



OPEN ACCESS

EDITED BY

Ambadas Rode,
Regional Centre for Biotechnology
(RCB), India

REVIEWED BY

Neel Sarovar Bhavesh,
International Centre for Genetic
Engineering and Biotechnology, India
Kyeong Kyu Kim,
Sungkyunkwan University, South Korea

*CORRESPONDENCE

Shrikant Kukreti,
skukreti@chemistry.du.ac.in

SPECIALTY SECTION

This article was submitted to Human
and Medical Genomics,
a section of the journal
Frontiers in Genetics

RECEIVED 01 June 2022

ACCEPTED 25 July 2022

PUBLISHED 05 September 2022

CITATION

Bansal A, Kaushik S and Kukreti S (2022),
Non-canonical DNA structures:
Diversity and disease association.
Front. Genet. 13:959258.
doi: 10.3389/fgene.2022.959258

COPYRIGHT

© 2022 Bansal, Kaushik and Kukreti. This
is an open-access article distributed
under the terms of the [Creative
Commons Attribution License \(CC BY\)](#).
The use, distribution or reproduction in
other forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the
original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution
or reproduction is permitted which does
not comply with these terms.

Non-canonical DNA structures: Diversity and disease association

Aparna Bansal^{1,2}, Shikha Kaushik^{1,3} and Shrikant Kukreti^{1*}

¹Nucleic Acid Research Lab, Department of Chemistry, University of Delhi, Delhi, India, ²Department of Chemistry, Hansraj College, University of Delhi, Delhi, India, ³Department of Chemistry, Rajdhani College, University of Delhi, New Delhi, India

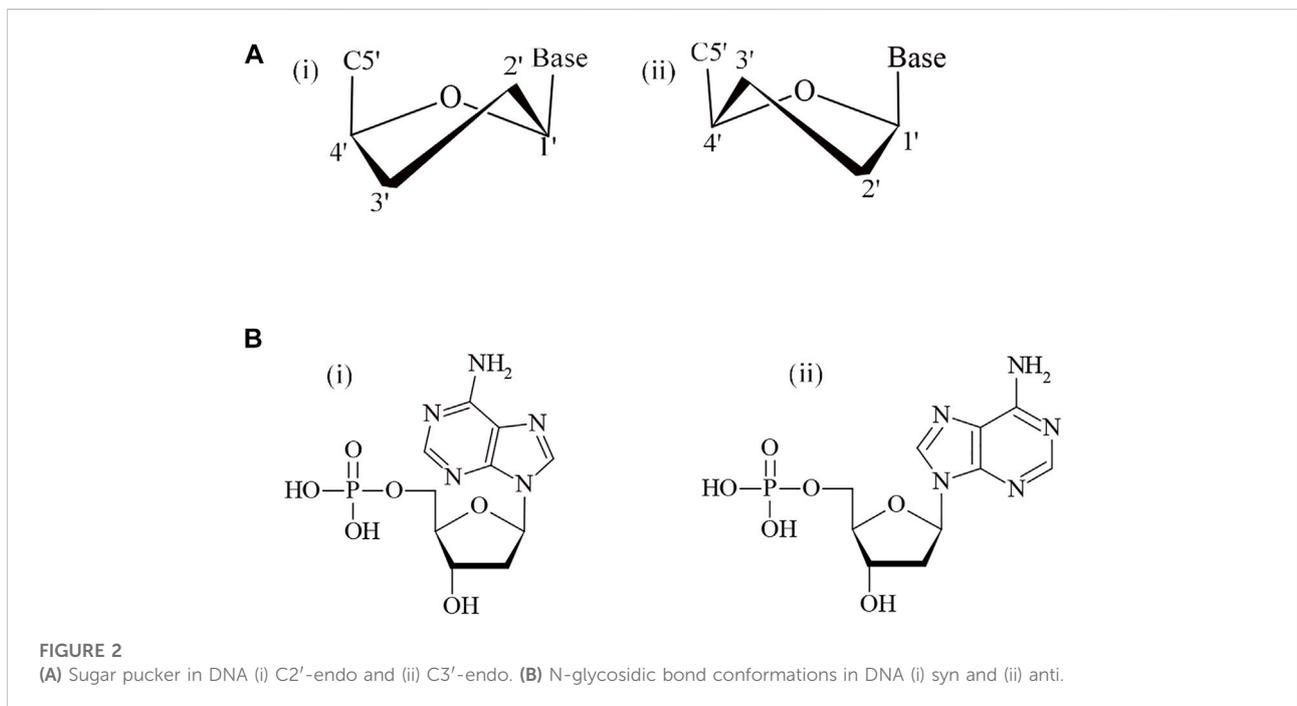
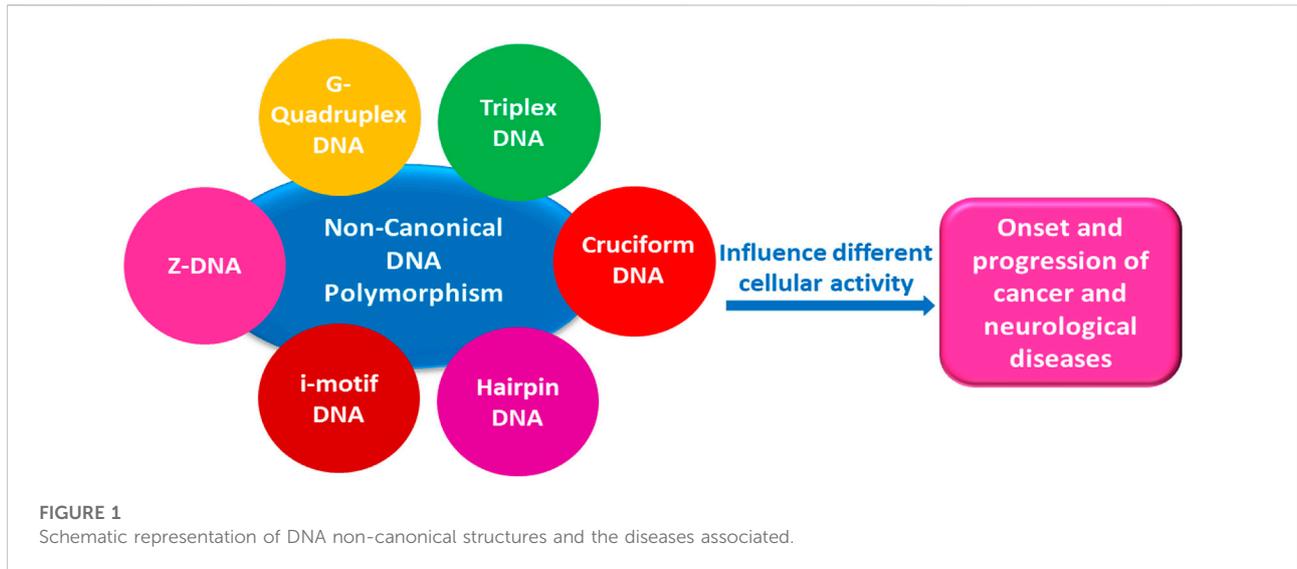
A complete understanding of DNA double-helical structure discovered by James Watson and Francis Crick in 1953, unveil the importance and significance of DNA. For the last seven decades, this has been a leading light in the course of the development of modern biology and biomedical science. Apart from the predominant B-form, experimental shreds of evidence have revealed the existence of a sequence-dependent structural diversity, unusual non-canonical structures like hairpin, cruciform, Z-DNA, multistranded structures such as DNA triplex, G-quadruplex, i-motif forms, etc. The diversity in the DNA structure depends on various factors such as base sequence, ions, superhelical stress, and ligands. In response to these various factors, the polymorphism of DNA regulates various genes *via* different processes like replication, transcription, translation, and recombination. However, altered levels of gene expression are associated with many human genetic diseases including neurological disorders and cancer. These non-B-DNA structures are expected to play a key role in determining genetic stability, DNA damage and repair *etc.* The present review is a modest attempt to summarize the available literature, illustrating the occurrence of non-canonical structures at the molecular level in response to the environment and interaction with ligands and proteins. This would provide an insight to understand the biological functions of these unusual DNA structures and their recognition as potential therapeutic targets for diverse genetic diseases.

KEYWORDS

non-canonical DNA, G-quadruplex, triplex, cruciform, Z-DNA

Introduction

The voyage of studying DNA architecture is continued since after its discovery by James Watson and Francis Crick in 1953. The first B-DNA structure proposed by Watson–Crick is proven as a milestone in the history of science and paved the way for exploring and understanding of all life processes. Even after completing almost seven decades, research in this field astonishes fast, and each new day emerges with gripping information. Structurally, DNA is a flexible molecule. Watson–Crick's rules of hybridization define secondary structures of nucleic acids. However, a variety of DNA and RNA structures do not rely on the simple A-T/U, GC base pairing, and disobey the Watson–Crick canon, and they are described as non-canonical. Nonetheless, the potential to adopt various non-canonical DNA structures depends on many factors like nucleotide sequence, hydration, ionic strength, and



ligand, etc. A-DNA and B-DNA are the most commonly studied form of DNA but other non-canonical structures like Z-DNA, hairpin/cruciform, triplex, G-quadruplex, and i-motif are also well established as given in Figure 1 (Pandya et al., 2021). The topological diversity of DNA is also attributed to various sugar and backbone conformational variables, the directionality of the glycosidic bonds and steric extensions, along with different base-pairing elasticities (Kaushik et al., 2011). In the right-handed B-DNA, the sugar-base linkage (N-glycosidic bond) is normally

anti and the sugar conformation is C2'-endo, whereas in the A-form DNA the sugar conformation is generally C3'-endo (Figure 2A). The Z-DNA, a structure favored by alternating GC-rich sequences, is different from A- and B-form DNAs as the purine (G) here adopts a syn orientation with C3'-endo pucker, whereas the pyrimidine residue (C) acquires anti-conformation and the sugar adopts the C2'-endo pucker (Figure 2B). The variation in sugar pucker and orientation of N-glycosidic bond affects the shape of duplex, widths, and depths of grooves,

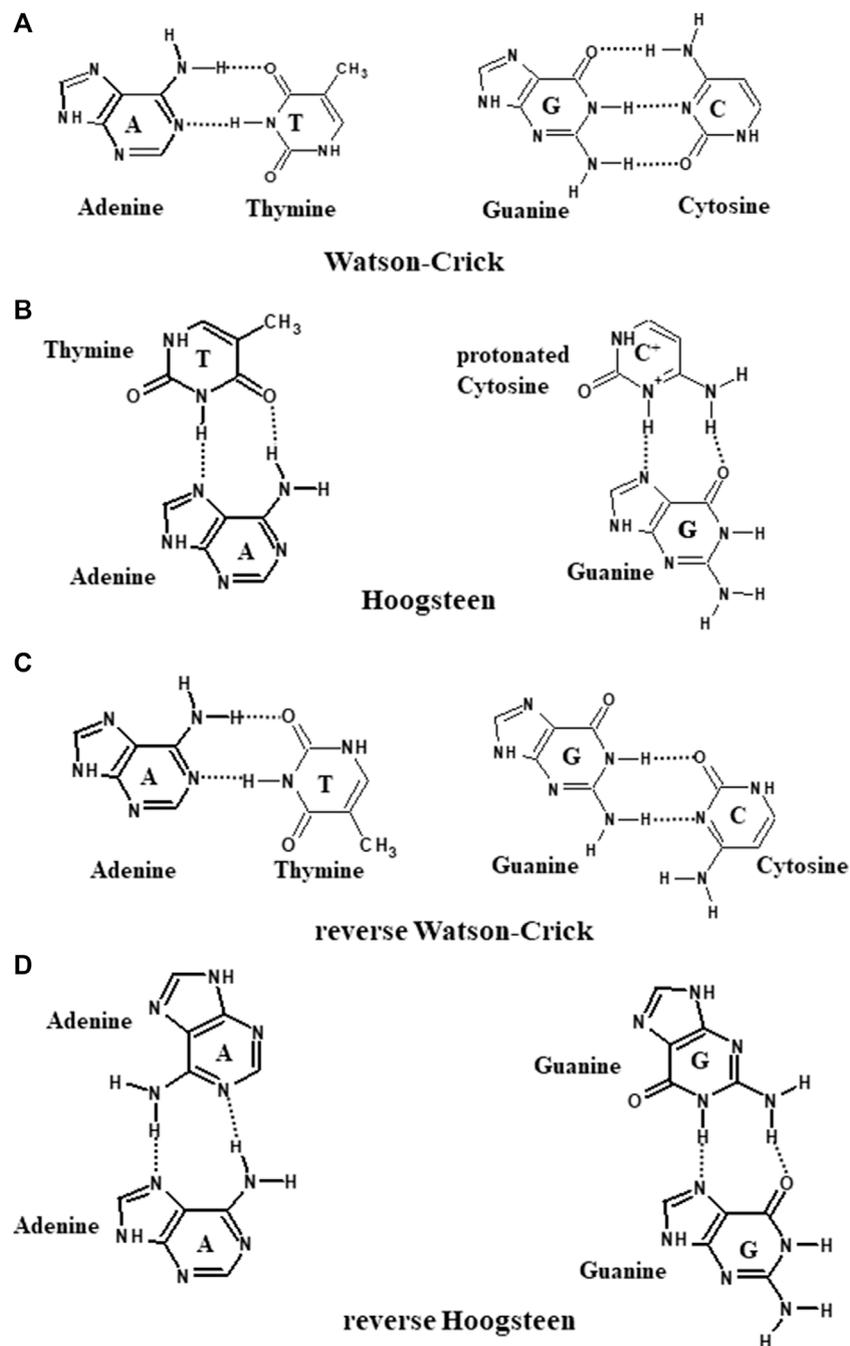


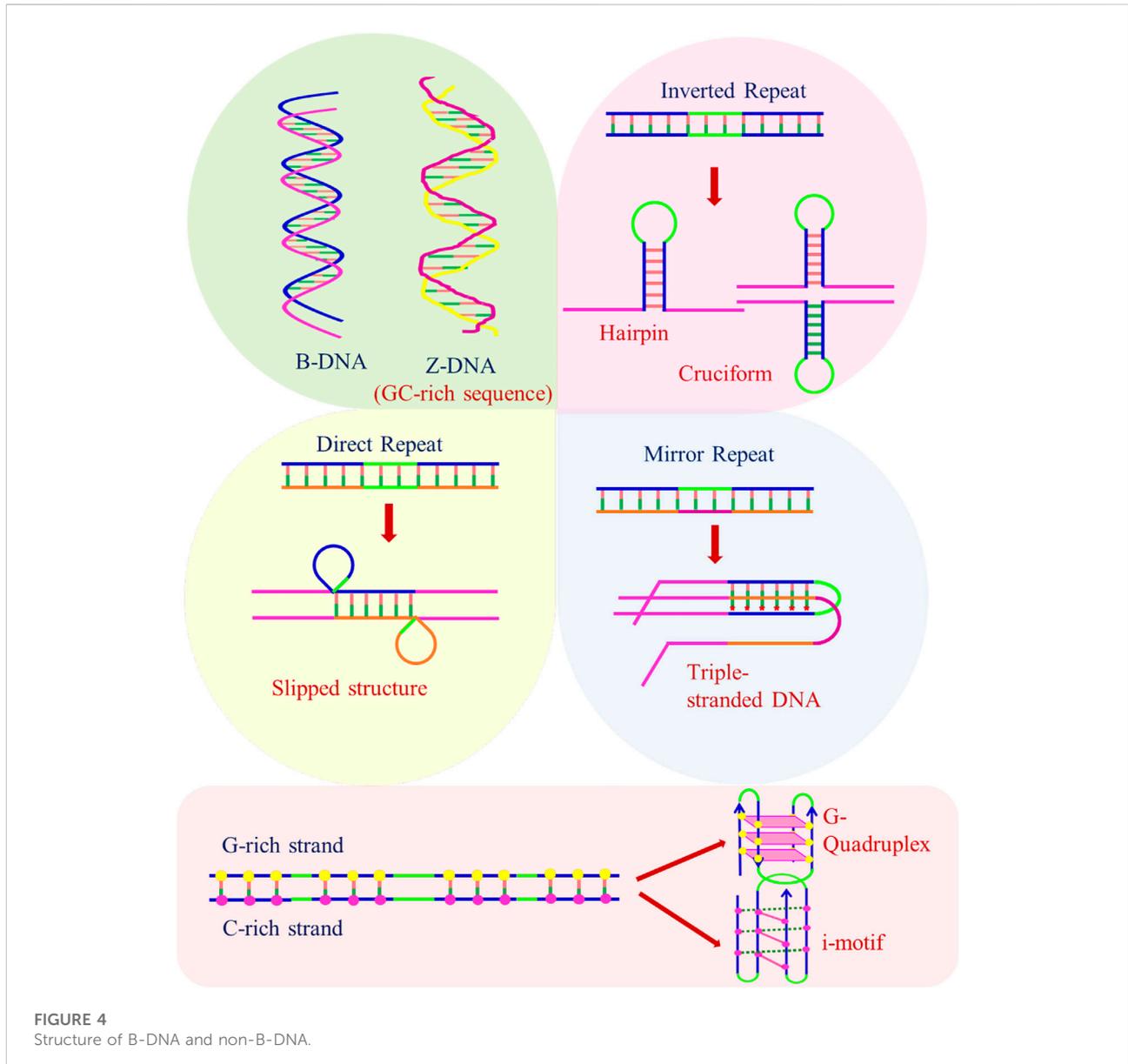
FIGURE 3

Schematic representation of (A) Watson–Crick, (B) Hoogsteen, (C) reverse Watson–Crick, and (D) reverse Hoogsteen hydrogen bonding patterns.

backbone hydration, and intrastrand phosphate–phosphate distances (Bothe et al., 2012; Egli 2019).

An assortment of possible hydrogen bonding patterns that contributes to the formation of a bouquet of DNA structures is shown in Figure 3 (Kaushik et al., 2016).

These alternative/non-canonical/non-B-DNA structures have been found to form at particular repetitive sequences such as direct repeats, mirror repeats, inverted repeats, and short tandem repeats and play very important biological roles. Z-DNA requires an alternating purine–pyrimidine/GC-rich



sequence, H-DNA requires a mirror repeat oligopurine sequence, cruciforms require inverted repeat sequences, G-quadruplexes require a contiguous stretch of guanines, while i-motif requires stretches of cytidines (Figure 4) (Georgakopoulos-Soares et al., 2018). A plethora of studies reveal that these non-B-DNA structures are associated with mutability and contribute to different types of cancer. Georgakopoulos-Soares et al. (2018) explored the structural motifs of these structures in the human genome and found that these non-B-DNA motifs are distributed non-uniformly and are extensively present at different chromosome regions. They can cause genetic instability in the presence/absence of DNA damage with/without replication and are also the hot spots of chromosomal breaks, homologous recombination, and gross chromosomal rearrangements

(Wells 2007; Rajeshwari 2012; Wang and Vasquez 2014; Georgakopoulos-Soares et al., 2018).

The human genome consists of several potential sequences, capable of forming numerous non-canonical structures such as hairpin, Z-DNA, triplex, G-quadruplexes, and i-motifs. The formation of such structures in disease-related genes inhibit or dysregulate the essential biological processes and are considered as the endogenous sources of genomic instability (Wang and Vasquez 2004; Wang et al., 2008). They are also involved in many cellular processes such as DNA combination, epigenetic regulation of chromatin, and regulation (inhibition or promotion) of gene expression by transcription and translation (Tateishi-Karimata and Sugimoto 2020, 2021). These non-canonical structures play a role in cancer progression by

perturbing the normal processes of central dogma. For example, DNA triplex and G-quadruplex structure formation may regulate the expression of cancer-related gene *via* these structures and thus inhibit transcriptional activity in the *c-myc* promoter region (Kim et al., 1998; Siddiqui-Jain et al., 2002). Several other studies have also shown that triplex-based approach can be employed to block transcription of various disease-causing genes. One interesting study on similar lines is DNA triplex-mediated inhibition of Ets2 transcription that results in growth inhibition and apoptosis in human prostate cancer cells (Carbone et al., 2004). Akhtar and Rajeshwari (2017) have reported that triplex formation selectively inhibits high-mobility group A1 proteins (HMGA1) expression and induce apoptosis in human cervical cancer. These studies have suggested that triplex-forming oligonucleotides (TFOs) that selectively target triplex-forming sequences in the promoter site of various oncogenes make attractive drug targets against anticancer therapeutics and would help the scientific community to synthesize/discover modulable ligands that show higher affinity for such triple-stranded structures and act as promising anticancer agents.

