-MYC which bind specifically to the GC rich double strand and the C-rich single strand of the NHE III_1. The heterogeneous ribonucleoprotein K (hnRNP K) is one such transcription factor which binds specifically to the C-rich single strand of c-MYC NHE III_1, activating the promoter. Hurley's group has hypothesised that formation of the i-motif structure is likely to affect the binding of hnRNP K and hence affect transcription. A similar effect has already been demonstrated for the G-quadruplex and G-rich strand specific proteins in the c-MYC NHE III_1.[81] Another protein, the human nonmetastatic 23 isoform 2 (NM23-H2) also activates c-MYC transcription via the NHE III_1.[79] It also binds to the G- and C-rich single strands; stabilisation of the G-quadruplex blocks its binding site and reduces affinity for the DNA, thus a similar affect could occur with stabilisation of the i-motif.[82]

3.3. The bcl-2 promoter i-motif

The P1 promoter region of the B-cell lymphoma-2 (bcl-2) oncogene has been shown to form an i-motif structure with large, 'class II' loop topology (8:5:7). The large loops are argued to be the reason for its high stability ($pH_T = 6.6$)[37,83] through formation of a thymine capping structure in the central loop.[83] Unlike other oncogenes, bcl-2 enables cell survival and proliferation via an anti-apoptotic function rather than affecting growth factors.[84,85] Interestingly although bcl-2 is over expressed in some cancers, in neurodegenerative diseases such as Alzheimer's and Parkinson's it may be under expressed, leaving the cells more vulnerable to apoptosis.[86] Again, there are particular proteins which, like in the case of c-MYC, are able to bind to the GC rich double stranded portion of the promoter sequence.[87,88] WT-1 is involved in the down regulation of bcl-2 and Hurley's group have suggested that formation of either G-quadruplex or i-motif structures could inhibit WT-1 and therefore up-regulate bcl-2 gene expression.[37] Their recent work shows the first example of modulation of gene expression via targeting i-motif DNA. When unwound, the bcl-2 promoter region can exist in a dynamic equilibrium between i-motif and hairpin structures.[89] Using two compounds, one selective for the bcl-2 i-motif (**IMC-48**) and one for a hairpin (**IMC-76**, Fig. 5), they were able to shift the equilibrium to each structure, respectively.

Stabilisation of the i-motif resulted in up-regulation of the expression of bcl-2 yet stabilisation of the hairpin, prevented i-motif formation and resulted in a down-regulation of bcl-2 expression (see also Section 4.4 and Fig. 6). In the deconvolution of the mechanism of these two compounds, a transcription factor protein specific for the bcl-2 i-motif was discovered.[90] The protein hnRNP LL, a member of the heterogenous ribonucleoprotein family, is analogous to the c-MYC protein hnRNP K and is a paralog of the protein hnRNP L which binds to bcl-2 mRNA.[91] This transcription factor binds specifically to the bcl-2 i-motif and acts to induce bcl-2 transcription by unfolding the i-motif into a single strand within the protein active site. Finally, Mao and Hurley developed a new population dynamics approach, using 'laser tweezers' to assess the dynamic equilibrium between populations under different conditions at the single molecule level.[92] They showed that the bcl-2 promoter sequence forms a variety of six hairpin and i-motif structures involving different numbers of nucleotides; the i-motif populations were favoured by lower pH, ligand **IMC-48** and the transcription factor hnRNP LL, whilst **IMC-76** favoured the hairpin populations. Although there has still been no direct visualisation of i-motifs in vivo, these recent experiments provide strong evidence for a biological function for the i-motif, with a dynamic nature akin to RNA.

3.4. Other promoter i-motif sequences

As well as c-MYC and bcl-2, several other oncogene promoter i-motif forming sequences have been studied in detail. Retinoblastoma (Rb) encodes a tumour suppressor protein and its regulation is disrupted in almost all cancers.[93] The oncogene has been shown to form an i-motif structure at the 5′ end of the gene.[94] Rb forms a 'class I' i-motif with a 2:4:2 loop topology ($pH_T = 5.9$).[37] Using systematic C to T substitutions it was found that three of the cytosines in the sequence were not critical for i-motif formation and that two different intercalation patterns were possible.[95]

The proximal promoter of the RET proto-oncogene has also been shown to form a stable 'class I' i-motif with a 2:3:2 loop structure ($pH_T = 6.4$).[37,96] The GC rich region of the RET promoter, where both the i-motif and G-quadruplex structures can form, is also the binding site of the Sp1 transcription factor which regulates RET expression.[97]

The vascular endothelial growth factor (VEGF) promoter region has been shown to form an i-motif in vitro at a pH equal or below 5.9.[98] It is a 'class I' i-motif, with a 2:3:2 loop structure ($pH_T = 5.8$).[37,98] The C-rich region upstream of the VEGF gene which forms the i-motif has been shown to be of key importance to VEGF transcription[99] and contains binding sites for the Sp1 transcription factor, indicative of potential sites for intervention with therapeutic agents.

The insulin minisatellite or insulin-linked polymorphic region (ILPR) is upstream of the gene coding for insulin. It was first shown that this region was able to form an i-motif via dimerisation of two $C–C^+$ hairpins,[31] but can also form an intramolecular i-motif.[100] Shortening in the length and sequence of the ILPR has been linked to development of insulin-dependent diabetes mellitus.[101] The ILPR intramolecular i-motif has a 'class I' structure with a 3:3:3 loop topology that is stable at pH 6 but also still detectable at pH 7.[100] Using 'laser tweezers' the ILPR i-motif was found to unfold via a partially folded structure at neutral pH; regardless, both structures have a force stability large enough to block the progress of RNA polymerase.[102] This intermediate structure has been characterised as a 3 stranded hairpin, consisting of 3 cytosine repeats, that is able to form an i-motif upon addition of a fourth strand or a longer 4 cytosine repeat sequence.[103] It has been shown that the i-motif and G-quadruplex of the ILPR are able to form within the double strand but are mutually exclusive, that is, the G-quadruplex and i-motif do not both form together in the double strand.[104]

The c-ki-ras i-motif was one of the first oncogene promoter i-motif structures to be identified.[105] It forms a stable i-motif structure ($T_m = 60$ °C) at pH 5.1 and has a transitional pH of 6.6. The cytosine rich region of the oncogene has been shown by systematic deletion experiments to be necessary for the genes transcriptional activation.[106] The sequence consists of 6 cytosine tracts, so is able to form several different i-motif structures but, as yet, no defined topology has been characterised.[105] The extra cytosines in the sequence could provide extra stability within loop or flanking regions but further investigation is required to fully characterise the structure.

The c-kit oncogene encodes a tyrosine kinase receptor and a sequence in the promoter region which has been shown to form an i-motif under mildly acidic conditions ($T_m = 33$ °C at pH 6.8) with a 2:5:1 loop topology.[107] In this study, the importance of the cytosine and guanine bases in the loop regions were investigated by systematic thymine substitution.[108] Interestingly, replacement of all loop bases with thymine had only a small effect at pH 5.0 ($\Delta T_m = +4$ °C) but significantly decreased the stability at higher pH. For example, at pH 6 the T_m was 13 °C less than in the native sequence and at pH 6.8, i-motif was not observed at all. This suggests the sequence of bases within the loops is important in providing stability at near neutral pHs, but less important at acidic pH. The characterised i-motif sequence is in a region 140 base pairs upstream of the transcription start site (TSS) but there is also another GC-rich region 87 base pairs upstream of the TSS which has been shown to form a G-quadruplex.[109] To our knowledge, there have been no investigations into whether the complimentary strand of this sequence is able to form an i-motif.

