Topics in Current Chemistry (2020) 378:26
https://doi.org/10.1007/s41061-020-0290-z

REVIEW

Biotechnological and Therapeutic Applications of Natural Nucleic Acid Structural Motifs

Jinwei Duan[1,2,3] · **Xing Wang**[2,3,4] · **Megan E. Kizer**[5,6]

Received: 31 October 2019 / Accepted: 11 February 2020 / Published online: 18 February 2020

Abstract

Genetic information and the blueprint of life are stored in the form of nucleic acids. The primary sequence of DNA, read from the canonical double helix, provides the code for RNA and protein synthesis. Yet these already-information-rich molecules have higher-order structures which play critical roles in transcription and translation. Uncovering the sequences, parameters, and conditions which govern the formation of these structural motifs has allowed researchers to study them and to utilize them in biotechnological and therapeutic applications in vitro and in vivo. This review covers both DNA and RNA structural motifs found naturally in biological systems including catalytic nucleic acids, non-coding RNA, aptamers, G-quadruplexes, i-motifs, and Holliday junctions. For each category, an overview of the structural characteristics, biological prevalence, and function will be discussed. The biotechnological and therapeutic applications of these structural motifs are highlighted. Future perspectives focus on the addition of proteins and unnatural modifications to enhance structural stability for greater applicability.

Keywords Nucleic acids · Structural motifs · G-quadruplex · i-Motif · Ribozymes · Holliday junction · Therapeutics · Nanotechnology

Chapter 8 was originally published as Duan, J., Wang, X. & Kizer, M. E. Topics in Current Chemistry (2020) 378: 26. https://doi.org/10.1007/s41061-020-0290-z.

✉ Jinwei Duan
duanjw@chd.edu.cn

✉ Xing Wang
xingw@illinois.edu

✉ Megan E. Kizer
mkizer@mit.edu

Extended author information available on the last page of the article

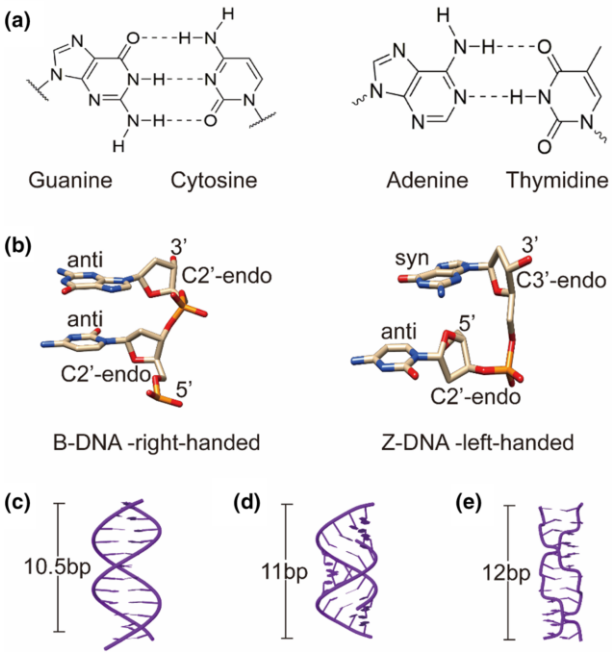

Fig. 1 Structure of DNA. **a** Watson–Crick base pairs. **b** *Anti*- and *syn*-conformations which differentiate Z- and B-DNA. Structural representation of B-DNA (**c**), A-DNA (**d**) and Z-DNA (**e**), with respective helical pitches

1 Introduction: Hierarchy of Interactions and Major Structural Motifs

The discovery of DNA structure in 1953 [1] has quickly evolved from the simplistic, Watson–Crick base-paired double helix to a more complex view. Soon, unique base–base or base–backbone interactions and higher-order nucleic acid structural states began to be uncovered [2–4]. Though initial structural knowledge provided an understanding of the basics of information storage from nucleic acid biomolecules, scientists soon realized there was much more at play than just sequence. It is now generally known that both sequence and structure of DNA and RNA are implicated in transcription, translation, and nuclear organization. This article highlights predominant nucleic acid structural motifs found in natural systems, their role in cellular function, and how they may be leveraged in therapeutic and biotechnological applications. To understand these higher-order structures, we must first understand the simplest forms of nucleic acids.

Nucleic acid structure can be broken down into three major hierarchical levels: primary, secondary, and tertiary. The primary form of DNA in nature is double stranded, while that of RNA is single stranded. Double stranded DNA (dsDNA) is formed through hybridization of bases, A–T and G–C (Fig. 1a), which promote an antiparallel arrangement of strands. DNA double helices can be found in one of three forms: B-DNA, A-DNA, and Z-DNA; they are structurally unique and their

characteristics are summarized in Table 1. B-DNA, being the predominant form physiologically, is a right-handed double helix with a major groove of 2.2 nm, a minor groove of 1.2 nm, and a helical pitch of 10.5 base pairs (bp) (Fig. 1c) [5]. DNA sequences are read from this structural form in the major groove, where a distinct pattern of hydrogen bonding accessible from the major groove is formed from each base pair. It is therefore this structural form which allows proteins to identify sequences for binding, transcription, and gene regulation. A-DNA is the double-stranded helix which forms under low humidity conditions. Its characteristics include a narrower and deeper major groove, a broader and shallower minor groove, and a helical pitch of 11 bp (Fig. 1d) [5]. This duplex form is still found in natural systems predominantly during DNA–protein complex interactions; it is also the form RNA adopts during hybridization (RNA–RNA). Interestingly, an RNA–DNA hybrid exhibits an intermediate of both B- and A-form helices in solution [6].

Much more distinct from the other two double-stranded forms is the left-handed Z-DNA. The purine and pyrimidine bases are usually in the *anti* position in right-handed DNA, but in left-handed DNA, the purine flips to the *syn* position (Fig. 1b) [5]. This rotation results in a puckering of the ribose sugar and generates a zig-zag pattern along the backbone, the basis for the name "Z-DNA". Structurally this form of DNA is more elongated, with a helical pitch of 12 bases per turn (Fig. 1e) [7]. Functionally, its exact role has not been unambiguously identified, though it has been implicated in regulating transcription events, following a moving polymerase. It is thought that the high energy conformer is able to form as a result of the negatively superhelical stress produced immediately after transcription and the Z-DNA that trails a polymerase can therefore block other polymerases from transcribing the same region. In other words, the Z-DNA provides control over the kinetics of transcription [5, 8]. Many proteins have been identified to bind Z-DNA with high affinity and specificity, further suggesting that they play important roles in transcriptional regulation, and therefore growth and development. The negatively superhelical stress also relieves topological strain from intertwining during recombination; chromosomal breakpoints in human tumors could arise from potential Z-DNA sequences. The secondary structures of DNA are only observed under specific conditions relevant for controlling biological processes and will be discussed in further detail in proceeding sections.

While DNA forms duplexes in its primary form, it is unusual that RNA forms long stretches of dsRNA. Instead, RNA retains single-stranded nature or folds back on itself for short, local dsRNA regions. When in duplex form, RNA resembles the

Table 1 Feature comparison of the different forms of biologically relevant duplex DNA

Duplex form	Handedness	Helical pitch (bp)	Diameter (Å)
B-DNA	Right	10.5	20
A-DNA	Right	11	23
Z-DNA	Left	12	18

A-DNA helix structural characteristics. The local dsRNA regions are considered secondary structural elements and include structures such as loops, hairpins, bulges, and pseudoknots (Fig. 2). Loops are flexible regions that connect two regions of secondary structure; they provide flexibility of the overall RNA molecule as well as flexibility in local and long-range base pairing [9]. Often, these structures confer specialized stability locally and globally. Hairpins are self-complementary sequences, connected by a loop which is usually at least four nucleotides (nt) long. The loop sequence UUCG shows special base-stacking interactions that promote stability of hairpin secondary structures. The hairpin structure has been heavily utilized in the biotechnological space in strand displacement-based methods of detection. Bulges provide local flexibility as well as enthalpic control over base pairing. A bulge in a region of dsRNA promotes melting to a greater extent than a fully duplexed strand; this is advantageous in hairpins and strand displacement events. When free distant ends of RNA fold over one another and hybridize with distant complementary sequences a pseudoknot structure is formed. Finally, pseudoknots result when one strand of RNA containing two stem–loops with one half of one stem intercalated between the two. This secondary structure is highly important for tertiary structure formation in vivo [10]. Formation of RNA knots in vivo has recently been accomplished to further understanding of RNA topology and development of RNA-based structural technologies [11].

DNA and RNA also form tertiary structures, though the greater flexibility imparted by the ribose sugar of RNA promotes a greater prevalence of non-Watson–Crick interactions which underscore these tertiary structures [12, 13].

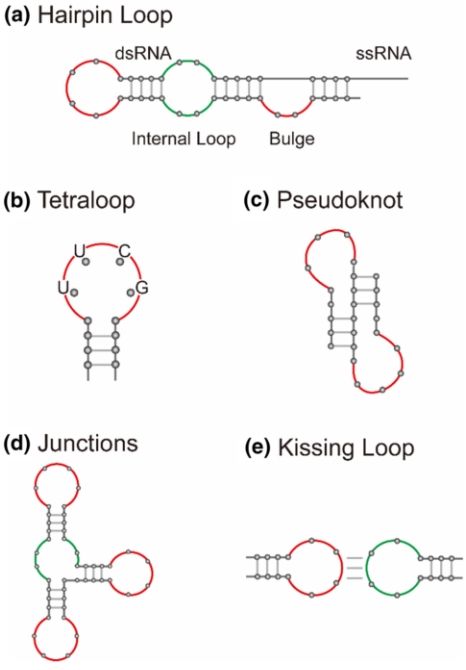

Fig. 2 Structural motifs of RNA. **a** Loops and bulges are intermittently observed along an RNA oligonucleotide, with single-stranded and double-stranded regions. **b** Certain nucleotide sequences in a loop confer special stability, such as the UUCG tetraloop. The **c** pseudoknot, **d** junctions, and **e** loop:loop secondary structures are prominent in tertiary RNA structures

Interactions such as G–U base pairs, base triples, and base–backbone interactions provide a much more diverse landscape of structural possibilities for RNA over DNA [13]. These tertiary structures are a result of multiple secondary structures which interact to fold the RNA into a highly compact and stable form [12]. Though not often mentioned, RNA can form quaternary structures by the assistance of proteins (e.g., ribosome). With this fundamental knowledge of nucleic acid structure, we are poised to fully understand higher-order natural structural motifs implicated in biological function and disease therapeutics.

2 RNA-Based Structural Technologies

2.1 Therapeutic Noncoding RNAs

RNA is well known for its crucial role in protein translation. But RNA plays equally important roles in many other biological processes. It, commonly called "noncoding RNA", protects and maintains the genome, processes RNA transcripts, is used as a defense mechanism against foreign DNA, exhibits enzymatic functions, and acts as structural elements which organize nuclear chromatin for gene expression [14, 15]. These diverse roles are linked by the primary, secondary, and tertiary structural elements which underscore proper RNA function. Biotechnological tools have leveraged the knowledge of sequence and structure information to control gene expression. The biochemical and structural background of dominant biotechnological tools is discussed in this section.

The well-characterized role of RNA in protein biosynthesis has been a long-standing target for controlling gene expression. Eukaryotes have a mechanism which allows for degradation of RNA (Fig. 3a). During transcription, a sense messenger RNA (mRNA) and antisense RNA (asRNA) strand are produced [14]. This asRNA hybridizes to the sense mRNA and the resulting dsRNA serves as a nucleation point for RNases, facilitating degradation. Specifically, an RNase called DICER degrades these transcripts into 20–25-nt-long dsRNAs containing a 2–3-base overhang at their 3′ end [16]. These are considered small interfering RNA (siRNA), where they are recognized by an RNAi-induced silencing complex (RISC) [17]. Current asRNA-based therapies rely on primary structure (or ssRNA) for effective control over biological processes. Introducing mRNA-targeted asRNA blocks the translation of the protein of interest, decreasing the diseased state cellular viability. In some cases, this interference does not directly block translation but rather blocks secondary RNA transcripts necessary for protein production. This mechanism is also imperative in the degradation of foreign RNA transcripts, resulting in RNA interference (RNAi) [18].