Recent studies have made known that the non-canonical structures formed by different repeat motifs of neurodegenerative disease-causing genes contribute in dysfunctioning of transcription and translation and thus causes toxic proteins accumulation. Also, intracellular phase separation promoting transcription and protein assembly are managed by these non-canonical structures that further direct the progression of neurodegenerative disease and cancer (Tateishi-Karimata and Sugimoto 2021). Interestingly, significant progress has been made to determine the therapeutic role of non-canonical structures made from chemically modified nucleobases. Reports divulge that chemically modified nucleobases like methyl, halogen, and aryl modifications at the C8 position of purines, and the C5 of pyrimidines can form DNA non-canonical structures. These modified sequences are associated with disease-inducing genes and play an important role in biological functions (Balasubramaniyam et al., 2021).

Herein, we have tried to summarize the facts illustrated in the literature about different non-canonical structures formed by DNA and their association with various human diseases at the level of genome stability, transcriptional and translational regulation *etc.* Throughout this review, we modestly abridge the current state of knowledge about the consequences of compromised genetic stability due to the presence of various DNA repeats at unlike locations in the genome.

Hairpin/cruciform secondary structures

Palindromic DNA sequences consist of inverted repeats (IRs), which are either adjacent (perfect palindrome) or separated by a spacer region (quasi-palindrome). These are

sequences with internal symmetry such that they can switch between inter-strand and intra-strand base pairing. Palindromic sequences play a vital role in regulating various biological processes and are widely distributed throughout the genome. It has been found that if the palindrome is of sufficient length, these repeats consequently form hairpins if only single-stranded DNA is involved and a cruciform structure with two hairpins forms when a double-stranded DNA is involved. Each hairpin is represented by a stem with complementary paired inverted repeats and a loop (Mitas 1997; SvetecMiklenić and Svetec 2021).

Several studies have illustrated that the hairpin structure is formed during DNA replication by the single-stranded lagging strand, while the cruciform structure is generated by the gradual extrusion of double-stranded DNA at the center of palindrome. Cruciform structures that are first proposed by Platt et al. (1955) are the best-explored species of unusual DNA. Its formation involves intra-strand hydrogen bond formation between complementary bases and disruption of intra-strand hydrogen bond of inverted repeats. It has been observed that there is a prerequisite of at least ten base pair at the center of symmetry of an inverted repeat palindromic sequence to undergo unwinding before nucleation. This process of intra-strand fusion and inter-strand fission is driven by the energy provided by negative DNA supercoiling. It has been found that relaxation of one negative supercoil occurs for every 10.5 base pairs of the inverted repeats in cruciform formation. The unwinding step in the cruciform formation depends on many factors like temperature, base composition, ionic strength, and superhelical density (Sinden 1994). Numerous studies have revealed the interaction of cruciform with various proteins, which recognize various structural features like DNA crossover, four-way junctions, and curved DNA. These structural transitions take place during DNA replication and transcription and lead to the formation of alternative DNA structures. It has been reported that there are some proteins that specifically binds to cruciform structure and contribute in various processes like DNA repair, transcription, and replication. Brazda et al. has reviewed the cruciform protein interaction remarkably. They have explained well the cruciform-protein binding and their functions and thus categorized them into 1) junction resolving enzymes, 2) transcription factors and DNA repair proteins, 3) replication machinery, and 4) chromatin-associated proteins. They also explain the onset and progression of diseases caused by dysregulation of these proteins (Brázda et al., 2011).

The various genomes explored by Miklenic et al. including the human revealed its complex architecture that comprises different sequences, which play important roles in varied biological processes. Palindromic sequences are a potent source of genetic instability as they cause chromosome breakage (double-strand break, DSB), which further lead to genetic recombination, resulting in translocation, deletion, or gene amplification (Bissler 1998; Saada et al., 2021;

TABLE 1 Diseases associated with hairpin and cruciform structures.

S.No.	Disease	Characteristics	References
1	Hereditary angioneurotic edema	Autosomal dominant disease caused by a mutation involving imperfect inverted repeats reduces the production of functional C1 inhibitor (regulatory protein in inflammation), which leads to tissue edema of the skin and mucosal surfaces	Bissler et al. (1997)
2	Triplet repeat mediated disease (CTG•CAG triplet repeat)	Myotonic dystrophy type-1 (DM1), an autosomal dominant neuromuscular disease, is caused by the expansion of CTG•CAG triplet repeat (>50 repeats) in the 3' UTR region of the DM protein kinase gene (DMPK) on chromosome 19	Bacolla et al. (2006) Izzo. et al. (2022)
3	Duchenne muscular dystrophy	X-linked recessive disease is mapped at dystrophin gene Xp21. ~one out of three thousand male new born suffer from muscular dystrophy	Robinson et al. (1997)
4	Osteogenesis imperfecta	Autosomal recessive disease characterized by brittle bones. The disease is associated with mutation (deletion, insertion, and point mutation) in the type-1 procollagen gene	Bateman et al., 1989, Lamandé 1995, Chessler et al., 1993
5	Antithrombin deficiency	Mutation in the antithrombin gene leads to the development of venous thromboembolism at a young age	Davis III et al. (1993)
6	Silent serum cholinesterase	Autosomal recessive phenotype manifests an absence of enzyme activity due to a deficiency of human serum cholinesterase	Nogueira et al. (1990)
7	Lesch-Nyhan syndrome	Neurological disease that involves mutation, which inactivates the human hypoxanthine phosphoribosyl transferase (HPRT) gene. This syndrome intellectual disability, self-mutilation, polyarthrosis, and an enhanced uric acid in serum	Gibbs et al. (1990)
8	Kearn-Sayre syndrome	This syndrome features retinal degeneration and a cardiac conduction block caused by a deletion in mitochondrion DNA	Remes et al., 1993
9	Biotinase deficiency	Autosomal recessive disorder with disability of recycling biotin. Thus, deficiency of biotin in the unavailability of functional biotinase shows clinical disorders like ataxia, seizures, and developmental delay coma	Pomponio et al. (1996)
10	Familial hypercholesterolemia	Autosomal dominant disease found in 1 out of 500 people caused by different types of mutation like deletion/insertion in the exon 8 of LDL receptor	Yamakawa-Kobayashi et al. (1994)

SvetećMiklenić and Sveteć 2021). Interestingly, the formation of hairpins and cruciform due to inter- and intra-strand base pairing of inverted repeats is associated with mutation and thus lead to several human genetic diseases, some of which are discussed in Table 1.

Moreover, genetic analysis has revealed that ~80% of all inverted repeats in the human genome are short (< 100 bp) and enrich at translocation breakpoint in cancer and stimulate the double-strand break. While, the abundance of long inverted repeats (>500 bp) is rare in the human genome, they also contribute to deletion, recombination, and gene amplification. Different studies on the mutagenic potential of short IRs have demonstrated the formation of cruciform at short IRs, which stimulate DSB by stalling RNA replication forks that cleave the structure causing deletion. The study also explained well about the relative pathway between short IRs and human cancer points and establishes the hypothesis that these non-canonicals are involved in genetic instability, etiology, disease, and evolution (Lu et al., 2015; SvetećMiklenić and Sveteć 2021).

As discussed, palindromic sequences are prone to cause chromosomal instability and lead to various diseases. Analysis revealed that breakpoints of palindrome-mediated translocations

present at the center of palindromic AT-rich repeats (PATRRs) in the human genome are responsible for frequent non-Robertsonian translocation, which results in Emanuel syndrome and non-recurrent translocation (Kato et al., 2012; Saada et al., 2021). Hence, numerous reports give an account for various diseases associated with palindrome-mediated large deletions, interchromosomal insertions, and translocation, for example, $\epsilon\gamma\Delta\beta$ thalassemia (Rooks et al., 2012), X-linked congenital hypertrichosis syndrome (Zhu et al., 2011), hereditary renal cell carcinoma (Kato et al., 2014) etc.

Z-DNA polymorphic form

Z-DNA is the double-stranded left-handed DNA conformation. The existence of this polymorphic form of DNA was unexpected and so its discovery was quite accidental in 1970. Using the X-ray diffraction pattern and CD spectra, Mitsui et al. (1970) suggested the occurrence of left-handed conformation in poly d (I-C)•poly d (I-C). Then, using the X-ray crystal structure of d (CpGpCpGpCpG), Alexander Rich and coworkers proposed

the existence of the only known left-handed helical conformation of DNA, that is, Z-DNA in 1979 (Wang et al., 1979). It can be formed with alternating purine-pyrimidine dinucleotide repeat sequences like (GC)_n. Interestingly, it has not found easy in case of (AT)_n. Under special conditions of high salt concentration or presence of at least 10 A•T base pairs embedded between GC or GT can form Z-DNA (Klysik et al., 1988; Nejedly et al., 1989). DNA supercoiling has also been found crucial in the formation and stabilization of Z-DNA. The presence of negative supercoiling minimizes the salt concentration required to drive Z-conformation. Furthermore, the level of supercoiling also affects the length of the alternating purine-pyrimidine tract required for Z-DNA formation (Sinden 1994). It has been reported that the BZ junction (B- and Z-DNA meeting point) forms at both sides of Z-DNA-forming sites is important for alleviating torsional stress and stabilizing the Z-DNA (Subramani et al., 2019).

Z-DNA has a zig-zag sugar-phosphate backbone with alternating syn-anti conformation. The bases are not toward the center of the helix-like B-DNA, they are positioned toward the outside of the helix. The small distance between the negatively charged phosphates leads to the repulsion of the phosphates and causes destabilization of the Z-DNA. Hence, the presence of cations like Na⁺, K⁺, Rb⁺, Cs⁺, Li⁺, Mg²⁺, and Mn²⁺ stabilize the structure by shielding the negative charge. Some polyamines like spermine, spermidine, and putrescine also stabilize Z-DNA. (Sinden 1994; Gueron et al., 2000; Drozdal et al., 2021).

Interestingly, research has evidenced the formation of Z-DNA and induces instability at the site of trinucleotide repeats (CAG, CGG, and GAC), which are associated with various neurodegenerative disorders like fragile X chromosome and skeletal dysplasia (Vorlickova et al., 2001; Renciuik et al., 2010; Khan et al., 2015). Also, Suram et al. (2002) showed the formation of Z-DNA signatures through CD spectra of severely affected Alzheimer's disease DNA while the B-DNA conformation in normal, young, and aged brain DNA .

The dynamic nature of Z-DNA has always intrigued scientists to unveil different facts about Z-DNA. The flipping of B-DNA into Z-DNA without strand cleavage by negative superhelical stress has made this left-handed conformation the dominion of biology (Schwartz et al., 1999). It is believed that Z-DNA regulates the level of supercoiling, and thus plays a very important role in transcription, gene expression eliciting immunogenic responses *etc.* Z-DNA-forming sequences play a vital role in controlling the cellular processes like recombination, translocation, and deletion (Wang and Vasquez 2007). Spontaneous chromosomal breakage at the genomic hot spot can cause translocation-related human disease. Wang et al. demonstrated that Z-DNA-forming sequences in human tumors are the hot spots of chromosomal breakpoints. These Z-DNA-forming sequences were found to be responsible for genetic instability as they caused gene translocation in leukemias

and lymphomas (Wang and Vasquez 2007). Furthermore, several other studies reported in the literature depicted the role of Z-DNA-forming sequences in chromosomal breakage and translocation in genes such as human BCL-2 (Adachi and Tsujimoto 1990; Seite et al., 1993), c-myc (Rimokh et al., 1991; Wölfl et al., 1995), and the SCL (Aplan et al., 1992), cancer-causing genes. Several immunoglobulin-related genes ETV6 are associated with blood-related cancers. Also, Crohn's disease, lupus erythematosus (SLE), amyotrophic lateral sclerosis (ALS), and polyradiculoneuritis are associated with the immunogenic behavior of Z-DNA (Ravichandran et al., 2019).

Z-DNA-forming DNA sequences generate large-scale deletions in mammalian cells (Wang et al., 2006). A correlation have been demonstrated between Z-DNA-forming DNA-segments and chromosome breakage hot spots, on tumor-related genes such as human c-myc, BCL-2, and the SCL gene, suggesting that Z-DNA may be implicated in chromosomal breakage and translocation. A study from Tsujimoto's group has observed that numerous alternating purine-pyrimidine fragments surround the breakpoints in the 5' flanking region of the BCL-2 gene. As per the Z-DNA-specific antibodies, the alternating purine-pyrimidine elements were constricted to the 5' hot spot region. The information indicates to a potential role for Z-DNA production in chromosome translocations in the 5' flanking region of the BCL-2 gene (Adachi and Tsujimoto 1990). Similarly, three sites in the human c-myc gene, Z-1 and Z-2 in the upstream promoter region and Z-3 at the juncture of the first intron and the second exon, have been found to generate Z-DNA in metabolically active nuclei. Z-DNA formation is only identified in all three elements by antibody binding when the gene is actively being transcribed. Negative supercoiling upstream of the gene is sufficient to stabilize the Z-DNA in Z-1, Z-2, and Z-3 site of the c-myc gene (Wölfl et al., 1995). Furthermore, Aplan et al. reported a T cell ALL (acute lymphoblastic leukemia) instance that additionally has a t (1; 3) (p34; p21) translocation, which damages the SCL locus and causes dysregulated SCL gene expression. A segment on chromosome 3 comprising alternating purine and pyrimidine repeats, forming a Z-DNA structure, is otherwise believed to be susceptible to recombination processes and is likewise disrupted by the t (1; 3) (Aplan et al., 1992).

Likewise, ADAM-12, a multidomain protein performs various functions like cell signaling, regulation of growth factors (heparin-binding EGF (HB-EGF), insulin-like growth factor binding protein-3 (IGFBP-3), and proteolysis. Ray et al. have demonstrated that there exists a highly conserved negative regulatory element (NRE) at the 5'-UTR of human ADAM-12 gene containing a stretch of dinucleotide repeat sequence, capable of forming Z-DNA. This NRE element was found to interact with Z-DNA-binding protein (hZαADAR1) and act as transcription repressor. The basic expressions of ADAM-12 are low in adults but it increases due to physiological conditions like during pregnancy in placenta. Interestingly, it is demonstrated

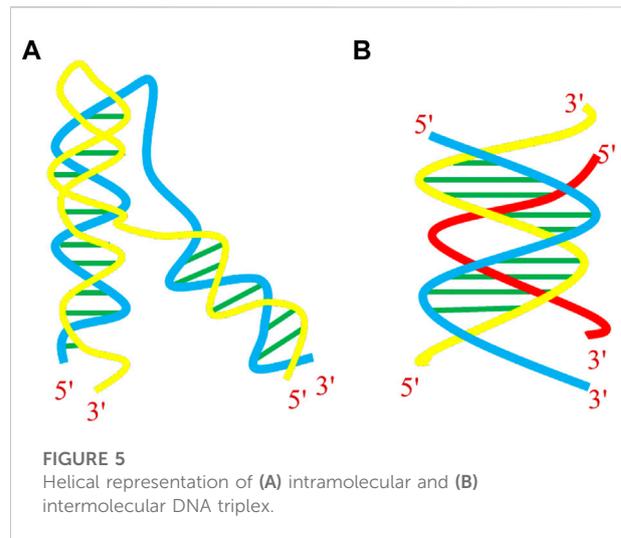
that under normal physiological conditions, NRE transcriptional repressor capable of forming Z-DNA is required for ADAM-12 gene expression (Mochizuki et al., 2007; Ray et al., 2011). Several other studies have shown that overexpression of ADAM-12 causes pathological conditions detected in many tumor tissues like breast, liver, brain, and bone (Tian et al., 2002; Lendeckel et al., 2005; Kveiborg et al., 2008)

Among all the Z-DNA-binding proteins derived from different sources, double-stranded RNA adenosine deaminase 1 (ADAR1), vaccinia virus protein (E3L), and DLM-1 (also known as Z-DNA-binding protein-1, ZBP-1) are the most studied Z-DNA-binding proteins (Wang and Vasquez 2007). It has been found that all the three proteins bind to Z-DNA and Z-RNA through the same Za domain. ADAR is an enzyme that deaminated the adenosine in ds RNA to produce inosine. This edit in the codon changes the protein sequence as the inosine is translated to guanosine and thus changes the expression of the gene they regulate (Soloman et al., 2017). E3L, an IFN resistance gene encoded by the vaccinia virus whose N-terminal domain has the sequence similarity to the Za domain of ADAR-1 and DLAM-1. It has been depicted that E3L protein is required to inhibit an IFN-primed virus-induced necroptosis (Koehler et al., 2017). DLM-1 is a tumor-associated gene highly expressed in lymphatic tissues (Fu et al., 1999). Human Mendelian genetic studies by Alan Herbert reported the occurrence of rare genetic diseases- dyschromatosis symmetrica hereditaria (DSH), Aicardi-Goutieres syndrome (AGS), and bilateral striatal necrosis (BSN) produced by a mutant allele of the Za domain (Herbert 2020).

Non-canonical multistranded structures

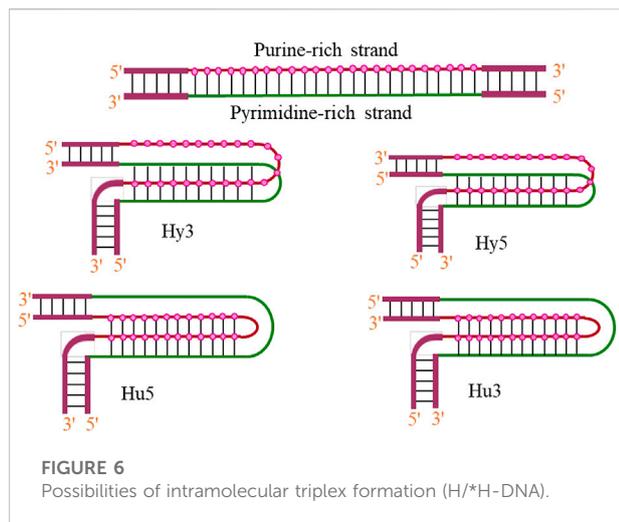
DNA triplex

Shortly after the discovery of DNA's double-helical structure, the capability of nucleic acids to form triple-helical structures (triplexes) was established (Felsenfeld et al., 1957). DNA triplex, another non-B-DNA structure is formed when a single-stranded triplex-forming oligonucleotide (TFO) binds to the major groove of polypurine•polypyrimidine (Pu•Py) segments of a target duplex in a sequence-specific manner. This spurring work was the starting point for several studies, showing that purine strand of Watson–Crick duplex could bind to the third strand containing either purines or pyrimidines. The specificity and stability of the triple-helical structures are met by Hoogsteen or reverse Hoogsteen hydrogen bonding patterns, different from the classical Watson–Crick hydrogen bonding of B-DNA. As a prerequisite, the formation of triple-helical structures requires a polypurine•polypyrimidine tract, and many databases have been developed to predict the genome-wide locations of these triplex-forming sequences. Studies have shown that the



mammalian genome contains numerous triplex-targeting sites (TTS), especially in 3'- untranslated region, 5'- of the promoter region, the gene linked with cell signaling and communication in the promoters, and within the transcribed regions of many genes (Bacolla et al., 2006; Buske et al., 2012; Jenjaroenpun et al., 2015). Literature is rich in reports that homopyrimidine and some homopurine tracts can form stable three-stranded complexes with corresponding Pu•Py sites by binding in the major groove of the target duplex (Sun and Helene 1993; Thuong and Helene 1993; Buchini and Leumann 2003). The third-strand binding to the DNA duplex can come from the same molecule, resulting in the formation of an intramolecular triplex or from a different molecule, termed an intermolecular triplex (Figure 5).

