



HAL
open science

i-Motif DNA: identification, formation, and cellular functions

Shentong Tao, Yonghang Run, David Monchaud, Wenli Zhang

► **To cite this version:**

Shentong Tao, Yonghang Run, David Monchaud, Wenli Zhang. i-Motif DNA: identification, formation, and cellular functions. Trends in Genetics, In press, 10.1016/j.tig.2024.05.011 . hal-04618449

HAL Id: hal-04618449

<https://hal.science/hal-04618449>

Submitted on 20 Jun 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

i-motif DNA: identification, formation, and cellular functions

Shentong Tao^{1,#}, Yonghan Run^{1,#}, David Monchaud², Wenli Zhang^{1,*}

¹State Key Laboratory of Crop Genetics and Germplasm Enhancement and Utilization, CIC-MCP, Nanjing Agricultural University, No.1 Weigang, Nanjing, Jiangsu 210095, China. ²Institut de Chimie Moleculaire, ICMUB CNRS UMR 6302, UBFC Dijon, France.

These authors contributed equally to this work. *Correspondence: wzhang25@njau.edu.cn (W.L. Zhang).

Abstract

An i-motif (iM) is a four-stranded (or quadruplex) DNA structure that folds from cytosine (C)-rich sequences. iMs can fold in many different conditions *in vitro*, which paves the way for their formation in living cells. Being suspected, iMs play key roles in various DNA transactions, notably in the regulation of genome stability, gene transcription and translation, DNA replication, telomere and centromere functions, and human diseases. We summarize here the different techniques used to assess the folding of iMs *in vitro* and provide an overview on both the internal and external factors that affect their formation and stability *in vivo*. We thus describe the possible biological relevance of iM and propose directions towards its use as target in biology.

Keywords

i-motifs, methodologies, base modifications, internal and external conditions, biological implications

Highlights

- The combination of different methodologies and molecular tools is useful to interrogate the dynamics of iMs in the genome.
- High throughput, sequencing-based methodologies can provide a global landscape of putative iM-forming sequences (PiMFSs) across the genome and help unveil their potential biological implications.
- The intrinsic nature of the sequence along with its epigenetic modifications and other environmental factors determines the folding and stability of iMs both *in vitro* or *in vivo*.
- Epigenetic modifications play complex roles in the control of the folding and stability of iMs both *in vitro* or *in vivo*, which is dependent on the position, number and types of modified bases.
- iMs are involved in various DNA transactions in all living organisms, being involved in the control of the genome stability, telomere and centromere biology, gene transcription and translation, DNA replication and repair.
- Beyond their possible use as therapeutic targets, and given their high sensitivity to external factors, iMs also find applications in biotechnology and nanotechnology, as biosensors for instance.

Glossary

- Non-B-DNA structures: alternative DNA structures that deviate from the canonical double helix structure of DNA (also known as duplex-DNA, B-DNA or Watson-Crick duplex).
- i-motifs (iMs): noncanonical quadruplex nucleic acid structures built with intercalated C-C+ base pairs, which fold from cytosine (C)-rich sequences with C stretches (also referred to as I-DNA).
- R-loop: a three-stranded nucleic acid structure consisting of an RNA-DNA hybrid and a displaced single-stranded DNA (also referred to as R-DNA).
- G-quadruplexes (G4s): noncanonical quadruplex nucleic acid structures built with G-quartets, which fold from guanine (G)-rich sequence with G stretches (also referred to as G-DNA).
- Putative i-motif forming sequences (PiMFSs): cytosine (C)-rich nucleotide sequences able to fold into an i-motif structure.
- Triplex: a three-stranded nucleic acid structure containing homopurine (hPu)-homopyrimidine (hPy) regions, also referred to as hinged DNA (or H-DNA).
- Z-DNA: a left-handed double helical nucleic acid structure folding from sequences containing alternating pyrimidines and purines, in a global zig-zag (Z) shape.
- Cruciform: a noncanonical DNA structure that folds from inverted repeats made of two hairpin-loop arms directly opposite to each other (also referred to as C-DNA, or 4-way DNA junction, FWJs).
- Hairpins: an intramolecular fold that forms a stem-loop structure formed by consecutive inverted repeats that associate to form a duplex (stem), separated by a single-stranded sequence (loop).
- Slipped-strand DNA: direct repeats paired between repeats in misaligned fashion leading to hairpin-like structures (also referred to as S-DNA, or 3-way DNA junction, TWJs).
- Cis-regulatory DNA elements (CREs): a type of DNA (typically of noncoding DNA) sequences containing intrinsic binding sites for trans-acting factors, essential for the regulation of gene transcription.
- iMab: an antibody fragment (single-chain variable fragment, scFv) specific for i-motif structures.
- iM-IP-seq: an iMab-based immunoprecipitation of iMs followed by next-generation sequencing for genome-wide identification of i-motifs.
- Cleavage Under Targets and Tagmentation (CUT&Tag): an enzyme-tethering strategy for genome-wide profiling of iMs using iMab combined with a protein A-Tn5 transposase fusion protein that generates fragment libraries processed by high-throughput sequencing.
- Chromatin Immunoprecipitation and Sequencing (ChIP-seq): a technique related to immunoprecipitation of iM in the chromatin context using iMab followed by high-throughput sequencing, which can be implemented with other specific antibodies for profiling transcript factors, functional proteins or histone modifications across the genome.
- G4-iM Grinder: an in silico tool for the characterization and selection of DNA sequences with potentials of folding G-quadruplexes (G4s) and i-motifs (iMs) based on ad hoc algorithms.
- iM-seeker: a machine learning toolkit used for predicting the folding status and strength of

- iMs through a Balanced Random Forest classifier trained on CUT&Tag sequencing data.
- G4hunter: a G-quadruplex predicting algorithm which takes into account both G-richness and G-skewness of given DNA sequences to provide G4-forming propensity scores (G4H scores).
 - Transposable elements (TEs): mobile genetic DNA elements, which can move within the genome once the chromatin conditions enable such motions.

Outstanding questions

- How can we develop a sensitive methodology to visualize or capture transient conformational changes of genomic iMs? Under favorable or unfavorable cellular conditions? How to detect hybrid iMs formed with DNA and RNA molecules?
- How the combination of multiple factors including the sequence per se along with epigenetic modifications and various environmental factors affects the folding/unfolding of iMs in vitro or in vivo? How DNA modifications and/or chromatin dynamics affect iM dynamics across the genome?
- The debate about the folding of iMs under physiological conditions being still open, what are the key determinants that control the folding dynamics, topology and stability of iMs in vivo?
- What are the roles that iMs play in the regulation of gene transcription during normal developmental processes or in response to cellular/environmental stress? Can iM structures instead of their primary sequences act as cis-regulatory elements or trans-factors binding hubs responsible for the regulation of gene transcription?
- Do the conditions affecting the formation and stability of DNA iMs affect the formation and stability of RNA iMs as well? And what are the biological implications of RNA iMs in vivo?
- Do iMs located in regulatory regions across the genome experience natural or artificial selection for controlling key phenotypical traits?

Introduction

i-motif is a special type of quadruplex DNA structure. An i-motif (iM), also known as an i-DNA, is a type of quadruplex (four-stranded) non-B DNA structures; it folds from cytosine (C)-rich DNA sequences. iM structures consist of antiparallel duplex ladders through intercalation of hemiprotonated C:C⁺ base-pair [1, 2]. The unique quadruplex structure is distinct from another typical quadruplex structure called guanine (G) quadruplex (G4) that forms from G-rich DNA context. A G4 is composed of stacked G-tetrad formed by Hoogsteen hydrogen bonds between four guanines [3]. Initially, there is a widely accepted view that iM folding is pH-dependent, which prefers to fold at acidic pH rather than at basic pH conditions. Recently, an iMab antibody-based pioneer study provides compelling evidence showing the existence of *in vivo* iMs under physiological conditions in human cells [4], which serves as a turning point to boost iM study and advance understanding of its biology in eukaryotic genomes. Accumulating evidence shows that iM folding can be affected by combined actions of various internal or external factors, which still need more detailed investigations.