The use of primary structure is also observed in other RNA species, which have the potential to be targeted in a similar manner as to mRNA. These include small RNAs such as piwi-interacting RNA (piRNA) [19], small nuclear RNA (snRNA) [20], small nucleolar RNA (snoRNA) [21, 22], and micro RNA (miRNA) [23, 24]. For example, the *lin-4* RNA from *Caenorhabditis elegans* forms a dsRNA with multiple bulges; after processing into its mature form, the miRNA sequences exhibit

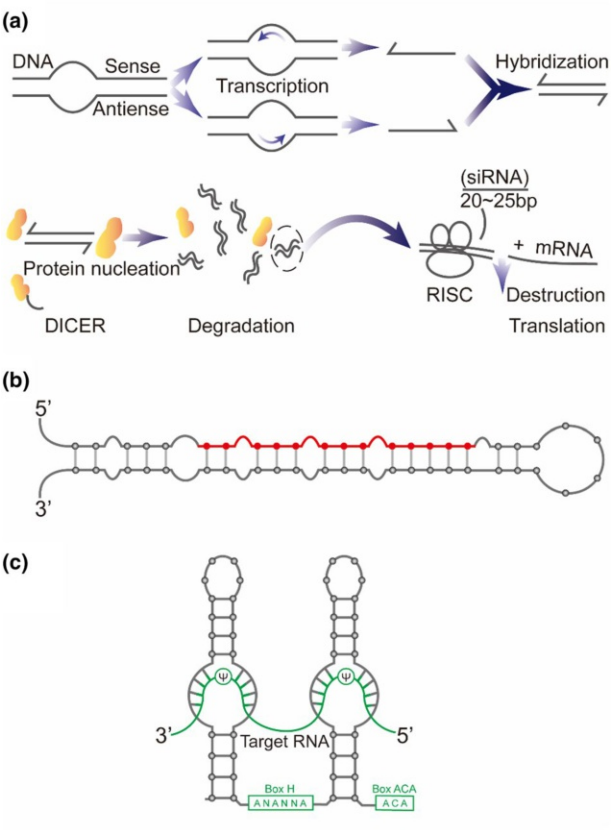

Fig. 3 Mechanism and structures of noncoding RNAs. **a** Antisense RNA binds mRNA and signals degradation. **b** The precursor structure of the miRNA *lin-4*. Mature *lin-4* miRNA is highlighted in red; regions of *lin-4* exhibit dsRNA with bulge regions to many portions of the *lin-14* target RNA, which degrades upon *lin-4* binding. **c** The structure of a snoRNA hybridized to its target rRNA. The H and ACA boxes are conserved sequences brought into proximity by two stem–loops. Base pairing within the stem bulge allows for rRNA specificity during pseudouridylation (Ψ)

complementarity to a cell fate determining protein, *lin-14,* exhibiting the ability to downregulate protein expression through RNAi mechanisms (Fig. 3b) [25]. SnoRNAs contain conserved sequence elements required for the conversion of specific ribosomal RNA (rRNA) uridines into pseudouridine [26]. These conserved sequences are always brought into close proximity to one another through the formation of two stem–loops. The secondary structure of two stem–loops containing bulges is necessary to allow for binding the rRNA substrate, where the modified pseudouridine base resides in the unpaired region of the bulge (Fig. 3c) [26]. These exemplify how an understanding of secondary structural elements in RNAi-related species can lead to the development of targeted nucleic acid therapies. Several RNAi-based therapies are in clinical trials and, most notably, one developed by Alnylam Pharmaceuticals to treat hereditary transthyretin-mediated (hATTR)

amyloidosis in adults was recently approved by the US Food and Drug and Administration (FDA) [27–29]. Furthermore, longer RNAs such as long noncoding RNAs (lncRNA) have just recently been implicated in ability to control gene expression and may represent a novel class of RNA as therapeutic targets [15, 30].

Prokaryotes exhibit similar protective strategies. The CRISPR system recognizes and removes foreign DNA, and since its discovery has burgeoned into a major therapeutic and biotechnology tool [31]. The strategic design of CRISPR RNA (crRNA), guide RNA (gRNA), and the combined single guide RNA (sgRNA) incorporates both primary sequence and secondary structure [32]. Recognition by the CRISPR-associated proteins involves specific secondary structures such as stem–loops, and targeting gene expression requires specific base pairing for each CRISPR system (PAM and NGG). The stem–loop secondary structure of the RNA makes key contacts with the Cas9 protein, allowing for highly efficient cleavage of the target DNA [33, 34]. This has been utilized in various genome editing contexts [35, 36].

2.2 Structural RNA Nanotechnology

Beyond its therapeutic applications of targeting gene expression, RNA has garnered great traction in the biotechnology space as a controllable nanomaterial for probes and tools [37]. Employing the naturally known interactions of RNA bases into higher-order structures, researchers have been able to develop large nanoscale RNA structures to perform specific tasks [11, 38]. This is executed in the same manner as DNA nanostructures (discussed in further detail in Sect. 5), where knowledge of the interactions between RNA sequences allows for the programmable formation of structures [39].

2.3 Catalytic Nucleic Acids: Ribozymes and DNAzymes

While primary structure underscores RNAi methods, RNA tertiary structure is utilized in quite a different mechanism for gene regulation and disease therapy. Certain RNAs in their tertiary form have been shown to catalyze reactions. These RNA-based enzymes, or ribozymes, are the only molecule known to catalyze their own cleavage and have revolutionized the way in which we view the primordial beginnings of the universe [40]. Discovered in the early 1980s, ribozymes have since become a heavily studied area of research, even developing novel ones [41]. Natural ribozymes are found in plants, bacteria, viruses, and lower eukaryotes, and mainly catalyze phosphate backbone cleavage (aside from RNase P which is discussed later). They exhibit the principal features of a protein enzyme, containing an active site, substrate, and cofactor binding site, yet they are 10^3-fold slower than proteins [42]. The chemical nature of RNA as an enzyme allows for substrate recognition through base pairing, conferring specificity for the ribonucleotide sequence to be removed.

They are categorized into two classes based on size and reaction mechanism. The large ribozymes include RNase P, group I and group II introns, while the small ribozymes include the hammerhead, hairpin, hepatitis delta virus (HDV),

and *Neurospora* Varkud satellite (VS) RNA [43]. Though they differ by reaction mechanism, they all require the presence of the divalent cation magnesium to facilitate tertiary structure formation and catalysis [44]. Group I and II introns catalyze a transesterification reaction to splice out introns from an RNA transcript. Group I introns range from a few hundred nt to around 3000; they have little sequence similarity across organisms, yet their secondary structures are highly conserved in four short regions which comprise a catalytic junction (boxed in Fig. 4a) [43]. In these regions, individual segments partially base pair with one another to confer the necessary structural requirements for strand cleavage. A total of ten paired segments form in these phylogenetically conserved secondary structures and retain catalytic active site function. Beyond secondary sequence, the group I introns contain large open reading frames (ORFs) and an internal guide sequence (IGS) which allows for the positioning of the target transcript and confers control over substrate specificity. The full secondary structural representation of the *Tetrahymena* ribozyme intron is shown in Fig. 4a. Group II introns, ranging only from several hundred to 2500 nt, are generally more elusive than group I introns since they are less widely distributed among organisms, and are rarely self-splicing in vitro because of extreme reaction condition requirements. Their secondary structure, generally across organisms, contains six helical domains (helices I–VI, Fig. 4b) where only helices I and V are indispensable. Specific intron binding sequences (IBS) and exon binding sequences (EBS) align the intron such that the 5′ and 3′ splice sites are in their proper locations.

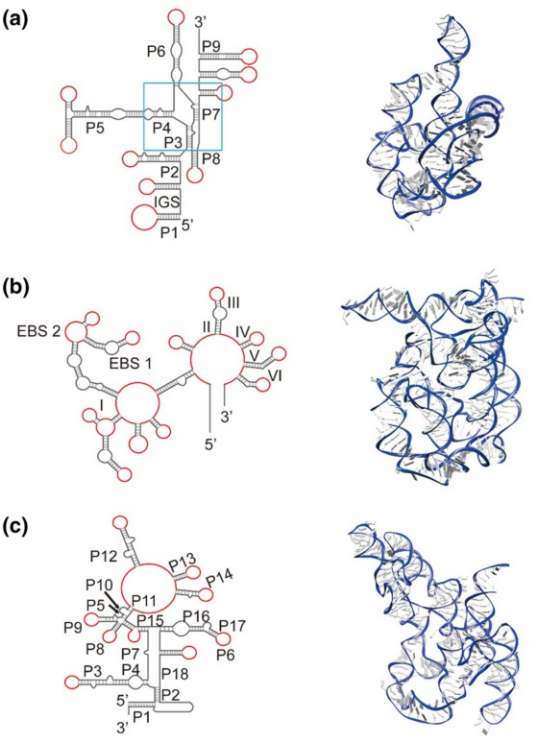

Fig. 4 Secondary and tertiary structures of the large class ribozymes: **a** *Tetrahymena* ribozyme, PDB file: 1GRZ; **b** Group IIC intron, PDB file: 3EOH; **c** RNase P, PDB file: 3Q1R

Any modification to these secondary structural motifs leads to diminished activity, revealing a common theme that conserved structural elements are highly important for proper tertiary structure formation, and therefore proper substrate binding and catalytic activity of ribozymes.

RNase P in all organisms processes the 5′ termini of transfer RNA (tRNA) precursors; it is unusual in that it is the only ribozyme that acts in a *trans*-active manner on multiple substrates, making it the only true naturally occurring RNA enzyme. It is also a unique ribozyme in that it is a ribonucleoprotein complex, where protein subunits are essential for proper formation of the full catalytic center. In eubacteria, the catalytic core is entirely made up of RNA, meanwhile in eukaryotes the protein content is much higher and RNA does not play the catalytic role. The general, common core structure of the catalytic RNase P RNA consists of 18 paired helices (Fig. 4c); different species contain extra stems and stem–loop structures to facilitate stability, and lower ionic strength requirements for catalysis. The 3′ half of the acceptor stem is thought to function as an external guide sequence (EGS) to position the substrate through tertiary structural features.

Small catalytic RNA species also include the hammerhead ribozyme, hairpin ribozyme, and VS ribozyme (Fig. 5). Similar to group I, the smaller ribozymes have conserved secondary structural motifs which enable function. The hammerhead ribozyme recognizes and cleaves an NUH sequence. The catalytic RNA contains three stem regions (Fig. 5a), of which sequences are conserved in stems I and III. Hammerhead ribozymes also contain single-stranded regions with highly conserved nucleotides, as well as three variable helical regions to effect self-cleavage. Mutational studies were performed to determine which nucleotides and structural motifs were requisite for catalysis. The hammerhead ribozyme is well characterized and utilized in many biotechnological applications. One major use is in the production of in vitro transcribed RNAs prepared for NMR studies; self-cleavage by the hammerhead ribozyme releases the desired transcript in high purity. They are also making strides in a therapeutic space, where a *trans*-cleaving hammerhead has been approved for phase II clinical trials against HIV-1 [45, 46]. The hairpin ribozyme, found in pathogenic plant satellite viruses, also requires conserved secondary structures for proper tertiary folding and catalysis. This ribozyme consists of four stem regions which comprise two major domains. The secondary structures of domain A (stem I–loop A–stem 2) and domain B (stem 3–loop B–stem 4), when lined up, resemble a hairpin (Fig. 5b) [47]. It recognizes the substrate of sequence RYN*CUG, where cleavage occurs at the *. Hairpin ribozymes from the tobacco ringspot virus, the arabis mosaic virus, and the chicory yellow mottle virus exhibit subtle nucleotide differences within helical regions that maintain the overall structure. Nucleotides are primarily conserved in the single-stranded regions, as well as the bulged motif between helices III and IV (Fig. 5b). The hairpin ribozyme has been modified to recognize and cleave mRNA of diseases including HIV-1, hepatitis B, and the Sindbis virus [46, 48]. The VS ribozyme mediates rolling-circle replication of plasmids from the *Neurospora* mitochondrion through formation of a multimeric, self-cleaving RNA [49]. During transcription of the VS plasmid, a dimeric RNA species is formed with each monomer containing one substrate domain and five catalytic domains (Fig. 5c). The substrate domain (helix 1) exhibits a stem–loop

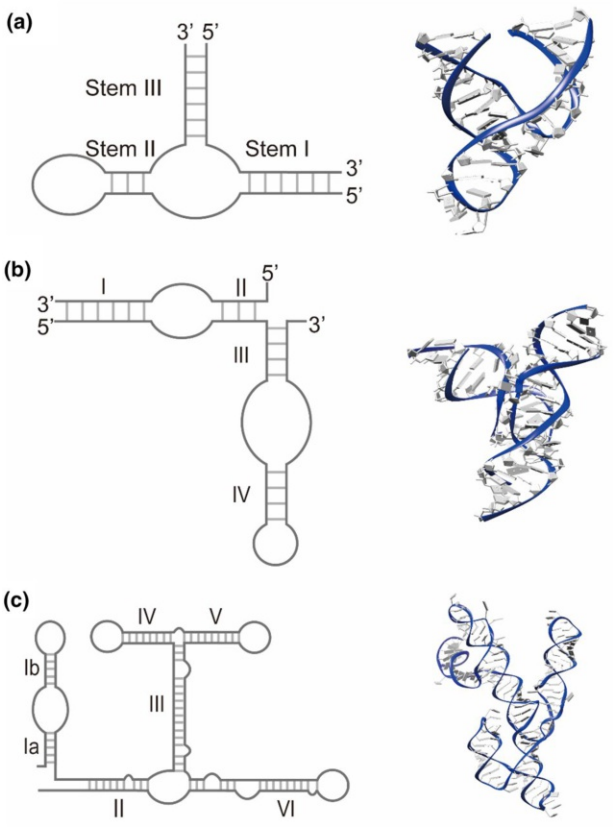

Fig. 5 Secondary and tertiary structures of the small class ribozymes: **a** the hammerhead ribozyme, PDB file: 299D; **b** hairpin ribozyme, PDB file: 1ZFV; and **c** the VS ribozyme (monomeric secondary and dimeric tertiary structures), PDB file: 4R4P

structure while the catalytic domains (helices 2–6) also exhibit stem–loop structures in addition to multiple tertiary contacts. These include three-way junctions and loop-loop kissing interactions to stabilize the catalytic domain. Globally, the catalytic helices come together in a tertiary structure such that the scissile phosphates are proximal to the catalytic nucleotides on the opposing monomer. The trend of secondary structure conservation for hairpin ribozyme catalysis emphasizes the fact that structural motifs, rather than individual nucleotides, are responsible for the function of catalytic RNAs. These RNA structures with known catalytic functionality can be leveraged to act as antivirals, therapeutics, and biotechnological agents.