Triplex formation arises from the fact that the base pairs of the duplex already involved in Watson–Crick hydrogen bonding, still have some hydrogen bond acceptor and donor sites, available for recognition of specific bases. Thanks to the hydrogen bonding ability of some nitrogenous bases. The third strand interacts through another surface of the Watson–Crick duplex, resulting in the formation of base triplets *via* the formation of Hoogsteen or reverse Hoogsteen hydrogen bonds. Intramolecular structures are formed when unwinding of homopurine•homopyrimidine mirror repeats within the DNA duplex allows one strand to fold back and be accommodated with the target duplex (Mirkin et al., 1987). Although, homopurine•homopyrimidine mirror repeats represent an ideal target for the formation of H-DNA *in vitro* (Wells et al., 1988; Frank-Kamenetskii and Mirkin 1995) nevertheless, these structures have also been reported to form within sequences those are neither homopurine•homopyrimidine nor mirror repeats (Dayn et al., 1992). Studies on DNA triplexes have shown that non-mirror repeat sequences can also form intramolecular triplexes (Klysik 1995), and such structures are often observed even with some mismatches (Belotserkovskii et al., 1990; Macaya et al., 1991).



Intramolecular triplexes that exist *in vivo* are known as H-DNA, and depending on whether the third strand is pyrimidine- or purine-rich, these structures are known as H-DNA or *H-DNA, respectively. Such structures are preferentially formed under negative superhelical stress, acidic pH, and in presence of divalent cations (Kohwi and Kohwi-Shigematsu 1988). Four different intramolecular isomers of triplex DNA (Hy3, Hy5, Hu3, and Hu5) have been reported in the literature as shown in Figure 6 and are named on the basis of which half-element serves as the third strand.

The interaction of a third strand to the major groove of the duplex DNA in a sequence-specific manner results in the formation of an intermolecular triplex, and these structures have been shown to exhibit immense potential as transcription inhibition agents and site-directed mutagens. Depending on the relative orientation and base composition of the third strand (TFO), triple helices can be broadly categorized into three categories. In pyrimidine motif (Py triplexes), the triplex-forming oligonucleotide (TFO) that binds to the target duplex consists of cytosine and thymine (C, T) and have parallel orientation to the central purine-rich strand of duplex DNA. The third strand interact with duplex *via* Hoogsteen hydrogen bonds and form T•A•T and C•G•C⁺ base triplets (• represents Watson–Crick hydrogen bonds and * depicts Hoogsteen hydrogen bonds). These triplexes are particularly stable at acidic pH, which facilitates protonation of cytosine at N3 position (Husler and Klump 1995; Asensio et al., 1998) (Figure 7). The formation of intermolecular pyrimidine motif triplex by a homopurine and homopyrimidine DNA strand at 1:1 ratio and low pH by nuclear magnetic resonance (NMR) has been demonstrated by Bhaumik et al. (1995). Another interesting work on similar lines demonstrated the formation of a novel palindromic triple-stranded structure by d-CTTCTCCTCTTC and d-GAAGAG. They used NMR and molecular dynamics for the structure

elucidation, and also discussed the effect of hydrogen ion concentration and temperature on triplex stability (Bhaumik et al., 1998). Since intermolecular pyrimidine motif triplexes are unstable at physiological pH, various approaches are employed to stabilize these structures (Kukreti et al., 1997, 1998). However, low pH is not a condition for intramolecular pyrimidine motif triplex formation, and such structures are found to be stable even at neutral pH (Plum and Breslauer 1995; Kaushik and Kukreti 2021). Purine motif triplexes are formed when a TFO consisting of guanines and adenines (G, A ODNs) interacts with the polypurine strand of target duplex in an antiparallel manner *via* reverse Hoogsteen interactions forming T•A•A and C•G•G base triplets. These structures are formed at or near physiological pH. Intermolecular purine motif triplex formation depends on divalent cations such as magnesium (Pilch et al., 1991; Frank-Kamenetskii and Mirkin 1995; Francois et al., 2000); however, their intramolecular counterparts can form even in the absence of divalent cations (Kaushik et al., 2011). In the third category, TFO consisting of thymines and guanines (T, G) interact *via* either Hoogsteen/reverse Hoogsteen bonding, and align in parallel or antiparallel manner to the purine strand of DNA duplex respectively to form a triplex (Figure 8).

Sequence considerations play an important role in determining the relative stability of DNA triplexes, especially under physiological conditions. These structures exhibit numerous biological applications, for instance they are involved in replication pausing, regulation of gene expression, manipulating gene silencing, gene-targeted mutagenesis, cell proliferation, and inhibition of virus propagation through triple-helix formation (Helene 1991; Praseuth et al., 1999; Krasilnikova and Mirkin 2004; Mukherjee and Vasquez 2011).

DNA triplexes and associated diseases

Under normal cellular conditions, the number of cells in an organism remains constant *via* cell division and proliferation. These are highly regulated processes in normal healthy tissue; however, genetic mutations can transform a healthy cell into diseased cancer cells. According to the “central dogma of molecular biology” (Figure 9), the genetic information stored in DNA duplex is first transcribed to messenger RNA (mRNA), and later transferred to the proteins *via* the process of translation. In general, gene expression is considered to be regulated by proteins such as transcription factors; however, genetic mutations interfere with these biological processes and affect protein production. Understanding the fundamental mechanisms involved in transcription and translation is of utmost medical importance because interference at this molecular level may lead to various human ailments such as cancer, heart diseases, neurological diseases, and various types of inflammation. Various environmental factors such as exposure to UV radiations and air pollutants, poor lifestyle (high alcohol

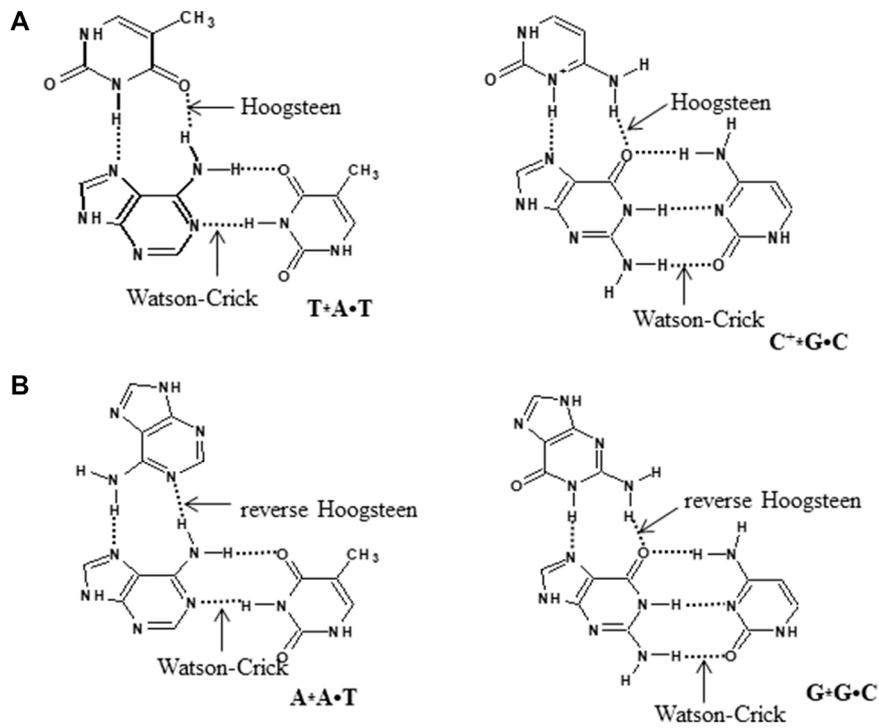
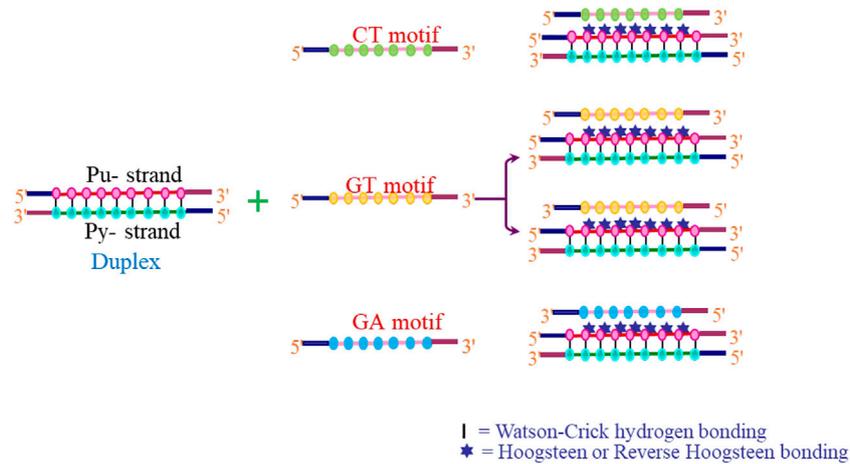
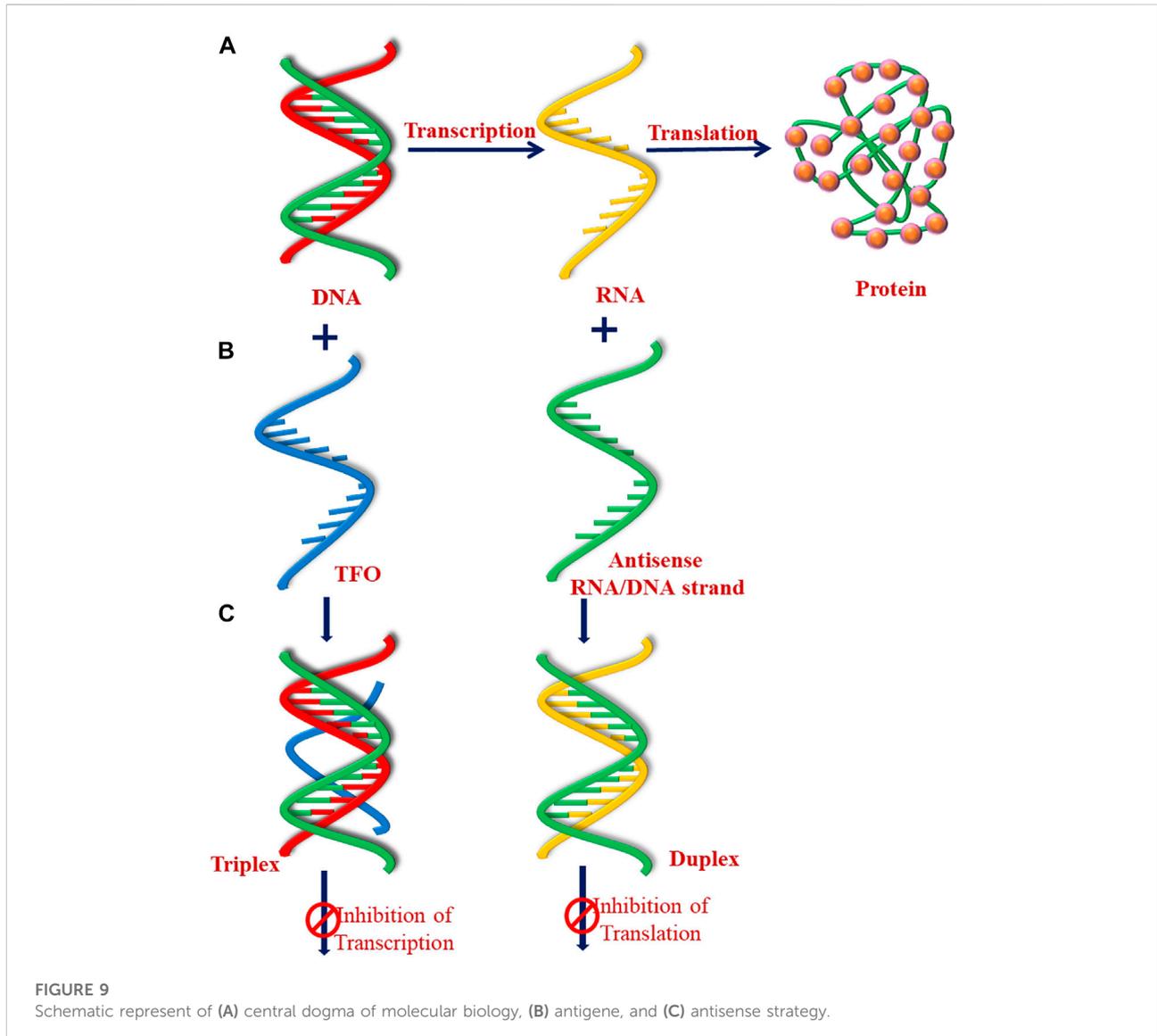


FIGURE 7
Hydrogen bonding pattern involved in DNA triplexes (A) pyrimidine motif and (B) purine motif.



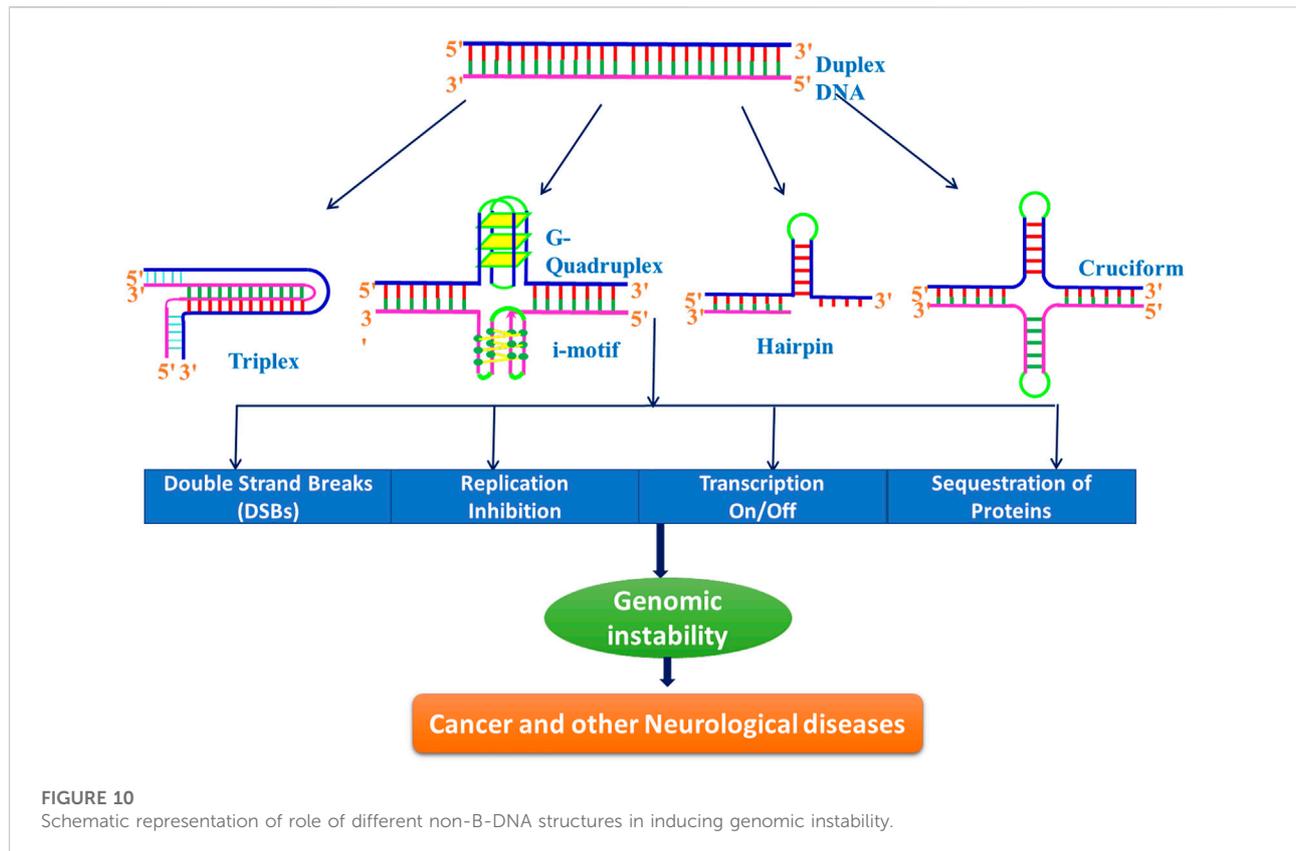
intake, smoking, and tobacco use), and certain microbial infections have also been reported to induce genetic mutations, and damage DNA (Sasco et al., 2004; Hopkinson et al., 2006).

Human genomes are rich in repetitive DNA sequences, known as short tandem repeats (STRs) or microsatellites. Although, larger repeats (tetranucleotides and pentanucleotides) also exist in human genome; the widely



studied sequences are trinucleotide repeats (TNRs) such as CGG, CAG, CTG, GAA, CCG, and CUG found, both in the coding and non-coding regions. Being polymorphic, they are more susceptible to alternations during various biological processes, and hence widely used in biological research (Fan and Chu 2007). The unstable microsatellite repeats have been identified as the cause of hereditary neurological disorders in humans in 1991. This discovery marked a turning point in the field of non-canonical DNA structures and firmly established that the formation of non-B-DNA structures in living cells induces genetic instability, leading to many human diseases (Figure 10) (Well et al., 1998; Bacolla and Wells 2004; Wang and Vasquez 2004). Trinucleotide repeat (TNR) is known to cause many rare, dominant, and neurological disorders such as Huntington's disease, Friedreich ataxia, fragile X syndrome,

myotonic dystrophy, and certain types of spinocerebellar ataxia. A total of 47 STR disease-causing genes have been reported to date. This list of neurological disorders associated with the expansion of repetitive sequences is growing rapidly, and recent addition to this is RFC1, VWA1, GIPC1, NOTCH2NLC, and LRP12 genes (Chintalaphani et al., 2021). Numerous potential nucleic acid tools such as triplex-forming oligonucleotides (TFOs), antisense oligonucleotides (ASOs), microRNAs (miRNAs), small interfering RNAs (siRNA), DNazymes, aptamers, and immunoregulatory oligonucleotides have been used to treat cancer and other neurological diseases (Meng et al., 2016; Bennett et al., 2019; Lin et al., 2020; Oh et al., 2020; Hill and Meisler, 2021; Wu et al., 2022). All of these are small oligonucleotides that can bind to specific sequences in the DNA, RNA, or proteins, and therefore, can inhibit transcription



(antigenic strategy), translation (antisense strategy), splicing, or protein activity.

The biological role of DNA triplex has been a mystery since its discovery in the late 1950s. Polypurine•polypyrimidine tracts are widespread in the human and other genomes, and these sites represent the potential to cause as well as a therapeutic target to treat many diseases. Several reports on triple-helical structures have shown a direct connection between these structures and various inherited as well as acquired human diseases, including cancer and other neurological disorders like FRDA, fragile X syndrome, muscular dystrophy, and spinocerebellar ataxia (Pollard et al., 2004; Raghavan et al., 2005; Wang and Vasquez, 2006, 2009; Liu et al., 2012). During the replication process, DNA double-strand must be unwound to a single-stranded state, and this phenomenon increases the possibility of the formation of various non-canonical DNA structures, which can affect replication. Replication errors cause mutations and recombination of genomic DNA, causing various serious diseases. Ample evidence have demonstrated the involvement of triplexes in mutagenesis, genetic instabilities, and DNA repair and recombination. The scientific findings have also indicated the role of DNA triplexes in a wide array of genetic rearrangements, such as deletions, translocations, inversion, insertion, and duplications. A new paradigm in disease etiology revealed rearrangements as

the genetic basis of approximately 50 human diseases, including follicular lymphomas, adrenoleukodystrophy, and spermatogenic failure. Studies have shown that the formation of a DNA triplex at lagging strand during replication results in fork collapse (Frick and Richardson 2001) and induce genomic instability *via* blocking the DNA replication and transcription elongation (Bétous et al., 2009; Liu et al., 2012). These structures have also been shown to involve in disease onset and progression. For example, triplex formation in the promoter region of *c-myc* oncogene hampers the process of transcription (Belotserkovskii et al., 2007), and *c-myc* translocations are often associated with leukemias and lymphomas (Haluska et al., 1988; Joos et al., 1992b; Saglio et al., 1993; Wang and Vasquez 2004). Burkitt's lymphoma in which t [8; 14] translocation occurs is also associated with triplex formation (Table 2).