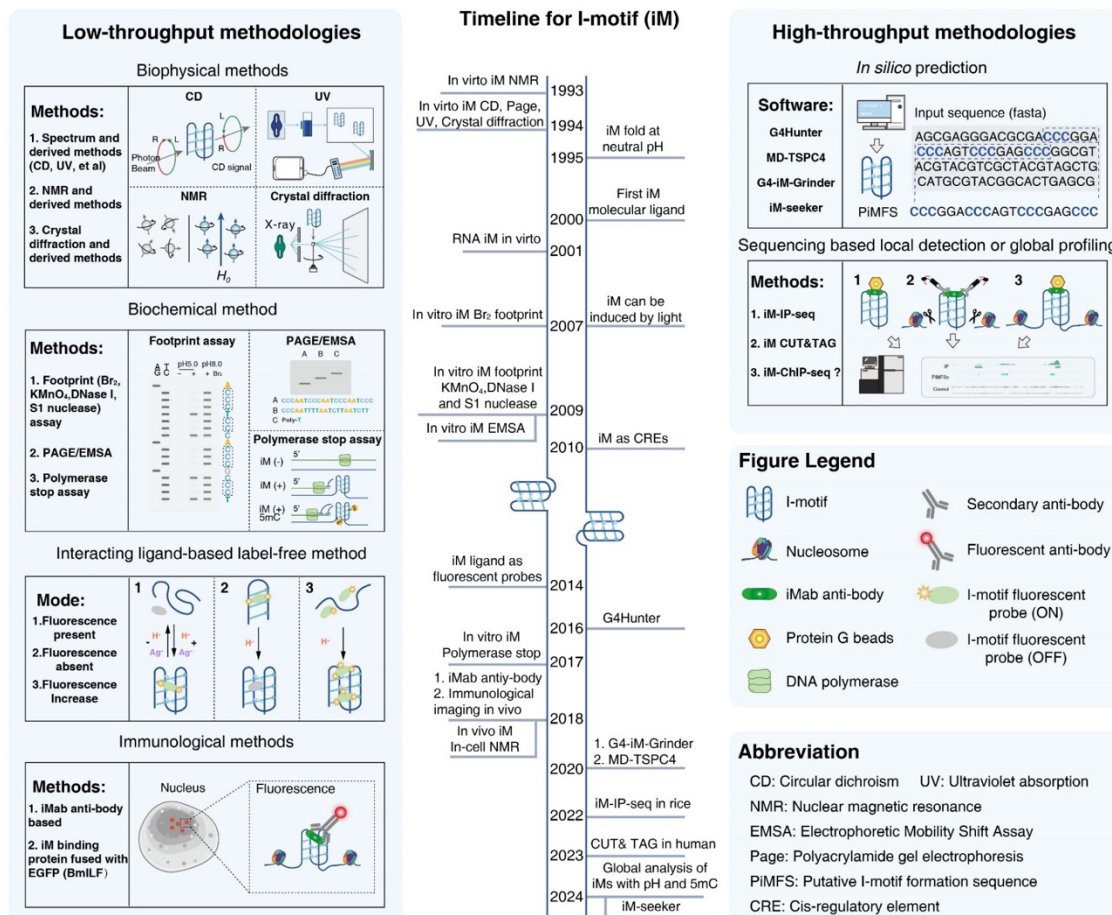


Figure 1. Summary of existing methodologies for identification of DNA iMs and timeline for important case studies. Existing methodologies include low- and high-throughput methods. Low-throughput methods can be divided into 4 categories as follows: (1) Biophysical methods, (2) Biochemical methods, (3) Interacting ligand-based label-free methods, and (4) Immuno-detection methods. High-throughput ones can be mainly classified into the following 2 categories: (1) In silico prediction, and (2) Sequencing-based global profiling. Time point for the first application of all methodologies and important case studies was summarized as timeline during 1993-2024. Abbreviations were present in the lower right corner of the figure.

Existing methodologies for the identification of DNA i-motifs. The direct detection of i-motifs (iMs) *in vitro* or *in vivo* is still technically challenging. To this end, the existing methodologies can be divided into six categories (Table 1 and **Figure 1**): biophysical, biochemical, immunological and ligand-based detection methods, along with *in silico* prediction and sequencing-based global profiling.

Table 1. Summary of existing methodologies for identification of DNA iMs

Categories	Subtypes	Methodologies	Ref.
Biophysical methods	Spectrometry	Circular Dichroism (CD)	[5-10]
		Surface-Enhanced Raman Spectroscopy (SERS)	
		Ensemble Force spectroscopy (EFS)	
		Data analysis strategy:	
		Deconvolution with Multivariate Curve Resolution-Alternating Least-Squares (MCR-ALS)	

		Time-Dependent DFT calculations Global fitting analysis	
	UV-based Spectrometry	UV absorption spectroscopies UV absorption melting curves Capillary Electrophoresis with an Ultraviolet absorption spectrophotometric detection (CE-UV)	[5, 8, 11]
	NMR	Nuclear Magnetic Resonance (NMR) In-cell NMR	[1, 12]
	Crystallography	Single crystal x-ray diffraction	[13, 14]
	Fluorescence based Spectrometry	Fluorescence Resonance Energy Transfer (FRET) Fluorescent Intercalator Displacement (FID)	[5, 15]
	Single-molecule based	Single-molecule fluorescence microscopy Fluorescence Resonance Energy Transfer (FRET)-based single-molecule analyses (smFRET) Nanopore-based single-molecule investigation and all-atom molecular dynamics simulations	[16, 17]
	Molecular Simulation	Computer simulations based on atomistic and coarse-grained force fields	[18]
	Thermal Analysis	Differential Scanning Calorimetry (DSC) (Microcalorimetry experiments)	[19]
Biochemical methods	Footprinting assays	Bromine, KMnO ₄ , DNase I or S1 nuclease	[20, 21]
	Mobility shift assay	Native polyacrylamide gel electrophoresis (PAGE) or electrophoretic mobility shift assay (EMSA)	[22]
	DNA polymerase stop assay	KOD-Plus DNA polymerase	[23]
Interacting ligand-based label-free detection	Presence of fluorescence signal	Thiazole orange (TO) as a fluorescent probe to indicate pH/Ag ⁺ triggered conformation changes of i-motif	[24]
	Loss of fluorescence signal	iM formation excludes binding of Thioflavin T (ThT) thereby quench its fluorescence signal derived from interacting with a hairpin structure	[25]
	Increase in fluorescence signal	iM can disrupt aggregation of PTCDI results in an increase in fluorescence signal	[26]
Immuno-detection	Antibody	iM-specific antibody (iMab)	[27]
	iM binding protein	BmILF or antibody specifically recognizing BmILF in <i>Bombyx mori</i> fused with EGFP	[28, 29]
Sequencing based	iM-IP-seq	iM-IP-seq in vitro iMs in rice (iMab antibody)	[30]
	CUT&TAG	CUT&TAG in vivo iMs in human (iMab antibody)	[4]
	ChIP-seq (?)	ChIP-seq of iMs in human (not officially published)	[4]
In silico prediction	G-quadruplex (G4)	G4hunter; G4-iM Grinder	[31, 32]
	iM	MD-TSPC4 (Predicting thermal stability) iM-seeker (A machine learning toolkit trained with CUT & TAG data)	[33-35]

Each of these methods has its own pros and cons. For instance, UV and CD spectroscopies are the most straightforward way to detect or validate iMs, and other secondary DNA structures [36], since the way the DNA folds results in differences of absorbance spectra, unique for each secondary structure [37], but these techniques are low-throughput. The difference in absorbance signature between folded and unfolded states upon temperature variations has been exploited to characterize these structures by their thermal difference spectra (TDS), which is to date the simplest (and one of the less expensive) method to interrogate secondary structures. CD is very insightful but spectropolarimeters are quite expensive as the underlying technology relies on polarized light (more precisely the difference of absorbance of left- *versus* right-handed polarized light directly originating in the asymmetry of the biomolecule), and the interpretation of CD signatures are global, that is, challenging if multiple structures co-exist in solution [38]. Another issue is related to the oligonucleotides used for these investigations: most often, biophysical studies rely on short synthetic oligonucleotides containing putative iM-forming sequences (PiMFSs) only, failing in recapitulating the possible influence of the flanking sequences and/or the presence of the complementary strand on the formation of iMs [39].

More structurally accurate methods such as NMR have proven useful for structural investigations, and can even be performed in-cell under functional conditions [40], which is an invaluable advantage but again, this technique is dependent on expensive machines and skillful manipulators. Capillary electrophoresis and PAGE are widely applied methods for distinguishing unfolded from partially or fully folded iMs according to peak shape and migration time [41], but again they are low-resolution methods.

Chemical biology investigations aiming at studying cellular iMs are currently being undertaken but they suffer from the lack of reliable iM-ligands. Indeed, some promising prototypes have been uncovered but their use is not unambiguously validated [42], which casts doubts about the reliability of the results' interpretation. Also, this handful of prototypes do not have desirable features (such as being manipulated *in situ* by biorthogonal techniques, such as click chemistry) for being amenable to iM cellular detection, which further limits their use. Additionally, ligand-based investigations in cells have known limitations, the first being that the molecule may promote the iM formation, thereby biasing their *bone fide* detection. To tackle this issue, classical immunodetection assays can be implemented but again, this approach suffers some caveats, the first being the reliability of the antibody used. As further described below, the only available anti-iM antibody (iMab) is currently at the heart of a controversy that still remains to be addressed. Additionally, immunofluorescence (IF) detection is a low-resolution approach, which cannot be reliably quantitative [43], its implementation relies on quite harsh cellular treatments (detergent, fixation agents) that may strongly impact the reliability of the iM detection.

In silico predictions underestimate the number of genomic iMs as they do not take non-canonical iMs with sequence variations into account [39, 44]. These predictive algorithms are yet high-throughput but the relevance and reliability of their outcomes must be experimentally validated, which is time-demanding.

Sequencing-based methods, such as ChIP-seq [45], CUT&Tag [4] and iM-IP-seq [30], are high-throughput and suited to global profiling of *in vitro* and *in vivo* iMs, but they are expensive

and rely on extensive data mining, which might be bias-prone. The cellular manipulations necessary to perform these investigations (sample fixation, DNA fragmentation, library preparation) might be bias-prone as well [44]. Additionally, the resolution of these techniques is still not high enough and, last but not least, they rely on the use of iMab, which is still a matter of debate since it can bind single-stranded DNA with at least two consecutive Cs in unfolding state [46], likely leading to overestimation of iMs.

Altogether, none of the aforementioned methods can be applied alone for asserting *in vitro* or *in vivo* iM formation and only a combination of several of these methods can provide reliable insights into the actual iM folding.

Impacts of pH alone or along with other factors on iM formation. Acidic pH is considered to be the key determinant for iM formation and stability since its initial discovery in 1993 [1, 2], which is largely due to the necessity of protonating cytosines (Cs) in order to create C-C⁺ base pairs as its basic building block [47]. Its dependency on pH conditions was exploited to create iM-based pH sensors or pH-dependent nano-devices [48]. The widely accepted notion was recently challenged by a study showing that iM may fold under normal cellular conditions [27]. Accumulating evidence now shows that many other internal and external factors can modulate iM formation and stability under a wide range of pH conditions (Table 2).