Though there are a set number of ribozymes known to nature, researchers have developed methods to generate novel ribozymes to catalyze a variety of different reactions. This relies on the SELEX method [50, 51], where a random oligonucleotide pool is evolved in vitro to perform a specific task such as catalysis or binding (also called aptamers; see next section). Some groups have even developed a

strategy where certain secondary elements are retained, while other regions are evolved through mutagenesis. This has resulted in ribozymes which are able to catalyze reactions which recognize a specific sequence, perform nucleophilic attacks on various chemical centers, add metal ions to complexes, and isomerize ring structures. Catalytic DNA species, or DNAzymes, are not known to be found in natural systems and have been evolved from the SELEX method as well [52, 53]. Initially evolved DNAzymes were developed to recreate their natural RNA counterparts, cleaving and ligating RNA species [52]. Since the initial discovery of the capability of DNA to perform similar reactions, other DNAzymes processing RNA or DNA have been developed including those that catalyze RNA cleavage, DNA cleavage, DNA depurination, RNA ligation, DNA phosphorylation, and thymine dimer cleavage [14]. Scientists have investigated these DNA-based enzymes not only for their greater stability but also in an effort to understand the beginnings and evolution of information storage and catalytic species.

Ribozymes have been heavily utilized in molecular biology, but have seen limited translation to the clinical setting. Most predominantly used is the hammerhead ribozyme since it is small, easy to incorporate, and well characterized. Of notable examples are hammerhead ribozymes which have been developed to inhibit HIV. Ribozymes are often incorporated into RNAi systems, so that the hammerhead will perform catalytic cleavage to release the therapeutic siRNA. Though the discussed RNA systems have great therapeutic potential, they suffer major limitations including efficient and targeted cell delivery, target transcript specificity, and RNA stability [46]. Research directives are focused on overcoming these limitations; some, such as nucleic acid modification, are discussed in Sect. 6.

3 G-Quadruplex

3.1 Sequence, Structure, and Topology

Guanine (G)-rich sequences of DNA and RNA can associate into G-quartets, stabilized by Hoogsteen base pairing; stacks of two or more of these G-quartets lead to a higher-order nucleic acid structural motif called a G-quadruplex (G4) (Fig. 6) [54]. These higher-order structures are highly prevalent in the genomic context and are generally understood to play diverse roles in DNA replication, transcription, and translation, as well as controlling gene expression and genome stability. They are abundant in cancer genomes, making them potential therapeutic targets. Additionally, their stable structure is prevalent in aptamers, making them a major structural force in DNA/RNA-based biotechnological ligands.

The sequence requirement for G4 formation is $G_3\ N_{1-7}\ G_3\ N_{1-7}\ G_3\ N_{1-7}\ G_3$, but is driven by environmental factors such as monovalent cations and molecular crowding. Monovalent cations are crucial for this structure to form; preferentially, K^+ ions occupy the interior of the G-quartet to mitigate ion interactions by the bases. Na^+ has also been shown to occupy this ion space, but may alter the overall structure. The general trend of cations which promote G4 formation follows $K^+ > Na^+ > NH_4^+ > Li^+$. It was also revealed that the loop region length affects G4

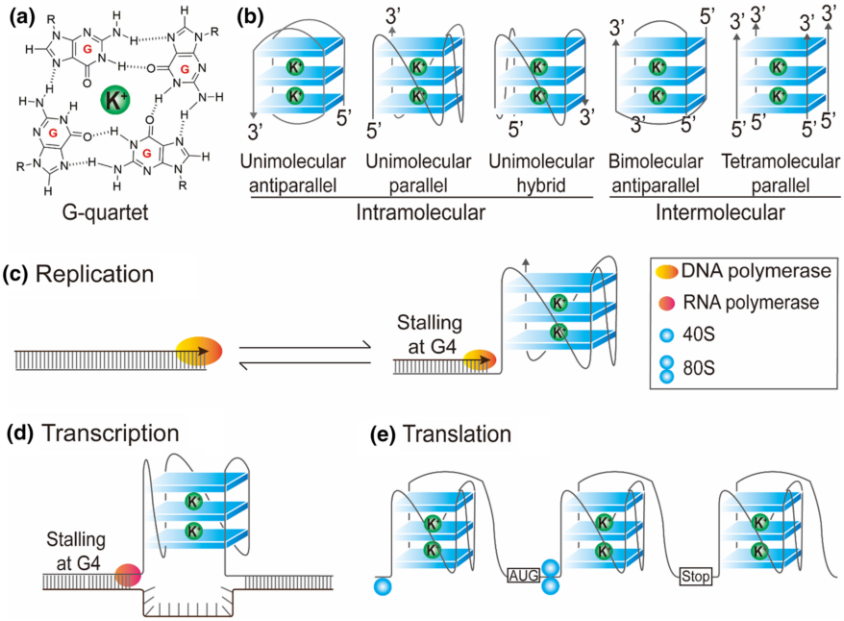

Fig. 6 G-Quadruplex structure and biological function. **a** Four guanine bases interact with a central K^+ ion to make one G-quartet. **b** Stacking of G-quartets forms G-quadruplexes (G4). The number of strands and their directionality dictate G4 topology. The presence of G4 structures in the genomic context regulates replication (**c**), transcription (**d**), and translation (**e**)

stability, and that G4 structures may also deviate from the general sequence requirements through incorporation of bulges. Structurally, the G4 is composed of four strands which can be unimolecular, bimolecular, or tetramolecular, and can have varying topologies governed by the parallel or antiparallel arrangements of the individual strands [54]. The different strand arrangements, schematically illustrated in Fig. 6, generate different topological features. G4 topology can crudely be determined through experimental methods such as monitoring of positive or negative changes in circular dichroism (CD) spectra at specific wavelengths. Complete structure determination is obtained though NMR studies. Many G4 structures have been confirmed in vitro, but fewer have been confirmed in vivo.

3.2 Biological Relevance

G4s are known to form at a variety of different regions in the cellular context including telomeres in eukaryotes, promoter regions, between introns and exons, 5′ untranslated regions (5′-UTR), and at DNA breakpoints. Stable G4 structures impede the progression of DNA polymerase and lead to replication stalling, DNA damage, and genomic instability. Still, the entire picture of G4 biological relevance is not complete. To clarify the elusive relevance of G4s, G-rich regions have been predicted from a number of algorithms, and G4-specific sequencing techniques and

detection probes have been developed to monitor said sequences. Over 300,000 sequences have been identified as potential G-quadruplex forming sequences (pG4s) [55]. And a more recent study identifies over 700,000 pG4s. Their prevalence is certainly great, but the spatiotemporal formation of pG4s and their relevance in vivo is a greater and more complex matter. Therefore, the pG4s must be unambiguously detected in vivo to validate these predictions.

G4 sequencing in vitro utilizes G4-induced DNA polymerase stalling followed by next-generation sequencing [56]. The sequencing is first performed under non-G4-promoting conditions then under G4-promoting conditions (presence of K^+). This type of sequencing has identified G4s in gene regulatory regions including 5′-UTR and splicing sites, and in cancer-related genes and in regions of somatic copy number alteration (SCNA) in cancer genomes.

Direct detection of G4 structures is mainly executed with small molecules or proteins which specifically bind G4s. G4-interacting molecules including pyridostatin (PDS) and its fluorescent analog PDS-α, as well as telomestatin have been utilized for imaging and therapeutics. Their interaction with G4s also traps the structure in the G4 formation, allowing a snapshot into the G4 but preventing the investigation of dynamics and regulation of G4 formation. A more spatiotemporal method of G4 mapping can be done with proteins which bind to naturally forming G4 as opposed to promoting G4 formation under specific conditions. The first antibody against G4s, a scFv antibody Sty49, was used to show that G4s form at telomeres [55]. Later two structure-specific antibodies were developed, BG4 and 1H6, which allowed for pull-down sequencing and immunofluorescence imaging of G4 location in eukaryotic cells. These proteins are paramount for G4 ChIP sequencing, where the G4-specific antibodies are used as probes. G4 ChIP-seq has identified double-stranded break sites induced by PDS, identification of protein interactions that promote formation of other G4 and other DNA structures, that G4s predominantly lie in regulatory, nucleosome-depleted chromatin regions that are highly transcribed.

Other dye- and protein-based probes have been developed to monitor G4 formation in vitro and in vivo. These include TSQ1, CyT, and anthrathiophenedione dyes, a G4-triggered fluorogenic hybridization probe, the zinc-finger protein GQ1, and ankyrin repeat binding proteins (DARPins) [55]. Helicases which resolve G4s have also been mapped to bind G4s, implicating that their biological functions are linked to G4 structures or genomic regions enriched in G4. One group developed a fluorophore-conjugated RHAU helicase peptide as a probe for G4 formation.

Overall, these methods have confirmed that G4s are prevalent in a genomic context and their formation is dynamic and regulated. This reasons that the expression of G4 structural motifs will vary depending on cell type and cell cycle progression. Yet there is much more work to be done to specifically determine their biological roles in various contexts.

3.3 Therapeutic and Biotechnological Applications

The G4 structure, as was just discussed, is heavily implicated in biological processes and predominantly prevalent in cancer genomics and is therefore a prime therapeutic

target. Telomeres are a particularly major target for G4 binding molecules since their abnormal processing in cancers ubiquitously confers cell immortality. Additionally, many oncogenes have pG4s or confirmed G4 structures in their promoter regions. Targeting these G4 structures with small molecules promotes the downregulation of expression of these genes, thereby reducing tumor viability. Since the G4s found in telomeres and oncogene promoter regions are so similar, it is imperative to obtain a selective binding G4 ligand for the particular target. Though small molecules have been developed which target G4s, selectivity of the particular oncogenic G4s still proves a challenging, yet crucial task.

Many ligands have been identified which bind to G4s in a therapeutic context. For telomeric G4s, these are chiefly 2,6-diaminoanthraquinone derivatives and telomestatin. Telomeric G4s may exist as dimers or multimers and some ligands have even been identified to selectively bind to multimeric G4s over monomeric G4s. These include a dinickel salophen dimer, berberine dimer, and telomestatin derivative tetramer which bind to dimeric G4s and m-TMPipEOPP which preferentially binds multimeric G4s. Interestingly, the junctions between the monomers in multimeric G4s have also been specifically recognized by ligands [57].

Oncogenic G4s are similarly targeted for therapeutic applications. Generally, rational and high-throughput screens have been employed to develop various molecular ligands to oncogenic-specific G4 structures. To enhance specificity for a particular oncogene, pyrrole-imidazole polyamide (PIP) molecules bind in a sequence-selective manner to duplex DNA and can be hybridized to G4-targeting ligands to enhance the overall selectivity. This was shown in the development of cyclic imidazole/lysine polyamide conjugated PIP (cIKP-PIP) [58]. Some specific oncogenes of interest include the promoter region G4 of *c-Myc,* which is specifically targeted by GQC-05 [59], crescent-shaped thiazole peptide (TH3) [60], four-leaf-clover-like molecule (IZCZ-3) [61], and others identified by a microarray screen of 20,000 small molecule binders. VEGF G4s are targeted by perylene monoimide derivative (PM2) [62], and a quinoline derivative (SYUIQ-FM05) [63]; these have encouraged studies of small molecule VEGF-G4-preferred ligands obtained through a low-volume screening approach [57]. BCL2 G4 structures are preferentially targeted through furo[2,3-*d*]pyridazin-4(5*H*)-one derivatives [64], as well as the fluorescent dye carbazole TO [65]. G4s in the promoter region of *c-Kit* oncogene (also known as mast/stem cell growth factor receptor Kit) are targeted by derivatives of isoalloxazine (*N,N*-dimethyl amine and *N,N*-dimethyl amine/2-fluorine substitutions) [66], naphthalene diimide [67], benzo[*a*]phenoxazine (BPO) [68], and carbazole derivatives [69]. Human telomerase reverse transcriptase (hTERT) is overexpressed in cancers and contains G4 tracts in the promoter region which form higher-order G4 structures; a dual-motif targeting small molecule, GTC365, binds to the G4 and mismatched duplex stem–loop structures [70]. *KRAS* G4 structures have recently been selectively targeted with triple-cation derivatives of indolo[3,2-*c*]quinolines (IQc) [71]. The *KRAS* oncogene is also downregulated through a unique decoy system where pyrene-modified oligonucleotides which form a more stable G4 attract essential transcription factors, preventing transcription from actually happening [72]. Topotecan was developed to selectively target *c-myb* G4 structures [73]. Beyond these recent examples, many other oncogenes have been identified to contain G-rich

promoter regions, suggesting that there may be selective small molecule inhibitors to genes such as *PDGFRB*, *PDGFA*, *STAT3*, and *FGFR2*.