Recently a study of the cytosine rich strands from the promoters of c-kit, PDGF-A, c-myb, HIF-1α and hTERT were identified as forming i-motif structures at pH 5.0. Out of these, PDGF-A and HIF-1α sequences would classically be termed 'class I' i-motifs as they have short loops, but both sequences were shown to form stable i-motifs at pH 7.0 with melting temperatures of 32 °C and 27 °C, respectively.[110] At first, this is surprising due to the small size of their loop regions, however, they have particularly long cytosine tracts (Table 1), enabling them to form up to 10 cytosine base pairs in the case of PDGF-A, which will add to their stability. Furthermore, the sequences used in the study had up to six flanking bases which may act in a similar stabilising way to long loop sequences, potentially end-capping the i-motif structure. Another study using sequences from n-MYC (Table 1) observed different structures depending on the pH.[67] n-MYC was found to have a transitional pH of 6.5; the presence of a long loop in the sequence allowed additional hairpin formation at higher pHs. Interestingly, the hairpin-loop structure was not favoured when the pH was reduced and the sequence adopts a different conformation. Nevertheless, the formation of an additional hairpin within the loop increases stability at near neutral pH.

The prevalence of genomic G-quadruplex forming sequences suggests there are likely to be many more complimentary cytosine sequences capable of forming i-motif structures which have not yet been characterised. The dynamic nature of the bcl-2 i-motif/hairpin structures and the interesting additional hairpin formation within one of the loops in n-MYC indicate that there are likely to be other similar examples in the genome with interesting structural dynamics. Whereas G-quadruplexes are stabilised by short loops, i-motifs, given their more dynamic nature, require longer sequences to form. It might be that there are examples of i-motif forming sequences in the genome which occur opposite sequences which are not conducive to formation of a stable G-quadruplex structure. However, further studies relating sequence to stability and dynamics of i-motif and potentially hairpin formation are

required to enable prediction of the likelihood of structure formation based on sequence.

3.5. Proteins which bind i-motif

The only i-motif binding protein to be examined in detail was the bcl-2 activating transcription factor hnRNP LL.[90] This was one of thirty five proteins identified by a pull-down assay, which bound to the bcl-2 i-motif and one of nine identified with a role in transcription. An electrophoretic mobility shift assay (EMSA) was used to show that hnRNP LL was specific for the i-motif at pH 6.8 versus a mutant single strand incapable of forming the i-motif and the bcl-2 double strand. Using surface plasmon resonance, the binding affinity of hnRNP LL for the bcl-2 i-motif was determined at pH 6.5 (K_d = 19.4 pM), the same sequence as a single strand showed a lower affinity (K_d = 69.8 pM). hnRNP LL is very similar to the RNA binding protein hnRNP L, with four recognition motifs for the sequences CCCGC and CGCCC. These sequences are found in the two lateral loops of the bcl-2 i-motif, suggesting that these form the binding site for the protein. The protein was shown to unfold the i-motif structure by CD and bromine footprinting experiments. These showed a different protection pattern to both the i-motif and the hairpin stabilised by **IMC-76**, indicating that the unfolded form in the protein is not the same as the hairpin. Lastly, the optimum distance between the protein recognition sequences was assessed by EMSA with sequences spaced by increasing numbers of thymines between 2 and 17. This demonstrated that the optimum distance was 13 bases, identical to the number in the natural i-motif sequence. Hurley's group have hypothesised that the role of the i-motif structure is to hold and present the protein recognition sequences in the most kinetically favourable conformation for binding which are then unfolded by the protein to the more thermodynamically stable single-strand. Their molecular population dynamics (MPD) experiments showed that the protein hnRNP LL favoured i-motif structure populations during a short (60 s) incubation period but favoured unfolding of the i-motif during longer incubation (180 s) with the protein remaining bound to the single stranded DNA.[92] This supports their suggestion that hnRNP LL acts by binding to the i-motif structure and unfolding it to enable transcription. Further investigation is required to fully understand the mechanism by which hnRNP LL activates transcription after binding to the i-motif sequence.

The discovery of the i-motif binding protein hnRNP LL indicates, at least in the case of bcl-2, a biological role of the i-motif as a protein recognition site for the activation of transcription. hnRNP K is another transcription factor protein in the same family as hnRNP LL and is known to bind to the C-rich region of the c-MYC promoter.[79] This protein contains similarly spaced recognition domains which recognise the TCCC sequences of c-MYC.[111] As with bcl-2, these recognition sequences are found in the lateral loops of the c-MYC i-motif and are spaced by the same number of bases as in bcl-2.[65] These similarities suggest that hnRNP K may bind to the c-MYC i-motif and activate transcription in an analogous manner. hnRNP K has also been shown to bind to the C-rich strand of human telomeric DNA which also contains CCCT sequences.[77] However, in formation of the human telomeric i-motif these sequences form the core structure rather than the lateral loops, thus an investigation into the difference in binding between the human telomeric and c-MYC i-motifs could be a useful comparison. Lacroix et al. examined the binding of hnRNP K to the human telomeric i-motif forming sequence at pH 6 to 9.2.[77] They showed binding of hnRNP K across this pH range but their results could not determine whether the protein bound to the i-motif and stabilised the structure at higher pH, or whether it bound to the single strand and unfolded the i-motif at acidic pH. Hurley's work on hnRNP LL suggests that the latter is likely be the case.

There are several other proteins which have been shown to bind to cytosine rich DNA. Of particular interest is the discovery of a highly cytosine specific protein in human HeLa cells.[76] The protein is not only specific for the human telomeric sequence $d(CCCTAA)_n$, but also sequences with at least four cytosine tracts rather than other similar cytosine rich sequences such as c-ki-ras or $d(C_{22})$.[112] This is key, as it seems to be very specific for the sequence which can form telomeric i-motif, however binding has only been assessed for the single strand at pH 7.4. At this pH, the human telomeric i-motif is mainly in the unfolded conformation, so the potential to bind to the i-motif structure still needs to be assessed. Maurice Guéron's group have also investigated the binding of proteins in yeast to the yeast, *tetrahymena* and vertebrate telomeric sequences at pH 6, where they would expect formation of i-motif.[113] EMSA experiments indicated formation of a protein DNA complex, however, this was still the case at pH 8 when the strand would normally be unfolded. It is not clear whether the protein binds to the i-motif and unfolds the structure or if it stabilises the structure at the higher pH. A variety of other C-rich sequence binding proteins have been identified which bind single stranded DNA.[114–117] Each of these is also worthy of further study to deduce whether they bind higher order DNA structures.

Regardless of whether the i-motif has a natural regulatory function in biology, it is possible that targeting the structure with small molecules may lead to potential therapeutics for genetic disease. For example, it has already been shown that induction of i-motif structure results in inhibition of telomerase activity,[74] and that modulation of bcl-2 secondary structure by ligands can affect gene expression.[89] Therefore, it is possible that similar stabilisation of other i-motifs in gene promoter regions may also alter gene expression by a similar mechanism. In order to test this hypothesis, a range of high affinity and highly stabilising i-motif ligands need to be developed. Such compounds could be used as chemical biological tools to probe the biological function of i-motif DNA structures.

4. i-Motif interacting ligands

4.1. Early i-motif ligands: TMPyP4 and BisA

The first example of an i-motif binding compound was published by Hurley and coworkers in 2000.[118] Using the human telomeric sequences $d(CCCAAT)_4$ and $d(AATCCC)_4$ they investigated the binding properties of the cationic porphyrin **TMPyP4** (Fig. 5), which was found to bind and promote the formation of i-motif at pH 4.5. Using UV titrations a dissociation constant of 45 μM was measured, which is weaker than the affinity for G-quadruplex (0.5 μM) and duplex (1.2 μM) structures.[119] The study indicated a highly cooperative effect upon binding, however, there was no significant change in the i-motif melting temperature on addition of ligand (ΔT_m <20 °C). Modelling experiments using the NMR structure suggested **TMPyP4** binds to the i-motif via stacking interactions at the ends of the structure. **TMPyP4** i-motif binding studies have also been documented with the intramolecular human telomeric i-motif and a mutant sequence where the adenines had been substituted by thymines to alter the loop interactions.[120] In this case, it was found that loop-interactions were important in the binding mode as binding properties were different for the native and the mutant sequence. The K_d was determined to be 1 μM, which is different to that originally calculated by Hurley and co-workers this difference in affinity could be due to the different binding modes. **TMPyP4** has already been shown to down regulate the expression of c-MYC through interaction with the G-quadruplex,[121] as well as inhibit telomerase and tumour growth.[122] In light of the stability of the c-MYC i-motif under