Table 2. Summary of impacts of internal and external factors along with pH on iM formation

Groups ^a	Factors	Conditions ^b	Effects	Sequences ^c	Ref.
Int	DNA sequence	Length and sequence of Loop	Long loop in 1st and 3rd loop contribute to stability of iM at neutral pH	(C)l-(N)m(T)n(N)m-CCC-(T)3-(C)l(N)m(T)n(N)m-CCC-3	[49]
		Length of Loop and C tracts	Sequences with a longer central spacer stabilize iM at neutral pH	C[3-6]D[1-6]C[3-6]D[1-15]C[3-6]D[1-6]C[3-6]	[50]
		Length of Loop and C tracts	d(C)n of length 15,19,23,27 (4n-1) fold into iM at neutral pH	Poly d(C) _n	[51]
		Length of Loop	longer loops prefer to form intramolecular i-motif structures	Different DNA C-rich sequence	[2]
		Composition of sequence	iM formation and stability is sensitive to their sequence and base composition	DNA sequence with variations in C-tracts/spacer length and arrangement	[52]
		Composition of loop sequence	Adenine bases within loops destabilize iM	(C ₃ -TA ₂) ₃ -CCC (C ₃ -TT ₂) ₃ -CCC	[53]

		Composition of sequence	Repetitive genomic sequence promote mini iM fold at neutral pH	d(TCGTTCCGT)	[54]	
	Capping interactions	Capping interactions	Capping interactions promote iM fold at neutral	d(TCGTTTCGT) d(TCGTTCCGT) d(CCGTTCCGT) and cyclic analog	[55, 56]	
	DNA supercoiling	Negative superhelicity	Promote iM fold at neutral pH	c-MYC promoter	[21]	
Ext	Temperature	Low temperature 4°C	Stabilize iM at neutral pH	HT	[5]	
	Pressure	High Pressure	iM was destabilized by high pressure when pH increased	HT	[57]	
	Cation	K ⁺ ; Na ⁺		Promote iM at acidic pH	HT	[58]
		Li ⁺		Destabilize iM at acidic pH	HT	[58]
		Ag ⁺		Stabilize iM at room temperature and neutral pH	HT	[59]
		Cu ⁺		destabilize iM without changing pH	HT	[60]
	Molecular crowding	PEG 8000		Stabilize iM at neutral pH	(CCT) _n repeat	[61]
		PEG 30% w/w PEG8000		Stabilize iM at neutral pH at PEG 30% w/w PEG8000 and destabilize when PEG increased	c-MYC promoter	[62]
		PEG-300		Stabilize iM at neutral pH	c-MYC promoter	[63]
	Nanomaterials	SWNT		Stabilize iM at neutral or even alkaline pH	HT	[64]
		GQDs		Stabilize iM at acidic pH promote iM fold at neutral or even alkaline pH	HT	[65]

Note: a, Abbreviations: Int, Internal factor; Ext, External factor. b, SWNT, single-wall carbon nanotubes; GQDs, graphene quantum dots; b, Sequence: HT, human telomeric sequence (CCC-[TAA-CCC]_n); c-MYC promoter (T-C₄-ACCTT-C₄-A-C₃-T-C₄-A-C₃-T-C₄-A).

The internal factors include the intrinsic nature of the DNA sequence itself (length of C tracts, nature and composition of the loops [50, 51]), capping interaction at nearby C-C⁺ stack [55, 56], and negative superhelicity [21]. For instance, for being stable at neutral pH, iMs must have long loops [49] and/or be folded under negative superhelicity [21]. Of course, the nature of the sequence is critical [50, 52, 66]: for instance, the repeated d(TCGTTCCGT) sequence, prevalent in the human genome, forms an iM at physiological that is stabilized by both C-C⁺

base pairs and G:C/G:T tetrad located in the minor groove [54]. The aforementioned iM-IP-seq techniques was used in plants (rice genome) and showed *ca.* 6,000 iMs formed uniquely at pH 5.5, *ca.* 20,000 iMs formed uniquely at pH 7.0, and *ca.* 39,000 iMs formed at both pH 5.5 and 7.0. These iMs exhibit distinct sequence features and genomic distributions, which advocated again for the critical roles of intrinsic DNA sequence and pH for iM formation [67].

Multiple external factors also affect formation of iMs, such as *i-* the temperature (at 4 °C, iMs are stable at different pH conditions [5, 57, 68, 69] because of the decrease in Brownian motion); *ii-* the pressure (iM was destabilized by high pressure when pH increased [57]); *iii-* the presence of cations (K⁺ favors iMs while Li⁺ does not [58], because of its smaller radius; Ag⁺ helps stabilize iMs [59] because of the formation of C-Ag(I)-C base pair [70], while Cu⁺ can destabilize i-motifs without changing the pH [60]); *iv-* the molecular crowding (for instance, PEG8000 stabilizes iMs at neutral or slightly basic pH [61-63]); and *v-* the nature of the solvent and co-solvent (the hydration state of iM being pivotal for its stability [63], as demonstrated for instance with glycerol, which stabilizes iMs by hydration while dimethoxymethane destabilizes them by dehydration).

Of note, alternative strategies have been implemented to artificially stabilize iMs at neutral or even alkaline pH conditions [64, 65], for instance by immobilizing them on the surface of nanoparticles (single-wall carbon nanotubes (SWNTs) or graphene quantum dots (GQDs), likely because the nanomaterials serve as a proton donor that facilitates the formation of C-C⁺ at non-acidic pH conditions [71]). The presence of a ligand might help iM formation but the situation is nowadays complex as the ideal ligand has yet to be found, and the casual relationship between iM formation and stabilization provided by the ligands that have been used to date is anything but straightforward [72-74]. More information about factors affecting iM stability can be found in recent review articles [75, 76].

These findings highlight again that iM formation is a multiparameter event, which originates in a combination of both internal and external factors, which are particularly difficult to decipher in a functional cellular context.

Impacts of DNA chemical modifications on iM formation. iMs fold from C-rich DNA sequences, known as hotspots for epigenetic modifications which could lead to various C derivatives including 5mC (5-methylcytosine), 4mC (N4-methylcytosine), 5BrC (5-bromocytosine), 5hmC (5-hydroxymethylcytosine), 5fC (5-formylcytosine), 5caC (5-carboxylcytosine), 5gmC (5-glyceryl-methylcytosine) and 4acC (N4-acetyldeoxycytosine) [77]. The impact of these modifications on iM formation and stability is a matter of debate. Recent studies showed that DNA methylation has complex effects on iM formation, depending on several factors such as the number and position of C modifications and their combination. For instance, the presence of 5mC can favor iM stability by increasing the base-pairing energy of the 5mC⁺-C and 5mC⁺-5mC base pairs, but this depends on the number of 5mC [78-80]. 5mC can stabilize the iM structure found in the VEGF sequence by increasing the transitional pH and the melting temperature [81]. Another example is the human telomeric iM: when it contains 5mC or 5BrC, this iM is stable under more acidic or basic pH values, respectively [79], because 5mC has a slightly higher pKa value while 5-BrC a dramatically lower pKa value. Genome-wide studies showed that hyper-methylation can decrease iM formation while hypo-methylation can promote it, at both pH 5.5 or 7.0 conditions [67]. However, the effects of C

modifications on the stability of human telomeric iM are complex, given that they depend on both the position and type (5mC, 5hmC, 5fC and 5caC) of the modifications. As an example, the telomeric iM is stable when it harbors one or two 5mCs, but more 5mCs (or one 5hmC) destabilize it [82], as do 5fC and 5caC [83], likely because of the electron withdrawing nature of their substituents [79]. These observations can be extended to telomeric iMs of *A. thaliana*, for which a complex effect of C modifications on its stability has been documented [84]. The effect of 5mC and 5hmC are partly due to their influence on the flexibility of DNA backbone and the thermal stability of the DNA [81, 85, 86] since, for instance, 5mC decreases the flexibility and increases the melting temperature of the DNA it belongs to.

The incorporation of nucleotide analogues within iM sequences dramatically affects its stability, but these modifications do not concern solely the nucleobase but also the sugar and the phosphate backbone (Table 3). It has been reported that sugar modifications exhibit position-dependent effects on iM stability: for instance, a single substitution by an unlocked nucleic acid (UNA) at the position C14 of human telomeric sequence leads to dramatic destabilization of human telomeric iM, while at the position A12 it significantly increases its stability [87]. When a D-ribose is replaced by a D-2-deoxyribose or a D-arabinose, the resulting iM is found to be more stable [88], which highlights the influence of the backbone nature. Similarly, iMs containing phosphorothioates and phosphodiester backbones are stable, while those displaying methylphosphonate [89, 90] and phosphoramidate backbones are not, even at acidic pH [91]. Another example is the presence of (2'S)-2'-deoxy-2'-C-methyl-cytidine units within the vertebrate telomeric iMs, which creates a C2'-endo conformation and induces charge distribution changes in the sugar, which improves inter-strand electrostatic interactions and thus, favors iM formation [92]. Similarly, 2'-deoxy-2'-fluoro-5-methyl-arabincytidine (5-Me-2'F-araC) was found to stabilize iMs at physiological pH [93].