In addition to being a therapeutic target, G4s are also highly prevalent in biotechnological applications such as biosensors and diagnostics. Aptamers, predominant therapeutic and diagnostic agents to a variety of substrates, exhibit G4s as a crucial structural feature [55]. G4 structures in an aptamer are highly advantageous as they provide thermodynamic and chemical stability, resist serum nucleases, reduce immunogenicity, and increase cellular uptake. Aptamers evolved to target a variety of different ligands using the SELEX method form G4s previous to or upon ligand binding. Some notable examples of G4-containing aptamers include the well-characterized thrombin-binding aptamer (TBA) which forms G4 upon interaction with thrombin [74]. Another catalytic aptamer that forms G4 upon ligand binding is PS2.M, which binds heme to form a G-quadruplex–hemin complex [75]. This unique aptamer requires G4 sequence for function and can be used to mimic the activity of horseradish peroxidase, providing a visual output for detection of various molecules. The spinach aptamer, which recognizes 3,5-difluoro-4-hydroxybenzylidene imidazolinone, contains a G4 motif [76]; since development of the spinach aptamer, derivatives which act as fluorescent binders to other molecules also contain G4 motifs. This provides insight into the structural requirements for fluorescent aptamer ligands.

One unique application of G4s involves the modulation of G4 formation to detect lead ions (Pb^{2+}) in solution [77]. A G4 tethered to a carbon nanotube electrode is linear in its native form. In the presence Pb^{2+} ions, a G4 structure is induced, resulting in a greater emission by intercalating agent ethyl green. Another creative diagnostic application includes G4 motifs that were utilized to detect low-abundance nucleic acid molecules such as pathogenic DNA [78, 79]. This employs a technique called quadruplex priming amplification (QPA) where the dissociation of dsDNA and formation of G4 occurs upon primer extension [79]. The G4 structure that forms during QPA is detected by incorporation of 2-aminopurine (2-AP) bases; 2-AP is quenched by neighboring bases in the linear form, but regains emissive properties in the G4 form. It is clear from the current research landscape that G4 structures are (1) highly abundant in the human genome, suggesting their role in biology and therapeutics; (2) predominant in various aptamer structures, suggesting that they play vital stabilizing roles; and (3) emerging in biosensor and diagnostic applications, signifying their potential in other disciplines.

4 i-Motif

4.1 Sequence, Structure, and Topology

The cytosine (C) complement to G-quadruplexes also forms unique higher-order structures. Initially discovered by Gehring et al. [80], C-rich oligonucleotides form an intercalated quadruple-helical tetramolecular structure, called intercalated or i-motifs. Similar to G-quadruplexes, the C-rich region preferentially forms under acidic conditions, but recently it has been shown that these sequences can form

i-motifs at neutral pH depending on the sequence and environmental conditions [81, 82]. The i-motif structure consists of two parallel-stranded duplexes intercalated in an antiparallel orientation (Fig. 7), generally composed of cytosine–cytosine base pair regions and loop regions [83]. The interacting C bases allow for the intercalated structure through a hemi-protonated $C:C^+$ base pair with three hydrogen bonds which confers significant stability [82]. The fundamental factor contributing to i-motif stability is the number of $C:C^+$ base pairs. A central positive charge and some π stacking further stabilize the interaction of the bases. Loop regions also add a special stability, such as the case of the G:T:G:T tetrad loop [82]. The loop regions define the two classes of i-motifs, where class I and II are characteristic of shorter and longer loops, respectively [84]. Longer loops, contrary to G4 motifs, are more stable than the shorter loops.

i-Motif structures also have topological features associated with their unique secondary structure. As a result of the spatial arrangement of the $C:C^+$ base pair, two

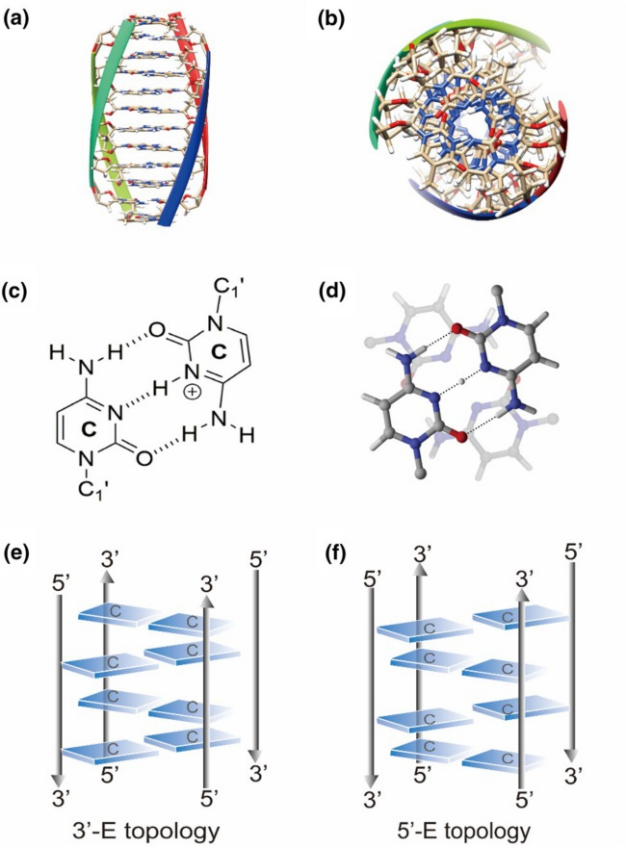

Fig. 7 i-Motif structure. **a** Side and **b** top view of the $d(TC_5)$ intermolecular i-motif. PDB file: 225D. **c** Chemical and **d** ball-and-stick representation of the $C:C^+$ base pair. i-Motifs exhibit either 3′ (**e**) or 5′ (**f**) topology depending on where the outmost $C:C^+$ base pair resides

topologies arise, 3E′ and 5E′ [83]. 3E′ forms when the outmost C:C^+ base pair is at the 3′ end of the base, and 5E′ forms when the terminal C:C^+ base pair is at the 5′ end. As a result of extended sugar–sugar contacts along the narrow grooves, the 3E′ topology imparts greater stability to the structure [82, 85].

4.2 Biological Relevance

It has been established that i-motifs can form in vivo; furthermore, that they are dependent on the cell cycle and the pH [81, 82, 86]. The environmental conditions which allow for this at physiological pH include salt concentration, negative superhelicity, and molecular crowding. Increasing NaCl concentration to 100 mM destabilizes the i-motif, but further increasing to 300 mM has no more destabilizing effect [85]. Negative superhelicity arises during unwinding of dsDNA during transcription events, and this also promotes i-motif formation. Finally, molecular crowding promotes i-motif formation by preventing Watson–Crick duplex formation and is likely the major factor which allows i-motifs to form in vivo. Uracil bases (U) from spontaneous deamination of cytosine significantly reduce the thermal stability of i-motifs; however, ribonucleic acid i-motif structures, with and without U, have been identified in vitro [87].

Understanding the conditions which promote formation at physiological conditions gives insight into the biological role i-motifs may play. Since G canonically base pairs with C, it is presumed that i-motifs form at the same genomic loci as G4s and therefore that they perform opposite gene regulatory functions. Also like G-rich regions, C-rich regions are found enriched at promoter regions, telomeres, and centromeres. However, not all C-rich regions form i-motifs as formation is dependent on the environmental conditions discussed. Promoter regions shown to form i-motifs include *c-Myc*, *BCL2*, and human acetyl-CoA carboxylase (*ACC1*). Whether or not the *c-Myc* expression is turned on is dependent on the extent of transcriptionally induced negative superhelicity which promotes i-motif formation. Regulation of *Bcl2* by stabilizing molecules identified in a screen shows that presence of the i-motif structure upregulates *Bcl2* expression while a hairpin structure represses transcription [85, 88]. *AAC1* i-motif formation under molecular crowding conditions upregulates expression. i-Motif foci are more prominent during the G1/S phase, further supporting the notion that i-motif structures play an opposing role to G4 structures in cell biology; G4 suppresses transcription while i-motif activates transcription [82].

Uncovering other i-motif-forming genomic loci requires specific ligands which can recognize these structural motifs. Unlike G4s, there are much fewer i-motif-specific peptide and small molecule ligands ideal for studying i-motifs. This makes it more difficult to identify the spatiotemporal expression of these C-rich higher-order structures in a cellular context. Though some notable ligands have been developed to further this understanding, they each have their disadvantages. A porphyrin ligand, TMPyP4, initially utilized to study G4 interactions showed independent binding to i-motif structures to induce an inhibitory effect on the NM23-H2 involved in transcriptional activation of the *c-Myc* gene [82, 89]. Molecules such as crystal

violet, bisacridine (BisA), and phenanthroline derivatives have exhibited binding to i-motifs; some were stabilizing probes, but were not selective [88, 90]. Metal-centered ligands such as the ruthenium complex $[Ru(bpy)_2(dppx)]^{2+}$ and the terbium aminoacid complex $[Tb_2(DLHVal)_4(H_2O)_8]Cl_6$ bind i-motifs and cause slight structural perturbations. Nanomaterials like carboxyl-modified single-walled carbon nanotubes (C-SWNTs) and graphene quantum dots (GQDs) have shown stabilization of i-motifs through respectively binding the end of the major groove and the internal TAA loop [91, 92]. Two small molecule ligands, IMC-48 and IMC-76, were identified to bind to the i-motif and hairpin conformation of C-rich regions, respectively. These two ligands were utilized to shift the equilibrium of the structures and as a result control expression of *BCL2* mRNA. Many proteins have been identified as poly C binding proteins (PCBP) and include heterogeneous nuclear riboprotein K (hnRNPK), a 39-kDa polypeptide from *Trypanosoma brucei*, and the BmlLF protein of *Bombyx mori*. Finally, one antibody exists against the i-motif, iMab. This antibody binds to C-rich, i-motif-forming DNA sequences over other DNA species (dsDNA, hairpins, and G4s) [86].

4.3 Therapeutic and Biotechnological Applications

In stark contrast to G4s as therapeutic targets, i-motifs are significantly lacking. Not many i-motifs have emerged as promising candidates for therapeutics since there are limited selective i-motif ligands, yet creative approaches are being pursued to incorporate i-motifs in therapeutics. The small molecule IMC-76 in concert with ellipticine (GQC-05) simultaneously targets the *BCL2* and *c-Myc* oncogene promoters. GQC-05 stabilizes the *c-Myc* G4 structure and turns off gene expression, while IMC-76 stabilizes the hairpin structure of *BCL2* and decreases mRNA levels [93]. The combinatorial approach to nucleic acid structure-targeted therapeutics resulted in high sensitivity of lymphoma cells to the chemotherapeutic drug cyclophosphamide. Further development and refinement of i-motif-specific ligands is paramount to their efficacy as cancer therapeutics.

The i-motif sees a much greater role in the biotechnology space; it is central in the design of nanotechnological systems for analytical and biomedical purposes, relying on the structural transition of i-motif sequences as a result of pH changes. The ease of changing structure is leveraged to provide desired outputs. Some notable examples of the i-motif switch (I-switch) in nanotechnology include monitoring pH, controlling DNA nanostructure assemblies, drug release platforms, and biosensors [83]. The first I-switch, developed by the Krishnan group, was designed to sense and report pH changes along endosomal maturation in living cells in culture and in vivo [94]. The addition of gold nanoparticles (AuNPs) to these switches enhances their ability to detect pH changes by a visible readout of aggregated AuNPs (Fig. 8a) [95]. Leveraging i-motif structural changes, their incorporation into DNA-based assemblies can provide sensitive readouts and key functionalities for various biotechnological applications [96]. Aside from detecting pH, i-switches can act as a controllable release mechanism, e.g., capping mesoporous silica nanoparticles that can open and close their pores to release cargo by changing pH [97]. Anchoring a

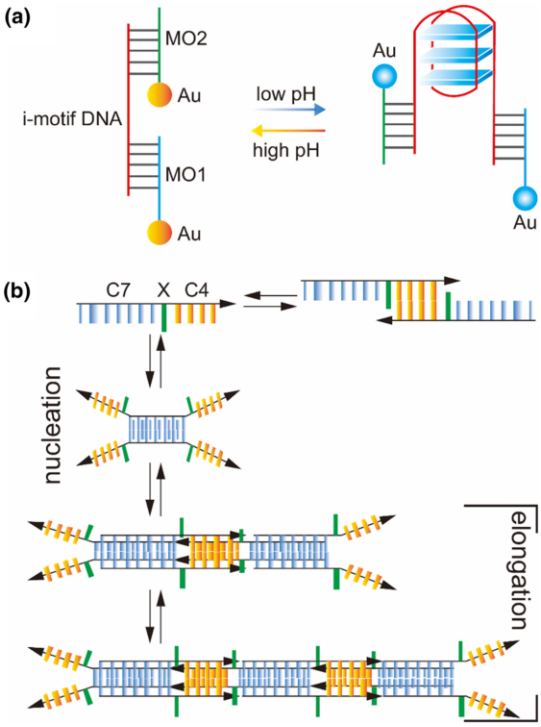

Fig. 8 i-Motif biotechnological applications. **a** Structural transition of a linear ↔ i-motif assembly under different pH results in aggregation of attached AuNPs and change in color. This i-motif molecular switch produces a visible color change to detect pH. **b** Association of i-motif-containing oligonucleotides into a 1D supramolecular "wire"

dense monolayer of i-motifs with polyA connectors to a flat gold surface also generates a nanocarrier; the i-motifs effectively cap the interior polyA, creating a lid over encapsulated ions or small molecules. i-Motifs have also been utilized in the controlled formation of DNA-based nanomaterials [96]. A 1D "wire" of i-motif-containing oligonucleotides was initially developed by Ghodke et al., where the designed DNA strands were annealed at low pH to allow for growth to persist in one direction (Fig. 8b) [98]. In a similar design, i-motifs were central to the formation of DNA pillars [99]. Allowing i-motif oligos to propagate in three dimensions generates a pH-responsive DNA hydrogel [100].