Autosomal dominant polycystic kidney disease (ADPKD)

ADPKD, one of the most common genetic disorders across the world is characterized by the formation of cysts within the kidneys (Dalgaard 1957). It affects approximately 1:800 individuals, and in the United States, this is responsible for up to 10% of cases of end-stage renal disease (Gabow 1993).

TABLE 2 Human genetic diseases associated with DNA triplexes.

Disease	Type of cell	Affected gene	Function	Reference
Autosomal dominant polycystic kidney disease (ADPKD)	Renal (kidney) cells	PKD1 and PKD2	Formation of polycystins	Van Raay et al. 1996; Bagheri et al. 2019
Friedreich's ataxia	COS-7	Frataxin gene	Regulating cellular iron homeostasis	Ohshima et al. (1998)
Cancer (Burkitt's lymphoma)	BL cell lines	c-myc	Regulating the transcription	Joos et al. 1992a; Wang and Vasquez 2004
Cancer (breast and ovarian cancer)	HeLa cells (cervical cancer cells)	BRCA1- associated proteins	Regulating the transcription	Wang et al. (2000)

ADPKD is not simply a kidney disease but other organ systems (multisystem disorder) also get affected by the formation of cysts. Cardiovascular impediments such as aortic aneurysms and valvular abnormalities have also been observed in an individual (Torres et al., 2007; Bansal et al., 2019).

Genetic analysis revealed that ADPKD is caused by mutations in one of two protein-coding genes, polycystic kidney disease 1, PKD1 [16p13.3], or PKD2 [4q21-23]. These genes make novel proteins, polycystins, which are essential for the proper health of the kidneys and other parts of the body. PKD1 is associated with more than 85% of the ADPKD cases, which is the most aggressive form of the disease; whereas PKD2 is responsible for 15% of patients who are found to be mutation positive (Rossetti et al., 2012).

The human PKD1 gene contains two long stretches of polypyrimidine tracts in intron-21 (2.5 kb) and -22 (602 bp), and the PKD1 intron-21 Pu•Py tract contains 97% pyrimidine (65% cytosine and 32% thymine) on the coding strand. This is one of the largest intragenic Pu•Py tracts found in the human genome that contain 23 mirror repeat sequences. Blaszkak et al. (1999) used primer extension reactions with chloroacetaldehyde and potassium permanganate modification to show the formation of triplex DNA within this Pu•Py tract. Using the same tract, Tiner et al. (2001) have shown the formation of intramolecular triplex by atomic force microscopy. The effect of PKD1 Pu•Py tracts on eukaryotic systems has been studied by Patel et al. (2004) using primer extension assay and an SV40 system. Results revealed that this sequence form an intramolecular DNA triplex and strongly interferes with DNA replication in the eukaryotic system. They have also demonstrated that there exists a correlation between replication blockade, and the potential length of the triplex as well as the superhelical tension of intramolecular triplex formation. Bacolla et al. (2001) have also studied the stability of the same tract in a bacterial model and have shown that this sequence causes a mutation in the PKD1 gene by stimulating repair and/or recombination functions. The inhibition of replication, generation of double-strand breaks, and rounds of DNA repair/reactivation of DNA forks are responsible for the

generation of a high new germline mutation rate in the PKD1 gene (Bissler 2007). All these findings have suggested the role of Pu•Py tracts in facilitating somatic mutations that lead to the development of ADPKD.

Friedreich's ataxia

Friedreich's ataxia [FRDA] is an autosomal recessive neurodegenerative disorder that affects peripheral nerves, the spinal cord, and the cerebellum portion of the brain. It is one of the most common forms of autosomal recessive ataxia that affects about 1 in every 50,000 individuals. FRDA is characterized by muscle weakness, spasticity in the lower limbs, bladder dysfunction, scoliosis, absent lower limb reflexes, and loss of position and vibration sense (Bidichandani and Delatycki 1993). This condition is named after German doctor Nikolaus Friedreich, who first described this disease in 1863. Friedreich ataxia is the first known example of an autosomal recessive genetic disease caused by the massive expansion of triplet [GAA] repeat. Mutation of FRDA gene called FXN [or X25], which is present on chromosome 9q21 has been identified to cause FRDA. FXN contains seven exons and encodes a mitochondrial protein, frataxin which plays an important role in regulating cellular iron homeostasis by binding iron (Rajeshwari 2012). The purine strand present in the expanded repeat region of the DNA folds back to form purine motif triplex (Py•Pu*Pu), which inhibits the gene expression of frataxin.

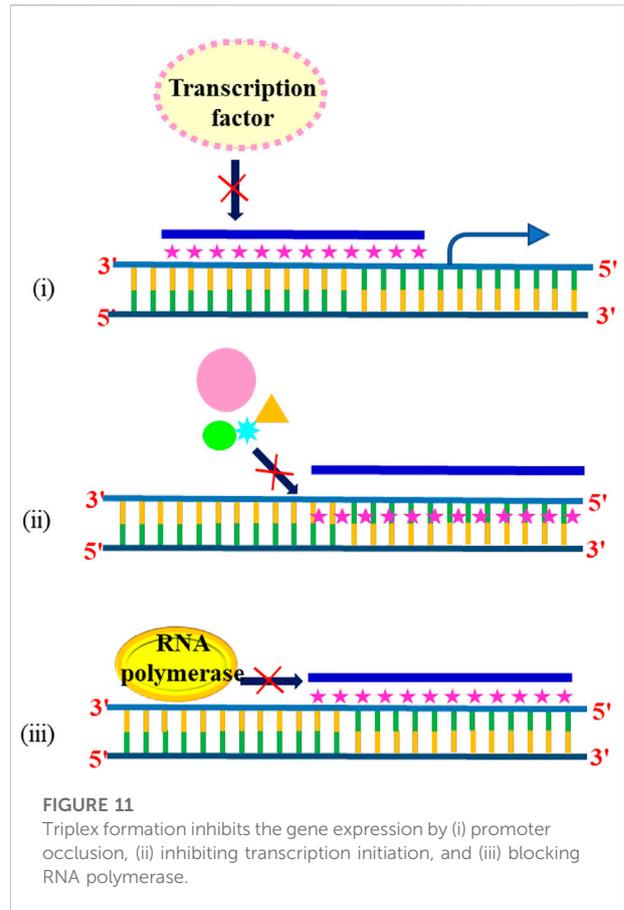
The expansion of the GAA•TTC tract in intron 1 leads to the transcriptional silencing, and the plausible mechanisms are [i] the formation of triplexes [Sticky DNA], which stalls RNA polymerase, [ii] the creation of a persistent DNA–RNA hybrid, or [iii] heterochromatin formation (Wells 2008). Delatycki et al. (1999) stated that the expansion of a GAA trinucleotide repeat sequence (TRS) in the first intron of the FXN gene is responsible for approximately 98% of FRDA cases, and the other 2% occurs due to point mutations in the FXN gene. A study on similar lines also indicated that 97% of cases of Friedreich ataxia are linked to GAA triplet repeat expansion (Lodi et al., 1999).

Follicular lymphoma

Analyses of the DNA sequences have revealed that the formation of some alternative DNA structures adversely affects DNA replication, repair, and recombination, and may lead to mutations. There are various pathways by which DNA can repair itself, and if not repaired, double-strand breaks (DSBs) can lead to cell death. Even in case, it is not repaired correctly, DSBs can potentiate chromosomal translocations, deletions, and fusion in the DNA.

The most common chromosomal translocation in human cancer is t [14; 18][q32; q21], and this translocation occurs in almost all follicular lymphomas. In most follicular lymphoma carriers, the break at chromosome 18 occurs within a small 150-bp region, termed as a major breakpoint region [Mbr], and three translocation hot spots have been identified within this 150-bp Mbr region. On the basis of gel assays, electron microscopy, and circular dichroism, Raghavan et al. (2005) have shown the formation of triplex DNA at the bcl-2 Mbr under physiological conditions, and this non-B-DNA structure within the bcl-2 Mbr leads to t[14;18] translocation.

It has already been discussed that sequences capable of forming alternative DNA structures induce genetic instability that resulted in generating mutagenic hot spots in the genome, and one such example is c-myc gene. c-myc is one of the most important transcriptional factors that plays a crucial role in regulating diverse array of cellular pathways (Nguyen et al., 2017). Chromosomal translocation t[8,14] of the c-myc gene to an immunoglobulin gene resulted in the development of Burkitt's lymphoma [BL]. In BL, translocation that involves the movement of part of chromosome 8 containing the c-myc gene onto the immunoglobulin [Ig] heavy chain gene locus on chromosome 14 [14q32], is the most common, occurs in approximately 80% of cases, while a significant minority of translocations also involves the immunoglobulin light chain genes at chromosome 2 [κ][2p12] or 22 [λ][22q11]. Previous studies have well-illustrated the connection between the formation of non-B-DNA structures within the c-myc promoter region and chromosomal translocation in the Burkett lymphoma (Wang and Vasquez 2004; Wilda et al., 2004; Del Mundo et al., 2017). The formation of antiparallel purine motif triplex in the c-myc promoter region has been shown by Umek and coworkers using electrophoretic mobility shift assays. They have also studied the formation and influence of an intramolecular [H-DNA] triplex on the strand invading capability of antigene LNA-ONs [locked nucleic acid-oligonucleotides] and suggested that the formation of such structures facilitates dsDNA strand-invasion of the antigene ONs (Umek et al., 2019). A study on similar lines has also shown that the formation of a triplex structure at the human c-myc promoter interferes with the transcription process (Belotserkovskii et al., 2007). They have also suggested the various plausible mechanisms of DNA triplex-induced transcription inhibition and potential biological implications.



DNA triple-helical structures as therapeutic tools

Gene products play a crucial role in regulating various cellular processes such as cell signaling, transcription, translation, and oncogenesis. Triplex formation targeting nucleic acids provides an interesting therapeutic approach that can be employed to control the processes involved in disease onset and progression. Two important mechanisms can prevent the progression of tumors in the normal cells, [i] activation of the tumor suppressor (TS) gene and [ii] inactivation of a proto-oncogene. Various genes such as brca1, p53, and Rb encode tumor suppressor proteins, which suppress tumor formation. Mutation in the TS gene or abnormal expression of TS proteins leads to oncogenesis. Also, an alteration (mutation) in the proto-oncogene results in an abnormality in the structure and function of a protein that can transform a normal cell into cancer cells. The relationship between non-canonical DNA structures such as triplexes and oncogenesis has been widely studied in recent years.

Triplex-specific effects have been the subject of intense research among scientists, and several reports have revealed

the effect of triplex DNA on the transcription of cancer-causing genes. The formation of DNA triplex structures inhibits the gene expression by any of the mechanisms: [i] promoter occlusion, also called transcriptional interference—where triplex formation at the promoter site does not allow transcription factors to bind (Mcshane 1992), [ii] by inhibiting transcription initiation (Duval-Valentin et al., 1992), or [iii] directly blocking progression of RNA polymerase (Scaggiante 1994) as shown in Figure 11. The first example of TFO-directed inhibition of gene expression was demonstrated by Postel and others in 1991. They have shown that triplex formation at the c-myc gene in HeLa cells reduced the c-myc expression, and the inhibition was found to be site as well as sequence-specific (Postel et al., 1991). Following this pioneering work, the effects of TFOs in transcription were studied on many genes, including Ha-ras, the HER-2/neu/c-erbB2 proto-oncogene, cyclin D1, and the alpha subunit of the interleukin-2 receptor (Ebbinghaus et al., 1993; Grigoriev et al., 1993; Kim and Miller, 1998).

The high-mobility group box 1 (hmgb1) gene encodes a protein that plays a crucial role in several cellular processes, and aberrant expression of hmgb1 has been reported to cause cancer. The selective transcriptional inhibition of the hmgb1 gene using triplex-based gene technology has been studied in a recent work (Lohani and Rajeshwari 2020). We have recently reported the formation of an intermolecular (Py-motif) triplex at 27-bp genomic homopurine-homopyrimidine tract present in the human DACH1 gene (Khan et al., 2021). This gene encodes for a chromatin-associated protein, DACH1, and expression of this gene is linked with various diseases including cancer. This work suggested that targeting the gene *via* triplex formation can help in elucidating the tissue-specific physiological function and will enhance the understanding of the disease pathophysiology from inhibition of DACH1 expression.

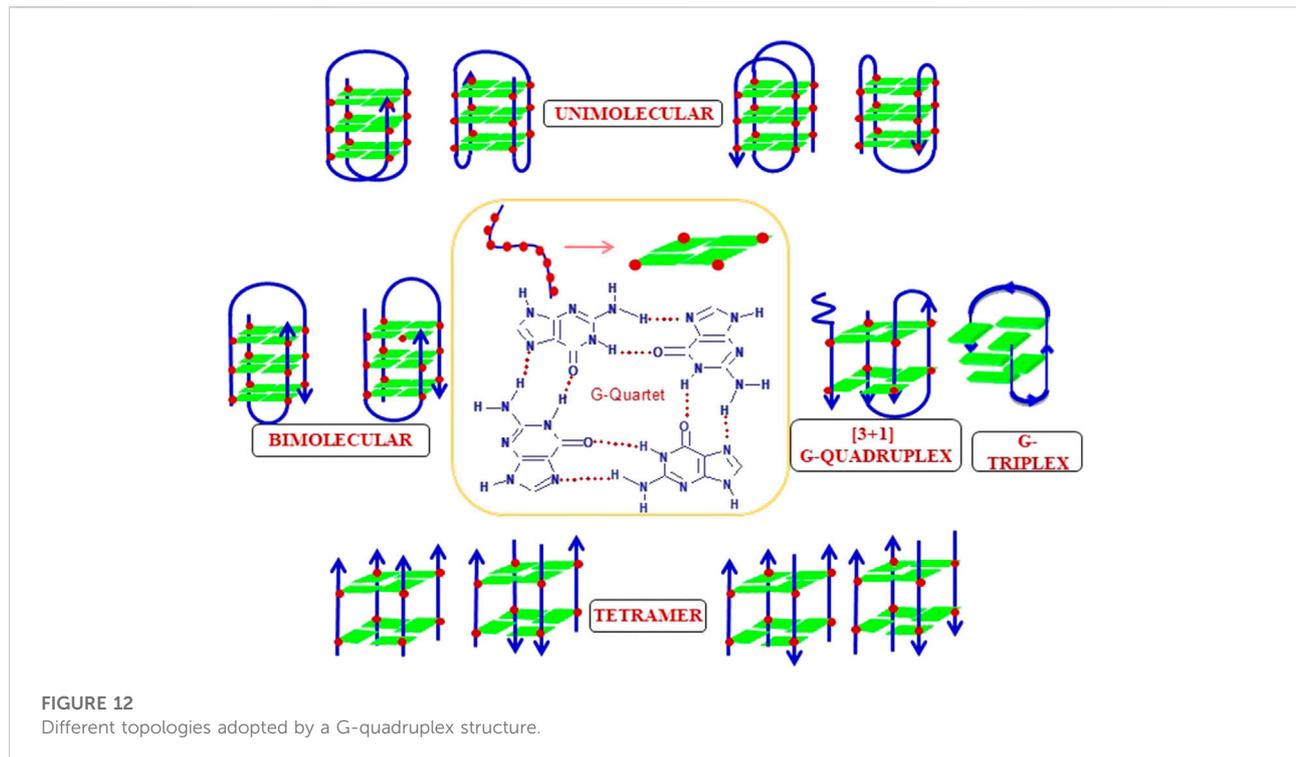
In addition to binding at the promoter site, another mechanism to hamper the prolongation of transcription is by blocking the progression of the RNA polymerase. Krasilnikov et al. (1997) have emphasized that there are fundamental differences in the way DNA polymerase responds to DNA triplex formation representing a barrier to polymerization, and it is attributed to the inability of polymerase to dissociate triplex structure. This work demonstrated that significantly higher temperatures were required to dissociate intramolecular triplexes than the corresponding duplex sequences, which leads to DNA polymerase arrest during replication. Another study on similar lines employed N3'-P5' phosphoramidate (np)-modified TFO to target a polypurine tract (PPT) in the coding region of two HIV-1 genes, revealed a 40–50% decrease in reporter RNA expression in both, transient and stable expression systems, implying the significance of specific polymerase arrest (Faria et al., 2000). DNA triplex-based therapeutic approach has also revealed remarkable results in the treatment of malignancies associated with MET

(mesenchymal epithelial transition factor) overexpression (Singhal 2011). These studies clearly described the importance of triplex-based strategies for specific polymerase arrest, thus inhibiting gene expression.

Another well-studied and potential therapeutic application of triplex technology is the use of TFOs to target site-specific mutations in the genome to correct/inactivate the target gene. Naturally occurring H-DNA-forming sequences have been shown to induce mutagenesis in cells and in mice, suggesting the mutagenic potential of triplex-forming structures *in vivo*. Studies have also revealed that TFOs can effectively induce targeted mutagenesis on chromosomal DNA in mammalian cells (Vasquez et al., 1999, 2000). Triplex-based gene therapy is also used as an effective approach for the treatment of autosomal dominant disorders, and the action of mutant/deleterious genes can be inactivated or corrected by stimulating recombination. Autosomal dominant retinitis pigmentosa (ADRP), a rare, inherited degenerative disease, is caused by deleterious protein expression that affects the photoreceptor cells of the retina, causing severe vision impairment, and progressive retinal degeneration ultimately leading to blindness. Although several genes associated with mutations in the RHO gene, encoding a G protein-coupled receptor involved in the manifestation of blindness, are the most common cause of ADRP (Dryja and Li 1995). Triplex-mediated DNA photo-crosslinking has been used for blocking the transcription of the human rhodopsin gene (Intody et al., 2000). Results demonstrated that directing DNA damage with psoralen-TFOs is an efficient and specific means of inhibiting the gene expression.

Homologous recombination (HR) plays an important role in DNA repair, replication, and several other aspects of chromosome maintenance. Among other strategies, sequence-specific DNA repair by TFOs is considered a promising tool to enhance recombinogenic frequencies. The recombinogenic potential of TFOs and TFO-directed psoralen inter-strand crosslinks [ICLS] in intermolecular, intramolecular, and the chromosomal context in mammalian cells has been studied by various researchers (Vasquez et al., 2001; Knauert 2006; Liu et al., 2008). Faruqi et al. (2000) have shown intermolecular triple-helix-induced homologous recombination in mammalian cells using a simian virus SV40-based shuttle vector, and the third-strand stimulated recombination was found to be dependent on the nucleotide excision repair [NER] pathway.

Although, triple-helical structures are formed with precise sequence specificity; their formation is a thermodynamically weaker and kinetically slower phenomenon than duplex formation. The inherent instability of triplexes has resulted in the development of various triplex-specific ligands, and TFOs have also been modified by attaching various intercalating agents to enhance triplex stability. The presence of divalent cations such as Mg²⁺ and Ca²⁺ as well



as polyamines such as spermidine, spermine, and putrescine also stabilize these structures.

The literature is rich in reports illustrating the stabilization of DNA triplexes with intercalators, and majority of these ligands were evaluated on intermolecular triplexes; however, such molecules have a profound effect on intramolecular triplexes as well. B [e]PI was the first molecule reported by Claude Helene's group, which preferentially and strongly stabilizes intermolecular triplexes rather than the underlying duplexes. It selectively stabilizes regions of T•A•T rather than C•G•C⁺ because a positive charge on protonated cytosine prevents its binding (Mergny et al., 1992). Existing studies have well demonstrated that targeting nucleic acid triplexes with small molecules has implications in the therapeutic and diagnostic field for the treatment of numerous diseases such as lupus, diabetes, hemophilia, Huntington's disease, and many other genetic disorders. Thus, triplex-related interactions using small molecules present a promising gene targeting strategy, well-reviewed by Fox (2000).