Table 3. Summary of impacts of DNA nucleotide analogs or derivatives on iM formation

Location	Types of Modifications ^a	Effects	Sequences ^b	Ref.
Sugar	UNA	C14 of the ribose ring of cytosine destabilize iM; A12 promote iM formation	HT	[87]
	LNA	Promote iM formation at lower pH	d(C ₅)	[94]
	HNA	Destabilize iM	h(A ₂ -C ₄) h(C ₄ -A ₂)	[10]
	araF-C	Promote and stabilize iM	d(T-C ₅)	[95]
	D-2-deoxyribose; D-arabinose riboadenosine (rA) linker	Stabilize iM	dV-5, rV-5, araV- 5, 2',5'-rV-5	[88]
Phosphate backbone	3'-S-phosphorothiolate linkages	Stabilize iM	d(T-C ₅)	[96]
	Phosphorothioates phosphodiester	Promote iM formation	C-rich	[89]
	methylphosphonate	Destabilize iM	C-rich	[90]

	phosphoramidate	Destabilize iM even at acidic pH	T(CCT) ₂ -T ₃ -(CCT) ₂	[91]
Sugar-Phosphate backbone	PNA	Stabilize iM	p(T-C ₈) p(C ₅ -T)	[97]
	(2'S)-2'-deoxy-2'-C-methyl-cytidine	Stabilize iM at neutral pH	HT	[92]
Base	5-propynyluracil	Promote and stabilize iM	T(CCT) ₂ -T ₃ -(CCT) ₂	[91]
	i-clamps	Promote and stabilize iM; 2 i-clamps on opposite strands stabilize iM	(C ₃ -AT) ₃ -C ₃	[98]
	T-vinyl	Promote iM formation	HT	[99]
Sugar-Base	dC; C ^m ; fC; fC ^m	C ^m and fC stabilized iM at neutral pH	HT	[93]

Note: a, UNA, Unlocked nucleic acid; LNA, Locked nucleic acid; HNA, hexose nucleic acid; PNA, Peptide nucleic acid; araF-C, 2'-deoxy-2'-fluoroarabincytidine; i-clamps, phenoxazine 2'-deoxynucleotide; dC, 2'-deoxycytidine; C^m, 2'-deoxy-5-methyl-cytidine (5-Me-dC); fC, 2'-deoxy-2'-fluoro-arabincytidine (2'F-araC); fC^m, 2'-deoxy-2'-fluoro-5-methyl-arabincytidine (5-Me-2'F-araC). b, Sequence: HT, human telomeric sequence (CCC-[TAA-CCC]_n); dV-5, rA[2',5'-dCCCCC]3',5'-dCCCCC; rV-5 araV- 5 and 2',5'-rV-5, modified dv-5. C-rich, see detailed sequence information in reference.

Collectively, these findings provide new insights into how the modifications of the sugar, the base and/or the phosphate backbone affect iM folding and stability, which opens interesting perspectives for chemically manipulating iM-forming sequences in order to make them stable enough for a broader panel of biotechnological applications.

Impacts of DNA modifications along with other factors on iM formation. As discussed above, an impressive number of factors affect iM formation and stability. A comprehensive understanding of these factors and of their combinations may open the way towards harnessing the biological functions of iMs and, beyond them, their potential medical and technological applications (**Figure 2**).

In cells, molecular crowding is important, as numerous biomacromolecules (nucleic acids, proteins and polysaccharides) act as cellular crowding factors affecting the stability of various types of DNA structures [100]. Crowding conditions were found to mitigate the influence of DNA methylation (5mC and 5hmC) on iM stability [63]. Several studies showed that 5mC and/or 5hmC do not significantly affect iM formation or stability under crowding conditions (or with different compositions of co-solvents) [101, 102]. These findings reflect the intricacy of the combination of C modifications and external factors [103] on iM stability. As an example, the presence of a 5-Me-2'F-araC alone or in combination with other C modifications (*i.e.*, dC, C^m, fC, fC^m; see full name of modifications in Table3) affects iM formation at physiological pH conditions, impacts of base pairs between modified cytosines on iM stabilization exhibiting a descending trend as fC-C^mH⁺ > fC^m-fCH⁺ > fC-fCH⁺ ≈ C^m-C^mH⁺ > fC^m-fC^mH⁺ ≈ fC^m-CH⁺ > C-CH⁺. For instance, the base pair between 5-Me-2'F-araC and 2'F-araC can stabilize iM at physiological pH [93], indicating a synergistic effect between sugar and base modifications on iM formation.

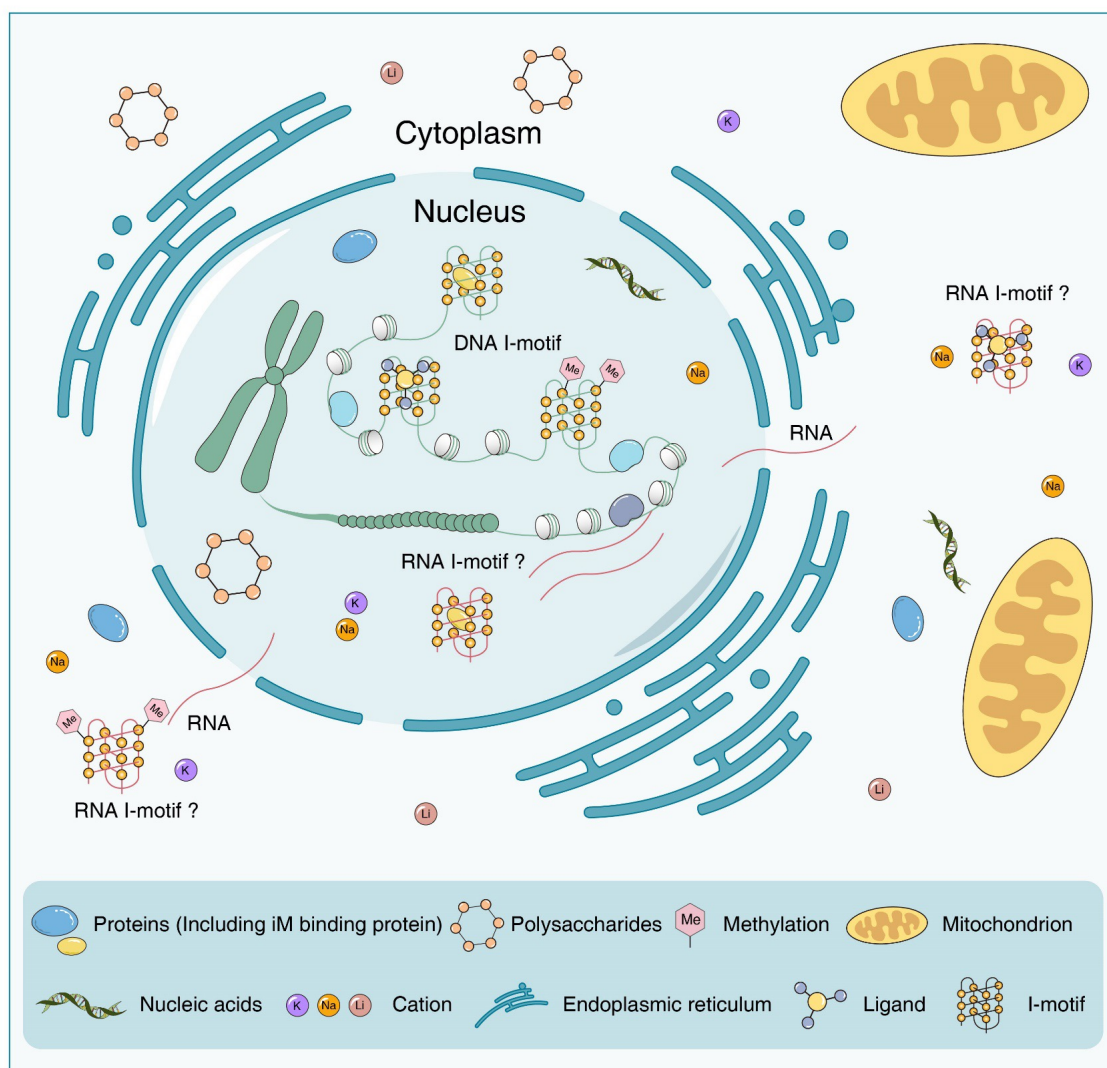


Figure 2. Potential combined effects of nucleic acid modification and environment factors on formation of in vivo iM structures. DNA methylation, cation, molecular ligand, protein and molecular crowding have been reported to promote and stabilize iM formation at physiological pH in vitro, and most of them can also be present in cellular environment. In particular, about 20-40% of the intracellular compartments contain various macromolecules such as nucleic acids, proteins and polysaccharides, which can serve as key cellular crowding factors.

On a genome-wide scale, recent findings show that distinct numbers of iMs were detected upon either hyper- or hypo-methylation treatment, under the same pH conditions. For instance, at pH 5.5, DNA hypomethylation treatment results in *ca.* 8,600 iMs gained (and *ca.* 1,700 iMs lost); at pH 7.0, DNA hypomethylation treatment results in *ca.* 7,000 iMs gained (and *ca.* 2,400 lost), while DNA hypermethylation treatment results in *ca.* 9,400 iMs gained (and *ca.* 4,200 lost). Once again, caution must be exercised as these results were obtained using iMab (see above); nevertheless, these results show how complex the effects of C methylation are complex, especially when combined with pH [67].