5 Nucleic Acid Junctions and Nanostructures

5.1 Biological Relevance and Structure

Helical junctions, or points where two nucleic acid helices cross over one another, are important in biological processes. For RNA, as discussed in Sect. 1, three-way junctions mainly serve an architectural and stabilizing role. Meanwhile for DNA, four-way junctions are key intermediates of homologous recombination events during meiosis; they are observed in various DNA repair events to maintain genomic integrity. The biological relevance of this junction was initially proposed by Robin Holliday in 1964 to describe the mechanism of DNA strand

exchange which results in genetic diversity. As such, these DNA-based four-way helical junctions are also known as Holliday junctions (HJ). In biology, HJs exhibit the ability to migrate along the DNA axis based on homologous sequence and are resolved by the action of enzymes which recognize the specific topological features of the crossover. Though most prominently studied in DNA, crossover events have also been observed in RNA of the Brome mosaic virus (BMV) [101]. Since HJ structures exhibit homology, they are mobile along the DNA strand, a phenomenon called branch migration (Fig. 9a). This initially rendered it difficult to characterize them in vitro, but the generation of "immobilized" junctions [102, 103] allowed for the characterization of the DNA crossover structures [104–106].

The two DNA helices can align parallel or antiparallel to one another, resulting in two different mechanisms of strand exchange: crossed and un-crossed (or "square") (Fig. 9b). But how these strands exchange in three-dimensional space was first determined through biochemical assays performed by Lilley [107]; the crossovers adopt a global X shape, where the two helical arms stack over one another. Lilley also showed that these crossovers require divalent cations such as Mg^{2+} to remain stable. In 1994, the first crystal structure of a Holliday junction complexed with a resolving enzyme was obtained [108]. We now have many

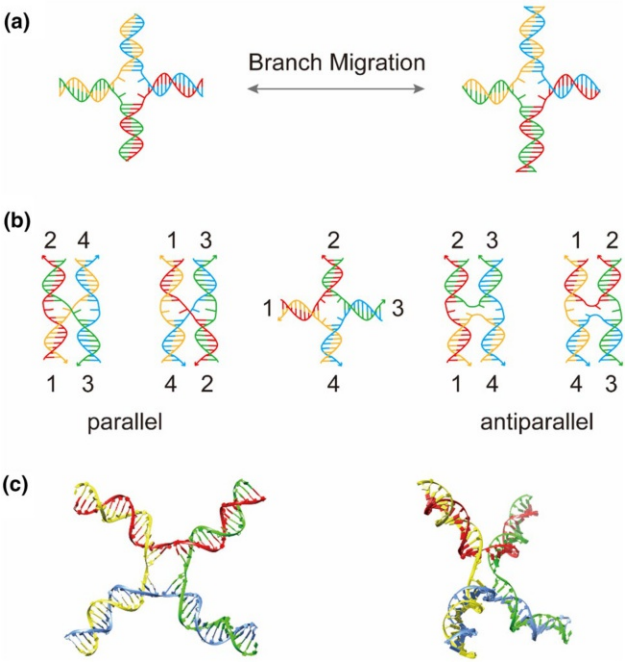

Fig. 9 Structure of a 4-way DNA junction (Holliday junction). **a** Homologous sequences of DNA can come together via reciprocal strand exchange to generate a 4-way junction. The junction is mobile along the DNA allowing for branch migration. **b** Strand exchange can occur from different helical alignments (parallel and antiparallel) as well as different strand arrangements (crossed or square). **c** Crystal structure of a Holliday junction [109] (PDB 2CRX) revealing an X-type tertiary structure

crystal structures of HJ-resolvases, providing a clear picture of these crossovers (Fig. 9c).

5.2 Therapeutic and Biotechnological Applications

The most prominent use of DNA junctions in biotechnology and therapeutics has been in the construction of DNA nanostructures. This significant discipline of the structural DNA nanotechnology field emerged from the ability to generate immobile (those which do not undergo branch migration) four-way helical junctions [102, 110]. Ned Seeman utilized these immobile junctions in the construction of repeating and periodic DNA branched junctions, resulting in DNA lattices and other higher-order DNA structures [102]. Some systems rely on structural motifs such as branched junctions [111] and stem–loops [112] to generate DNA-based hydrogels. The judicious incorporation of these elements and the base-pairing design allows for hybridization chain reactions to propagate and generate a long, complex, and entangled mesh of DNA, or hydrogel. Such hydrogels have broad biomedical applications for controlling cell growth or development as well as controllable drug release [113, 114].

In another design approach, the alignment of multiple, parallel HJ motifs along two DNA duplexes generates paranemic crossover (PX) structures [115]; these have been used in nanostructure formation, computation, and nanomachines [116]. The expansion of HJ motifs into an in vitro assembly of multiple repeating units pioneered the DNA nanostructure field and gave rise to the use of DNA as a spatially addressable material [102]. Later, others began developing different types of DNA nanostructures based on the same concept. Paul Rothemund utilized multiple branched junctions as structural motifs to fold single-stranded bacteriophage DNA, M13mp18, creating the DNA origami technique (Fig. 10a) [117, 118]. Peng Yin developed the self-assembly of DNA tiles and bricks to further expand the rapidly growing number of structures that can be made using short DNA oligonucleotides (Fig. 10b, c) [119–121].

Strategic base pairing and formation of crossover junctions underscores the DNA nanotechnology field, allowing researchers to develop specific and precise shapes on the nanometer scale. The ever-growing chemical modifications available to DNA and the inherent specificity of DNA allows for the precise placement of various ligands [121]. Incorporation of stimuli-responsive species imparts control to DNA nanostructures, promoting mechanical motions, specific outputs, or timed-release. These exquisite properties and characteristics make DNA nanostructures advantageous tools for a variety of biotechnological and therapeutic applications, the major categories being molecular tools, biosensors, and cell delivery agents [127].

The precise placement of ligands has enabled development of “molecular pegboards” as tools to study individual molecular interactions, distance requirements, and enzymatic cascades. Recent efforts have utilized DNA origami and other DNA-based structures to understand physical determinants for energy transfer [128–130]. Investigation and utilization of enzymatic cascades have also benefitted greatly from the properties of templated DNA nanostructures (Fig. 10d) [122, 131, 132]. One

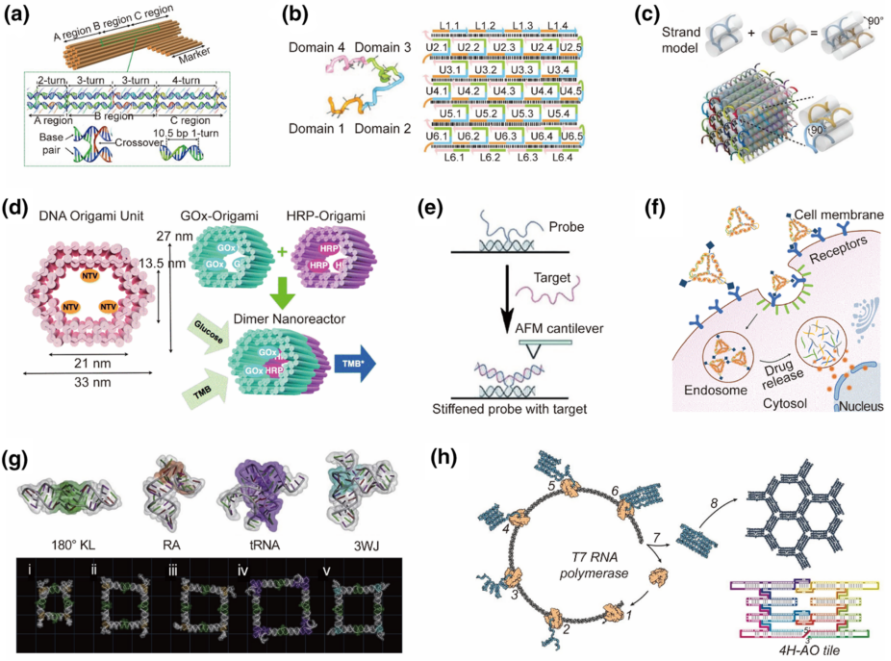

Fig. 10 DNA and RNA nanostructures and their applications. **a** The DNA origami technique judiciously places crossover junctions to fold ssDNA into nanoscale objects. Adapted with permission from Ref. [118]. Copyright 2017, The Japan Society of Applied Physics. **b** Single-stranded oligonucleotides containing four variable domains repeatedly assemble by crossovers, leading to the design of DNA tiles which generate various shapes and patterns. Adapted with permission from Ref. [120]. Copyright 2012, Springer Nature. **c** DNA bricks form a "LEGO-like" interaction of two DNA strands with complementary head/tail regions. The perpendicular arrangement of complementary DNA strands generates three parallel helices for each pair of bricks, allowing the structure to be built into controlled, three-dimensional shapes. Adapted with permission from Ref. [119]. Copyright 2012, American Association for the Advancement of Science. **d** Enzymatic reactions are studied using DNA origami. A DNA origami unit containing NeutrAvidin (NTV) sites was designed to encapsulate biotinylated glucose oxidase (GOx) or horseradish peroxidase (HRP) enzymes. The nanoreactors dimerize through base pair interactions and their close proximity allows for the enzymatic cascade to ensue, producing TMB*. Adapted with permission from Ref. [122]. Copyright 2015, Royal Society of Chemistry. **e** DNA origami enables single-molecule analysis. A single-stranded probe placed on the origami surface is indistinguishable by atomic force microscopy (AFM). Upon hybridization with a single target RNA strand, a V-shaped junction is formed and is easily visible by AFM. Adapted with permission from Ref. [123]. Copyright 2012, Wiley-VCH Verlag GmbH and Co. **f** DNA nanostructures act as drug delivery systems. Drug-loaded DNA nanostructures can be appended with cell-targeting ligands. The nanostructures enter the cells via receptor-mediated endocytosis where they are then degraded to release the drug cargo. Adapted with permission from Ref. [124]. Copyright 2018, American Chemical Society. **g** RNA structural motifs are utilized to controllably build polygons. The top row shows the 3D structures of representative structural motifs, PDB codes from left to right: 180° KL (1JJM), RA (1JJ2), tRNA (4TNA), and 3WJ (4V4Q). The bottom row shows various polygonal shapes made with the representative structural motifs. The colors in the polygons correspond to the regions colored in the top row. Adapted with permission from Ref. [125]. Copyright, 2018 Elsevier. **h** RNA origami can be folded co-transcriptionally to produce RNA tiles. As the RNA is transcribed from the DNA template, the programmed 180° and 120° KL motifs form to produce 11 helical subdomains. Adapted with permission from Ref. [126]. Copyright 2014, American Association for the Advancement of Science

particular system incorporates a four-way junction and two crossover motifs to strategically place glucose oxidase (GOx) and horseradish peroxidase (HRP) enzymes which are controlled by a strand displacement mechanism [133]. Origami surfaces are also an advantageous platform for single-molecule analysis (Fig. 10e) [123], and the development of a DNA origami "frame" allowed for the investigation of various nucleic acid interactions [127]. At the same time, origami structures can be designed to mimic more intractable biological structures, such as transmembrane pores, to understand the key determinants for molecular transport [134, 135]. In a unique approach, a DNA-based copy-print system confers two-dimensional information from a DNA origami sheet to the surface of AuNPs [136]. DNA origami-based nanoimprinting lithography (DONIL) [136] shows great promise for self-assembly of nanostructures to be applied in biomedical applications for precisely tailored optical and electronic properties. Together, the variety of applications of DNA origami as molecular tools provides a wealth of opportunities for the translation of these structures in diverse disciplines related to biomedical research.

Aside from providing a platform for fundamental understanding of biological phenomena as just described, these controllable origami structures are incredibly advantageous for therapeutic applications such as biosensing and drug delivery. The addition of i-motifs, fluorophores, AuNPs, and other responsive modules and outputs are all major players in the development of DNA origami biosensors [137, 138]. Recently, we have constructed a star-shaped DNA architecture, called "DNA star" that contains five 4-arm junctions at each of the inner pentagon vertices. At these junctions, 10 dengue envelope protein domain III (ED3)-binding aptamers are placed in a precise 2D pattern which mirror the spatial arrangement of ED3 clusters on the dengue viral surface. The resulting polyvalent, spatial pattern-matching interactions provide high dengue-binding avidity, which leads to potent viral inhibition. Hybridization of fluorescent output pairs renders the DNA star into a sensitive dengue biosensor [139]. In another recent work, a supramolecular DNA origami precisely immobilized AuNPs to localize single dye molecules for the generation of surface-enhanced Raman scattering (SERS) [140]. This provides a new approach to single-molecule studies using the well-characterized DNA origami as a platform, showing the breadth of disciplines where DNA origami-based advances are still being uncovered. Finally, DNA-based materials have been intercalated with anticancer drugs such as doxorubicin, and directly conjugated with disease-targeting ligands for specific, efficient, and controllable drug delivery (Fig. 10f) [124].