G-quadruplex structures

The structural dynamics of DNA make it capable of adopting alternate non-B-DNA structures. Among all the five nucleosides found in DNA and RNA, guanosine is the only one capable of

undergoing self-association in the solution to form G-G base pairs using Hoogsteen hydrogen bonding and thus form higher order non-canonical structures known as G-quartet/G-quadruplex as shown in Figure 12. The G-quadruplexes are very well known for their polymorphic behavior due to inter- or intramolecular folding of G-rich strands (Burge et al., 2006; Patel et al., 2007; Neidle 2009; Phan 2010; Agarwala et al., 2015; Bansal and Kukreti 2020). Also, many factors like the number of stacked G-quartet, strand orientation, loop length, and presence of counterions, which contribute to the formation and stabilization of G-quadruplexes. Multiple studies have delineated the rules to predict the stability of different polymorphic forms of G-quadruplex with different repeats, loop length, presence of ions *etc.* Using different techniques like circular dichroism, gel electrophoresis, and thermal melting, Rachwal and coworkers explained the stability of intramolecular G-quadruplex formed by the oligonucleotide sequence of the type d(GnT)₄ and d(GnT₂)₄, where $n = 3-7$ in the presence of Na⁺ and K⁺. They revealed the formation of intramolecular G-quadruplex with both the sequences. Their stability increases with increase in the sequence length ($n = 7 > n = 6 > n = 5$ and so on) in case of d(GnT₂)₄, but the trend was not regular in case of d(GnT)₄. The major conformation depicted in case of K⁺ was found to be parallel while the presence of Na⁺ favors antiparallel conformation (Rachwal et al., 2007). Furthermore, the role of loop length in the folding pattern of G-quadruplex has been

demonstrated by Hezel et al. They showed that in the presence of single residue loop, the only possibility for intramolecular G-quadruplexes is to fold into parallel conformation. When loop length is more (>2 residue) or single thymine with other loop residues, the G-quadruplexes have the potential to fold into both parallel and antiparallel topology. Also, the loop length contributes in determining the stability of G-quadruplexes. They also demonstrated the melting studies of various nucleotide sequences with different loop length and unveil that short loop length shows more stability as they are energetically more stable (Hazel et al., 2004). The structure and stability of the polymorphic forms of G-quadruplexes have always been the area to explore since after its discovery. Numerous studies have documented in the literature unveiling about structural heterogeneity of G-quadruplexes under different parameters (reviewed by Dolinnaya et al., 2016), like, sequence length, loop length, and cations also determine the stability and the folding of G-quadruplex. The interaction between the ions and G-quadruplex is based on their arrangement along the G-quartet plane or between the planes depending on the size, charge, interactions with the loop, groove *etc.*, and thus determine the stability and the structure of the G-quadruplex (reviewed by Bhattacharyya et al., 2016). Maximum studies have involved Na⁺ and K⁺ as these two are the most physiologically relevant ions, but the role of some divalent ions have also been examined. In an early study by Hardin et al. (1992), the order of ions in determining the stability of cations was suggested as K⁺ > Ca²⁺ > Na⁺ > Mg²⁺ > Li⁺ and K⁺ > Rb⁺ > Cs⁺. Furthermore, study by Vanczel and Sen 1993, the order of the ions from IA and IIA were given as Sr²⁺ > Ba²⁺ > Ca²⁺ > Mg²⁺ and K⁺ > Rb⁺ > Na⁺ > Li⁺ = Cs⁺.

Another very important factor which decides the mode of interaction of any ion to G-quartet is its ionic radius and hydration energy. For example, K⁺ and NH⁴⁺ have ionic radii 1.33 and 1.48 Å, respectively, which is very large to fit within the plane of a G-quartet, while Na⁺ is very small to move within the plane (Laughlan et al., 1994). Moreover, the hydration energy is inversely related to ionic radii. It is reported that free energy for binding Na⁺ is more favorable than that of K⁺ but the energy required to coordinate K⁺ within G-quadruplex is far less than the energy required to dehydrate Na⁺. Thus, the net difference between the two energies determines the cation selectivity for G-quadruplex. Hence, higher energetic cost of Na⁺ in comparison to K⁺, makes K⁺ more preferred for G-quadruplex binding (Gu et al., 2002; Sacca et al., 2005).

The elegant studies by Hosur's group in deciphering the structural diversity of G-quadruplexes came out with some novel structures like A-tetrad and T-tetrad (Patel and Hosur 1999; Patel, Koti and Hosur 1999). It has been demonstrated that double repeat (d-TGGTGGC) of *Saccharomyces cerevisiae* telomere forms a parallel quadruplex, comprising both G-tetrad and T-tetrad in the presence of K⁺ ions. This observation indicated that like guanine, thymine also carries

the potential to adopt diverse structures. The presence of single T can be adjusted easily without disturbing the structure, while presence of more than one T can be arranged only in the formation of loop in a G-quadruplex. Likewise, presence of single adenine (A) in the two truncated human telomere sequence, d-AG₃T and d-TAG₃T were studied in K⁺ and revealed different structural arrangements. It has been shown that both the sequences form parallel G-quadruplexes; however, an additional interesting feature of A-tetrad formation at 5'-end was also observed in d-AG₃T and not in the d-TAG₃T (Patel and Hosur 1999; Patel, Koti and Hosur 1999).

Kettani et al. (1998) came forth with a new sequence-specific structure formation demonstrated in a dodecanucleotide sequenced (G-G-G-C-T₄-G-G-G-C) found in adeno-associated human parvovirus. The studied dodecamer was found to form an antiparallel quadruplex in the presence of Na⁺, with two G-tetrad flanked by GCGC tetrad connecting T₄ lateral loop (Kettani et al., 1998). Similarly, another sequence from human telomeric variant with four CTAGGG repeats was shown to form a novel antiparallel G-quadruplex with two G-tetrads sandwiched between G•C•G•C tetrad and G•C base pair (Lim et al., 2009).

Mounting evidence of presence of contiguous stretches of G's are found not only in the telomeric region of eukaryotic chromosomes but also in other genomic locations (Sen and Gilbert 1988; Fry and Loeb 1994; SanthanaMariappan et al., 1996), promoter regions of several human proto-oncogenes including retinoblastoma gene susceptibility (Murchie and Lilley 1992) and c-myc (Simonsson et al., 1998), mutation and recombination hot spots (Simonsson 2001), and UTR's (untranslated region) (Kumari et al., 2007). Moreover, using the high resolution sequencing based method, the occurrence and the distribution of these G4 motifs have been reported in the human genome by Chambers group. Based on mounting evidence, they also reported the association of G-quadruplex with oncogenes, somatic genes related to cancer development, and thus represent a promising approach for cancer intervention (Chambers et al., 2015). In another interesting study by Schaffitzal et al., the actual existence of G-quadruplex in the nucleus was experimented by high-affinity binder taken from human combinatorial antibody library. Several single-chain antibody fragments (ScFv) were taken to detect their binding, especially with G-quadruplex formed by *Stylonychia* telomeric repeats. It has been found that one of the ScFv selected (Sty3) had affinity for parallel G-quadruplex 1,000 times more than antiparallel conformation, while another fragment Sty49 showed the same affinity for both parallel and antiparallel quadruplex (Schaffitzal et al., 2001). Their landmark studies also revealed the importance of G-quadruplex in the telomere functioning as no staining of the replication band was found in the interaction of macronuclei of *Stylonychia* telomere with Sty49, which proved G-quadruplex resolved during replication (Schaffitzal et al., 2001). An elegant review by Datta et al. has nicely described

the G-quadruplex biology in ribosomal DNA (rDNA). It has been demonstrated that G4 targeting in ribosomes and other loci play a vital role in regulation of biological processes and genomic stability and also provide a useful approach to combat with cancer (Datta et al., 2021).

Thus, G-quadruplexes has emerged as a potential therapeutic insinuation due to their potential to adopt diverse topologies, and various biological functions like gene regulation, DNA replication, transcription, translation, telomere maintenance (Siddiqui-Jain et al., 2002; Phan et al., 2005; Gunaratnam et al., 2009; Verma et al., 2009; Brooks et al., 2010; Brosh 2013), and therapeutic targets (Balasubramanian and Neidle 2009; Cosconati et al., 2010; Rhodes and Lipps 2015). G-quadruplexes found near the origin of the replication, and may regulate its initiation. Furthermore, replication of DNA G-motifs require the unfolding of DNA by helicases, an important class of proteins that play an important role to correct endogenous- or exogenous-induced DNA damage or replication error. In the absence of or in deficiency of these helicases-dependent pathways, the genomic integrity is compromised, which results in mutagenesis, cellular senescence and death, carcinogenesis, and neurological disease. Mutation in helicases acting on G-quadruplex structures (BLM, WRN, and FANCF) cause genetic instability.

Diseases associated with replication helicases involving G4-motifs

1. α -thalassemia intellectual disability syndrome is a genetic condition characterized by intellectual disability, muscle weakness, short height, *etc.* It is caused by the mutation in ATRX protein (ATP dependent helicase) encoded by ATRX gene, which is common in tumors of the central nervous system. It activates the maintenance of telomere by the telomerase-independent ALT (alternate lengthening telomere) pathway. The ATRX is found at telomeres where their target is a variable number of tandem repeats of G-rich sequences found at α -globin gene (CGCGGGGCGGGG)_n, which forms G-quadruplexes. The absence or the deficiency of ATRX causes downregulating α -globin gene expression and causes α -thalassemia (Law et al., 2010; Sarkies et al., 2010; Shay et al., 2012)
2. Werner syndrome is caused by mutation of RecQ3 helicase (WRN), which manifests in premature aging like graying and thinning hair, bilateral cataract formation, atherosclerosis, and cancer predisposition. WRN is a G4 helicase that prevents the shortening of the telomeres, which form G-quadruplex structures. In WRN deficient cells, telomeres sequence is obscured due to impaired replication of G-rich strand (Damerla et al., 2012; Goto et al., 2013).
3. Fanconi anemia is a genetic human cancer susceptibility disorder. FANCF is a structure-specific DNA helicase that

dissociates DNA G-quadruplexes. Mutation in FANCF gene causes 'Fanconi Anemia,' with the symptoms of bone marrow failure, predisposition to cancer, *etc.* (London et al., 2008; Wu et al., 2008).

4. Bloom syndrome is a rare autosomal recessive disease characterized by genomic instability, premature aging, primordial dwarfism, and predisposition to cancer, pulmonary diseases, *etc.* It is caused by BLM, the gene encodes a protein homologous to RecQ helicases and unwinds G4 DNA, and removes quadruplexes from switch regions to enable replication of switched B cells and robust immune response (Sun et al., 1998).

Diseases associated with transcription helicases with G4-motifs

The mapping of the human genome revealed that the presence of G4-motifs near the promoter site helps to regulate the transcription by binding with the two most important transcription-associated helicases XPB and XPD. It has been determined that both the helicases are linked with 1.8% of the G4-motifs found in the human genome. This association of XPB and XPD with G4-motifs characterize specific signaling and regulatory pathways. These helicases (XPB and XPD) have been found important for nucleotide excision repair, and their specific mutant allele are associated with three human genetic disorders—xeroderma pigmentosum, trichothiodystrophy, and Cockayne syndrome. Xeroderma pigmentosum is characterized by damage from UV radiation, skin cancer, and melanoma. Trichothiodystrophy shows the symptoms of distinctive developmental defects, brittle hair, and physical and but no cancer predisposition, while developmental defects and progressive neurological degeneration are the symptoms identified for Cockayne syndrome (Gray et al., 2014).

Neurodegenerative disorders due to G4 expansion repeats

The fragile X syndrome is an intellectual disability disorder caused by expansion of more than 200 repeats of CGG at 5' untranslated region of FMR1 gene. If the repeat number is between 55 and 200, it is referred to as a permutation allele that does not cause FXS phenotype but is prone to increase in repeat length during meiosis. Though there are runs of 2 G's, they have been found to form G-quadruplex. The symptoms include behavioral problems and anticipation. The sternness of the disease and the age of onset of the disease decrease from one generation to the next generation due to increase in the number of repeats. (Santoro et al., 2012; Maizels 2015).

Another example of G4 expansion is GGGGCC repeat in C9ORF72 gene at chromosome 9, whose sequence motif has been

shown to form G-quadruplex (Haeusler et al., 2014), associated with two neurodegenerative diseases, namely, amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD) (DeJesus-Hernandez et al., 2011; Renton et al., 2011; Rademakers 2012). The symptoms of ALS include stiff muscles, difficulty in speaking, breathing, swallowing, and respiratory failure, while FTD shows the symptoms of dementia, compulsive behavior but not loss of memory like Alzheimer's disease (Maizal 2015). The number of GGGGCC repeats reported pathological is 30 but can be expanded further in some cases.

The two other neurodegenerative diseases, that is, primary progressive aphasia (Simón-Sánchez et al., 2012) and depressive pseudo depressive dementia (Bieniel et al., 2014) have also been found to be associated with GGGGCC repeat expansions but the reason for identical pathology with distinct symptoms is yet to explore.

G-quadruplexes and their therapeutic effects

Thus an understanding of the diverse roles of G-quadruplexes in the biological system not only provides critical insights into fundamental molecular mechanisms that control cellular processes but also provides opportunities to identify novel therapeutic targets to treat injury and disease. Hence, an abundance of diverse G-quadruplex topologies and their biological implications have been well reported. The literature is rich in the evidence, showing the prevalence of G-quadruplex in the cellular system and its involvement in various diseases like neurological disease, cancer, and aging. This has proven these exotic structures a promising target for a therapeutic drug design (Teng et al., 2021). The structural heterogeneity of G4 DNA and their contribution in regulating biological processes replication, transcription of disease-related genes (c-myc, BCL-2, KRAS, and c-KIT), telomere maintenance etc. has been summarized (Teng et al., 2021). Similarly, Sanchez-Martin and others gave a nice detail of interaction of ligands with DNA and RNA G-quadruplexes, in cancer therapy (Sanchez-Martin et al., 2021).

The c-myc promoter is one of the most studied gene to understand the structure and biology of G-quadruplexes. The structural topology exhibited by c-myc G-quadruplex has been depicted as parallel stranded (Phan et al., 2004) c-myc, an oncogene, control the cell growth as well as apoptosis. Its overexpression leads to human malignancies like colon, breast, prostate, and cervical carcinomas. It has been found that nuclease hypersensitivity elements-III (NHEIII₁) and the G-quadruplex structure are important for c-myc transcriptional silencing (Siddiqui-Jain et al., 2002; Thakur et al., 2009; Brooks and Hurley 2010). Thus G-quadruplex stabilizing ligand reduces c-myc expression

and are called antitumorigenic (Grand et al., 2002; Ou et al., 2007).

Likewise, BCL-2 promoter, a mitochondrial membrane protein acts as inhibitor of cell apoptosis. When over expressed, it leads to different kind of human tumors like B cell and T cell lymphomas, breast, prostate, and cervical carcinomas (reviewed by Yang and Okamoto 2010). BCL-2 is a GC-rich promoter with multiple transcriptional start site located within a nuclease hypersensitive site. The G-rich strand found in this region has been shown to form intramolecular mixed parallel/antiparallel G-quadruplex with both reversal side loop and lateral loop (Dai et al., 2006). Similarly, the G-quadruplex formation in other human promoter gene like VEGF, KRAS, and C-KIT has been nicely reviewed by Yang and Okamoto (2010).

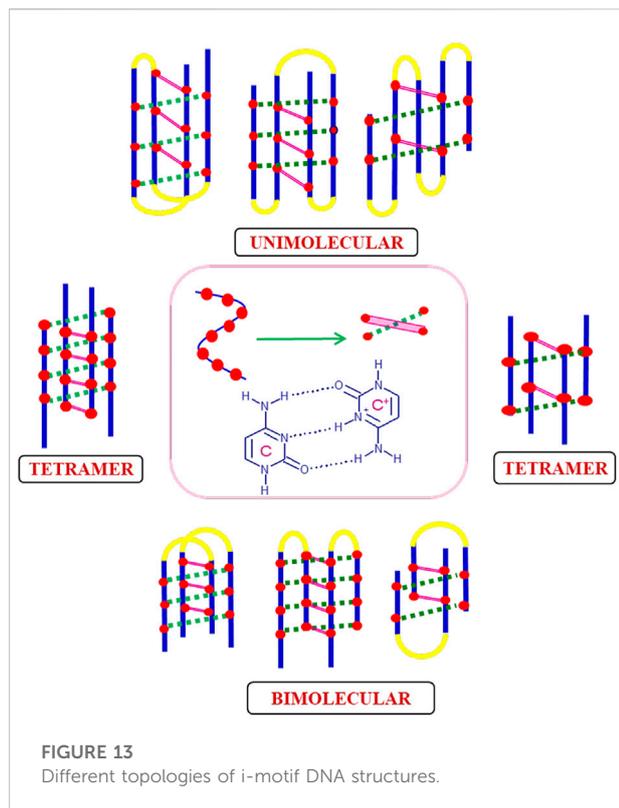
A large number of G4 ligands like porphyrin derivatives, acridine derivatives, pyridine derivatives, and natural alkaloids have been tailored to act as anticancer therapeutics by stabilizing G-quadruplexes and have been summarized elegantly by Teng et al. For instance, the potential G4 ligands reported for c-myc includes TMPy4, BRACO-19, and pyridostatin, which suppresses the growth and proliferation of breast cancer cells, retinoblastoma cells, and melanoma cells (Mikami-Terao et al., 2009; Rapozzi et al., 2014; Asamitsu et al., 2019; Konieczna et al., 2019).

Furthermore, multiple telomeric G4 targeting ligands have been reported, which inhibit telomerase activity and thus tumor growth. Telomestatin, a telomeric G4 binding ligand is a natural compound obtained from *Streptomyces anulatus*, which inhibit telomerase activity and shorten telomere length and show antiproliferative effects in acute leukemia and myeloma (Kim et al., 2003; Shamma et al., 2004).

Berberine has also been reported as telomere binding ligands. It possess anti-inflammatory activity for various chronic disease and anticancer properties against human cancer cells (Yu-Ru et al., 2020). Another application of berberine has been reported in a recent report where it has been demonstrated to target the promoter site also in DUX4 gene inhibiting muscle fibrosis. Furthermore, berberine derivatives have also been developed to improve the efficacy. Ber8, a 9-substituted berberine derivative shows strong interaction with telomeric G-quadruplex and accelerate DNA damage in multiple cancer cells (Ciszewski et al., 2020).

Thus multiple studies have flooded the literature with diverse families of G4-interacting compounds, with improved affinity and specificity. Numerous other potential of G-quadruplex-interacting drugs including quarfloxin (Ki et al. JMC 2003), 307A (Lemarteleur et al., 2004), RHPS4 (Leonetti et al., 2004), perylene such as PIPER (Han et al., 1999), quindoline derivative (Zhou et al., 2005), berberrubine (Saha et al., 2019), and quercetin (Vinnarasi et al., 2019) have been documented.

Exploring various G-quadruplex stabilizing drugs have been proven significant for the development of anticancer drugs. For



example, ligands like berberrubine (Saha et al., 2019), berberine (Suganthi et al., 2018), quercetin (Vinnarasi et al., 2019), and telomestatin (Kim et al., 2003) have been documented as DNA G-quadruplex stabilizing drugs possessing anticancer properties.

Moreover, the gene therapy of various diseases has also been studied by targeting G-quadruplexes by compounds like ranitidine chloride and benzophenanthridine derivative. The binding was studied with G-rich DNA sequences of HIV-1 promoter (Mitrasinovic et al., 2018). The *in vitro* interaction study of folic acid conjugated meso-silica nanoparticle with G-quadruplexes has provided a new paradigm for understanding anticancer activity (PoshtehShirani et al., 2018). Furthermore, another novel enantioselective approach of ligand binding with DNA G-quadruplex has taken the DNA-binding studies to an established approach (Park et al., 2018).