Potential biological implications of iMs. iMs are now suspected to be involved in several key DNA transactions including the regulation of DNA replication and transcription, telomere and

centromere functions, recombination and repair, and transposable element (TE) activity [104, 105] (**Figure 3**). An illustrative example is the role of an iM-forming sequence found in the promoter region of the *BCL2* gene, which modulates its transcription. This effect, which is mediated by the binding of hnRNPLL in lymphoma cell lines, could be upregulated by IMC-48, a small molecule that stabilizes iMs and favors hnRNPLL binding, and downregulated by IMC-76, a small molecule that destabilizes iMs and hampers hnRNPLL binding [106, 107]. Such a transcriptional regulation has been reported for several human genes, including *ACCI* [108], *HIV-1* [109], *PDGFR-β* [110], *VEGF* [111], *c-MYC* [112], *KRAS* [113], *RET* [20], *HARS* [114], *Rb* [115] and *hESC* [116]. This regulatory activity is not restricted to humans, as the activation of the *BmPOUM2* gene, responsible for developmental processes during metamorphosis of silkworms, is governed by the folding of a promoter iM, which recruits BmILF [28].

In agreement with this activity, iMs are overrepresented in G1/S phase of the cell cycle, corresponding to gene active transcription [27]. On a genome-wide level (using iMab), *ca.* 25,000 *in vivo* iMs were found to be strongly enriched in promoter regions in human cells, at 94% in human embryonic kidney (HEK293T) cells and 84% in well-differentiated liposarcoma (WDLPS) cells [4, 117], which indicates that iMs may play significant roles in the regulation of transcription in human cells, in agreement with previous results [118] which can partly be explained by their putative role as *cis*-regulatory elements (CREs) [119].

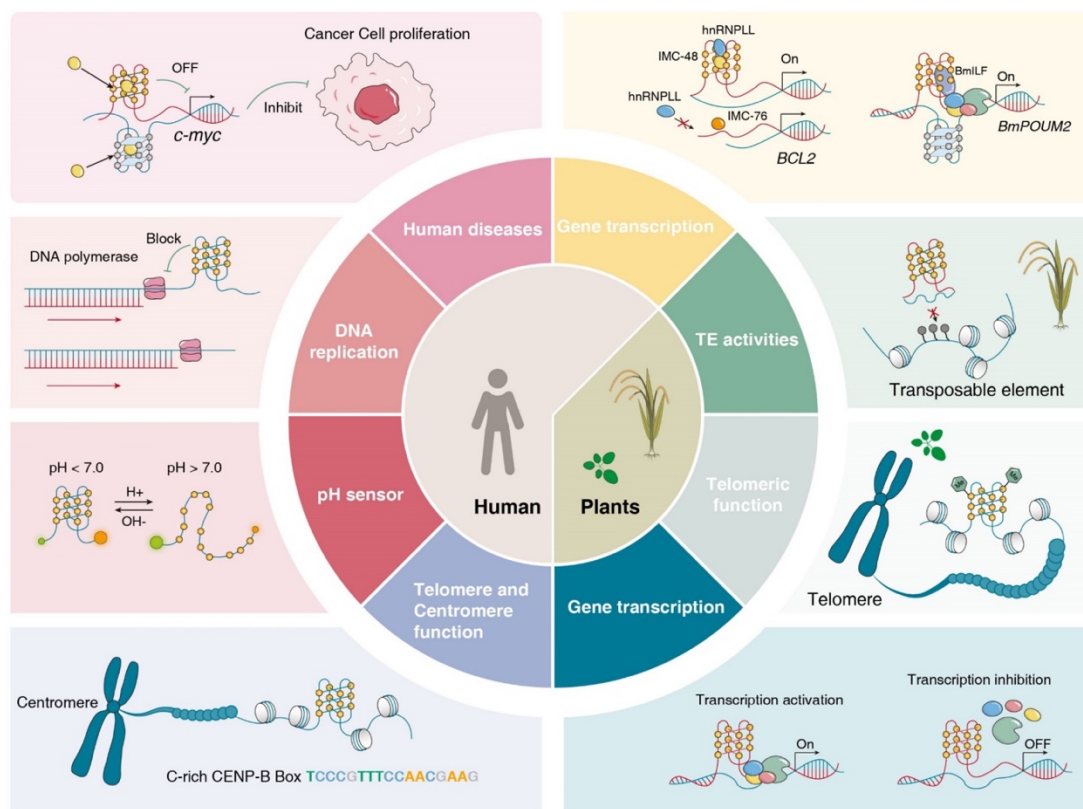


Figure 3. Potential biological implications of iMs in humans and plants. They include gene transcription, DNA replication, genome stability, pH sensor, telomere and centromere functions in humans, human diseases, TE activities and telomeric functions in plants (*Arabidopsis* and *Oryza sativa*). Black arrow on DNA indicates the transcription start site (TSS) for each gene. On and OFF on the TSS indicates transcription activation and inhibition.

As iMs are located in key chromosomal regions (promoters of oncogenes, centromeres and telomeres), they could be used as potential targets for drug development in humans [74, 120-123]. As discussed above, the transcription of *BCL2* is governed by iMs and can be modulated by IMC-48 and IMC-76 but also by the acridone derivative A22, found to alleviate liver damage of non-alcoholic fatty liver diseases (NAFLDs, including the non-alcoholic steatohepatitis NASH) [124]. The binding of the bisacridine derivative B05 to the promoter iM of *c-Kit* gene downregulates its expression, thereby preventing cell proliferation and metastasis [125]. Similarly, targeting the promoter iM of *KRAS* gene by unmodified peptide nucleic acid (PNA) and gamma-modified PNA (γ PNA) oligomers was found to inhibit its transcription [126], as well as the binding of the oleanolic acid 3c to the promoter iM of *VEGF*, which impedes its expression in MCF-7 cells, suppressing cancer cell proliferation [127]. The transcription of *c-myc* gene can be modulated by negative superhelicity-induced iM in its promoter [21] and the binding of the bisacridine A9 suppresses its transcription, preventing cancer cell proliferation as well [128]. The downregulation of *PDGFR- β* and *KRAS* gene originates in an opposite effect, that is, the destabilization of the corresponding promoter iM by the acridone-naphthalimide WZZ02 and the acridone C4, respectively [127, 129]. These results highlight how complicated it could be to rationalize the cellular activity of iM-targeting agents, and again how important it is to exercise caution when interpreting cellular outcomes triggered by ligands, as more information about their affinity, selectivity and binding mode must be obtained.

Another type of cellular activity for iMs was recently uncovered: just like telomeric G4s, telomeric iMs might prevent the extension telomeric DNA by blocking the activity of telomerase, which jeopardizes telomere maintenance, triggering telomere uncapping and loss of telomere-binding proteins [130]. The acridine dimer DI26 can destabilize telomeric iMs, thereby affecting telomere biology by downregulating *TERRA* production [131]. Also, the C-rich DNA sequences in the CENP-A and CENP-B boxes within both human and *Drosophila melanogaster* centromeres can form iMs, which may have a role during chromosome separation [132-134], suggesting that iMs may be key players in centromere location and functions [135]. Alternatively, iMs (again just like G4s) might act as roadblocks that hamper proper polymerase processivity, thereby triggering genomic instability [136].

iMs are also involved in the regulation of gene transcription, telomeric functions and TE activities in plants. Telomeric iMs can be stabilized by 5mC methylation in *Arabidopsis* [84]. In rice, a global study reveals that *ca.* 25,000 *in vitro* iMs positively correlate with active histone markers, and are associated with CREs like enhancers, which implies their regulatory roles in gene transcription [30]. Also, TEs are highly associated with PiMFSs and iMs, suggesting again their potential roles in the modulation of TE activities in rice [30].

Collectively, here-above-described results underline how complicated the biology of iMs could be, which emphasizes the need of investing massive efforts to better understand the roles they may play in the regulation of DNA transactions and genomic stability.

Concluding remarks and future perspectives.

In eukaryotic and prokaryotic cells, DNA is largely found into the so-called B-form (or Watson-Crick duplex), which is the most stable form of storage of the genetic information and

proved invaluable for the faithful transmission of this information from generations to generations. Accumulating evidence now shows that DNA may transiently fold into a broad variety of non-B forms, whose nature is encoded in the sequences involved. For instance, inverted DNA repeats fold into four-way DNA junction (or cruciform DNA, C-DNA), direct repeats into three-way DNA junction (or slipped-strand DNA, S-DNA), mirror repeats into triplex-DNA (or hinged DNA, H-DNA), and short tandem repeats into either G-quadruplexes (or G4s, G-DNA) or i-motifs (iMs or I-DNA) depending on whether they are G- or C-rich, respectively [137]. Importantly, compelling evidence now shows that these non-B DNA structures are key modulators in a wide panel of biological processes and certainly involved in many dysregulations underlying human genetic diseases [104, 105] [138, 139].

As compared to the G4 field, which is highly active for more than 30 years, the interest for iMs is now emerging. Initially studied because they fold from C-rich sequences that are complementary to G-rich, that is, G4-forming sequences, iMs are now scrutinized more avidly because it is now established that their formation and stability are compatible with physiological conditions. These considerations have sparked interest from the community and molecular tools and technologies are now being designed and used to further our knowledge about iM biology. For instance, the recently developed anti-iM antibody iMab was used to assess the existence of iMs in both human and rice genome, both qualitatively (immunofluorescence detection) or quantitatively (ChIP-seq, CUT&Tag and iM-IP-seq) [27]. The functional relevance of iMab is currently being investigated, and this is also true for the panel of iM ligands that have been studied over the past years. To date, the main limitation to a greater development of the iM field is the lack of widely accepted molecular tools (antibodies or chemicals). However, some nice advances have been made, notably regarding the *in vitro* validations of iM formation under near physiological conditions [140]. With these progresses, in both human and plants, a review that summarizes the conditions affecting iM formation and stability, iM detection and its biological implications was needed.