As a more conformationally dynamic and complex nucleic acid species, RNA has also been employed for the generation of nanostructures [125]. The rise of RNA as a nanostructure scaffold has emerged from the assembly of packaging RNA (pRNA) into dimers, trimers, and hexamers, and the observation that kissing loops promote RNA structural formation [141]. Since these formative discoveries, great efforts were made to generate more complex and chemically defined RNA nanostructures. Precise programming of secondary and tertiary interactions including kissing loops, hand-in-hand, foot-to-foot, crossovers, and junctions leads to the engineering of higher-order RNA structures such as nanoparticles, tiles, lattices, polyhedra, and origami [141, 142]. Specific structural motifs have been utilized to build various polygonal nanostructures (Fig. 10g), and are fully summarized with in a recent review by

Ohno et al. [125]. Exemplar advancements in RNA as structural tools involve the design of single-stranded RNA to fold back over on itself without the formation of knots [143], and the unique design of an RNA which concomitant with transcription folds back on itself to form structural elements requisite for tile formation (Fig. 10h) [126]. RNA nanoparticles are utilized to contain other therapeutically active RNA species such as siRNA, ribozymes, and aptamers [37]; meanwhile, other structural RNAs are utilized as molecular scaffolds to enhance bacterial metabolic activity, control mammalian cell surface interactions, and cell fate signals through localization and aggregation of distinct molecular targets [125].

6 Conclusions and Future Outlooks

6.1 Proteins Join the Game

Proteins existing naturally in biological systems specifically bind DNA on the basis of sequence. These include transcription activator-like (TAL) effector proteins and the CRISPR-associated proteins. TAL proteins produced by plant pathogenic bacteria recognize sequence-specific promoter regions of host cells; meanwhile, the protein recognition region allows for protein engineering to develop those which bind user-defined sequences. These proteins have been cleverly applied to fold DNA species into DNA–protein hybrid shapes, similar to the DNA origami method [144]. Bringing proteins into the fold of megadalton, multilayer DNA structures promotes structural rigidity for more robust applications. This natural, sequence-specific DNA–protein interaction can also be leveraged by the CRISPR-associated proteins and can lead to greater nucleic acid-based tools for biotechnological applications.

RNA nanostructures have also been constructed through a combination of RNA–protein (RNP) interactions. Various proteins are known to interact with specific RNA structural motifs, e.g., the L7 protein binds a specific target called "box C/D". Tailoring the RNA sequence and judicious choice of RNA-binding protein result in formation of complex which can generate triangular and square RNP structures [125].

6.2 Unnatural Modifications for Stability and Applicability

Extensive reports of unnatural modifications in the nucleic acid structural motifs discussed above have revealed our ability to chemically manipulate these naturally found structures for prolonged stability or to promote a particular equilibrium/structure formation in vivo. G4 and i-motif both have the ability to incorporate nucleic acid modifications to increase stability. While most sugar modifications destabilize i-motifs, modification of the phosphates promotes i-motif formation [82]. Phosphorothioate allows formation at neutral pH and Rp-chirality exhibits greater stabilization of the i-motif [82]. Peptide nucleic acids (PNA) and locked nucleic acids (LNA) have been investigated to form these higher-order structures [54]. While they have been shown to form G4 and i-motifs, PNA tends to do so at narrower pH ranges.

Though RNA has greater conformational flexibility which confers a greater molecular design space, it is also more susceptible to degradation, a tremendous disadvantage for applicability. Yet the increasing number and incorporation of nucleic acid modifications have enabled the use of RNA as a molecular scaffold amenable to applications in vivo [141].

Funding Xing Wang was funded by HT Materials Corporation. Jinwei Duan was funded by Natural Science Foundation of Shaanxi Province (Nos. 2017JQ2035, 2018JQ2023) and Fundamental Research Funds for the Central Universities (No. 300102129101).

Compliance with Ethical Standards

Conflict of interest On behalf of all authors, the corresponding authors state that there is no conflict of interest.

References

1. Watson JD, Crick FHC (1953) Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. Nature 171(4356):737–738. https://doi.org/10.1038/171737a0
2. Das J, Mukherjee S, Mitra A, Bhattacharyya D (2006) Non-canonical base pairs and higher order structures in nucleic acids: crystal structure database analysis. J Biomol Struct Dyn 24(2):149–161. https://doi.org/10.1080/07391102.2006.10507108
3. Lipps HJ, Gruissem W, Prescott DM (1982) Higher order DNA structure in macronuclear chromatin of the hypotrichous ciliate *Oxytricha nova*. Proc Natl Acad Sci 79(8):2495. https://doi.org/10.1073/pnas.79.8.2495
4. Guéron M, Leroy J-L (2000) The i-motif in nucleic acids. Curr Opin Struct Biol 10(3):326–331. https://doi.org/10.1016/S0959-440X(00)00091-9
5. Watson J, Baker TA, Bell SP, Gann A, Levine M, Losick R (2014) The structures of DNA and RNA. In: Molecular biology of the gene. Pearson, London
6. Salazar M, Fedoroff OY, Miller JM, Ribeiro NS, Reid BR (1993) The DNA strand in DNA–RNA hybrid duplexes is neither B-form nor A-form in solution. Biochemistry 32(16):4207–4215. https://doi.org/10.1021/bi00067a007
7. Rich A, Nordheim A, Wang AHJ (1984) The chemistry and biology of left-handed Z-DNA. Annu Rev Biochem 53(1):791–846. https://doi.org/10.1146/annurev.bi.53.070184.004043
8. Herbert A, Rich A (1998) Left-handed Z-DNA: structure and function. In: Bradbury EM, Pongor S (eds) Structural biology and functional genomics, vol 71. Kluwer, Dordrecht. https://doi.org/10.1007/978-94-011-4631-9_3
9. Conn GL, Draper DE (1998) RNA structure. Curr Opin Struct Biol 8(3):278–285. https://doi.org/10.1016/S0959-440X(98)80059-6
10. Staple DW, Butcher SE (2005) Pseudoknots: RNA structures with diverse functions. PLoS Biol 3(6):e213–e213. https://doi.org/10.1371/journal.pbio.0030213
11. Qi X, Zhang F, Su Z, Jiang S, Han D, Ding B, Liu Y, Chiu W, Yin P, Yan H (2018) Programming molecular topologies from single-stranded nucleic acids. Nat Commun 9(1):4579. https://doi.org/10.1038/s41467-018-07039-7
12. Batey RT, Rambo RP, Doudna JA (1999) Tertiary motifs in RNA structure and folding. Angew Chem Int Ed 38(16):2326–2343. https://doi.org/10.1002/(SICI)1521-3773(19990816)38:16%3c2326:AID-ANIE2326%3e3.0.CO;2-3
13. Butcher SE, Pyle AM (2011) The molecular interactions that stabilize RNA tertiary structure: RNA motifs, patterns, and networks. Acc Chem Res 44(12):1302–1311. https://doi.org/10.1021/ar200098t

14. Clark DP, Pazdernik NJ (2016) RNA-based technologies, chapter 5. In: Clark DP, Pazdernik NJ (eds) Biotechnology, 2nd edn. Academic Cell, Boston, pp 131–179. https://doi.org/10.1016/B978-0-12-385015-7.00005-3
15. Mercer TR, Mattick JS (2013) Structure and function of long noncoding RNAs in epigenetic regulation. Nat Struct Mol Biol 20:300. https://doi.org/10.1038/nsmb.2480
16. Jaskiewicz L, Filipowicz W (2008) Role of dicer in posttranscriptional RNA silencing. In: Paddison PJ, Vogt PK (eds) RNA interference. Springer, Berlin, pp 77–97. https://doi.org/10.1007/978-3-540-75157-1_4
17. Houseley J, Tollervey D (2009) The many pathways of RNA degradation. Cell 136(4):763–776. https://doi.org/10.1016/j.cell.2009.01.019
18. Tuschl T (2001) RNA interference and small interfering RNAs. ChemBioChem 2(4):239–245. https://doi.org/10.1002/1439-7633(20010401)2:4%3c239:AID-CBIC239%3e3.0.CO;2-R
19. Ozata DM, Gainetdinov I, Zoch A, O'Carroll D, Zamore PD (2019) PIWI-interacting RNAs: small RNAs with big functions. Nat Rev Genet 20(2):89–108. https://doi.org/10.1038/s41576-018-0073-3
20. Paul CP, Good PD, Winer I, Engelke DR (2002) Effective expression of small interfering RNA in human cells. Nat Biotechnol 20(5):505–508. https://doi.org/10.1038/nbt0502-505
21. Aftab MN, He H, Skogerbø G, Chen R (2008) Microarray analysis of ncRNA expression patterns in *Caenorhabditis elegans* after RNAi against snoRNA associated proteins. BMC Genom 9(1):278. https://doi.org/10.1186/1471-2164-9-278
22. Kiss T (2002) Small nucleolar RNAs: an abundant group of noncoding RNAs with diverse cellular functions. Cell 109(2):145–148. https://doi.org/10.1016/S0092-8674(02)00718-3
23. Fang W, Bartel DP (2015) The menu of features that define primary microRNAs and enable de novo design of microRNA genes. Mol Cell 60(1):131–145. https://doi.org/10.1016/j.molcel.2015.08.015
24. Hammond SM (2006) RNAi, microRNAs, and human disease. Cancer Chemother Pharmacol 58(1):63–68. https://doi.org/10.1007/s00280-006-0318-2
25. Lee RC, Feinbaum RL, Ambros V (1993) The *C. elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell 75(5):843–854. https://doi.org/10.1016/0092-8674(93)90529-Y
26. Ganot P, Bortolin M-L, Kiss T (1997) Site-specific pseudouridine formation in preribosomal RNA is guided by small nucleolar RNAs. Cell 89(5):799–809. https://doi.org/10.1016/S0092-8674(00)80263-9
27. Kim D, Rossi J (2008) RNAi mechanisms and applications. Biotechniques 44(5):613–616. https://doi.org/10.2144/000112792
28. Setten RL, Rossi JJ, Han S-p (2019) The current state and future directions of RNAi-based therapeutics. Nat Rev Drug Discov 18(6):421–446. https://doi.org/10.1038/s41573-019-0017-4
29. Weng Y, Xiao H, Zhang J, Liang X-J, Huang Y (2019) RNAi therapeutic and its innovative biotechnological evolution. Biotechnol Adv 37(5):801–825. https://doi.org/10.1016/j.biotechadv.2019.04.012
30. Fernandes JCR, Acuña SM, Aoki JI, Floeter-Winter LM, Muxel SM (2019) Long non-coding RNAs in the regulation of gene expression: physiology and disease. Noncoding RNA 5(1):17. https://doi.org/10.3390/ncrna5010017
31. Hsu Patrick D, Lander Eric S, Zhang F (2014) Development and applications of CRISPR-Cas9 for genome engineering. Cell 157(6):1262–1278. https://doi.org/10.1016/j.cell.2014.05.010
32. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337(6096):816. https://doi.org/10.1126/science.1225829
33. Jiang F, Taylor DW, Chen JS, Kornfeld JE, Zhou K, Thompson AJ, Nogales E, Doudna JA (2016) Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. Science 351(6275):867. https://doi.org/10.1126/science.aad8282
34. Nishimasu H, Ran FA, Hsu Patrick D, Konermann S, Shehata Soraya I, Dohmae N, Ishitani R, Zhang F, Nureki O (2014) Crystal structure of Cas9 in complex with guide RNA and target DNA. Cell 156(5):935–949. https://doi.org/10.1016/j.cell.2014.02.001
35. Dominguez AA, Lim WA, Qi LS (2015) Beyond editing: repurposing CRISPR–Cas9 for precision genome regulation and interrogation. Nat Rev Mol Cell Biol 17:5. https://doi.org/10.1038/nrm.2015.2