Aptamers

Aptamers are single-stranded (ss) DNA or RNA molecules consist of 20–80 nucleotides that fold into a three-dimensional conformation and bind to specific targets. The word aptamer is derived from the Latin word *aptus*, meaning to fit, and the Greek word *meros*, meaning part. Aptamers are also termed as chemical antibodies, as their process of molecular recognition is identical of antibody–antigen interaction and can be synthesized by an

in vitro evolution method, known as “Systematic Evolution of Ligands by Exponential Enrichment” (SELEX) (Ku et al., 2015; Yüce et al., 2018). Aptamers can specifically recognize and interact with a wide variety of targets such as small molecules, peptides, proteins (receptors), antigens, biomarkers, viruses, whole cells, and tissues *via* hydrogen bonding, electrostatic interactions, stacking interactions, and van der Waals forces (Hayashi et al., 2014; Kumar Kulabhusan et al., 2020). Their ability to bind specifically to the targets ranging from ions to the whole cells, as well as other special attributes such as ease in synthesis, low immunogenicity, excellent tissue permeability, good stability, and modifiability make them a potential candidate for diagnostic and therapeutic applications in medical, clinical, and immune hematological fields (Yang et al., 2018; Guan and Zhang 2020).

A comprehensively described state-of-the-art knowledge on the DNA-based aptamers targeting thrombin, their optimized analogs, and their biological performances in therapeutic applications is reviewed (Riccardi et al., 2021). Several known aptamers consist of G-rich sequences capable with significant biological activity are able to fold into peculiar G-quadruplex (G4) structures. Accordingly, a pentadecanucleotide, the best characterized thrombin-binding aptamer (TBA) so far, prone to form G-quadruplex, which is able to specifically recognize the protein exosite I, inhibiting the conversion of soluble fibrinogen into insoluble fibrin strands. Studies revealed that in thrombin-bound or unbound forms, TBA is able to form an antiparallel, intramolecular G4 structure, consisting of two G-tetrads connected through three edge-loops, that is, a central TGT and two lateral TT loops. The plethora of studies carried out on TBA in almost 30 years from its discovery revealed that loop residues T4, T9, T13, and G8 are critical to preserve the G4 structure, whereas T3, T7, and T12 are more flexible moieties and are particularly involved in the thrombin inhibition process (Adachi, and Nakamura 2019; Riccardi et al., 2021).

The process of making aptamer-integrated DNA nanostructure is an important challenging issue as its integrity is affected by many factors like pH, temperature, concentration of metal ion, and nuclease degradation. Several methods have been formulated to improve the stability, affinity, and specificity of these structures. For instance, ligation of 5'-phosphate group and 3'-OH group of two aptamers construct a circular bivalent aptamer, with enhanced enzymatic resistance to Exo I (Kuai et al., 2017).

Several aptamer-based DNA nanostructures have been devised to sense cells. Zuo and others developed tetrahedron-based apta-sensor to detect cancer cells in HeLa and MCF7 cells. With the help of multibranch hybridization chain reaction (HCR), gold nanoparticles *etc.*, high sensitivity, efficient signals could be achieved to capture and detect down 24 cancer cells (Zhou et al., 2014). Thus, the integration of aptamers with DNA nanostructures open a window to examine the specificity of

various biomolecule present on/within the cell surface (Huang et al., 2021). Furthermore, some aptamers bind to specific receptors on the cell membrane as a probe to visualize it. For example, Wang et al. designed an aptamer probe capable of targeting membrane proteins of living cancer cells with clear imaging (Shi et al., 2011).

Aptamers binds to the misfolded (target) proteins and impede their accumulation in the central nervous system; hence, they have also been reported as treatment option for numerous neurodegenerative diseases including Alzheimer's disease, Parkinson's disease, Huntington's disease, and multiple sclerosis (Ozturk et al., 2021).

With the aim of lowering side effects and high therapeutic effects, the targeted drug delivery system (DDS) has become one of the most important paradigms for the treatment of diseases. Hence, aptamer-based DNA nanostructure and aptamer drug conjugates have been emerged out promising candidate for biological and biomedical applications (Huang et al., 2021).

i-motif structures

i-motif is another exotic example of diverse non-canonical DNA structures. Various aspects of this structurally and biologically relevant non-canonical DNA structural entity have elegantly been reviewed (Abou Assi et al., 2018). The complementary strand of guanine-rich sequences tends to form an interesting and novel four-stranded structure, termed the i-motif DNA (intercalated-motif DNA), consisting of two parallel cytosine-rich strands forming a homoduplex that is intercalated in an antiparallel orientation. Protonation at N3 of cytosine present in the C-rich sequence is required for the formation of an i-motif structure; however, stable i-motif structures can be formed even at neutral pH. Intriguingly, a novel motif structure, called AC-motif has been reported where the conventional C⁺:C base pairs intercalate with protonated adenine cytosine (A⁺:C) in the presence of Mg²⁺ at physiological pH (Hur et al., 2021). The pH-dependent formation of i-motif questioned the existence of this structure *in vivo* at physiological pH; however, the prevalence of the C-rich regions, especially within the human telomere and promoter region of the oncogenes, near or within the regulatory region of eukaryotic genomes, recommended its biological importance (Manzini et al., 1994; Huppert and Balasubramanian 2007; Brooks et al., 2010). Though these structures are thought to have regulatory functions, their *in vivo* existence has so far remained elusive. Recently, i-motifs have been detected in the nuclei of human cells using antibody fragments that recognize their structure specifically (Zeraati 2018). They also discovered that these structures mostly form at a particular point in the cell's cycle—the late G1 phase, and hence are cell and pH-dependent. Using NMR, the stability of i-motifs inside living mammalian cells was demonstrated by Dzatko et al. (2018)

confirming the existence of these exotic structures in the cellular environment. Biological processes may lead to localized changes in pH that can stabilize pH-dependent DNA structures, such as i-motif and Py-motif triplexes.

Depending on the number of cytosine bases, solution conditions, and the loop length, i-motif structures with different molecularity like uni-, bi-, and tetramolecular have been reported as shown in Figure 13. The stability of these structures is also governed by modifying the sugar backbone of DNA; for instance, 2'-fluoro arabinonucleic acid modifications increase i-motif stability whereas substitution of cytidines with threoninol destabilizes i-motifs (Pérez-Rentero et al., 2015; Abou Assi et al., 2017).

The first intermolecular i-motif, for the sequenced (TCCCCC) was characterized by Gehring et al. (1993) under acidic conditions. Han et al. (1998) have shown that d (5mCCT₃CCT₃ACCT₃CC) forms an i-motif structure that is stable even at neutral pH. Recently, the co-existence of i-motif and B-DNA, resulting in the formation of i-motif/duplex interfaces (or junctions) has been shown by NMR methods at neutral pH (Serrano-Chacón et al., 2021). Researchers have shown that i-motif is found to be structurally dynamic adopting different structures/conformations, ranging from a bimolecular complex to tetramolecular forms over a wide pH range. The structural status of the truncated double repeat of human telomeric sequenced (CCCTAACCC), and its extended version, d (CCCTAACCC₂TAA) has been investigated by Kaushik et al. The truncated version was shown to adopt a bimolecular i-motif structure, whereas its complete double repeat (12-mer) sequence exists in two (bimolecular and tetramolecular) forms as revealed by pH-dependent UV-melting, gel-electrophoretic, and circular dichroism studies (Kaushik et al., 2010). A study on similar lines using laser tweezers techniques, the C-rich human ILPR sequence demonstrated that partially folded and i-motif forms can co-exist and can inhibit transcription catalyzed by RNA polymerase (Dhakal et al., 2010).

Several studies relating to i-motif provided numerous breakthroughs, which posit this structure as a target for therapeutic development, their ability to regulate nuclear processes like they can block DNA replication (Catasti et al., 1997; Brown and Kemdrick 2021). The credibility of i-motif as a biologically relevant structure for regulating gene expression is approved with the proximity of i-motif elements to the transcription start site. Depending on the specific promoter region, i-motif sequence, and the involved transcription factors show their capability to act as activators or repressors of transcription. There are reports of the i-motif structure formation, in the bone development proteins, sequence-specific DNA-binding domain transcription factors, and different gene encoding oncogene in the human genome (Kang et al., 2014; Kendrick et al., 2014; Delbridge 2016; Wright et al., 2017).

TABLE 3 Role of non-canonical structures in progression of cancer and neurological diseases.

Gene name	Expected structure	Location in genome	Biological role in disease	Reference
ADAM-12	Z-DNA	5'-UTR	Negative regulatory element (NRE) at the 5'-UTR of ADAM-12 act as transcription repressor and regulate ADAM-12 expression	Mochizuki et al. (2007)
SCL	Z-DNA	Alternating Pu-Py on chromosome 3	Translocation causes dysregulation of SCL gene function and cause human tumor	Aplan et al. (1992)
BCL-2	Z-DNA	Alternating Pu-Py region at 5'- hot spot region	Causes chromosome translocation in 5'-flanking region in the BCL-2 gene and dysregulation of BCL-2 gene function	Adachi and Tsujimoto (1990)
c-myc	Z-DNA	Promoter	Causes cancer	Wölf et al. (1995)
FMR1	G-quadruplex	CGG at 5'-UTR	Causes fragile X syndrome due to expansion of CGG repeats	Maizels (2015)
C9ORF72	G-quadruplex	GGGGCC repeat in chromosome 9	Causes different neurodegenerative diseases, amyotrophic lateral sclerosis (ALS), and frontotemporal dementia (FTD)	Haeusler et al. 2014, DeJesus-Hernande et al. 2011
c-myc	G-quadruplex	Promoter	Overexpression leads to human malignancies like colon, breast, prostate, and cervical carcinomas	Brooks and Hurley 2010, Siddiqui-Jain et al. 2002
BCL-2	G-quadruplex	Mitochondrial membrane protein	Over expression leads to different kind of human tumors like B cell and T cell lymphomas, breast, prostate, and cervical carcinomas	Yang and Okamoto (2010)
VEGF	G-quadruplex	Promoter	Cancer	Teng et al. (2021)
KRAS	G-quadruplex	Promoter	Cancer	Yang and Okamoto (2010)
c-KIT	G-quadruplex	Promoter	Cancer	Yang and Okamoto (2010)
c-myc	Triplex	Promoter	The intramolecular triplex (H-DNA) formed interferes with transcription	Belotserkovskii et al. (2007)
Ha-ras	Triplex	Promoter	Triplex formation at Ha-ras gene specifically inhibits Sp1 binding and transcription	Mayfield et al. (1994)
HER-2/neu	Triplex	Promoter	Formation if intermolecular purine motif prevents protein binding resulting in transcription inhibition	Ebbinghaus et al. (1993)
Cyclin D1	Triplex	Promoter	Transcription inhibition is attained <i>via</i> antiparallel triplex formation using phosphorothioate TFOs	Kim and Miller (1998)
FXN	Triplex	Intron	Segment of the sequence containing GAA triplet repeat linked with Friedreich's ataxia is shown to form a purine motif triplex	Rajeswari M.R. 2012

Research reports have highlighted various small molecules elucidating the function of i-motif as a therapeutic target and also facilitating drug discovery. For example, the porphyrin, TMPyP4, the first i-motif interactive molecule increases the stability of c-myc i-motif and prevents binding of hnRNPK (Bialis et al., 2007). However, it decreases the stability of BmPOUM2 i-motif and results in the downregulation of transcription (Niu et al., 2018). Similarly, in a recent report, the binding effect of flavonoids with BCL-2 i-motif was investigated where luteolin and quercetin was found to interact with BCL-2 i-motif with great affinity, giving the insight to understand the inhibitory condition of human cancer at the molecular level (Yang et al., 2020). Poly-C-binding proteins (PCBPs) consisting of hnRNPK (heterogeneous nuclear ribonucleoprotein K), CPI-4, and CP-KL proteins have been shown to interact with C-rich sequences and contribute a fundamental role in regulating gene expression (Yoga et al., 2012). Furthermore, carboxylated single-walled carbon nanotubes (SWNTs) were shown to selectively stabilize the telomeric i-motif and inhibit telomerase. The stability of i-motif by SWNTs results in uncapping of telomeres, DNA damage, and apoptosis or senescence. The mechanism of stabilization of i-motif by SWNTs is associated with the

proton exchange between the carboxyl group on the SWNT and i-motif cytosine, which results in hemiprotonation of the sequence and thus the folding of a stable structure (Chen et al., 2012; Wolski et al., 2019). Thus, the literature is rich in reports about different ligands that stabilize or destabilize i-motifs, depending on the nature of the interaction (Sheng et al., 2017; Pagano et al., 2018).

The ongoing research on the formation, stability, and function of i-motif has recently highlighted the role of this non-canonical structure in Parkinson's disease. On examining the α -syn protein that forms oligomers and amyloid fibrils linked to cell death of neurons. It has been determined that α -syn protein binds to both telomeric G-quadruplexes and i-motif but neuronal protein promotes the folding and stability of the only i-motif structure, not the G-quadruplex. Unfortunately, the mechanism driving Parkinson's disease is not completely defined yet; however, an attempt to solve the puzzle of the interplay between the α -syn protein and i-motif in the regulatory regions of the important genes of this disease provides an insight into the pathological defects of other neurological diseases like bipolar disorder and schizophrenia (Knop et al., 2020).

The literature is scanty in reports on the 3D structure of any i-motif/ligand complex. Attempts for such studies will pave the way for the development of drugs based on i-motif recognition. Though the present knowledge strongly suggests the *in vivo* existence of i-motifs, it is believed that these structural states are transient in the cell. Therefore, more *in vivo* studies are needed to confirm i-motif formation in various cell cycle stages or at different transcriptionally active sites within the chromosomes.

Conclusion and future outlook

The biological chemistry of DNA and RNA is facing a new aspect today. In this review, we strived to understand and summarize the current knowledge on various non-canonical structures formed by different DNA repeats and their connection with various human diseases (summarized in Table 3). Biological processes such as replication, transcription, reverse transcription, translational regulation, and genome stability are governed by non-canonical DNA structures formed within these disease-related genes. For almost last seven decades, biochemists have characterized and elucidated the various conformations of non-B-DNA structures like hairpin, cruciform, triplex, tetraplex, Z-DNA, and i-motif. These structures play a crucial role in almost all biologically imperative processes. It is now well understood that non-B-DNA-forming sequences induce the genetic instability and cause many rare, dominant, hereditary, and neurological disorders such as xeroderma pigmentosum, trichothiodystrophy, and Cockayne syndrome, Alzheimer's disease, Parkinson's disease, Huntington's disease, Friedreich ataxia (FRDA), fragile X syndrome (FRAX), myotonic dystrophy, and certain types of spinocerebellar ataxia. Some of these structures also act as a signpost of an oncogene, and a regulator for the expression of oncogenes at the transcription level. The stabilization of these structures is important to inhibit/downregulate the expression of a gene associated with human diseases like cancer. This emerging area has implications for understanding the challenges posed by alternative DNA structures and the mutagenic mechanisms through which they alter genome stability. A recent interesting study on G-quadruplex and i-motif structures reveals that their formation in the human cells is interdependent. These non-canonical structures can be discriminately stabilized or destabilized, acting as gene regulatory controllers. The finding may have important implications as the interdependency of structures points toward the interplay as an important gene regulatory switch (King et al., 2020).

There is still much to know about non-canonical DNA/RNA structures. The dynamic and transient *in vivo* nature of these diverse non-canonical/non-B-DNA structures in the cellular molecularly crowded environment has always been a challenge for targeting or recognizing these structures. Even though a minor population of conformations is not often represented in structural

studies, they may contribute to stability, folding, and functions. Conformations like Z-DNA would be of special interest as they may alter the readout of sequence information from the genome. The structure and function of Z-DNA will continue to create many surprises. Long Z-DNA containing genes, augmented for disease-causing mutations are yet to be fully explored. Structure-specific proteins/peptides and drugs/ligands have to be discovered to recognize specific non-canonical DNA/RNA structures to elucidate their roles in different biological processes. Since these structures have appeared as potential therapeutic targets for various genetic diseases, searching for the proteins or ligands, which could discriminate the DNA/RNA sequence and their actual structures would help the therapeutic strategy and drug discovery.

The structure and function of non-canonical DNA will continue to spawn many surprises. An insightful future goal is to screen such small molecules, which act as promising drug candidates for the treatment of the ailment by downregulating the expression of the disease-causing gene. Hope the aforementioned discussions foster a new interest in the paradigm of DNA non-canonical structural polymorphism, stabilization, and their implication in understanding the role in different biological processes.

Author contributions

AB contributed in writing and finalizing the manuscript. SK (second author) contributed in writing, and preparation of figures. SK contributed in concept design and overall supervision and finalizing the manuscript.