This is why we have summarized herein the most recent results regarding iMs, with a notable emphasis on the conditions that affect their formation *in vitro* and *in vivo* (in cells). There are several urgent questions that need more investment in the near future (see Outstanding Questions). For instance, how DNA modifications and/or chromatin organization affect in-cell dynamics of iMs across the genome? What are the iMs' functions in the regulation of gene transcription during normal developmental processes and in response to cellular stress or environmental cues? To address these questions, the iM field is in need of sensitive methodologies that could allow for monitoring real-time, and in cells, the dynamics of iM folding and unfolding. Massive efforts are thus needed to better portray the functions of cellular iMs and their roles in normal *versus* pathological conditions.

Acknowledgments

The authors acknowledge financial support from the National Key R&D Program of China (2023YFD1200800, 2022YFA1303400). The National Natural Science Foundation of China (U23A20179, 32201782, and 32070561). The authors also give thanks to the Bioinformatics Center, Nanjing Agricultural University, for providing computing facilities for data processing and analyses.

Declaration of Interests

The authors have no interests to declare.

References

- 1 Gehring, K., *et al.* (1993) A tetrameric DNA structure with protonated cytosine-cytosine base pairs. *Nature* 363, 561-565
- 2 Mergny, J.-L., *et al.* (1995) Intramolecular folding of pyrimidine oligodeoxynucleotides into an i-DNA motif. *J. Am. Chem. Soc.* 117, 8887-8898
- 3 Balasubramanian, S., *et al.* (2011) Targeting G-quadruplexes in gene promoters: a novel anticancer strategy? *Nat. Rev. Drug Discov.* 10, 261-275
- 4 Zanin, I., *et al.* (2023) Genome-wide mapping of i-motifs reveals their association with transcription regulation in live human cells. *Nucleic Acids Res.* 51, 8309-8321
- 5 Zhou, J., *et al.* (2010) Formation of i-motif structure at neutral and slightly alkaline pH. *Mol. Biosyst.* 6, 580-586
- 6 Li, Y., *et al.* (2018) Label-Free Detection of Tetramolecular i-Motifs by Surface-Enhanced Raman Spectroscopy. *Anal. Chem.* 90, 2996-3000
- 7 Pokhrel, P., *et al.* (2022) Ensemble Force Spectroscopy by Shear Forces. *J. Vis. Exp.*, e63741
- 8 Bchara, L., *et al.* (2023) Rapid and Highly Efficient Separation of i-Motif DNA Species by CE-UV and Multivariate Curve Resolution. *Anal. Chem.* 95, 15189-15198
- 9 Improta, R. (2023) Shedding Light on the Photophysics and Photochemistry of I-Motifs Using Quantum Mechanical Calculations. *Int. J. Mol. Sci.* 24, 12614
- 10 Ghezzi, M., *et al.* (2023) Modulation of the tetrameric I-motif folding of C-rich Tetrahymena telomeric sequences by hexitol nucleic acid (HNA) modifications. *Biochimie* 214, 112-122
- 11 Skolakova, P., *et al.* (2019) Systematic investigation of sequence requirements for DNA i-motif formation. *Nucleic Acids Res.* 47, 2177-2189
- 12 Víšková, P., *et al.* (2024) In-cell NMR suggests that DNA i-motif levels are strongly depleted in living human cells. *Nat. Commun.* 15, 1992
- 13 Chen, L., *et al.* (1994) Crystal structure of a four-stranded intercalated DNA: d(C4). *Biochemistry* 33, 13540-13546
- 14 Dzatko, S., *et al.* (2018) Evaluation of the stability of DNA i-Motifs in the nuclei of living mammalian cells. *Angew. Chem. Int. Ed. Engl.* 57, 2165-2169
- 15 Sheng, Q., *et al.* (2017) Identification of new DNA i-motif binding ligands through a fluorescent intercalator displacement assay. *Org. Biomol. Chem.* 15, 5669-5673
- 16 Megalathan, A., *et al.* (2021) Single-Molecule Analysis of Nanocircle-Embedded I-Motifs under Crowding. *J. Phys. Chem. B* 125, 2193-2201
- 17 Xi, D., *et al.* (2021) Nanopore-Based Single-Molecule Investigation of DNA Sequences with Potential to Form i-Motif Structures. *ACS Sens.* 6, 2691-2699
- 18 Panczyk, T., *et al.* (2022) Stability and Existence of Noncanonical I-motif DNA Structures in Computer Simulations Based on Atomistic and Coarse-Grained Force Fields. *Molecules* 27, 4915
- 19 Skolakova, P., *et al.* (2023) DNA i-motif formation at neutral pH is driven by kinetic partitioning. *Nucleic Acids Res.* 51, 2950-2962
- 20 Guo, K., *et al.* (2007) Formation of pseudosymmetrical G-quadruplex and i-motif structures in the proximal promoter region of the RET oncogene. *J. Am. Chem. Soc.* 129, 10220-10228
- 21 Sun, D. and Hurley, L.H. (2009) 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, 2863-2874
- 22 Kendrick, S., *et al.* (2009) The i-motif in the bcl-2 P1 promoter forms an unexpectedly stable structure with a unique 8: 5: 7 loop folding pattern. *J. Am. Chem. Soc.* 131, 17667-17676
 - 23 Kshirsagar, R., *et al.* (2017) Probing the potential role of non-B DNA structures at yeast meiosis-specific DNA double-strand breaks. *Biophys. J.* 112, 2056-2074
 - 24 Kang, B.H., *et al.* (2016) Thiazole orange as a fluorescent probe: Label-free and selective detection of silver ions based on the structural change of i-motif DNA at neutral pH. *Talanta* 156-157, 141-146
 - 25 Lee, I.J., *et al.* (2015) Probing structural changes of self assembled i-motif DNA. *Chem. Commun. (Camb)* 51, 3747-3749
 - 26 Fu, B., *et al.* (2015) Label-free detection of pH based on the i-motif using an aggregation-caused quenching strategy. *Chem. Commun. (Camb)* 51, 16960-16963
 - 27 Zeraati, M., *et al.* (2018) I-motif DNA structures are formed in the nuclei of human cells. *Nat. Chem.* 10, 631-637
 - 28 Niu, K., *et al.* (2018) BmILF and i-motif structure are involved in transcriptional regulation of BmPOUM2 in *Bombyx mori*. *Nucleic Acids Res.* 46, 1710-1723
 - 29 Tang, W., *et al.* (2020) In vivo visualization of the i-motif DNA secondary structure in the *Bombyx mori* testis. *Epigenetics Chromatin* 13, 12
 - 30 Ma, X., *et al.* (2022) Genome-wide characterization of i-motifs and their potential roles in the stability and evolution of transposable elements in rice. *Nucleic Acids Res.* 50, 3226-3238
 - 31 Bedrat, A., *et al.* (2016) Re-evaluation of G-quadruplex propensity with G4Hunter. *Nucleic Acids Res.* 44, 1746-1759
 - 32 Belmonte-Reche, E. and Morales, J.C. (2020) G4-iM Grinder: when size and frequency matter. G-Quadruplex, i-Motif and higher order structure search and analysis tool. *NAR Genom. Bioinform.* 2, lqz005
 - 33 Shamim, A., *et al.* (2020) MD-TSPC4: Computational Method for Predicting the Thermal Stability of I-Motif. *Int. J. Mol. Sci.* 22, 61
 - 34 Yang, B., *et al.* (2024) Prediction of DNA i-motifs via machine learning. *Nucleic Acids Res.* 52, 2188-2197
 - 35 Yu, H., *et al.* (2024) iM-Seeker: a webserver for DNA i-motifs prediction and scoring via automated machine learning. *Nucleic Acids Res.*, gkae315
 - 36 Mergny, J.L., *et al.* (2005) Thermal difference spectra: a specific signature for nucleic acid structures. *Nucleic Acids Res.* 33, e138
 - 37 Subramani, V.K. and Kim, K.K. (2023) Characterization of Z-DNA Using Circular Dichroism. In *Methods Mol. Biol.* (Kim, K.K. and Subramani, V.K., eds), pp. 33-51, Springer US
 - 38 Kwok, C.K. and Merrick, C.J. (2017) G-quadruplexes: prediction, characterization, and biological application. *Trends Biotechnol.* 35, 997-1013
 - 39 Lyu, K., *et al.* (2021) RNA G-quadruplexes (rG4s): genomics and biological functions. *Nucleic Acids Res.* 49, 5426-5450
 - 40 Dai, J., *et al.* (2009) A direct and nondestructive approach to determine the folding structure of the I-motif DNA secondary structure by NMR. *J. Am. Chem. Soc.* 131, 6102-6104
 - 41 Wang, B. and Chatterton, E. (2021) Identifying i-motif formation using capillary electrophoresis. *Electrophoresis* 42, 1300-1305