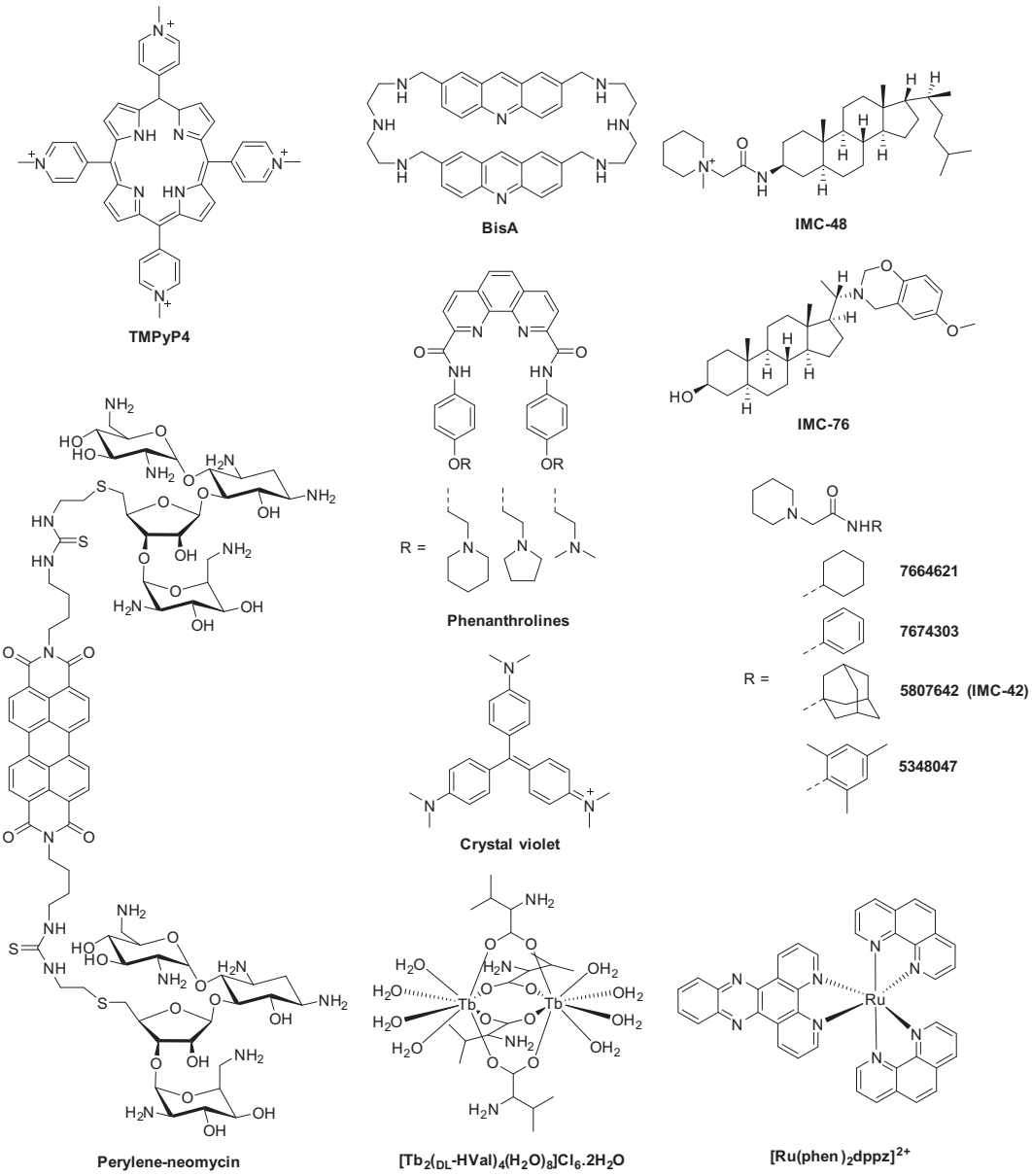

Figure 5. Structures of known i-motif binding ligands.

negative supercoiling,[65] there is scope to investigate whether its interaction with i-motif can lead to similar biological outcomes. Given the promiscuous nature of **TmPyP4**, which can bind many different DNA structures, it could be that the effects observed in the expression of c-MYC could be due to a synergistic effect between G-quadruplex and i-motif. **TmPyP4** could be used as a lead for further development to improve selectivity between i-motif and other DNA secondary structures. However, given its issues with specificity, in its current form it should be used with caution in purely i-motif based experiments.

The first use of FRET experiments to study ligand i-motif interactions was by Alberti et al. They studied the interaction of a cyclic acridine dimer (Fig. 5) with human telomeric G-quadruplex and i-motif DNA structures.[123] Using FRET to monitor DNA melting, they found that the **BisA** caused a significant stabilisation of i-motif at pH 6.8 (ΔT_m = +33 °C). The bis-acridine also stabilised the G-quadruplex (ΔT_m = +15 °C) and inhibited telomerase in vitro with an IC_{50} of 0.75 μM. A control compound consisting of a single acridine unit was found to have no effect on i-motif or G-quadruplex stability, nor was it able to inhibit telomerase. To date, no other studies with this compound have been published and the further work focused on development of the bis-acridines as G-quadruplex, rather than i-motif, ligands.

4.2. Carboxyl-modified single-walled carbon nanotubes

One of the most studied i-motif ligands are carboxyl-modified single-walled carbon nanotubes (SWNTs) which can inhibit duplex and induce formation of the human telomeric i-motif, even at pH 8. The SWNTs stabilise i-motif structure, at pH 5.5 with a stabilisation temperature of +22 °C at 10 μg mL^{-1}.[124] Due to their size, it has been proposed that the nanotubes bind to the 5′-end in the major groove of the i-motif. This mode of binding would stabilise the structure by forming favourable electrostatic interactions between the hemiprotonated cytosine core and the carboxyl groups. Similar carboxyl-modified multi-walled carbon nanotubes (MWNTs) have no effect on i-motif stability and are not able to bind to the i-motif due to their larger size.[74] Under molecular crowding conditions,

SWNTs can also induce i-motif formation at physiological pH.[125] A study into the biomedical effects of carboxylated single-wall carbon nanotubes on telomerase and telomeres by Qu and co-workers provided the first evidence that SWNTs could inhibit telomerase activity and interfere with the telomere functions in cancer cells.[74] SWNTs have also been shown to accelerate S1 nuclease cleavage rate[126] and inhibit the activity of telomerase in vitro, with an IC_{50} of 0.27 ± 0.02 μg mL^{-1}.[74,127] SWNTs are i-motif specific and do not bind either G-quadruplex or duplex DNA,[124] but the observed inhibition of telomerase may not be due the formation of the i-motif structure directly.[74] SWNTs favour duplex dissociation, so formation of i-motif will facilitate the induction of G-quadruplex structure[124] which is already known to inhibit telomerase.[73] The localisation of the SWNTs in telomeric regions of the nucleus was confirmed using fluorescently labelled SWNTs co-localised with the telomere binding protein TRF1, a known marker for interphase telomeres.[128] SWNTs were shown to inhibit telomerase in living K562 and HeLa cells using a standard TRAP assay with IC_{50} values of 10.2 ± 0.8 and 7.5 ± 0.45 μg mL^{-1}, respectively.[74] DNA damage response, displacement of telomere-binding proteins and telomere uncapping were all shown to occur in cells as a consequence of treatment with SWNTs.[74] The dysfunctional telomere causes DNA damage response and activates the DNA repair pathways resulting in p16 and p21-mediated cell cycle arrest, apoptosis and senescence. Furthermore, the SWNTs do not induce acute toxicity; treatment of both K562 and HeLa cells with carboxyl-modified SWNTs at concentrations up to 100 μg mL^{-1} did not give rise to evidence of acute toxic effects after 3 and 6 days of treatment. Although not ideal for development as a drug, the SWNTs demonstrated the first proof-of-concept intervention of chemical biology by targeting i-motif DNA structures. This work inspired investigation into carboxyl-modified graphene quantum dots (GQDs). These have also been shown to promote and stabilise i-motif formation at pH 8.0 by end-stacking interactions with the loop regions.[129] GQDs show a significant concentration dependent stabilisation of i-motif DNA structures with maximum ΔT_ms of 20 °C and 16 °C for the human telomeric and c-MYC i-motifs, respectively. Given the stabilisation effect on i-motif structure in acidic conditions, maintenance of consistent pH during experiments is extremely important. With both the carboxyl-modified single-walled carbon nanotubes and modified graphene quantum dots, the number of carboxylic acid groups per molecule is not clear.[130,131] Nevertheless, carboxyl-modified MWNTs which also contain acidic groups, did not have an effect on the stabilisation of i-motif structure, so it is not the presence of acidic groups alone which cause these effects.[124]

4.3. Other i-motif ligands

Phenanthroline compounds (Fig. 5) have been shown to be weakly stabilising of human telomeric i-motif with 10 equivalents of ligand increasing the thermal stability by 7–10 °C at pH 5.5. Binding affinities were measured between 4 and 8 μM by fluorescence titrations.[132] These compounds are not specific for i-motif however, and have slightly higher affinity for G-quadruplex (1.6–2.5 μM) and also interact with DNA duplex (6.7–12 μM).