Funding

This research was funded by IOE, University of Delhi.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Abou Assi, H., El-Khoury, R., González, C., and Damha, M. J. (2017). 2'-Fluoroarabinonucleic acid modification traps G-quadruplex and i-motif structures in human telomeric DNA. *Nucleic Acids Res.* 45 (20), 12055–11546. doi:10.1093/nar/gkx962
- Abou Assi, H., Garavís, M., González, C., and Damha, M. J. (2018). i-Motif DNA: structural features and significance to cell biology. *Nucleic Acids Res.* 46 (16), 8038–8056. doi:10.1093/nar/gky735
- Adachi, M., and Tsujimoto, Y. (1990). Potential Z-DNA elements surround the breakpoints of chromosome translocation within the 5' flanking region of bcl-2 gene. *Oncogene* 5 (11), 1653–1657.
- Adachi, T., and Nakamura, Y. (2019). Aptamers: A review of their chemical properties and modifications for therapeutic application. *Molecules* 24 (23), 4229. doi:10.3390/molecules24234229
- Agarwala, P., Pandey, S., and Maiti, S. (2015). The tale of RNA G-quadruplex. *Org. Biomol. Chem.* 13 (20), 5570–5585. doi:10.1039/c4ob02681k
- Akhter, M. Z., and Rajeswari, M. R. (2017). Triplex forming oligonucleotides targeted to hmgal selectively inhibit its expression and induce apoptosis in human cervical cancer. *J. Biomol. Struct. Dyn.* 35 (4), 689–703. doi:10.1080/07391102.2016.1160257
- Aplan, P. D., Raimondi, S. C., and Kirsch, I. R. (1992). Disruption of the SCL gene by a (1; 3) translocation in a patient with T cell acute lymphoblastic leukemia. *J. Exp. Med.* 176 (5), 1303–1310. doi:10.1084/jem.176.5.1303
- Asamitsu, S., Obata, S., Yu, Z., Bando, T., and Sugiyama, H. (2019). Recent progress of targeted G-quadruplex-preferred ligands toward cancer therapy. *Molecules* 24 (3), 429. doi:10.3390/molecules24030429
- Asensio, J. L., Lane, A. N., Dhese, J., Bergqvist, S., and Brown, T. (1998). The contribution of cytosine protonation to the stability of parallel DNA triple helices. *J. Mol. Biol.* 275 (5), 811–822. doi:10.1006/jmbi.1997.1520
- Bacolla, A., Jaworski, A., Connors, T. D., and Wells, R. D. (2001). Pkd1 unusual DNA conformations are recognized by nucleotide excision repair. *J. Biol. Chem.* 276 (21), 18597–18604. doi:10.1074/jbc.M100845200
- Bacolla, A., and Wells, R. D. (2004). Non-B DNA conformations, genomic rearrangements, and human disease. *J. Biol. Chem.* 279 (46), 47411–47414. doi:10.1074/jbc.R400028200
- Bacolla, A., Wojciechowska, M., Kosmider, B., Larson, J. E., and Wells, R. D. (2006). The involvement of non-B DNA structures in gross chromosomal rearrangements. *DNA repair* 5 (9–10), 1161–1170. doi:10.1016/j.dnarep.2006.05.032
- Bagheri, M., Makhdoomi, K., Afshari, A. T., Nikibakhsh, A. A., and Rad, I. A. (2019). Examining the role of polymorphisms in exon 25 of the PKD1 gene in the pathogenesis of autosomal dominant polycystic kidney disease in Iranian patients. *Rep. Biochem. Mol. Biol.* 8 (2), 102–110.
- Balasubramanian, S., and Neidle, S. (2009). G-quadruplex nucleic acids as therapeutic targets. *Curr. Opin. Chem. Biol.* 13 (3), 345–353. doi:10.1016/j.cbpa.2009.04.637
- Balasubramaniyam, T., Oh, K. I., Jin, H. S., Ahn, H. B., Kim, B. S., and Lee, J. H. (2021). Non-canonical helical structure of nucleic acids containing base-modified nucleotides. *Int. J. Mol. Sci.* 22 (17), 9552. doi:10.3390/ijms22179552
- Bansal, A., Kaushik, S., Ahmed, S., and Kukreti, S. (2019). Autosomal dominant polycystic kidney disease: A review. *J. Biomed. Ther. Sci.* 6 (1), 15–23.
- Bansal, A., and Kukreti, S. (2020). The four repeat Giardia lamblia telomere forms tetramolecular G-quadruplex with antiparallel topology. *J. Biomol. Struct. Dyn.* 38 (7), 1975–1983. doi:10.1080/07391102.2019.1623074
- Bateman, J. F., Lamande, S. R., Dahl, H. H., Chan, D., Mascara, T., and Cole, W. G. (1989). A frameshift mutation results in a truncated nonfunctional carboxyl-terminal pro α 1 (I) propeptide of type I collagen in osteogenesis imperfecta. *J. Biol. Chem.* 264 (19), 10960–10964. doi:10.1016/s0021-9258(18)60412-0
- Bauer, L., Tlučková, K., Tóthová, P., and Víglašký, V. (2011). G-quadruplex motifs arranged in tandem occurring in telomeric repeats and the insulin-linked polymorphic region. *Biochemistry* 50 (35), 7484–7492. doi:10.1021/bi1003235
- Belotserkovskii, B. P., De Silva, E., Tornaletti, S., Wang, G., Vasquez, K. M., and Hanawalt, P. C. (2007). A triplex-forming sequence from the human c-MYC promoter interferes with DNA transcription. *J. Biol. Chem.* 282 (44), 32433–32441. doi:10.1074/jbc.M704618200
- Belotserkovskii, B. P., Veselkov, A. G., Filippov, S. A., Dobrynin, V. N., Mirkin, S. M., and Frank-Kamenetskii, M. D. (1990). Formation of intramolecular triplex in homopurine-homopyrimidine mirror repeats with point substitutions. *Nucleic Acids Res.* 18 (22), 6621–6624. doi:10.1093/nar/18.22.6621
- Bennett, C. F., Krainer, A. R., and Cleveland, D. W. (2019). Antisense oligonucleotide therapies for neurodegenerative diseases. *Annu. Rev. Neurosci.* 42, 385–406. doi:10.1146/annurev-neuro-070918-050501
- Bétou, R., Rey, L., Wang, G., Pillaire, M. J., Puget, N., Selves, J., et al. (2009). Role of TLS DNA polymerases η and κ in processing naturally occurring structured DNA in human cells. *Mol. Carcinog. Publ. Coop. Univ. Tex. MD Anderson Cancer Cent.* 48 (4), 369–378. doi:10.1002/mc.20509
- Bhattacharyya, D., Mirihana Arachchilage, G., and Basu, S. (2016). Metal cations in G-quadruplex folding and stability. *Front. Chem.* 4, 38. doi:10.3389/fchem.2016.00038
- Bhaumik, S. R., Chary, K. V., Govil, G., Liu, K., and Miles, H. T. (1995). NMR characterization of a triple stranded complex formed by homo-purine and homopyrimidine DNA strands at 1: 1 molar ratio and acidic pH. *Nucleic Acids Res.* 23 (20), 4116–4121. doi:10.1093/nar/23.20.4116
- Bhaumik, S. R., Chary, K. V., Govil, G., Liu, K., and Todd Miles, H. (1998). A novel palindromic triple-stranded structure formed by homopyrimidine dodecamer d-CTTCTCCTCTTC and homopurine hexamer d-GAAGAG. *Nucleic Acids Res.* 26 (12), 2981–2988. doi:10.1093/nar/26.12.2981
- Bialis, T., Dexheimer, T., Gleason-Guzman, M., Yang, D., and Hurley, L. (2007). Transcriptional consequences of targeting the i-motif structure of the c-Myc promoter with TMPyP4. *Cancer Res.* 67, 3169.
- Bidichandani, S. I., and Delatycki, M. B. (1993). "Friedreich ataxia," in *GeneReviews*[®]. Editors R. A. Pagon, M. P. Adam, H. H. Ardinger, T. D. Bird, C. R. Dolan, C.-T. Fong, et al. (Seattle: University of Washington).
- Bieniek, K. F., Van Blitterswijk, M., Baker, M. C., Petrucelli, L., Rademakers, R., and Dickson, D. W. (2014). Expanded C9ORF72 hexanucleotide repeat in depressive pseudodementia. *JAMA Neurol.* 71 (6), 775–781. doi:10.1001/jamaneurol.2013.6368
- Bissler, J. J., Aulak, K. S., Donaldson, V. H., Rosen, F. S., Cicardi, M., Harrison, R. A., et al. (1997). Molecular defects in hereditary angioneurotic edema. *Proc. Assoc. Am. Physicians* 109 (2), 164–173.
- Bissler, J. J. (1998). DNA inverted repeats and human disease. *Front. Biosci.* 3 (4), d408–d418. doi:10.2741/a284
- Bissler, J. J. (2007). Triplex DNA and human disease. *Front. Biosci.* 12, 4536–4546. doi:10.2741/2408
- Blaszak, R. T., Potaman, V., Sinden, R. R., and Bissler, J. J. (1999). DNA structural transitions within the PKD1 gene. *Nucleic Acids Res.* 27 (13), 2610–2617. doi:10.1093/nar/27.13.2610
- Bothe, J. R., Lowenhaupt, K., and Al-Hashimi, H. M. (2012). Incorporation of CC steps into Z-DNA: Interplay between B-Z junction and Z-DNA helical formation. *Biochemistry* 51 (34), 6871–6879. doi:10.1021/bi300785b
- Brázda, V., Laister, R. C., Jagelská, E. B., and Arrowsmith, C. (2011). Cruciform structures are a common DNA feature important for regulating biological processes. *BMC Mol. Biol.* 12 (1), 33–16. doi:10.1186/1471-2199-12-33
- Brooks, T. A., and Hurley, L. H. (2010). Targeting MYC expression through G-quadruplexes. *Genes Cancer* 1 (6), 641–649. doi:10.1177/1947601910377493
- Brooks, T. A., Kendrick, S., and Hurley, L. (2010). Making sense of G-quadruplex and i-motif functions in oncogene promoters. *FEBS J.* 277 (17), 3459–3469. doi:10.1111/j.1742-4658.2010.07759.x
- Brosh, R. M. (2013). DNA helicases involved in DNA repair and their roles in cancer. *Nat. Rev. Cancer* 13 (8), 542–558. doi:10.1038/nrc3560
- Brown, S. L., and Kendrick, S. (2021). The i-motif as a molecular target: More than a complementary DNA secondary structure. *Pharmaceuticals* 14 (2), 96. doi:10.3390/ph14020096
- Buchini, S., and Leumann, C. J. (2003). Recent improvements in antigenic technology. *Curr. Opin. Chem. Biol.* 7 (6), 717–726. doi:10.1016/j.cbpa.2003.10.007
- Burge, S., Parkinson, G. N., Hazel, P., Todd, A. K., and Neidle, S. (2006). Quadruplex DNA: Sequence, topology and structure. *Nucleic Acids Res.* 34 (19), 5402–5415. doi:10.1093/nar/gkl655
- Buske, F. A., Bauer, D. C., Mattick, J. S., and Bailey, T. L. (2012). Triplexator: Detecting nucleic acid triple helices in genomic and transcriptomic data. *Genome Res.* 22 (7), 1372–1381. doi:10.1101/gr.130237.111
- Carbone, G. M., Napoli, S., Valentini, A., Cavalli, F., Watson, D. K., and Catapano, C. V. (2004). Triplex DNA-mediated downregulation of Ets2 expression results in growth inhibition and apoptosis in human prostate cancer cells. *Nucleic Acids Res.* 32 (14), 4358–4367. doi:10.1093/nar/gkh744
- Catasti, P., Chen, X., Deaven, L. L., Moyzis, R. K., Bradbury, E. M., and Gupta, G. (1997). Cytosine-rich strands of the insulin minisatellite adopt hairpins with intercalated Cytosine+ Cytosine pairs. *J. Mol. Biol.* 272 (3), 369–382. doi:10.1006/jmbi.1997.1248
- Chambers, V. S., Marsico, G., Boutell, J. M., Di Antonio, M., Smith, G. P., and Balasubramanian, S. (2015). High-throughput sequencing of DNA G-quadruplex structures in the human genome. *Nat. Biotechnol.* 33 (8), 877–881. doi:10.1038/nbt.3295

- Chen, Y., Qu, K., Zhao, C., Wu, L., Ren, J., Wang, J., Qu, X., et al. (2012). Insights into the biomedical effects of carboxylated single-wall carbon nanotubes on telomerase and telomeres. *Nat. Commun.* 3 (1), 1074–1087. doi:10.1038/ncomms2091
- Chessler, S. D., Wallis, G. A., and Byers, P. H. (1993). Mutations in the carboxyl-terminal propeptide of the pro alpha 1 (I) chain of type I collagen result in defective chain association and produce lethal osteogenesis imperfecta. *J. Biol. Chem.* 268 (24), 18218–18225. doi:10.1016/s0021-9258(17)46833-5
- Chintalaphani, S. R., Pineda, S. S., Deveson, I. W., and Kumar, K. R. (2021). An update on the neurological short tandem repeat expansion disorders and the emergence of long-read sequencing diagnostics. *Acta Neuropathol. Commun.* 9 (1), 98–20. doi:10.1186/s40478-021-01201-x
- Ciszewski, L., Lu-Nguyen, N., Slater, A., Brennan, A., Williams, H. E. L., Dickson, G., et al. (2020). G-quadruplex ligands mediate downregulation of DUX4 expression. *Nucleic Acids Res.* 48 (8), 4179–4194. doi:10.1093/nar/gkaa146
- Cosconati, S., Marinelli, L., Trotta, R., Virno, A., De Tito, S., Romagnoli, R., et al. (2010). Structural and conformational requisites in DNA quadruplex groove binding: Another piece to the puzzle. *J. Am. Chem. Soc.* 132 (18), 6425–6433. doi:10.1021/ja1003872
- Dai, J., Dexheimer, T. S., Chen, D., Carver, M., Ambrus, A., Jones, R. A., et al. (2006). An intramolecular G-quadruplex structure with mixed parallel/antiparallel G-strands formed in the human BCL-2 promoter region in solution. *J. Am. Chem. Soc.* 128 (4), 1096–1098. doi:10.1021/ja055636a
- Dalgaard, O. Z. (1957). Bilateral polycystic disease of the kidneys: A follow-up of two hundred and eightyfourpatients and their families. *Acta Med. Scand. Suppl.* 328, 1–255.
- Damerla, R. R., Knickelbein, K. E., Strutt, S., Liu, F. J., Wang, H., and Opresko, P. L. (2012). Werner syndrome protein suppresses the formation of large deletions during the replication of human telomeric sequences. *Cell Cycle* 11, 3036–3044. doi:10.4161/cc.21399
- Datta, A., Pollock, K. J., Kormuth, K. A., and Brosh Jr, R. M. (2021). G-quadruplex assembly by ribosomal DNA: Emerging roles in disease pathogenesis and cancer biology. *Cytogenet and Genome Res.* 161 (6–7), 285–296. doi:10.1159/000516394
- Davis, A. E., III, Bissler, J. J., and Aulak, K. S. (1993). Genetic defects in the C1 inhibitor gene. *Complement. Today* 1, 133–150.
- Dayn, A., Samadashwily, G. M., and Mirkin, S. M. (1992). Intramolecular DNA triplexes: Unusual sequence requirements and influence on DNA polymerization. *Proc. Natl. Acad. Sci. U. S. A.* 89 (23), 11406–11410. doi:10.1073/pnas.89.23.11406
- DeJesus-Hernandez, M., Mackenzie, I. R., Boeve, B. F., Boxer, A. L., Baker, M., Rutherford, N. J., et al. (2011). Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron* 72 (2), 245–256. doi:10.1016/j.neuron.2011.09.011
- Del Mundo, I. M. A., Zewail-Foote, M., Kerwin, S. M., and Vasquez, K. M. (2017). Alternative DNA structure formation in the mutagenic human c-MYC promoter. *Nucleic Acids Res.* 45 (8), 4929–4943. doi:10.1093/nar/gkx100
- Delatycki, M. B., Knight, M., Koenig, M., Cossée, M., Williamson, R., and Forrest, S. M. (1999). G130V, a common FRDA point mutation, appears to have arisen from a common founder. *Hum. Genet.* 105 (4), 343–346. doi:10.1007/s004399900142
- Delbridge, A. R., Grabow, S., Strasser, A., and Vaux, D. L. (2016). Thirty years of BCL-2: Translating cell death discoveries into novel cancer therapies. *Nat. Rev. Cancer* 16 (2), 99–109. doi:10.1038/nrc.2015.17
- Dhakal, S., Schonhoff, J. D., Koirala, D., Yu, Z., Basu, S., and Mao, H. (2010). Coexistence of an ILPR i-motif and a partially folded structure with comparable mechanical stability revealed at the single-molecule level. *J. Am. Chem. Soc.* 132 (26), 8991–8997. doi:10.1021/ja100944j
- Dolinnaya, N. G., Oglloblina, A. M., and Yakubovskaya, M. G. (2016). Structure, properties, and biological relevance of the DNA and RNA G-quadruplexes: Overview 50 years after their discovery. *Biochemistry.* 81 (13), 1602–1649. doi:10.1134/S0006297916130034
- Drozdal, P., Gilski, M., and Jaskolski, M. (2021). Crystal structure of Z-DNA in complex with the polyamine putrescine and potassium cations at ultra-high resolution. *Acta Crystallogr. B Struct. Sci. Cryst. Eng. Mat.* 77 (3), 331–338. doi:10.1107/S2052520621002663
- Dryja, T. P., and Li, T. (1995). Molecular genetics of retinitis pigmentosa. *Hum. Mol. Genet.* 4 (1), 1739–1743. doi:10.1093/hmg/4.suppl_1.1739
- Duval-Valentin, G., Thuong, N. T., and Helene, C. (1992). Specific inhibition of transcription by triple helix-forming oligonucleotides. *Proc. Natl. Acad. Sci. U. S. A.* 89 (2), 504–508. doi:10.1073/pnas.89.2.504
- Dzatkow, S., Krafcikova, M., Hänsel-Hertsch, R., Fessl, T., Fiala, R., Loja, T., et al. (2018). Evaluation of the stability of DNA i-Motifs in the nuclei of living mammalian cells. *Angew. Chem. Int. Ed. Engl.* 57 (8), 2165–2169. doi:10.1002/anie.201712284
- Ebbinghaus, S. W., Gee, J. E., Rodu, B., Mayfield, C. A., Sanders, G., and Miller, D. M. (1993). Triplex formation inhibits HER-2/neu transcription *in vitro*. *J. Clin. Invest.* 92 (5), 2433–2439. doi:10.1172/JCI116850
- Egli, M. (2019). “Sugar pucker and nucleic acid structure,” in *The excitement of discovery: Selected papers of alexander rich: A tribute to alexander rich*, 309–315.
- Fan, H., and Chu, J. Y. (2007). A brief review of short tandem repeat mutation. *Genomics. proteomics Bioinforma.* 5 (1), 7–14. doi:10.1073/pnas.97.8.3862
- Faria, M., Wood, C. D., Perrouault, L., Nelson, J. S., Winter, A., White, M. R. H., et al. (2000). Targeted inhibition of transcription elongation in cells mediated by triplex-forming oligonucleotides. *Proc. Natl. Acad. Sci. U. S. A.* 97 (8), 3862–3867.
- Faruqi, A. F., Datta, H. J., Carroll, D., Seidman, M. M., and Glazer, P. M. (2000). Triple-helix formation induces recombination in mammalian cells via a nucleotide excision repair-dependent pathway. *Mol. Cell. Biol.* 20 (3), 990–1000. doi:10.1128/MCB.20.3.990-1000.2000
- Felsenfeld, G., Davies, D. R., and Rich, A. (1957). Formation of a three-stranded polynucleotide molecule. *J. Am. Chem. Soc.* 79 (8), 2023–2024. doi:10.1021/ja01565a074
- Fox, K. R. (2000). Targeting DNA with triplexes. *Curr. Med. Chem.* 7 (1), 17–37. doi:10.2174/0929867003375506
- François, J. C., Lacoste, J., Lacroix, L., and Mergny, J. L. (2000). 313. Cambridge: Academic Press, 74–95. Design of antisense and triplex-forming oligonucleotides *Methods Enzym.*
- Frank-Kamenetskii, M. D., and Mirkin, S. M. (1995). Triplex DNA structures. *Annu. Rev. Biochem.* 64 (1), 65–95. doi:10.1146/annurev.bi.64.070195.000433
- Frick, D. N., and Richardson, C. C. (2001). DNA primases. *Annu. Rev. Biochem.* 70 (1), 39–80. doi:10.1146/annurev.biochem.70.1.39
- Fry, M., and Loeb, L. A. (1994). The fragile X syndrome d (CGG) n nucleotide repeats form a stable tetrahedral structure. *Proc. Natl. Acad. Sci. U. S. A.* 91 (11), 4950–4954. doi:10.1073/pnas.91.11.4950
- Fu, Y., Comella, N., Tognazzi, K., Brown, L. F., Dvorak, H. F., and Kocher, O. (1999). Cloning of DLM-1, a novel gene that is up-regulated in activated macrophages, using RNA differential display. *Gene* 240 (1), 157–163. doi:10.1016/s0378-1119(99)00419-9
- Gabow, P. A. (1993). Autosomal dominant polycystic kidney disease. *N. Engl. J. Med.* 329 (5), 332–342. doi:10.1056/NEJM199307293290508
- Gehring, K., Leroy, J. L., and Gueron, M. (1993). A tetrameric DNA structure with protonated cytosine-cytosine base pairs. *Nature* 363 (6429), 561–565. doi:10.1038/363561a0
- Georgakopoulos-Soares, I., Morganello, S., Jain, N., Hemberg, M., and Nik-Zainal, S. (2018). Noncanonical secondary structures arising from non-B DNA motifs are determinants of mutagenesis. *Genome Res.* 28 (9), 1264–1271. doi:10.1101/gr.231688.117
- Gibbs, R. A., Nguyen, P. N., Edwards, A. L., Civitello, A. B., and Caskey, C. T. (1990). Multiplex DNA deletion detection and exon sequencing of the hypoxanthine phosphoribosyltransferase gene in Lesch-Nyhan families. *Genomics* 7 (2), 235–244. doi:10.1016/0888-7543(90)90545-6
- Goto, M., Ishikawa, Y., Sugimoto, M., and Furuichi, Y. (2013). Werner syndrome: A changing pattern of clinical manifestations in Japan (1917~ 2008). *Biosci. Trends* 7 (1), 13–22.
- Grand, C. L., Han, H., Muñoz, R. M., Weitman, S., Von Hoff, D. D., Hurley, L. H., et al. (2002). The cationic porphyrin TMPyP4 down-regulates c-MYC and human telomerase reverse transcriptase expression and inhibits tumor growth *in vivo*. *Mol. Cancer Ther.* 1 (8), 565–573.
- Gray, L. T., Vallur, A. C., Eddy, J., and Maizels, N. (2014). G quadruplexes are genome-wide targets of transcriptional helicases XPB and XPD. *Nat. Chem. Biol.* 10 (4), 313–318. doi:10.1038/nchembio.1475
- Grigoriev, M., Praseuth, D., Guieysse, A. L., Robin, P., Thuong, N. T., Hélène, C., et al. (1993). Inhibition of interleukin-2 receptor alpha-subunit gene expression by oligonucleotide-directed triple helix formation. *C. R. Acad. Sci. III* 316 (5), 492–495.
- Gu, J., and Leszczynski, J. (2002). Origin of Na⁺/K⁺ selectivity of the guanine tetraplexes in water: The theoretical rationale. *J. Phys. Chem. A* 106 (3), 529–532. doi:10.1021/jp012739g
- Guan, B., and Zhang, X. (2020). Aptamers as versatile ligands for biomedical and pharmaceutical applications. *Int. J. Nanomedicine* 15, 1059–1071. doi:10.2147/IJN.S237544
- Gueron, M., Demaret, J. P., and Filoche, M. (2000). A unified theory of the BZ transition of DNA in high and low concentrations of multivalent ions. *Biophys. J.* 78 (2), 1070–1083. doi:10.1016/s0006-3495(00)76665-3