- 42 Bonnet, H., *et al.* (2022) Assessment of presumed small-molecule ligands of telomeric i-DNA by biolayer interferometry (BLI). *Chem Commun (Camb)* 58, 5116-5119
- 43 Manna, S. and Srivatsan, S.G. (2018) Fluorescence-based tools to probe G-quadruplexes in cell-free and cellular environments. *RSC Adv.* 8, 25673-25694
- 44 Varshney, D., *et al.* (2020) The regulation and functions of DNA and RNA G-quadruplexes. *Nat. Rev. Mol. Cell. Biol.* 21, 459-474
- 45 Peña Martínez, C.D., *et al.* (2022) Human genomic DNA is widely interspersed with i-motif structures. *bioRxiv*, 2022.2004.2014.488274
- 46 Boissieras, J., *et al.* (2023) iMab Antibody Binds Single-Stranded Cytosine-Rich Sequences and Unfolds DNA i-Motifs. *bioRxiv*, 2023.2011.2021.568054
- 47 Choi, J., *et al.* (2011) pH-induced intramolecular folding dynamics of i-motif DNA. *J. Am. Chem. Soc.* 133, 16146-16153
- 48 Dembska, A., *et al.* (2017) pH-Sensing fluorescence oligonucleotide probes based on an i-motif scaffold: a review. *Anal. Methods* 9, 6092-6106
- 49 Fujii, T. and Sugimoto, N. (2015) Loop nucleotides impact the stability of intrastrand i-motif structures at neutral pH. *Phys. Chem. Chem. Phys.* 17, 16719-16722
- 50 Cheng, M., *et al.* (2021) Thermal and pH Stabilities of i-DNA: Confronting in vitro Experiments with Models and In-Cell NMR Data. *Angew. Chem. Int. Ed. Engl.* 60, 10286-10294
- 51 Fleming, A.M., *et al.* (2017) 4n-1 Is a "Sweet Spot" in DNA i-Motif Folding of 2'-Deoxycytidine Homopolymers. *J. Am. Chem. Soc.* 139, 4682-4689
- 52 McKim, M., *et al.* (2016) Loop Sequence Context Influences the Formation and Stability of the i-Motif for DNA Oligomers of Sequence (CCCXXX)₄, where X = A and/or T, under Slightly Acidic Conditions. *J. Phys. Chem. B* 120, 7652-7661
- 53 Fernandez, S., *et al.* (2011) Influence of pH, temperature and the cationic porphyrin TMPyP4 on the stability of the i-motif formed by the 5'-(C3TA2)₄-3' sequence of the human telomere. *Int. J. Biol. Macromol.* 49, 729-736
- 54 Mir, B., *et al.* (2017) Prevalent Sequences in the Human Genome Can Form Mini i-Motif Structures at Physiological pH. *J. Am. Chem. Soc.* 139, 13985-13988
- 55 Escaja, N., *et al.* (2012) A minimal i-motif stabilized by minor groove G:T:G:T tetrads. *Nucleic Acids Res.* 40, 11737-11747
- 56 Serrano-Chacon, I., *et al.* (2023) pH-Dependent Capping Interactions Induce Large-Scale Structural Transitions in i-Motifs. *J. Am. Chem. Soc.* 145, 3696-3705
- 57 Lepper, C.P., *et al.* (2019) Effects of Pressure and pH on the Physical Stability of an I-Motif DNA Structure. *Chemphyschem* 20, 1567-1571
- 58 Kim, S.E., *et al.* (2014) Destabilization of i-motif by submolar concentrations of a monovalent cation. *J. Phys. Chem. B* 118, 4753-4760
- 59 Day, H.A., *et al.* (2013) Silver cations fold i-motif at neutral pH. *Chem. Commun. (Camb)* 49, 7696-7698
- 60 Day, H.A., *et al.* (2015) Reversible DNA i-motif to hairpin switching induced by copper (II) cations. *Chem. Commun. (Camb)* 51, 14099-14102
- 61 Rajendran, A., *et al.* (2010) Molecular crowding of the cosolutes induces an intramolecular i-motif structure of triplet repeat DNA oligomers at neutral pH. *Chem. Commun. (Camb)* 46, 1299-1301
- 62 Cui, J., *et al.* (2013) The effect of molecular crowding on the stability of human c-MYC promoter sequence I-motif at neutral pH. *Molecules* 18, 12751-12767

- 63 Bhavsar-Jog, Y.P., *et al.* (2014) Epigenetic modification, dehydration, and molecular crowding effects on the thermodynamics of i-motif structure formation from C-rich DNA. *Biochemistry* 53, 1586-1594
- 64 Li, X., *et al.* (2006) Carboxyl-modified single-walled carbon nanotubes selectively induce human telomeric i-motif formation. *Proc. Natl. Acad. Sci. U. S. A.* 103, 19658-19663
- 65 Chen, X., *et al.* (2013) Stabilization and induction of oligonucleotide i-motif structure via graphene quantum dots. *Acs Nano.* 7, 531-537
- 66 Iaccarino, N., *et al.* (2021) Effects of Sequence and Base Composition on the CD and TDS Profiles of i-DNA. *Angew. Chem. Int. Ed. Engl.* 60, 10295-10303
- 67 Feng, Y., *et al.* (2024) The roles of DNA methylation on pH dependent i-motif (iM) formation in rice. *Nucleic Acids Res.* 52, 1243-1257
- 68 Wang, Y. and Patel, D.J. (1993) Solution structure of the human telomeric repeat d [AG₃ (T₂AG₃)₃] G-tetraplex. *Structure* 1, 263-282
- 69 Takahashi, S. and Sugimoto, N. (2015) Pressure-dependent formation of i-motif and G-quadruplex DNA structures. *Phys. Chem. Chem. Phys.* 17, 31004-31010
- 70 Ono, A., *et al.* (2011) Binding of metal ions by pyrimidine base pairs in DNA duplexes. *Chem. Soc. Rev.* 40, 5855-5866
- 71 Wolski, P., *et al.* (2019) Interaction of Human Telomeric i-Motif DNA with Single-Walled Carbon Nanotubes: Insights from Molecular Dynamics Simulations. *J. Phys. Chem. B* 123, 10343-10353
- 72 Sedghi Masoud, S. and Nagasawa, K. (2018) i-Motif-Binding Ligands and Their Effects on the Structure and Biological Functions of i-Motif. *Chem. Pharm. Bull. (Tokyo)* 66, 1091-1103
- 73 Day, H.A., *et al.* (2014) i-Motif DNA: structure, stability and targeting with ligands. *Bioorg. Med. Chem.* 22, 4407-4418
- 74 Luo, X., *et al.* (2023) Emerging roles of i-motif in gene expression and disease treatment. *Front Pharmacol.* 14, 1136251
- 75 Mergny, J.-L., *et al.* (2023) Quadruplexes Are Everywhere... On the Other Strand Too: The i-Motif. In *Handbook of Chemical Biology of Nucleic Acids*, pp. 1-25, Springer Nature Singapore
- 76 Abou Assi, H., *et al.* (2018) i-Motif DNA: structural features and significance to cell biology. *Nucleic Acids Res.* 46, 8038-8056
- 77 Meng, W.Y., *et al.* (2024) Epigenetic marks or not? The discovery of novel DNA modifications in eukaryotes. *J. Biol. Chem.* 300, 106791
- 78 Seidu, Y.S., *et al.* (2021) Influence of 5-Methylation and the 2'-and 3'-Hydroxy Substituents on the Base Pairing Energies of Protonated Cytidine Nucleoside Analogue Base Pairs: Implications for the Stabilities of i-Motif Structures. *J. Phys. Chem. A* 125, 5939-5955
- 79 Lannes, L., *et al.* (2015) Tuning the pH Response of i-Motif DNA Oligonucleotides. *Chembiochem* 16, 1647-1656
- 80 Yang, B., *et al.* (2013) Base-pairing energies of proton-bound homodimers determined by guided ion beam tandem mass spectrometry: application to cytosine and 5-substituted cytosines. *Anal. Chem.* 85, 11000-11006
- 81 Kimura, K., *et al.* (2022) Stabilization of VEGF i-motif structure by CpG methylation. *Biochem. Biophys. Res. Commun.* 594, 88-92
- 82 Xu, B., *et al.* (2015) Regulation of telomeric i-motif stability by 5-methylcytosine and 5-hydroxymethylcytosine modification. *Org. Biomol. Chem.* 13, 5646-5651
- 83 Wright, E.P., *et al.* (2020) Epigenetic modification of cytosines fine tunes the stability of i-motif