A neomycin–perylene conjugate (Fig. 5) was tested as a specific ligand for human telomeric G-quadruplex. During the investigations the conjugate was also compared with i-motif, duplex and other G-quadruplexes using competition dialysis experiments. The bound concentration of the ligand was measured by fluorescence as ~1.2 μM for i-motif, weaker in comparison to telomeric G-quadruplex (~4 μM)[133] but higher than duplex and other G-quadruplex structures in the study. However, the competition dialysis experiments were all performed at pH 7. At neutral pH, the human telomeric i-motif is normally unfolded and it is therefore likely that the measured binding concentration is for the unfolded single strand rather than the i-motif structure.

Whilst developing a fluorescent, DNA-based logic gate, the binding of the fluorescent dye crystal violet (Fig. 5) with i-motif DNA was investigated.[134] Using UV titrations against a 29 base i-motif sequence $d(C_5T_3)_3C_5$ a K_d of 0.83 μM was measured. When comparing the affinity of crystal violet with two other shorter i-motif sequences, crystal violet showed weaker affinity and binding to G-quadruplex sequences was found to be even weaker still (28 and 53 μM for the two sequences tested).[135] Molecular docking simulations suggested that crystal violet binds to the i-motif sequence $(ACCCT)_4$ via end stacking at the terminus of the structure with a calculated binding energy of −38.44 kcal mol^{-1}.

Several metal complexes have been investigated as potential i-motif ligands. In particular two different amino acid complexes of Terbium, **$[Tb_2(_{DL}\text{-Cys})_4(H_2O)_8]Cl_2$** and **$[Tb_2(_{DL}\text{-HVal})_4(H_2O)_8]Cl_6 \cdot 2H_2O$**, (Fig. 5) were found to bind to i-motif DNA at pH 5.5 (K_d = 22 μM and 30 μM, respectively); however, they also marginally destabilise the structure (ΔT_m = −0.5 and −4.0 °C, respectively).[136] The binding data for the complexes with G-quadruplex DNA show similar values at pH 7.1, (K_d = 26 and 23 μM and ΔT_m = −4 and −3 °C, respectively), indicating that these complexes are not i-motif specific. Ruthenium complexes have also been investigated; for example, the complex **$[Ru(bpy)_2(dppz)]^{2+}$** (Fig. 5) was shown to bind to i-motif DNA with low micromolar affinity (5.6 μM), yet again the complex did not affect the melting temperature of i-motif. More importantly, the ligands show a much higher affinity for G-quadruplex DNA (K_d = 0.7 μM).[137] Similar results were obtained for the complex **$[Ru(phen)_2(dppz)]^{2+}$** (Fig. 5) which also showed low micromolar K_ds (1.5 and 2.1 μM).[138]

Although not strictly a free ligand, several investigations have modified loop bases with pyrene, an intercalator which has been used as a fluorescent probe for i-motif structure. Initially, by modifying the terminal adenine base with a pyrene moiety, it was found that the pyrene modified structure was more stable than the unmodified form, increasing the melting temperature by 11.4 °C. Upon decreasing pH, the DNA folds into an i-motif and the fluorescence of the pyrene moiety is quenched.[139] Incorporation of two pyrene modified adenines, one at the end and one at the midpoint of a dimeric i-motif, resulted in further i-motif stabilisation with an increase in melting temperature of 23.5 °C for the modified sequence at pH 5.0.[140] Replacement of adenines in the central loop portion of an intramolecular i-motif with a pyrene modified unlocked nucleic acid caused destabilisation,[141] one adenine substitution reduced the melting temperature by 8.9 °C. Substitution of an adenine in the central loop 2 with the pyrene intercalator also resulted in destabilisation (ΔT_m between −8 °C and −4.5 °C, depending on the position) at pH 5.2 but had little effect at pH 6.2.[142] This suggests that it is difficult for the structure to accommodate a ligand inside the central loop region of i-motifs with short loop sequences (3 bases).[143,144] These results also imply that free pyrene or ligands based on the pyrene structure may also warrant investigation into their potential as i-motif ligands.

4.4. Bcl-2 specific ligands; IMC-48 and IMC-76

As discussed in Section 3.3, Hurley, Hecht and co-workers recently showed that the bcl-2 i-motif sequence exists in a dynamic equilibrium between an i-motif and a hairpin structure.[89] Using a FRET-based high throughput assay, they screened a library of 1990 compounds against the bcl-2 i-motif at a ligand concentration of 5 μM. From 14 initial hits, they identified two compounds: **IMC-48** as an i-motif stabilising ligand and **IMC-76** as an i-motif destabilising ligand (Fig. 5). Further FRET screening against a mutant bcl-2 sequence incapable of i-motif formation, two other promoter i-motifs (c-MYC and VEGF) as well as the