36. Donohoue PD, Barrangou R, May AP (2018) Advances in industrial biotechnology using CRISPR-Cas systems. Trends Biotechnol 36(2):134–146. https://doi.org/10.1016/j.tibtech.2017.07.007
37. Guo P (2010) The emerging field of RNA nanotechnology. Nat Nanotechnol 5:833. https://doi.org/10.1038/nnano.2010.231
38. Li M, Zheng M, Wu S, Tian C, Liu D, Weizmann Y, Jiang W, Wang G, Mao C (2018) In vivo production of RNA nanostructures via programmed folding of single-stranded RNAs. Nat Commun 9(1):2196. https://doi.org/10.1038/s41467-018-04652-4
39. Grabow WW, Jaeger L (2014) RNA self-assembly and RNA nanotechnology. Acc Chem Res 47(6):1871–1880. https://doi.org/10.1021/ar500076k
40. Cech TR (2002) Ribozymes, the first 20 years. Biochem Soc Trans 30(6):1162. https://doi.org/10.1042/bst0301162
41. Cech TR (1990) Self-splicing and enzymatic activity of an intervening sequence RNA from *Tetrahymena*. Biosci Rep 10(3):239. https://doi.org/10.1007/BF01117241
42. Doudna JA, Lorsch JR (2005) Ribozyme catalysis: not different, just worse. Nat Struct Mol Biol 12(5):395–402. https://doi.org/10.1038/nsmb932
43. Tanner NK (1999) Ribozymes: the characteristics and properties of catalytic RNAs. FEMS Microbiol Rev 23(3):257–275. https://doi.org/10.1111/j.1574-6976.1999.tb00399.x
44. Walter NG, Engelke DR (2002) Ribozymes: catalytic RNAs that cut things, make things, and do odd and useful jobs. Biologist (London) 49(5):199–203
45. Fedoruk-Wyszomirska A, Szymański M, Głodowicz P, Gabryelska M, Wyszko E, Estrin William J, Barciszewski J (2015) Inhibition of HIV-1 gp41 expression with hammerhead ribozymes. Biochem J 471(1):53. https://doi.org/10.1042/BJ20150398
46. James HA, Gibson I (1998) The therapeutic potential of ribozymes. Blood 91(2):371
47. Fedor MJ (2000) Structure and function of the hairpin ribozyme. J Mol Biol 297(2):269–291. https://doi.org/10.1006/jmbi.2000.3560
48. Famulok M, Hartig JS, Mayer G (2007) Functional aptamers and aptazymes in biotechnology, diagnostics, and therapy. Chem Rev 107(9):3715–3743. https://doi.org/10.1021/cr0306743
49. Suslov NB, DasGupta S, Huang H, Fuller JR, Lilley DMJ, Rice PA, Piccirilli JA (2015) Crystal structure of the Varkud satellite ribozyme. Nat Chem Biol 11:840. https://doi.org/10.1038/nchembio.1929. https://www.nature.com/articles/nchembio.1929#supplementary-information
50. Tuerk C, Gold L (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science 249(4968):505. https://doi.org/10.1126/science.2200121
51. Ellington AD, Szostak JW (1990) In vitro selection of RNA molecules that bind specific ligands. Nature 346(6287):818–822. https://doi.org/10.1038/346818a0
52. Breaker RR, Joyce GF (1994) A DNA enzyme that cleaves RNA. Chem Biol 1(4):223–229. https://doi.org/10.1016/1074-5521(94)90014-0
53. Cuenoud B, Szostak JW (1995) A DNA metalloenzyme with DNA ligase activity. Nature 375(6532):611–614. https://doi.org/10.1038/375611a0
54. Burge S, Parkinson GN, Hazel P, Todd AK, Neidle S (2006) Quadruplex DNA: sequence, topology and structure. Nucleic Acids Res 34(19):5402–5415. https://doi.org/10.1093/nar/gkl655
55. Kwok CK, Merrick CJ (2017) G-Quadruplexes: prediction, characterization, and biological application. Trends Biotechnol 35(10):997–1013. https://doi.org/10.1016/j.tibtech.2017.06.012
56. Lam EYN, Beraldi D, Tannahill D, Balasubramanian S (2013) G-quadruplex structures are stable and detectable in human genomic DNA. Nat Commun 4:1796. https://doi.org/10.1038/ncomms2792. https://www.nature.com/articles/ncomms2792#supplementary-information
57. Asamitsu S, Obata S, Yu Z, Bando T, Sugiyama H (2019) Recent progress of targeted G-quadruplex-preferred ligands toward cancer therapy. Molecules 24(3):429. https://doi.org/10.3390/molecules24030429
58. Asamitsu S, Li Y, Bando T, Sugiyama H (2016) Ligand-mediated G-quadruplex induction in a double-stranded DNA context by cyclic imidazole/lysine polyamide. ChemBioChem 17(14):1317–1322. https://doi.org/10.1002/cbic.201600198
59. Brown RV, Danford FL, Gokhale V, Hurley LH, Brooks TA (2011) Demonstration that drug-targeted down-regulation of MYC in non-Hodgkins lymphoma is directly mediated through the promoter G-quadruplex. J Biol Chem 286(47):41018–41027. https://doi.org/10.1074/jbc.M111.274720

60. Dutta D, Debnath M, Müller D, Paul R, Das T, Bessi I, Schwalbe H, Dash J (2018) Cell penetrating thiazole peptides inhibit c-MYC expression via site-specific targeting of c-MYC G-quadruplex. Nucleic Acids Res 46(11):5355–5365. https://doi.org/10.1093/nar/gky385
61. Hu M-H, Wang Y-Q, Yu Z-Y, Hu L-N, Ou T-M, Chen S-B, Huang Z-S, Tan J-H (2018) Discovery of a new four-leaf clover-like ligand as a potent c-MYC transcription inhibitor specifically targeting the promoter G-quadruplex. J Med Chem 61(6):2447–2459. https://doi.org/10.1021/acs.jmedchem.7b01697
62. Taka T, Joonlasak K, Huang L, Randall Lee T, Chang S-WT, Tuntiwechapikul W (2012) Down-regulation of the human VEGF gene expression by perylene monoimide derivatives. Bioorg Med Chem Lett 22(1):518–522. https://doi.org/10.1016/j.bmcl.2011.10.089
63. Wu Y, Zan L-P, Wang X-D, Lu Y-J, Ou T-M, Lin J, Huang Z-S, Gu L-Q (2014) Stabilization of VEGF G-quadruplex and inhibition of angiogenesis by quindoline derivatives. Biochim Biophys Acta 9:2970–2977. https://doi.org/10.1016/j.bbagen.2014.06.002
64. Amato J, Pagano A, Capasso D, Di Gaetano S, Giustiniano M, Novellino E, Randazzo A, Pagano B (2018) Targeting the BCL2 gene promoter G-quadruplex with a new class of furopyridazinone-based molecules. ChemMedChem 13(5):406–410. https://doi.org/10.1002/cmdc.201700749
65. Gu Y, Lin D, Tang Y, Fei X, Wang C, Zhang B, Zhou J (2018) A light-up probe targeting for Bcl-2 2345 G-quadruplex DNA with carbazole TO. Spectrochim Acta Part A Mol Biomol Spectrosc 191:180–188. https://doi.org/10.1016/j.saa.2017.10.012
66. Bejugam M, Sewitz S, Shirude PS, Rodriguez R, Shahid R, Balasubramanian S (2007) Trisubstituted isoalloxazines as a new class of G-quadruplex binding ligands: small molecule regulation of c-kit oncogene expression. J Am Chem Soc 129(43):12926–12927. https://doi.org/10.1021/ja075881p
67. Gunaratnam M, Collie GW, Reszka AP, Todd AK, Parkinson GN, Neidle S (2018) A naphthalene diimide G-quadruplex ligand inhibits cell growth and down-regulates BCL-2 expression in an imatinib-resistant gastrointestinal cancer cell line. Bioorg Med Chem 26(11):2958–2964. https://doi.org/10.1016/j.bmc.2018.04.050
68. McLuckie KIE, Waller ZAE, Sanders DA, Alves D, Rodriguez R, Dash J, McKenzie GJ, Venkitaraman AR, Balasubramanian S (2011) G-quadruplex-binding benzo[a]phenoxazines down-regulate c-KIT expression in human gastric carcinoma cells. J Am Chem Soc 133(8):2658–2663. https://doi.org/10.1021/ja109474c
69. Głuszyńska A, Juskowiak B, Kuta-Siejkowska M, Hoffmann M, Haider S (2018) Carbazole derivatives' binding to c-KIT G-quadruplex DNA. Molecules 23(5):1134. https://doi.org/10.3390/molecules23051134
70. Kang H-J, Cui Y, Yin H, Scheid A, Hendricks WPD, Schmidt J, Sekulic A, Kong D, Trent JM, Gokhale V, Mao H, Hurley LH (2016) A pharmacological chaperone molecule induces cancer cell death by restoring tertiary DNA structures in mutant hTERT promoters. J Am Chem Soc 138(41):13673–13692. https://doi.org/10.1021/jacs.6b07598
71. Lavrado J, Brito H, Borralho PM, Ohnmacht SA, Kim N-S, Leitão C, Pisco S, Gunaratnam M, Rodrigues CMP, Moreira R, Neidle S, Paulo A (2015) KRAS oncogene repression in colon cancer cell lines by G-quadruplex binding indolo[3,2-c]quinolines. Sci Rep 5:9696. https://doi.org/10.1038/srep09696
72. Cogoi S, Paramasivam M, Filichev V, Géci I, Pedersen EB, Xodo LE (2009) Identification of a new G-quadruplex motif in the KRAS promoter and design of pyrene-modified G4-decoys with antiproliferative activity in pancreatic cancer cells. J Med Chem 52(2):564–568. https://doi.org/10.1021/jm800874t
73. Li F, Zhou J, Xu M, Yuan G (2018) Exploration of G-quadruplex function in c-Myb gene and its transcriptional regulation by topotecan. Int J Biol Macromol 107:1474–1479. https://doi.org/10.1016/j.ijbiomac.2017.10.010
74. Schultze P, Macaya RF, Feigon J (1994) Three-dimensional solution structure of the thrombin-binding DNA aptamer d(GGTTGGTGTGGTTGG). J Mol Biol 235(5):1532–1547. https://doi.org/10.1006/jmbi.1994.1105
75. Travascio P, Li Y, Sen D (1998) DNA-enhanced peroxidase activity of a DNA aptamer–hemin complex. Chem Biol 5(9):505–517. https://doi.org/10.1016/S1074-5521(98)90006-0
76. Paige JS, Wu KY, Jaffrey SR (2011) RNA mimics of green fluorescent protein. Science 333(6042):642. https://doi.org/10.1126/science.1207339

77. Ebrahimi M, Raoof JB, Ojani R (2017) Design of a novel electrochemical biosensor based on intramolecular G-quadruplex DNA for selective determination of lead(II) ions. Anal Bioanal Chem 409(20):4729–4739. https://doi.org/10.1007/s00216-017-0416-5
78. Gogichaishvili S, Lomidze L, Kankia B (2014) Quadruplex priming amplification combined with nicking enzyme for diagnostics. Anal Biochem 466:44–48. https://doi.org/10.1016/j.ab.2014.08.025
79. Taylor A, Joseph A, Okyere R, Gogichaishvili S, Musier-Forsyth K, Kankia B (2013) Isothermal quadruplex priming amplification for DNA-based diagnostics. Biophys Chem 171:1–8. https://doi.org/10.1016/j.bpc.2012.11.001
80. Gehring K, Leroy JL, Guéron M (1993) A tetrameric DNA structure with protonated cytosine–cytosine base pairs. Nature 363(6429):561–565. https://doi.org/10.1038/363561a0
81. Wright EP, Huppert JL, Zoë ZAE (2017) Identification of multiple genomic DNA sequences which form i-motif structures at neutral pH. Nucleic Acids Res. https://doi.org/10.1093/nar/gkx090
82. Abou Assi H, Garavís M, González C, Damha MJ (2018) i-Motif DNA: structural features and significance to cell biology. Nucleic Acids Res 46(16):8038–8056. https://doi.org/10.1093/nar/gky735
83. Benabou S, Aviñó A, Eritja R, González C, Gargallo R (2014) Fundamental aspects of the nucleic acid i-motif structures. RSC Adv 4(51):26956–26980. https://doi.org/10.1039/c4ra02129k
84. Gurung SP, Schwarz C, Hall JP, Cardin CJ, Brazier JA (2015) The importance of loop length on the stability of i-motif structures. Chem Commun 51(26):5630–5632. https://doi.org/10.1039/c4cc07279k
85. Day HA, Pavlou P, Waller ZAE (2014) i-Motif DNA: structure, stability and targeting with ligands. Bioorg Med Chem 22(16):4407–4418. https://doi.org/10.1016/j.bmc.2014.05.047
86. Zeraati M, Langley DB, Schofield P, Moye AL, Rouet R, Hughes WE, Bryan TM, Dinger ME, Christ D (2018) I-motif DNA structures are formed in the nuclei of human cells. Nat Chem 10(6):631–637. https://doi.org/10.1038/s41557-018-0046-3
87. Snoussi K, Nonin-Lecomte S, Leroy JL (2001) The RNA i-motif. J Mol Biol 309(1):139–153. https://doi.org/10.1006/jmbi.2001.4618
88. Sedghi Masoud S, Nagasawa K (2018) i-Motif-binding ligands and their effects on the structure and biological functions of i-motif. Chem Pharm Bull 66(12):1091–1103. https://doi.org/10.1248/cpb.c18-00720
89. Dexheimer TS, Carey SS, Zuohe S, Gokhale VM, Hu X, Murata LB, Maes EM, Weichsel A, Sun D, Meuillet EJ, Montfort WR, Hurley LH (2009) NM23-H2 may play an indirect role in transcriptional activation of c-myc gene expression but does not cleave the nuclease hypersensitive element III(1). Mol Cancer Ther 8(5):1363–1377. https://doi.org/10.1158/1535-7163.MCT-08-1093
90. Zhang XY, Luo HQ, Li NB (2014) Crystal violet as an i-motif structure probe for reversible and label-free pH-driven electrochemical switch. Anal Biochem 455:55–59. https://doi.org/10.1016/j.ab.2014.03.015
91. Li X, Peng Y, Ren J, Qu X (2006) Carboxyl-modified single-walled carbon nanotubes selectively induce human telomeric i-motif formation. Proc Natl Acad Sci 103(52):19658. https://doi.org/10.1073/pnas.0607245103
92. Chen X, Zhou X, Han T, Wu J, Zhang J, Guo S (2013) Stabilization and induction of oligonucleotide i-motif structure via graphene quantum dots. ACS Nano 7(1):531–537. https://doi.org/10.1021/nn304673a
93. Kendrick S, Muranyi A, Gokhale V, Hurley LH, Rimsza LM (2017) Simultaneous drug targeting of the promoter MYC G-quadruplex and BCL2 i-motif in diffuse large B-cell lymphoma delays tumor growth. J Med Chem 60(15):6587–6597. https://doi.org/10.1021/acs.jmedchem.7b00298
94. Modi S, Swetha MG, Goswami D, Gupta GD, Mayor S, Krishnan Y (2009) A DNA nanomachine that maps spatial and temporal pH changes inside living cells. Nat Nanotechnol 4(5):325–330. https://doi.org/10.1038/nnano.2009.83
95. Zhao Y, Cao L, Ouyang J, Wang M, Wang K, Xia X-H (2013) Reversible plasmonic probe sensitive for pH in micro/nanospaces based on i-motif-modulated morpholino-gold nanoparticle assembly. Anal Chem 85(2):1053–1057. https://doi.org/10.1021/ac302915a
96. Dong Y, Yang Z, Liu D (2014) DNA nanotechnology based on i-motif structures. Acc Chem Res 47(6):1853–1860. https://doi.org/10.1021/ar500073a
97. Chen C, Pu F, Huang Z, Liu Z, Ren J, Qu X (2010) Stimuli-responsive controlled-release system using quadruplex DNA-capped silica nanocontainers. Nucleic Acids Res 39(4):1638–1644. https://doi.org/10.1093/nar/gkq893