- DNA. *Nucleic Acids Res.* 48, 55-62
- 84 Školáková, P., *et al.* (2020) 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. Gen. Subj.* 1864, 129651
- 85 Ngo, T.T., *et al.* (2016) Effects of cytosine modifications on DNA flexibility and nucleosome mechanical stability. *Nat. Commun.* 7, 10813
- 86 Oshikawa, D., *et al.* (2022) CpG Methylation Altered the Stability and Structure of the i-Motifs Located in the CpG Islands. *Int. J. Mol. Sci.* 23, 6467
- 87 Pasternak, A. and Wengel, J. (2011) Modulation of i-motif thermodynamic stability by the introduction of UNA (unlocked nucleic acid) monomers. *Bioorg. Med. Chem. Lett.* 21, 752-755
- 88 Robidoux, S. and Damha, M.J. (1997) D-2-deoxyribose and D-arabinose, but not D-ribose, stabilize the cytosine tetrad (i-DNA) structure. *J. Biomol. Struct. Dyn.* 15, 529-535
- 89 Kanehara, H., *et al.* (1997) Spectroscopic evidence for the formation of four-stranded solution structure of oligodeoxycytidine phosphorothioate. *Biochemistry* 36, 1790-1797
- 90 Mergny, J.-L. and Lacroix, L. (1998) Kinetics and thermodynamics of i-DNA formation: phosphodiester versus modified oligodeoxynucleotides. *Nucleic Acids Res.* 26, 4797-4803
- 91 Lacroix, L. and Mergny, J.-L. (2000) Chemical modification of pyrimidine TFOs: effect on i-motif and triple helix formation. *Arch. Biochem. Biophys.* 381, 153-163
- 92 Avino, A., *et al.* (2017) Stabilization of Telomeric I-Motif Structures by (2'S)-2'-Deoxy-2'-C-Methylcytidine Residues. *Chembiochem* 18, 1123-1128
- 93 Abou Assi, H., *et al.* (2018) Probing Synergistic Effects of DNA Methylation and 2'-beta-Fluorination on i-Motif Stability. *Chemistry* 24, 471-477
- 94 Kumar, N., *et al.* (2007) i-Motif formation with locked nucleic acid (LNA). *Angew. Chem. Int. Ed. Engl.* 46, 9220-9222
- 95 El-Khoury, R., *et al.* (2023) i-Motif folding intermediates with zero-nucleotide loops are trapped by 2'-fluoroarabincytidine via F \cdots H and O \cdots H hydrogen bonds. *Commun. Chem.* 6, 31
- 96 Brazier, J.A., *et al.* (2006) Stabilization of the DNA I-Motif Structure by Incorporation of 3'-S-Phosphorothiolate Linkages. *Angew. Chem. Int. Ed Engl.* 118, 120-123
- 97 Krishnan-Ghosh, Y., *et al.* (2005) PNA forms an i-motif. *Chem. Commun. (Camb)*, 5278-5280
- 98 Tsvetkov, V.B., *et al.* (2018) i-Clamp phenoxazine for the fine tuning of DNA i-motif stability. *Nucleic Acids Res.* 46, 2751-2764
- 99 Kikuta, K., *et al.* (2015) Stabilization of the i-motif structure by the intra-strand cross-link formation. *Bioorg. Med. Chem. Lett.* 25, 3307-3310
- 100 Miyoshi, D. and Sugimoto, N. (2008) Molecular crowding effects on structure and stability of DNA. *Biochimie* 90, 1040-1051
- 101 Morgan, R.K., *et al.* (2018) Effects of 5-Hydroxymethylcytosine Epigenetic Modification on the Stability and Molecular Recognition of VEGF i-Motif and G-Quadruplex Structures. *J. Nucleic Acids* 2018, 9281286
- 102 Matsumoto, S., *et al.* (2021) Effect of DNA modifications on the transition between canonical and non-canonical DNA structures in CpG islands during senescence. *RSC Adv.* 11, 37205-37217
- 103 Tsuruta, M., *et al.* (2021) Combined Effects of Methylated Cytosine and Molecular Crowding on the Thermodynamic Stability of DNA Duplexes. *Int. J. Mol. Sci.* 22, 947
- 104 Duardo, R.C., *et al.* (2023) Non-B DNA structures as a booster of genome instability. *Biochimie* 214, 176-192

- 105 Wang, G. and Vasquez, K.M. (2023) Dynamic alternative DNA structures in biology and disease. *Nat. Rev. Genet.* 24, 211-234
- 106 Kendrick, S., *et al.* (2014) 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, 4161-4171
- 107 Kang, H.J., *et al.* (2014) 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, 4172-4185
- 108 Kaulage, M.H., *et al.* (2018) Structural Characterization of i-Motif Structure in the Human Acetyl-CoA Carboxylase 1 Gene Promoters and Their Role in the Regulation of Gene Expression. *Chembiochem* 19, 1078-1087
- 109 Ruggiero, E., *et al.* (2019) A dynamic i-motif with a duplex stem-loop in the long terminal repeat promoter of the HIV-1 proviral genome modulates viral transcription. *Nucleic Acids Res.* 47, 11057-11068
- 110 Brown, R.V., *et al.* (2017) The Consequences of Overlapping G-Quadruplexes and i-Motifs in the Platelet-Derived Growth Factor Receptor beta 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, 7456-7475
- 111 Guo, K., *et al.* (2008) Intramolecularly folded G-quadruplex and i-motif structures in the proximal promoter of the vascular endothelial growth factor gene. *Nucleic Acids Res.* 36, 4598-4608
- 112 Saha, P., *et al.* (2020) In situ formation of transcriptional modulators using non-canonical DNA i-motifs. *Chem. Sci.* 11, 2058-2067
- 113 Kaiser, C.E., *et al.* (2017) 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, 8522-8536
- 114 Miglietta, G., *et al.* (2015) GC-elements controlling HRAS transcription form i-motif structures unfolded by heterogeneous ribonucleoprotein particle A1. *Sci. Rep.* 5, 18097
- 115 Xu, Y. and Sugiyama, H. (2006) Formation of the G-quadruplex and i-motif structures in retinoblastoma susceptibility genes (Rb). *Nucleic Acids Res.* 34, 949-954
- 116 Bhavsar-Jog, Y.P., *et al.* (2019) Co-Localization of DNA i-Motif-Forming Sequences and 5-Hydroxymethyl-cytosines in Human Embryonic Stem Cells. *Molecules* 24, 3619
- 117 Zhao, P., *et al.* (2024) Identification and functional characterization of i-motifs in plants. *Trends Plant Sci.* S1360-1385(24)00030-X
- 118 Todorov, G. and Cunha, C. (2019) Hypothesis: Regulation of neuroplasticity may involve I-motif and G-quadruplex DNA formation modulated by epigenetic mechanisms. *Med. Hypotheses* 127, 129-135
- 119 Kendrick, S. and Hurley, L.H. (2010) The role of G-quadruplex/i-motif secondary structures as cis-acting regulatory elements. *Pure Appl. Chem.* 82, 1609-1621
- 120 Amato, J., *et al.* (2014) Noncanonical DNA secondary structures as drug targets: the prospect of the i-motif. *ChemMedChem* 9, 2026-2030
- 121 Pagano, A., *et al.* (2018) Common G-Quadruplex Binding Agents Found to Interact With i-Motif-Forming DNA: Unexpected Multi-Target-Directed Compounds. *Front. Chem.* 6, 281
- 122 Duan, J., *et al.* (2020) Biotechnological and Therapeutic Applications of Natural Nucleic Acid

- Structural Motifs. *Top. Curr. Chem. (Cham)* 378, 26
- 123 Debnath, M., *et al.* (2019) Chemical Regulation of DNA i-Motifs for Nanobiotechnology and Therapeutics. *Angew. Chem. Int. Ed. Engl.* 58, 2942-2957
- 124 Li, X., *et al.* (2020) Upregulation of BCL-2 by acridone derivative through gene promoter i-motif for alleviating liver damage of NAFLD/NASH. *Nucleic Acids Res.* 48, 8255-8268
- 125 Gong, X., *et al.* (2023) Regulation of c-Kit gene transcription selectively by bisacridine derivative through promoter dual i-motif structures. *Biochim. Biophys. Acta. Gene Regul. Mech.* 1866, 194912
- 126 Sarkar, S., *et al.* (2023) Targeting a KRAS i-motif forming sequence by unmodified and gamma-modified peptide nucleic acid oligomers. *Biopolymers* 114, e23529
- 127 Zeng, H., *et al.* (2021) Synthesis and Biological Evaluation of Oleanolic Acid Derivatives as Selective Vascular Endothelial Growth Factor Promoter i-Motif Ligands. *Int. J. Mol. Sci.* 22, 1711
- 128 Kuang, G., *et al.* (2020) Syntheses and Evaluation of New Bisacridine Derivatives for Dual Binding of G-Quadruplex and i-Motif in Regulating Oncogene c-myc Expression. *J. Med. Chem.* 63, 9136-9153
- 129 Wei, Z., *et al.* (2023) Syntheses and evaluation of acridone derivatives as anticancer agents targeting Kras promoter i-motif structure. *Bioorg. Chem.* 136, 106526
- 130 El-Khoury, R., *et al.* (2023) Telomeric i-motifs and C-strands inhibit parallel G-quadruplex extension by telomerase. *Nucleic Acids Res.* 51, 10395-10410
- 131 Kang, S., *et al.* (2021) Transcriptional regulation of telomeric repeat-containing RNA by acridine derivatives. *RNA Biol.* 18, 2261-2277
- 132 Gallego, J., *et al.* (1999) The folding of centromeric DNA strands into intercalated structures: a physicochemical and computational study. *J. Mol. Biol.* 285, 1039-1052
- 133 Gallego, J., *et al.* (1997) Centromeric pyrimidine strands fold into an intercalated motif by forming a double hairpin with a novel T:G:G:T tetrad: solution structure of the d(TCCCGTTTCCA) dimer. *J. Mol. Biol.* 273, 840-856
- 134 Garavis, M., *et al.* (2015) The structure of an endogenous Drosophila centromere reveals the prevalence of tandemly repeated sequences able to form i-motifs. *Sci. Rep.* 5, 13307
- 135 Henikoff, S., *et al.* (2017) Remarkable Evolutionary Plasticity of Centromeric Chromatin. *Cold Spring Harb. Symp. Quant. Biol.* 82, 71-82
- 136 Takahashi, S., *et al.* (2017) Topological impact of noncanonical DNA structures on Klenow fragment of DNA polymerase. *Proc. Natl. Acad. Sci. U. S. A.* 114, 9605-9610
- 137 McQuaid, K.T., *et al.* (2022) Interactions of small molecules with DNA junctions. *Nucleic Acids Res.* 50, 12636-12656
- 138 Matos-Rodrigues, G., *et al.* (2023) Detection of alternative DNA structures and its implications for human disease. *Mol. Cell* 83, 3622-3641
- 139 Mellor, C., *et al.* (2022) Creation and resolution of non-B-DNA structural impediments during replication. *Crit. Rev. Biochem. Mol. Biol.* 57, 412-442
- 140 Wright, E.P., *et al.* (2017) Identification of multiple genomic DNA sequences which form i-motif structures at neutral pH. *Nucleic Acids Res.* 45, 2951-2959