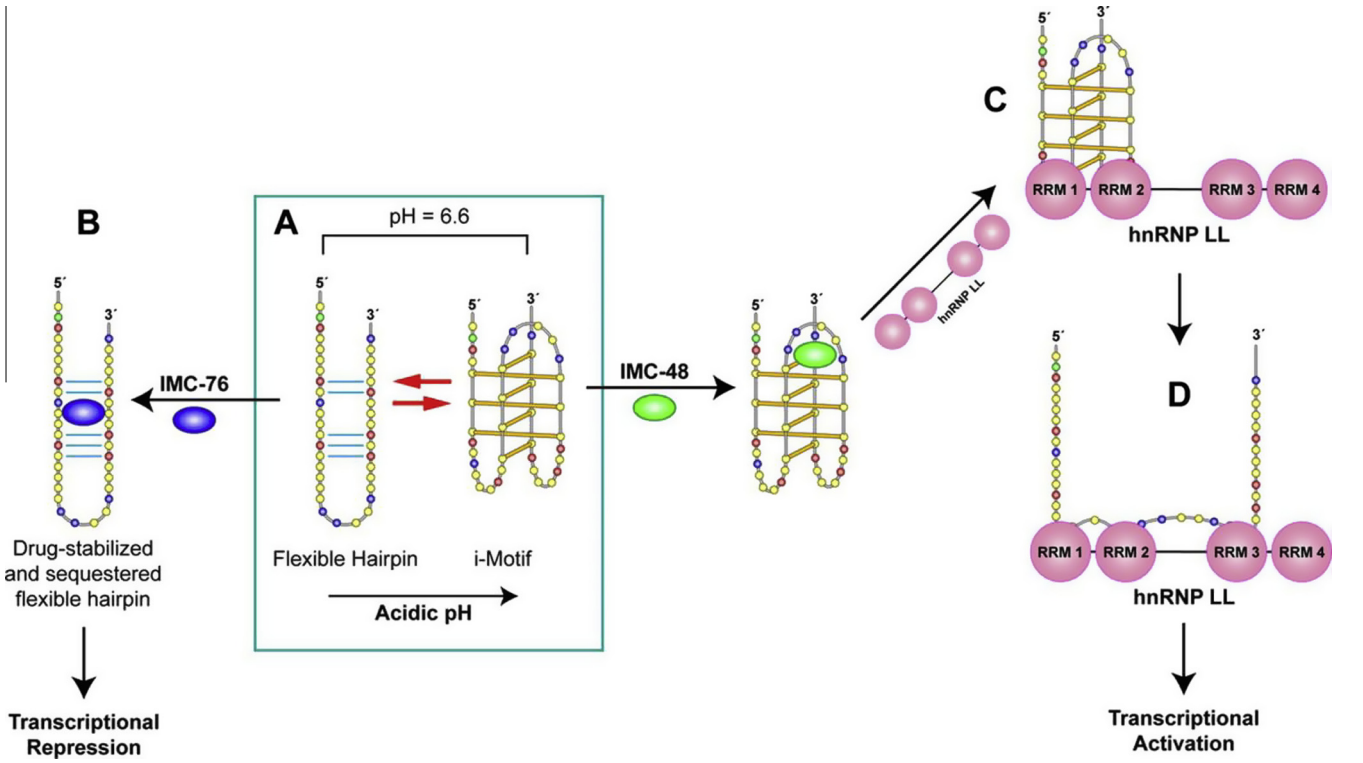

Figure 6. Proposed model of conformational transitions and biological consequences that occur following binding of **IMC-76**, **IMC-48**, and hnRNP LL to the C-rich strand in the promoter of bcl-2: (A) The different conformational states of the bcl-2 promoter, influenced by pH. Acidic pH favours i-motif whereas pH 6.6 has a mixture of hairpin and i-motif. (B) Addition of **IMC-76** stabilises the hairpin and results in transcriptional repression. (C) Addition of **IMC-48** stabilises the i-motif, the RNA recognition motifs (RRM) 1 and 2 of hnRNP LL bind to loops II and V in the i-motif. (D) hnRNP LL–driven changes form an alternative conformation of the C-rich strand which results in transcriptional activation of bcl-2.

bcl-2 G-quadruplex and duplex structures demonstrated the selectivity of both compounds for the bcl-2 i-motif. The affinity of each compound was measured by fluorescence, which indicated low/sub micromolar dissociation constants (**IMC-48** K_d = 0.49 μM, **IMC-76** K_d = 1.01 μM). However, neither of the compounds was found to significantly stabilise i-motif by CD melting. By monitoring the difference in chemical shift of the imino protons between the two structures, ^{1}H NMR was used to show that **IMC-48** favoured formation of the i-motif structure whereas **IMC-76** favoured the hairpin. Furthermore, an antagonistic effect between the two ligands was determined. Introduction of increasing equivalents of either **IMC-48** with a 2:1 **IMC-76**:bcl-2 complex or **IMC-76** with a 2:1 **IMC-48**:bcl-2 complex resulted in formation of the i-motif structure and hairpin structure, respectively. The hypothesised binding site for **IMC-48** is within the central loop region of the i-motif, due to the availability of stacking interactions with thymines, forming a capping structure within the central loop. To test this hypothesis, the central loop of the c-MYC i-motif was replaced with that of bcl-2. **IMC-48** had no effect on the natural c-MYC i-motif, however the hybrid c-MYC/bcl-2 i-motif showed similar changes in fluorescence to that of natural bcl-2, suggesting that the central loop is critical to ligand binding. Analogues of **IMC-48** (Fig. 5) each showed similar effects to **IMC-48** in a FRET assay and one compound (**IMC-42** Fig. 5) was tested and shown to increase bcl-2 expression in MCF-7 cells to a similar extent as **IMC-48**.

In cells with high bcl-2 expression, the hairpin stabilising compound **IMC-76** markedly decreases bcl-2 mRNA levels. By contrast, in cells with low bcl-2 expression, the i-motif stabilising compound **IMC-48** increases both the bcl-2 mRNA and protein expression. These compounds are selective for bcl-2; experiments with c-MYC and VEGF, both of which contain i-motifs in their promoters,[78,98] resulted in no significant change in gene expression. **IMC-76** was also used to induce chemosensitivity to etoposide and cyclophosphamide, known chemotherapeutic agents, in the correspondingly resistant B95.8 and GRANTA-519 cells. This resulted in a 2.5- and 1.9-fold increase in caspase-3 activity, respectively, an indicator of cell apoptosis.[145] The antagonistic effects of **IMC-48** and **IMC-76** which were initially shown by NMR were also demonstrated in cells. Treatment with **IMC-48** and cyclophosphamide did not increase caspase-3 activity and the induction of apoptosis by **IMC-76** was inhibited in the presence of equal amounts of **IMC-48**. Finally, the effect of **IMC-76** on bcl-2 expression was demonstrated in vivo using mice with GRANTA-519 lymphoma xenografts. A 20% decrease in bcl-2 mRNA was observed and combined treatment with etoposide resulted in 65% inhibition of tumour growth.

As discussed, the Hurley group showed that bcl-2 expression is driven by the binding of a transcription factor to the i-motif structure which then facilitates unfolding into a single-stranded form ready for transcription.[90] Using EMSA, they showed that **IMC-76** is able to compete with hnRNP LL, resulting in a reduction in the amount of hnRNP LL-bcl-2 i-motif complex present. On the other hand **IMC-48** which stabilises the i-motif enabled an increase in complex formation with a decrease in the amount of both hairpin and uncomplexed bcl-2 i-motif structures. The groups MPD experiments further showed that **IMC-76** favoured the hairpin structures of bcl-2 whilst **IMC-48** favoured i-motif structures during a short incubation period (60 s).[92] In fact comparison between populations favoured by hnRNP LL and **IMC-48** showed 80% similarity but like hnRNP LL, **IMC-48** favoured unfolding of the i-motif after longer incubation (180 s). This suggests that **IMC-48** can also activate transcription by unfolding the i-motif to the single strand. It is therefore hypothesised that the interaction of either **IMC-48** and

IMC-76 with bcl-2 has a biological effect by modulating the amount of i-motif available to bind to hnRNP LL which activates transcription (Fig. 6).

4.5. Perspectives and challenges in the design of i-motif ligands

The past decade has observed the rise of G-quadruplexes from peculiar DNA secondary structures to a potential drug design target which forms in cells and can be targeted with compounds to modulate gene expression. There are now many cytosine rich sequences from gene promoter regions which have been characterised as capable of forming i-motif structures. The work started by Xiaogang Qu and Laurence Hurley lays the foundation into the investigation of i-motifs as a target for chemical intervention of cell biology. Examples of the effects of targeting i-motifs in the telomeres and gene promoters now indicate there is just cause for dedicating resource and time to answering further questions about what role i-motifs can potentially play in biology and drug design. Whether i-motifs exist in vivo and have a natural regulatory function, is a separate question to whether it is possible to target, stabilise and induce i-motif formation in biology. Hurley's work in particular indicates that the dynamics of i-motif structures are important and the effects of ligand and protein binding can result in different biological effects. This evidence suggests that the DNA i-motif structures are actually acting more like RNA, similar to riboswitches. What influences the dynamics of i-motif structures? How important is the composition of the sequence in their stability and dynamics? There is evidence to suggest that formation of i-motifs and G-quadruplex are independent of each other; is this consistent with all examples, or isolated to the cases identified already? How many potential i-motif forming sequences are present in the human genome and do they exist in vivo? Now the methodologies are in place, the time is now ideal to investigate these questions.

Using chemistry to learn about biological function, one needs an arsenal of small molecules, proteins and antibodies to act as chemical biological probes. Compared to the wealth of knowledge on the design of G-quadruplex interacting ligands, the i-motif literature is comparatively scant with well-documented examples of ligands. One of the main challenges for discovery of i-motif binding ligands is the dependence on acidic pH for i-motif formation in vitro. Whilst there have been several investigations showing i-motif formation at neutral pH under conditions of negative superhelicity,[65] molecular crowding[58–60] or in the presence of silver ions.[68] These have not yet been applied to in vitro ligand screening assays, all of which rely on acidic pH to form the i-motif structure, which can skew ligand screening results. Another challenge facing researchers interested in targeting higher order DNA structures, is the design of ligands which are specific. There are not only many different types of higher order DNA structures but also multiple sub-classes of each type which could be present in vivo. Relatively few i-motif binding ligands have been identified and, with the exception of the carboxyl-modified SWNTs and **IMC-48**, all published examples are not i-motif specific, and in most cases have a greater affinity for G-quadruplex and/or duplex DNA. There is not enough structural understanding of i-motif binding sites and how ligand binding affects the conformation and dynamics. Although there are published crystal structures of intermolecular i-motifs,[26,30,146–148] structures of intramolecular i-motifs are also absent in the literature. Further ligand-DNA structural studies by crystallography and NMR spectroscopy would aid ligand design. There is significant room for not only development of compounds to target i-motif but to better understand the potential ligand binding modes.