98. Ghodke HB, Krishnan R, Vignesh K, Kumar GVP, Narayana C, Krishnan Y (2007) The I-tetraplex building block: rational design and controlled fabrication of robust 1D DNA scaffolds through non-Watson–Crick interactions. Angew Chem Int Ed 46(15):2646–2649. https://doi.org/10.1002/anie.200604461
99. Yang Y, Zhou C, Zhang T, Cheng E, Yang Z, Liu D (2012) DNA pillars constructed from an i-motif stem and duplex branches. Small 8(4):552–556. https://doi.org/10.1002/smll.201102061
100. Cheng E, Xing Y, Chen P, Yang Y, Sun Y, Zhou D, Xu L, Fan Q, Liu D (2009) A pH-triggered, fast-responding DNA hydrogel. Angew Chem Int Ed 48(41):7660–7663. https://doi.org/10.1002/anie.200902538
101. Bruyere A, Wantroba M, Flasinski S, Dzianott A, Bujarski JJ (2000) Frequent homologous recombination events between molecules of one RNA component in a multipartite RNA virus. J Virol 74(9):4214–4219. https://doi.org/10.1128/jvi.74.9.4214-4219.2000
102. Seeman NC (1982) Nucleic acid junctions and lattices. J Theor Biol 99(2):237–247. https://doi.org/10.1016/0022-5193(82)90002-9
103. Kallenbach NR, Ma R-I, Seeman NC (1983) An immobile nucleic acid junction constructed from oligonucleotides. Nature 305(5937):829–831. https://doi.org/10.1038/305829a0
104. Ma R-I, Kallenbach NR, Sheardy RD, Petrillo ML, Seeman NC (1986) Three-arm nucleic acid junctions are flexible. Nucleic Acids Res 14(24):9745–9753. https://doi.org/10.1093/nar/14.24.9745
105. Wang X, Seeman NC (2007) Assembly and characterization of 8-arm and 12-arm DNA branched junctions. J Am Chem Soc 129(26):8169–8176. https://doi.org/10.1021/ja0693441
106. Wang Y, Mueller JE, Kemper B, Seeman NC (1991) Assembly and characterization of five-arm and six-arm DNA branched junctions. Biochemistry 30(23):5667–5674. https://doi.org/10.1021/bi00237a005
107. Lilley DMJ (2000) Structures of helical junctions in nucleic acids. Q Rev Biophys 33(2):109–159. https://doi.org/10.1017/s0033583500003590
108. Ariyoshi M, Vassylyev DG, Iwasaki H, Nakamura H, Shinagawa H, Morikawa K (1994) Atomic structure of the RuvC resolvase: a Holliday junction-specific endonuclease from *E. coli*. Cell 78(6):1063–1072. https://doi.org/10.1016/0092-8674(94)90280-1
109. Gopaul DN, Guo F, Van Duyne GD (1998) Structure of the Holliday junction intermediate in Cre–loxP site-specific recombination. EMBO J 17(14):4175–4187. https://doi.org/10.1093/emboj/17.14.4175
110. Seeman NC (2010) Nanomaterials based on DNA. Annu Rev Biochem 79:65–87. https://doi.org/10.1146/annurev-biochem-060308-102244
111. Um SH, Lee JB, Park N, Kwon SY, Umbach CC, Luo D (2006) Enzyme-catalysed assembly of DNA hydrogel. Nat Mater 5(10):797–801. https://doi.org/10.1038/nmat1741
112. Wang J, Chao J, Liu H, Su S, Wang L, Huang W, Willner I, Fan C (2017) Clamped hybridization chain reactions for the self-assembly of patterned DNA hydrogels. Angew Chem Int Ed 56(8):2171–2175. https://doi.org/10.1002/anie.201610125
113. Gačanin J, Synatschke CV, Weil T (2019) Biomedical applications of DNA-based hydrogels. Adv Funct Mater. https://doi.org/10.1002/adfm.201906253
114. Wang D, Hu Y, Liu P, Luo D (2017) Bioresponsive DNA hydrogels: beyond the conventional stimuli responsiveness. Acc Chem Res 50(4):733–739. https://doi.org/10.1021/acs.accounts.6b00581
115. Shen Z, Yan H, Wang T, Seeman NC (2004) Paranemic crossover DNA: a generalized Holliday structure with applications in nanotechnology. J Am Chem Soc 126(6):1666–1674. https://doi.org/10.1021/ja038381e
116. Wang X, Chandrasekaran AR, Shen Z, Ohayon YP, Wang T, Kizer ME, Sha R, Mao C, Yan H, Zhang X, Liao S, Ding B, Chakraborty B, Jonoska N, Niu D, Gu H, Chao J, Gao X, Li Y, Ciengshin T, Seeman NC (2019) Paranemic crossover DNA: there and back again. Chem Rev 119(10):6273–6289. https://doi.org/10.1021/acs.chemrev.8b00207
117. Rothemund PWK (2006) Folding DNA to create nanoscale shapes and patterns. Nature 440(7082):297–302. https://doi.org/10.1038/nature04586
118. Ma Z, Kawai K, Hirai Y, Tsuchiya T, Tabata O (2017) Tuning porosity and radial mechanical properties of DNA origami nanotubes via crossover design. Jpn J Appl Phys 56(6S1):06GJ02. https://doi.org/10.7567/jjap.56.06gj02
119. Ke Y, Ong LL, Shih WM, Yin P (2012) Three-dimensional structures self-assembled from DNA bricks. Science 338(6111):1177. https://doi.org/10.1126/science.1227268

120. Wei B, Dai M, Yin P (2012) Complex shapes self-assembled from single-stranded DNA tiles. Nature 485(7400):623–626. https://doi.org/10.1038/nature11075
121. Wang P, Meyer TA, Pan V, Dutta PK, Ke Y (2017) The beauty and utility of DNA origami. Chem 2(3):359–382. https://doi.org/10.1016/j.chempr.2017.02.009
122. Linko V, Eerikäinen M, Kostiainen MA (2015) A modular DNA origami-based enzyme cascade nanoreactor. Chem Commun 51(25):5351–5354. https://doi.org/10.1039/C4CC08472A
123. Rajendran A, Endo M, Sugiyama H (2012) Single-molecule analysis using DNA origami. Angew Chem Int Ed 51(4):874–890. https://doi.org/10.1002/anie.201102113
124. Hu Q, Li H, Wang L, Gu H, Fan C (2019) DNA nanotechnology-enabled drug delivery systems. Chem Rev 119(10):6459–6506. https://doi.org/10.1021/acs.chemrev.7b00663
125. Ohno H, Akamine S, Saito H (2019) RNA nanostructures and scaffolds for biotechnology applications. Curr Opin Biotechnol 58:53–61. https://doi.org/10.1016/j.copbio.2018.11.006
126. Geary C, Rothemund PWK, Andersen ES (2014) A single-stranded architecture for cotranscriptional folding of RNA nanostructures. Science 345(6198):799. https://doi.org/10.1126/science.1253920
127. Chandrasekaran AR, Anderson N, Kizer M, Halvorsen K, Wang X (2016) Beyond the fold: emerging biological applications of DNA origami. ChemBioChem 17(12):1081–1089. https://doi.org/10.1002/cbic.201600038
128. Albinsson B, Hannestad JK, Börjesson K (2012) Functionalized DNA nanostructures for light harvesting and charge separation. Coord Chem Rev 256(21):2399–2413. https://doi.org/10.1016/j.ccr.2012.02.024
129. Anderson NT, Dinolfo PH, Wang X (2018) Synthesis and characterization of porphyrin–DNA constructs for the self-assembly of modular energy transfer arrays. J Mater Chem C 6(10):2452–2459. https://doi.org/10.1039/C7TC05272C
130. Anderson NT, Ren S, Chao J, Dinolfo PH, Wang X (2019) Exploiting plasmon-mediated energy transfer to enhance end-to-end efficiency in a DNA origami energy transfer array. ACS Appl Nano Mater 2(9):5563–5572. https://doi.org/10.1021/acsanm.9b01137
131. Grossi G, Jaekel A, Andersen ES, Saccà B (2017) Enzyme-functionalized DNA nanostructures as tools for organizing and controlling enzymatic reactions. MRS Bull 42(12):920–924. https://doi.org/10.1557/mrs.2017.269
132. Fu J, Li T (2017) Spatial organization of enzyme cascade on a DNA origami nanostructure. In: Ke Y, Wang P (eds) 3D DNA nanostructure: methods and protocols. Springer, New York, pp 153–164. https://doi.org/10.1007/978-1-4939-6454-3_11
133. Xin L, Zhou C, Yang Z, Liu D (2013) Regulation of an enzyme cascade reaction by a DNA machine. Small 9(18):3088–3091. https://doi.org/10.1002/smll.201300019
134. Krishnan S, Ziegler D, Arnaut V, Martin TG, Kapsner K, Henneberg K, Bausch AR, Dietz H, Simmel FC (2016) Molecular transport through large-diameter DNA nanopores. Nat Commun 7(1):12787. https://doi.org/10.1038/ncomms12787
135. Wang D, Zhang Y, Wang M, Dong Y, Zhou C, Isbell MA, Yang Z, Liu H, Liu D (2016) A switchable DNA origami nanochannel for regulating molecular transport at the nanometer scale. Nanoscale 8(7):3944–3948. https://doi.org/10.1039/C5NR08206D
136. Zhang Y, Chao J, Liu H, Wang F, Su S, Liu B, Zhang L, Shi J, Wang L, Huang W, Wang L, Fan C (2016) Transfer of two-dimensional oligonucleotide patterns onto stereocontrolled plasmonic nanostructures through DNA-origami-based nanoimprinting lithography. Angew Chem Int Ed 55(28):8036–8040. https://doi.org/10.1002/anie.201512022
137. Selnihhin D, Sparvath SM, Preus S, Birkedal V, Andersen ES (2018) Multifluorophore DNA origami beacon as a biosensing platform. ACS Nano 12(6):5699–5708. https://doi.org/10.1021/acsnano.8b01510
138. Tinnefeld P, Acuna GP, Wei Q, Ozcan A, Vietz C, Lalkens B, Trofymchuk K, Close CM, Inan H, Ochmann S, Grabenhorst L, Glembockyte V (2019) DNA origami nanotools for single-molecule biosensing and superresolution microscopy. In: Biophotonics congress: optics in the life sciences congress 2019 (BODA,BRAIN,NTM,OMA,OMP), Tucson, 2019/04/15. OSA Technical Digest. Optical Society of America, p AW5E.5. https://doi.org/10.1364/oma.2019.aw5e.5
139. Kwon PS, Ren S, Kwon S-J, Kizer ME, Kuo L, Xie M, Zhu D, Zhou F, Zhang F, Kim D, Fraser K, Kramer LD, Seeman NC, Dordick JS, Linhardt RJ, Chao J, Wang X (2020) Designer DNA architecture offers precise and multivalent spatial pattern-recognition for viral sensing and inhibition. Nat Chem 12(1):26–35. https://doi.org/10.1038/s41557-019-0369-8

140. Fang W, Jia S, Chao J, Wang L, Duan X, Liu H, Li Q, Zuo X, Wang L, Wang L, Liu N, Fan C (2019) Quantizing single-molecule surface-enhanced Raman scattering with DNA origami metamolecules. Sci Adv 5(9):eaau4506. https://doi.org/10.1126/sciadv.aau4506
141. Jasinski D, Haque F, Binzel DW, Guo P (2017) Advancement of the emerging field of RNA nanotechnology. ACS Nano 11(2):1142–1164. https://doi.org/10.1021/acsnano.6b05737
142. Leontis NB, Westhof E (2014) Self-assembled RNA nanostructures. Science 345(6198):732. https://doi.org/10.1126/science.1257989
143. Han D, Qi X, Myhrvold C, Wang B, Dai M, Jiang S, Bates M, Liu Y, An B, Zhang F, Yan H, Yin P (2017) Single-stranded DNA and RNA origami. Science 358(6369):eaao2648. https://doi.org/10.1126/science.aao2648
144. Praetorius F, Dietz H (2017) Self-assembly of genetically encoded DNA–protein hybrid nanoscale shapes. Science 355(6331):eaam5488. https://doi.org/10.1126/science.aam5488

Affiliations

Jinwei Duan[1,2,3] · Xing Wang[2,3,4] · Megan E. Kizer[5,6]

1 Department of Chemistry and Materials Science, College of Sciences, Chang'an University, Xi'an 710064, Shaanxi, People's Republic of China

2 Department of Chemistry, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

3 Holonyak Micro and Nanotechnology Laboratory, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

4 Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

5 Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

6 Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA 02139, USA