Within this review we have highlighted factors which affect i-motif stability, discussed the biological implications of a range of i-motif forming sequences and finally reviewed the main compounds which have been found to target the i-motif. The G-quadruplex has recently been proven to exist in cells and has been accepted as a potential drug target for development of cancer therapeutics. Given the increasing evidence supporting a hypothesis that i-motifs can form under physiological conditions and can be targeted with small molecule ligands to affect biology, the time is right to commit to investigations into their potential as therapeutic targets for genetic disease.

Acknowledgements

We thank Novartis for a studentship (HD) and Professor Laurence Hurley for kind permission to reproduce Figure 6.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.bmc.2014.05.047. These data include MOL files and InChiKeys of the most important compounds described in this article.

References and notes

1. Watson, J.; Crick, F. *Nature* **1953**, *171*, 737.
2. Choi, J.; Majima, T. *Chem. Soc. Rev.* **2011**, *40*, 5893.
3. Huppert, J. L. *FEBS J.* **2010**, *277*, 3452.
4. Rodriguez, R.; Miller, K. M.; Forment, J. V.; Bradshaw, C. R.; Nikan, M.; Britton, S.; Oelschlaegel, T.; Xhemalce, B.; Balasubramanian, S.; Jackson, S. P. *Nat. Chem. Biol.* **2012**, *8*, 301.
5. Biffi, G.; Tannahill, D.; McCafferty, J.; Balasubramanian, S. *Nat. Chem.* **2013**, *5*, 182.
6. Balasubramanian, S.; Hurley, L. H.; Neidle, S. *Nat. Rev. Drug Disc.* **2011**, *10*, 261.
7. Gehring, K.; Leroy, J.-L.; Guéron, M. *Nature* **1993**, *363*, 561.
8. Liu, D.; Balasubramanian, S. *Angew. Chem., Int. Ed.* **2003**, *42*, 5734.
9. Wang, Z.-G.; Elbaz, J.; Willner, I. *Nano Lett.* **2011**, *11*, 304.
10. Wang, W.; Liu, H.; Liu, D.; Xu, Y.; Yang, Y.; Zhou, D. *Langmuir* **2007**, *23*, 11956.
11. Sharma, J.; Chhabra, R.; Yan, H.; Liu, Y. *Chem. Commun. (Camb).* **2007**, 477.
12. Elbaz, J.; Wang, Z.-G.; Orbach, R.; Willner, I. *Nano Lett.* **2009**, *9*, 4510.
13. Yang, Y.; Liu, G.; Liu, H.; Li, D.; Fan, C.; Liu, D. *Nano Lett.* **2010**, *10*, 1393.
14. Li, T.; Ackermann, D.; Hall, A. M.; Famulok, M. *J. Am. Chem. Soc.* **2012**, *134*, 3508.
15. Modi, S.; Swetha, M. G.; Goswami, D.; Gupta, G. D.; Mayor, S.; Krishnan, Y. *Nat. Nanotechnol.* **2009**, *4*, 325.
16. Surana, S.; Bhat, J. M.; Koushika, S. P.; Krishnan, Y. *Nat. Commun.* **2011**, 2. article number 340.
17. Han, X.; Leroy, J.; Guéron, M. *J. Mol. Biol.* **1998**, *278*, 949.
18. Phan, A. T.; Guéron, M.; Leroy, J. L. *J. Mol. Biol.* **2000**, *299*, 123.
19. Marsh, R. E.; Bierstedt, R.; Eichhorn, E. L. *Acta Crystallogr.* **1962**, *15*, 310.
20. Ahmed, S.; Henderson, E. *Nucleic Acids Res.* **1992**, *20*, 507.
21. Leroy, J. L.; Gehring, K.; Kettani, A.; Guéron, M. *Biochemistry* **1993**, *32*, 6019.
22. Kanaori, K.; Maeda, A.; Kanehara, H.; Tajima, K.; Makino, K. *Biochemistry* **1998**, *37*, 12979.
23. Kanaori, K.; Shibayama, N.; Gohda, K.; Tajima, K.; Makino, K. *Nucleic Acids Res.* **2001**, *29*, 831.
24. Guéron, M.; Leroy, J. L. *Curr. Opin. Struct. Biol.* **2000**, *10*, 326.
25. Nonin-Lecomte, S.; Leroy, J. L. *J. Mol. Biol.* **2001**, *309*, 491.
26. Kang, C.; Berger, I.; Lockshin, C.; Ratliff, R.; Moyzis, R.; Rich, A. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 3874.
27. Weil, J.; Min, T.; Yang, C.; Wang, S.; Sutherland, C.; Sinha, N.; Kang, C. *Acta Crystallogr. Sect., D Biol. Crystallogr.* **1999**, *55*, 422.
28. Canalia, M.; Leroy, J.-L. *J. Am. Chem. Soc.* **2009**, *131*, 12870.
29. Leroy, J. L. *J. Mol. Biol.* **2003**, *333*, 125.
30. Cai, L.; Chen, L.; Raghavan, S.; Ratliff, R.; Moyzis, R.; Rich, A. *Nucleic Acids Res.* **1998**, *26*, 4696.
31. Catasti, P.; Chen, X.; Deaven, L. L.; Moyzis, R. K.; Bradbury, E. M.; Gupta, G. *J. Mol. Biol.* **1997**, *272*, 369.
32. Nonin, S.; Phan, A. T.; Leroy, J.-L. *Structure* **1997**, *5*, 1231.
33. Gallego, J.; Chou, S. H.; Reid, B. R. *J. Mol. Biol.* **1997**, *273*, 840.
34. Escaja, N.; Viladoms, J.; Garavís, M.; Villasante, A.; Pedroso, E.; González, C. *Nucleic Acids Res.* **2012**, *40*, 11737.
35. Lieblein, A. L.; Buck, J.; Schlepckow, K.; Fürtig, B.; Schwalbe, H. *Angew. Chem., Int. Ed.* **2012**, *51*, 250.
36. Yang, Y.; Sun, Y.; Yang, Y.; Xing, Y.; Zhang, T.; Wang, Z.; Yang, Z.; Liu, D. *Macromolecules* **2012**, *45*, 2643.
37. Brooks, T. A.; Kendrick, S.; Hurley, L. *FEBS J.* **2010**, *277*, 3459.
38. Malliavin, T. E.; Gau, J.; Snoussi, K.; Leroy, J.-L. *Biophys. J.* **2003**, *84*, 3838.

39. Leroy, J.-L.; Snoussi, K.; Guéron, M. *Magn. Reson. Chem.* **2001**, *39*, S171.
40. Kanehara, H.; Mizuguchi, M.; Tajima, K.; Kanaori, K.; Makino, K. *Biochemistry* **1997**, *36*, 1790.
41. Mergny, J. L.; Lacroix, L. *Nucleic Acids Res.* **1998**, *26*, 4797.
42. Lacroix, L.; Mergny, J. L. *Arch. Biochem. Biophys.* **2000**, *381*, 153.
43. Kanaori, K.; Sakamoto, S.; Yoshida, H.; Guga, P.; Stec, W.; Tajima, K.; Makino, K. *Biochemistry* **2004**, *43*, 5672.
44. Brazier, J. A.; Fisher, J.; Cosstick, R. *Angew. Chem., Int. Ed.* **2005**, *45*, 114.
45. Lacroix, L.; Mergny, J. L.; Leroy, J. L.; Hélène, C. *Biochemistry* **1996**, *35*, 8715.
46. Collin, D.; Gehring, K. *J. Am. Chem. Soc.* **1998**, *120*, 4069.
47. Robidoux, S.; Damha, M. J. *J. Biomol. Struct. Dyn.* **1997**, *15*, 529.
48. Chakraborty, S.; Krishnan, Y. *Biochimie* **2008**, *90*, 1088.
49. Fenna, C. P.; Wilkinson, V. J.; Arnold, J. R. P.; Cosstick, R.; Fisher, J. *Chem. Commun.* **2008**, 3567.
50. Diederichsen, U. *Angew. Chem., Int. Ed.* **1998**, *37*, 2273.
51. Krishnan-Ghosh, Y.; Stephens, E.; Balasubramanian, S. *Chem. Commun.* **2005**, 5278.
52. Sharma, N. K.; Ganesh, K. N. *Chem. Commun.* **2005**, 4330.
53. Modi, S.; Wani, A. H.; Krishnan, Y. *Nucleic Acids Res.* **2006**, *34*, 4354.
54. Chakraborty, S.; Modi, S.; Krishnan, Y. *Chem. Commun.* **2008**, 70.
55. Mergny, J.; Lacroix, J. L.; Han, X.; Leroy, J.; Helene, C. *J. Am. Chem. Soc.* **1995**, *117*, 8887.
56. Bhavsar-jog, Y. P.; Dornshuld, E.; Van Brooks, T. A.; Tschumper, G. S.; Wadkins, R. M. *Biochemistry* **2014**, *53*, 1586.
57. Zhou, J.; Wei, C.; Jia, G.; Wang, X.; Feng, Z.; Li, C. *Mol. Biosyst.* **2010**, *6*, 580.
58. Miyoshi, D.; Matsumura, S.; Nakano, S.-I.; Sugimoto, N. *J. Am. Chem. Soc.* **2004**, *126*, 165.
59. Cui, J.; Waltman, P.; Le, V. H.; Lewis, E. A. *Molecules* **2013**, *18*, 12751.
60. Rajendran, A.; Nakano, S.; Sugimoto, N. *Chem. Commun.* **2010**, 1299.
61. Phan, A. T.; Mergny, J.-L. *Nucleic Acids Res.* **2002**, *30*, 4618.
62. Li, W.; Wu, P.; Ohmichi, T.; Sugimoto, N. *FEBS Lett.* **2002**, *526*, 77.
63. Li, W.; Miyoshi, D.; Nakano, S.; Sugimoto, N. *Biochemistry* **2003**, *42*, 11736.
64. König, S. L. B.; Huppert, J. L.; Sigel, R. K. O.; Evans, A. C. *Nucleic Acids Res.* **2013**, *41*, 7453.
65. Sun, D.; Hurley, L. H. *J. Med. Chem.* **2009**, *52*, 2863.
66. Saxena, S.; Bansal, A.; Kukreti, S. *Arch. Biochem. Biophys.* **2008**, *471*, 95.
67. Benabou, S.; Ferreira, R.; Aviñó, A.; González, C.; Lyonnais, S.; Solà, M.; Eritja, R.; Jaumot, J.; Gargallo, R. *Biochim. Biophys. Acta* **2014**, *41*, 1840.
68. Day, H. A.; Huguin, C.; Waller, Z. A. E. *Chem. Commun.* **2013**, 7696.
69. Goncharova, I. *Spectrochim. Acta. A. Mol. Biomol. Spectrosc.* **2014**, *118*, 221.
70. Huppert, J. L.; Balasubramanian, S. *Nucleic Acids Res.* **2007**, *35*, 406.
71. Mergny, J.-L.; Riou, J.-F.; Mailliet, P.; Teulade-Fichou, M.-P.; Gilson, E. *Nucleic Acids Res.* **2002**, *30*, 839.
72. Kim, N. W.; Piatyszek, M. A.; Prowse, K. R.; Harley, C. B.; West, M. D.; Ho, P. L. C.; Coviello, G. M.; Wright, W. E.; Weinrich, S. L.; Shay, J. W. *Science* **2011**, *1994*, 266.
73. Balasubramanian, S.; Neidle, S. *Curr. Opin. Chem. Biol.* **2009**, *13*, 345.
74. Chen, Y.; Qu, K.; Zhao, C.; Wu, L.; Ren, J.; Wang, J.; Qu, X. *Nat. Commun.* **2012**, 3. article number 1074.
75. Leroy, J. L.; Guéron, M.; Mergny, J. L.; Hélène, C. *Nucleic Acids Res.* **1994**, *22*, 1600.
76. Marsich, E.; Piccini, A.; Xodo, L. E.; Manzini, G. *Nucleic Acids Res.* **1996**, *24*, 4029.
77. Lacroix, L.; Liénard, H.; Labourier, E.; Djavaheri-Mergny, M.; Lacoste, J.; Leffers, H.; Tazi, J.; Hélène, C.; Mergny, J. L. *Nucleic Acids Res.* **2000**, *28*, 1564.
78. Simonsson, T.; Pribylova, M.; Vorlickova, M. *Biochem. Biophys. Res. Commun.* **2000**, *278*, 158.
79. González, V.; Hurley, L. H. *Annu. Rev. Pharmacol. Toxicol.* **2010**, *50*, 111.
80. Berberich, S. J.; Postel, E. H. *Oncogene* **1995**, *10*, 2343.
81. Brooks, T. A.; Hurley, L. H. *Genes Cancer* **2010**, *1*, 641.
82. Dexheimer, T. S.; Carey, S. S.; Zuohe, S.; Gokhale, V. M.; Hu, X.; Murata, L. B.; Maes, E. M.; Weichsel, A.; Sun, D.; Meuillet, E. J.; Montfort, W. R.; Hurley, L. H. *Mol Cancer Ther.* **2009**, *8*, 1363.
83. Kendrick, S.; Akiyama, Y.; Hecht, S. M.; Hurley, L. H. *J. Am. Chem. Soc.* **2009**, *131*, 17667.
84. Vaux, D. L.; Cory, S.; Adams, J. M. *Nature* **1988**, *335*, 440.
85. Joensuu, H.; Pyikkanen, L.; Toikkanen, S. *Am. J. Pathol.* **1994**, *145*, 1191.
86. Bar-Am, O.; Weinreb, O.; Amit, T.; Youdim, M. B. H. *FASEB J.* **2005**.
87. Heckman, C.; Mochon, E.; Arcinas, M.; Boxer, L. M. *J. Biol. Chem.* **1997**, *272*, 19609.
88. Heckman, C. A.; Mehew, J. W.; Boxer, L. M. *Oncogene* **2002**, *21*, 3898.
89. Kendrick, S.; Kang, H.-J.; Alam, M.; Madathil, M.; Agrawal, P.; Gokhale, V.; Yang, D.; Hecht, S. M.; Hurley, L. H. *J. Am. Chem. Soc.* **2014**, *136*, 4161.
90. Kang, H.-J.; Kendrick, S.; Hecht, S. M.; Hurley, L. H. *J. Am. Chem. Soc.* **2014**, *136*, 4172.
91. Lee, D.-H.; Lim, M.-H.; Youn, D.-Y.; Jung, S. E.; Ahn, Y. S.; Tsujimoto, Y.; Lee, J.-H. *Biochem. Biophys. Res. Commun.* **2009**, *382*, 583.
92. Cui, Y.; Koirala, D.; Kang, H.; Dhakal, S.; Yangyuoru, P.; Hurley, L. H.; Mao, H. *Nucleic Acids Res.* **2014**, 1.
93. Nevins, J. R. *Hum. Mol. Genet.* **2001**, *10*, 699.
94. Xu, Y.; Sugiyama, H. *Nucleic Acids Res.* **2005**, *49*, 177.
95. Xu, Y.; Sugiyama, H. *Nucleic Acids Res.* **2006**, *34*, 949.
96. Guo, K.; Pourpak, A.; Beetz-Rogers, K.; Gokhale, V.; Sun, D.; Hurley, L. H. *J. Am. Chem. Soc.* **2007**, *129*, 10220.
97. Kodama, Y.; Asai, N.; Kawai, K.; Jijiwa, M.; Murakumo, Y.; Ichihara, M.; Takahashi, M. *Cancer Sci.* **2005**, *96*, 143.
98. Guo, K.; Gokhale, V.; Hurley, L. H.; Sun, D. *Nucleic Acids Res.* **2008**, *36*, 4598.
99. Finkenzeller, G.; Sparacio, A.; Technau, A.; Marmé, D.; Siemeister, G. *Oncogene* **1997**, *15*, 669.
100. Jolad, V. V.; Murad, F. K.; Arnold, J. R. P.; Fisher, J. *Org. Biomol. Chem.* **2005**, *3*, 2234.
101. Kennedy, G. C.; German, M. S.; Rutter, W. J. *Nat. Genet.* **1995**, *9*, 293.
102. Dhakal, S.; Schonhoft, J. D.; Koirala, D.; Yu, Z.; Basu, S.; Mao, H. *J. Am. Chem. Soc.* **2010**, *132*, 8991.
103. Dhakal, S.; Lafontaine, J. L.; Yu, Z.; Koirala, D.; Mao, H. *PLoS ONE* **2012**, *7*, e39271.
104. Dhakal, S.; Yu, Z.; Konik, R.; Cui, Y.; Koirala, D.; Mao, H. *Biophys. J.* **2012**, *102*, 2575.
105. Manzini, G.; Yathindra, N.; Xodo, L. E. *Nucleic Acids Res.* **1994**, *22*, 4634.
106. Hoffman, E. K.; Trusko, S. P.; Murphy, M.; George, D. L. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 2705.
107. Bucek, P.; Jaumot, J.; Aviñó, A.; Eritja, R.; Gargallo, R. *Chem. A Eur. J.* **2009**, *15*, 12663.
108. Bucek, P.; Gargallo, R.; Kudrev, A. *Anal. Chim. Acta* **2010**, *683*, 69.
109. Phan, A. T.; Kuryavyi, V.; Burge, S.; Neidle, S.; Patel, D. J. *J. Am. Chem. Soc.* **2007**, *129*, 4386.
110. Brazier, J. A.; Shah, A.; Brown, G. D. *Chem. Commun.* **2012**, 10739.
111. Backe, P. H.; Messias, A. C.; Ravelli, R. B. G.; Sattler, M.; Cusack, S. *Structure* **2005**, *13*, 1055.
112. Marsich, E.; Xodo, L. E.; Manzini, G. *Eur. J. Biochem.* **1998**, *258*, 93.
113. Cornuel, J. F.; Moraillon, A.; Guéron, M. *Biochimie* **2002**, *84*, 279.
114. Eid, J. E.; Sollner-Webb, B. *Mol. Cell. Biol.* **1995**, *15*, 389.
115. Sarig, G.; Weisman-shomer, P.; Erlitzki, R.; Fry, M. *J. Biol. Chem.* **1997**, *272*, 4474.
116. Cortés, A.; Huertas, D.; Fanti, L.; Pimpinelli, S.; Marsellach, F. X.; Piña, B.; Azorín, F. *EMBO J.* **1999**, *18*, 3820.
117. Cortés, A.; Azorín, F. *Mol. Cell. Biol.* **2000**, *20*, 3860.
118. Fedoroff, O. Y.; Rangan, A.; Chemeris, V. V.; Hurley, L. H. *Biochemistry* **2000**, *39*, 15083.
119. Martino, L.; Pagano, B.; Fotticchia, I.; Neidle, S.; Giancola, C. *J. Phys. Chem. B* **2009**, *113*, 14779.
120. Fernández, S.; Eritja, R.; Aviñó, A.; Jaumot, J.; Gargallo, R. *Int. J. Biol. Macromol.* **2011**, *49*, 729.
121. Siddiqui-Jain, A.; Grand, C. L.; Bearss, D. J.; Hurley, L. H. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 11593.
122. Grand, C. L.; Han, H.; Muñoz, R. M.; Weitman, S.; Von Hoff, D. D.; Hurley, L. H.; Bearss, D. J. *Mol. Cancer Ther.* **2002**, *1*, 565.
123. Alberti, P.; Ren, J.; Teulade-Fichou, M.-P.; Guittat, L.; Riou, J.-F.; Chaires, J. B.; Helene, C.; Vigneron, J.; Lehn, J.-M.; Mergny, J.-L. *J. Biomol. Struct. Dyn.* **2001**, *19*, 505.
124. Li, X.; Peng, Y.; Ren, J.; Qu, X. *Proc. Natl. Acad. Sci.* **2006**, *103*, 19658.
125. Zhao, C.; Ren, J.; Qu, X. *Chem. A Eur. J.* **2008**, *14*, 5435.
126. Peng, Y.; Li, X.; Ren, J.; Qu, X. *Chem. Commun.* **2007**, 5176.
127. Gomez, D.; Mergny, J.; Riou, J. *Cancer Res.* **2002**, *62*, 3365.
128. DeLange, T. *Science* **2009**, *326*, 948.
129. Chen, X.; Zhou, X.; Han, T.; Wu, J.; Zhang, J.; Guo, S. *ACS Nano* **2013**, *7*, 531.
130. Li, X.; Peng, Y.; Qu, X. *Nucleic Acids Res.* **2006**, *34*, 3670.
131. Zhou, X.; Zhang, Y.; Wang, C.; Wu, X.; Yang, Y.; Zheng, B.; Wu, H.; Guo, S.; Zhang, J. *ACS Nano* **2012**, *6*, 6592.
132. Wang, L.; Wu, Y.; Chen, T.; Wei, C. *Int. J. Biol. Macromol.* **2013**, *52*, 1.
133. Xue, L.; Ranjan, N.; Arya, D. P. *Biochemistry* **2011**, *50*, 2838.
134. Ma, D.-L.; Kwan, M. H.-T.; Chan, D. S.-H.; Lee, P.; Yang, H.; Ma, V. P.-Y.; Bai, L.-P.; Jiang, Z.-H.; Leung, C.-H. *Analyst* **2011**, *136*, 2692.
135. Kong, D.-M.; Ma, Y.-E.; Wu, J.; Shen, H.-X. *Chem. A Eur. J.* **2009**, *15*, 901.
136. Xu, H.; Zhang, H.; Qu, X. *J. Inorg. Biochem.* **2006**, *100*, 1646.
137. Shi, S.; Geng, X.; Zhao, J.; Yao, T.; Wang, C.; Yang, D.; Zheng, L.; Ji, L. *Biochimie* **2010**, *92*, 370.
138. Shi, S.; Zhao, J.; Geng, X.; Yao, T.; Huang, H.; Liu, T.; Zheng, L.; Li, Z.; Yang, D.; Ji, L. *Dalt. Trans.* **2010**, *39*, 2490.
139. Lee, I. J.; Yi, J. W.; Kim, B. H. *Chem. Commun.* **2009**, 5383.
140. Lee, I. J.; Kim, B. H. *Chem. Commun.* **2012**, 2074.
141. Perlíková, P.; Karlsen, K. K.; Pedersen, E. B.; Wengel, J. *ChemBioChem* **2014**, *15*, 146.
142. El-Sayed, A. A.; Pedersen, E. B.; Khaireldin, N. A. *Nucleosides. Nucleotides Nucleic Acids* **2012**, *31*, 872.
143. Park, J. W.; Seo, Y. J.; Kim, B. H. *Chem. Commun. (Camb).* **2014**, *50*, 52.
144. Lee, I. J.; Park, M.; Joo, T.; Kim, B. H. *Mol. Biosyst.* **2012**, *8*, 486.
145. Cohen, G. M. *Biochem. J.* **1997**, *326*, 1.
146. Weil, J.; Min, T.; Yang, C.; Wang, S.; Sutherland, C.; Sinha, N.; Kang, C. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **1999**, *55*, 422.
147. Chen, L.; Cai, L.; Zhang, X.; Rich, A. *Biochemistry* **1994**, *33*, 13540.
148. Kang, C. H.; Berger, I.; Lockshin, C.; Ratliff, R.; Moyzis, R.; Rich, A. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 11636.