- Gunaratnam, M., Swank, S., Haider, S. M., Galesa, K., Reszka, A. P., Beltran, M., et al. (2009). Targeting human gastrointestinal stromal tumor cells with a quadruplex-binding small molecule. *J. Med. Chem.* 52 (12), 3774–3783. doi:10.1021/jm900424a
- Haeusler, A. R., Donnelly, C. J., Periz, G., Simko, E. A., Shaw, P. G., Kim, M. S., et al. (2014). C9orf72 nucleotide repeat structures initiate molecular cascades of disease. *Nature* 507 (7491), 195–200. doi:10.1038/nature13124
- Haluska, F. G., Tsujimoto, Y., and Croce, C. M. (1988). The t (8; 14) breakpoint of the EW 36 undifferentiated lymphoma cell line lies 5' of MYC in a region prone to involvement in endemic Burkitt's lymphomas. *Nucleic Acids Res.* 16 (5), 2077–2085. doi:10.1093/nar/16.5.2077
- Han, H., Cliff, C. L., and Hurley, L. H. (1999). Accelerated assembly of G-quadruplex structures by a small molecule. *Biochemistry* 38 (22), 6981–6986. doi:10.1021/bi9905922
- Han, X., Leroy, J. L., and Guéron, M. (1998). An intramolecular i-motif: The solution structure and base-pair opening kinetics of d (5mCCT3CCT3ACCT3CC). *J. Mol. Biol.* 278 (5), 949–965. doi:10.1006/jmbi.1998.1740
- Hardin, C. C., Watson, T., Corregan, M., and Bailey, C. (1992). Cation-dependent transition between the quadruplex and Watson-Crick hairpin forms of d (CGCG3GCG). *Biochemistry* 31 (3), 833–841. doi:10.1021/bi00118a028
- Hayashi, T., Oshima, H., Mashima, T., Nagata, T., Katahira, M., and Kinoshita, M. (2014). Binding of an RNA aptamer and a partial peptide of a prion protein: Crucial importance of water entropy in molecular recognition. *Nucleic Acids Res.* 42 (11), 6861–6875. doi:10.1093/nar/gku382
- Hazel, P., Huppert, J., Balasubramanian, S., and Neidle, S. (2004). Loop-length-dependent folding of G-quadruplexes. *J. Am. Chem. Soc.* 126 (50), 16405–16415. doi:10.1021/ja045154j
- Helene, C. (1991). The anti-gene strategy: Control of gene expression by triplex-forming-oligonucleotides. *Anticancer. Drug Des.* 6 (6), 569–584.
- Herbert, A. (2020). Mendelian disease caused by variants affecting recognition of Z-DNA and Z-RNA by the Za domain of the double-stranded RNA editing enzyme ADAR. *Eur. J. Hum. Genet.* 28 (1), 114–117. doi:10.1038/s41431-019-0458-6
- Hill, S. F., and Meisler, M. H. (2021). Antisense oligonucleotide therapy for neurodevelopmental disorders. *Dev. Neurosci.* 43 (3–4), 247–252. doi:10.1159/000517686
- Hopkinson, J. B., Wright, D. N., McDonald, J. W., and Corner, J. L. (2006). The prevalence of concern about weight loss and change in eating habits in people with advanced cancer. *J. Pain Symptom Manage.* 32 (4), 322–331. doi:10.1016/j.jpainsymman.2006.05.012
- Huang, Z., Qiu, L., Zhang, T., and Tan, W. (2021). Integrating DNA nanotechnology with aptamers for biological and biomedical applications. *Matter* 4 (2), 461–489. doi:10.1016/j.matt.2020.11.002
- Huppert, J. L., and Balasubramanian, S. (2007). G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res.* 35 (2), 406–413. doi:10.1093/nar/gkl1057
- Hur, J. H., Kang, C. Y., Lee, S., Parveen, N., Yu, J., Shamim, A., et al. (2021). AC-Motif: A DNA motif containing adenine and cytosine repeat plays a role in gene regulation. *Nucleic Acids Res.* 49 (17), 10150–10165. doi:10.1093/nar/gkab728
- Husler, P. L., and Klump, H. H. (1995). Prediction of pH-dependent properties of DNA triple helices. *Arch. Biochem. Biophys.* 317 (1), 46–56. doi:10.1006/abbi.1995.1134
- Intody, Z., Perkins, B. D., Wilson, J. H., and Wensel, T. G. (2000). Blocking transcription of the human rhodopsin gene by triplex-mediated DNA photocrosslinking. *Nucleic Acids Res.* 28 (21), 4283–4290. doi:10.1093/nar/28.21.4283
- Izzo, M., Battistini, J., Provenzano, C., Martelli, F., Cardinali, B., and Falcone, G. (2022). Molecular therapies for myotonic dystrophy type 1: From small drugs to gene editing. *Int. J. Mol. Sci.* 23 (9), 4622. doi:10.3390/ijms23094622
- Jenjaroenpun, P., Chew, C. S., Yong, T. P., Choowongkamon, K., Thammasorn, W., and Kuznetsov, V. A. (2015). The TTSM database: A catalog of triplex target DNA sites associated with genes and regulatory elements in the human genome. *Nucleic Acids Res.* 43 (1), D110–D116. doi:10.1093/nar/gku970
- Joos, S., Falk, M. H., Lichter, P., Haluska, F. G., Henglein, B., Lenoir, G. M., et al. (1992a). Variable breakpoints in Burkitt lymphoma cells with chromosomal t (8; 14) translocation separate c-myc and the IgH locus up to several hundred kb. *Hum. Mol. Genet.* 1 (8), 625–632. doi:10.1093/hmg/1.8.625
- Joos, S., Haluska, F. G., Falk, M. H., Henglein, B., Hameister, H., Croce, C. M., et al. (1992b). Mapping chromosomal breakpoints of Burkitt's t (8; 14) translocations far upstream of c-myc. *J. Am. Chem. Soc.* 114 (23), 6547–6552. doi:10.1021/ja4109352
- Kang, H. J., Kendrick, S., Hecht, S. M., and Hurley, L. H. (2014). The transcriptional complex between the BCL2 i-motif and hnRNP LL is a molecular switch for control of gene expression that can be modulated by small molecules. *J. Am. Chem. Soc.* 136 (11), 4172–4185.
- Kato, T., Franconi, C. P., Sheridan, M. B., Hacker, A. M., Inagakai, H., Glover, T. W., et al. (2014). Analysis of the t (3; 8) of hereditary renal cell carcinoma: A palindrome-mediated translocation. *Cancer Genet.* 207 (4), 133–140. doi:10.1016/j.cancergen.2014.03.004
- Kato, T., Kurahashi, H., and Emanuel, B. S. (2012). Chromosomal translocations and palindromic AT-rich repeats. *Curr. Opin. Genet. Dev.* 22 (3), 221–228. doi:10.1016/j.gde.2012.02.004
- Kaushik, M., Kaushik, S., Bansal, A., Saxena, S., and Kukreti, S. (2011). Structural diversity and specific recognition of four stranded G-quadruplex DNA. *Curr. Mol. Med.* 11 (9), 744–769. doi:10.2174/156652411798062421
- Kaushik, M., Kaushik, S., Roy, K., Singh, A., Mahendru, S., Kumar, M., et al. (2016). A bouquet of DNA structures: Emerging diversity. *Biochem. Biophys. Rep.* 5, 388–395. doi:10.1016/j.bbrep.2016.01.013
- Kaushik, M., Prasad, M., Kaushik, S., Singh, A., and Kukreti, S. (2010). Structural transition from dimeric to tetrameric i-motif, caused by the presence of TAA at the 3'-end of human telomeric C-rich sequence. *Biopolymers* 93 (2), 150–160. doi:10.1002/bip.21313
- Kaushik, S., Kaushik, M., Svinarchuk, F., Malvy, C., Femandjian, S., and Kukreti, S. (2011). Presence of divalent cation is not mandatory for the formation of intramolecular purine-motif triplex containing human c-jun protooncogene target. *Biochemistry* 50 (19), 4132–4142. doi:10.1021/bi1012589
- Kaushik, S., and Kukreti, S. (2021). Formation of a DNA triple helical structure at BOLF1 gene of human herpesvirus 4 (HH4) genome. *J. Biomol. Struct. Dyn.* 39 (9), 3324–3335. doi:10.1080/07391102.2020.1764390
- Kendrick, S., Kang, H. J., Alam, M. P., Madathil, M. M., Agrawal, P., Gokhale, V., et al. (2014). The dynamic character of the BCL2 promoter i-motif provides a mechanism for modulation of gene expression by compounds that bind selectively to the alternative DNA hairpin structure. *J. Am. Chem. Soc.* 136 (11), 4161–4171. doi:10.1021/ja410934b
- Kettani, A., Bouaziz, S., Gorin, A., Zhao, H., Jones, R. A., and Patel, D. J. (1998). Solution structure of a Na cation stabilized DNA quadruplex containing G-G-G-G and G-C-G-C tetrads formed by GGGC repeats observed in adeno-associated viral DNA. *J. Mol. Biol.* 282 (3), 619–636. doi:10.1006/jmbi.1998.2030
- Khan, N., Kolimi, N., and Rathinavelan, T. (2015). Twisting right to left: A...A mismatch in a CAG trinucleotide repeat overexpansion provokes left-handed Z-DNA conformation. A mismatch in a CAG trinucleotide repeat overexpansion provokes left-handed Z-DNA conformation. *PLoS Comput. Biol.* 11 (4), e1004162. doi:10.1371/journal.pcbi.1004162
- Khan, S., Singh, A., Nain, N., Gulati, S., and Kukreti, S. (2021). Sequence-specific recognition of a coding segment of human DACH1 gene via short pyrimidine/purine oligonucleotides. *RSC Adv.* 11 (63), 40011–40021. doi:10.1039/d1ra06604h
- Kim, H. G., and Miller, D. M. (1998). A novel triplex-forming oligonucleotide targeted to human cyclin D1 (bcl-1, proto-oncogene) promoter inhibits transcription in HeLa cells. *Biochemistry* 37 (8), 2666–2672. doi:10.1021/bi972399i
- Kim, H. G., Reddoch, J. F., Mayfield, C., Ebbinghaus, S., Vigneshwaran, N., Thomas, S., et al. (1998). Inhibition of transcription of the human c-myc protooncogene by intermolecular triplex. *Biochemistry* 37 (8), 2299–2304. doi:10.1021/bi971819i
- Kim, M. Y., Duan, W., Gleason-Guzman, M., and Hurley, L. H. (2003). Design, synthesis, and biological evaluation of a series of fluoroquinolone derivatives with contrasting dual mechanisms of action against topoisomerase II and G-quadruplexes. *J. Med. Chem.* 46 (4), 571–583. doi:10.1021/jm0203377
- Kim, M. Y., Gleason-Guzman, M., Izbicka, E., Nishioka, D., and Hurley, L. H. (2003). The different biological effects of telomestatin and TMPyP4 can be attributed to their selectivity for interaction with intramolecular or intermolecular G-quadruplex structures. *Cancer Res.* 63 (12), 3247–3256.
- King, J. J., Irving, K. L., Evans, C. W., Chikhale, R. V., Becker, R., Morris, C. J., et al. (2020). DNA G-quadruplex and i-motif structure formation is interdependent in human cells. *J. Am. Chem. Soc.* 142 (49), 20600–20604. doi:10.1021/jacs.0c11708
- Klysik, J. (1995). An intramolecular triplex structure from non-mirror repeated sequence containing BothPy: Pu-PyandPu: Pu-Py triads. *J. Mol. Biol.* 245 (5), 499–507. doi:10.1006/jmbi.1994.0041
- Klysik, J., Zacharias, W., Galazka, G., Kwinkowski, M., Uznanski, B., and Okruszek, A. (1988). Structural interconversion of alternating purine-pyrimidine inverted repeats cloned in supercoiled plasmids. *Nucleic Acids Res.* 16 (14), 6915–6933. doi:10.1093/nar/16.14.6915
- Knauer, M. P., Kalish, J. M., Hegan, D. C., and Glazer, P. M. (2006). Triplex-stimulated intermolecular recombination at a single-copy genomic target. *Mol. Ther.* 14 (3), 392–400. doi:10.1016/j.ythme.2006.03.020

- Knop, J. M., Mukherjee, S. K., Oliva, R., Möbitz, S., and Winter, R. (2020). Remodeling of the conformational dynamics of noncanonical DNA structures by monomeric and aggregated α -synuclein. *J. Am. Chem. Soc.* 142 (43), 18299–18303. doi:10.1021/jacs.0c07192
- Koehler, H., Cotsmire, S., Langland, J., Kibler, K. V., Kalman, D., Upton, J. W., et al. (2017). Inhibition of DAI-dependent necroptosis by the Z-DNA binding domain of the vaccinia virus innate immune evasion protein, E3. *Proc. Natl. Acad. Sci. U. S. A.* 114 (43), 11506–11511. doi:10.1073/pnas.1700999114
- Kohwi, Y., and Kohwi-Shigematsu, T. (1988). Magnesium ion-dependent triple-helix structure formed by homopurine-homopyrimidine sequences in supercoiled plasmid DNA. *Proc. Natl. Acad. Sci. U. S. A.* 85 (11), 3781–3785. doi:10.1073/pnas.85.11.3781
- Konieczna, N., Romaniuk-Drapała, A., Lisiak, N., Totoń, E., Paszel-Jaworska, A., Kaczmarek, M., et al. (2019). Telomerase inhibitor TMPyP4 alters adhesion and migration of breast-cancer cells MCF7 and MDA-MB-231. *Int. J. Mol. Sci.* 20 (11), 2670. doi:10.3390/ijms20112670
- Krasilnikov, A. S., Panyutin, I. G., Samadashwily, G. M., Cox, R., Lazurkin, Y. S., and Mirkin, S. M. (1997). Mechanisms of triplex-caused polymerization arrest. *Nucleic Acids Res.* 25 (7), 1339–1346. doi:10.1093/nar/25.7.1339
- Krasilnikova, M. M., and Mirkin, S. M. (2004). Replication stalling at Friedreich's ataxia (GAA) n repeats *in vivo*. *Mol. Cell. Biol.* 24 (6), 2286–2295. doi:10.1128/MCB.24.6.2286-2295.2004
- Ku, T. H., Zhang, T., Luo, H., Yen, T. M., Chen, P. W., Han, Y., et al. (2015). Nucleic acid aptamers: An emerging tool for biotechnology and biomedical sensing. *Sensors* 15 (7), 16281–16313. doi:10.3390/s150716281
- Kuai, H., Zhao, Z., Mo, L., Liu, H., Hu, X., Fu, T., et al. (2017). Circular bivalent aptamers enable *in vivo* stability and recognition. *J. Am. Chem. Soc.* 139 (27), 9128–9131. doi:10.1021/jacs.7b04547
- Kukreti, S., Sun, J. S., Garestier, T., and Hélène, C. (1997). Extension of the range of DNA sequences available for triple helix formation: Stabilization of mismatched triplexes by acridine-containing oligonucleotides. *Nucleic Acids Res.* 25 (21), 4264–4270. doi:10.1093/nar/25.21.4264
- Kukreti, S., Sun, J. S., Garestier, T., Helene, C., Loakes, D., Brown, D. M., et al. (1998). Triple helices formed at oligopyrimidine-oligopurine sequences with base pair inversions: Effect of a triplex-specific ligand on stability and selectivity. *Nucleic Acids Res.* 26 (9), 2179–2183. doi:10.1093/nar/26.9.2179
- Kumar Kulabhusan, P., Hussain, B., and Yüce, M. (2020). Current perspectives on aptamers as diagnostic tools and therapeutic agents. *Pharmaceutics* 12 (7), 646. doi:10.3390/pharmaceutics12070646
- Kumari, S., Bugaut, A., Huppert, J. L., and Balasubramanian, S. (2007). An RNA G-quadruplex in the 5' UTR of the NRAS proto-oncogene modulates translation. *Nat. Chem. Biol.* 3 (4), 218–221. doi:10.1038/nchembio864
- Kveiborg, M., Albrechtsen, R., Couchman, J. R., and Wewer, U. M. (2008). Cellular roles of ADAM12 in health and disease. *Int. J. Biochem. Cell Biol.* 40 (9), 1685–1702. doi:10.1016/j.biocel.2008.01.025
- Lamandé, S. R., Chessler, S. D., Golub, S. B., Byers, P. H., Chan, C., Cole, W. G., et al. (1995). Endoplasmic reticulum-mediated quality control of type I collagen production by cells from osteogenesis imperfecta patients with mutations in the pro α 1(I) chain carboxyl-terminal propeptide which impair subunit assembly. *J. Biol. Chem.* 270 (15), 8642–8649. doi:10.1074/jbc.270.15.8642
- Laughlan, G., Murchie, A. I., Norman, D. G., Moore, M. H., Moody, P. C., Lilley, D. M., et al. (1994). The high-resolution crystal structure of a parallel-stranded guanine tetraplex. *Science* 265 (5171), 520–524. doi:10.1126/science.8036494
- Law, M. J., Lower, K. M., Voon, H. P., Hughes, J. R., Garrick, D., Viprasak, V., et al. (2010). ATR-X syndrome protein targets tandem repeats and influences allele-specific expression in a size-dependent manner. *Cell* 143, 367–378. doi:10.1016/j.cell.2010.09.023
- Lemarteleur, T., Gomez, D., Paterski, R., Mandine, E., Mailliet, P., and Riou, J. F. (2004). Stabilization of the c-myc gene promoter quadruplex by specific ligands' inhibitors of telomerase. *Biochem. Biophys. Res. Commun.* 323 (3), 802–808. doi:10.1016/j.bbrc.2004.08.150
- Lendeckel, U., Kohl, J., Arndt, M., Carl-McGrath, S., Donat, H., and Röcken, C. (2005). Increased expression of ADAM family members in human breast cancer and breast cancer cell lines. *J. Cancer Res. Clin. Oncol.* 131 (1), 41–48. doi:10.1007/s00432-004-0619-y
- Leonetti, C., Amodei, S., D'Angelo, C., Rizzo, A., Benassi, B., Antonelli, A., et al. (2004). Biological activity of the G-quadruplex ligand RHPS4 (3, 11-difluoro-6, 8, 13-trimethyl-8H-quinolone [4, 3, 2-k] acridinium methosulfate) is associated with telomere capping alteration. *Mol. Pharmacol.* 66 (5), 1138–1146. doi:10.1124/mol.104.001537
- Lim, K. W., Alberti, P., Guedin, A., Lacroix, L., Riou, J. F., Royle, N. J., et al. (2009). Sequence variant (CTAGGG) n in the human telomere favors a G-quadruplex structure containing a G-C-G-C tetrad. *Nucleic Acids Res.* 37 (18), 6239–6248. doi:10.1093/nar/gkp630
- Lin, Y. X., Wang, Y., Blake, S., Yu, M., Mei, L., Wang, H., et al. (2020). RNA nanotechnology-mediated cancer immunotherapy. *Theranostics* 10 (1), 281–299. doi:10.7150/thno.35568
- Liu, G., Myers, S., Chen, X., Bissler, J. J., Sinden, R. R., and Leffak, M. (2012). Replication fork stalling and checkpoint activation by a PKD1 locus mirror repeat polypurine-polypyrimidine (Pu-Py) tract. *J. Biol. Chem.* 287 (40), 33412–33423. doi:10.1074/jbc.M112.402503
- Liu, Y., Nairn, R. S., and Vasquez, K. M. (2008). Processing of triplex-directed psoralen DNA interstrand crosslinks by recombination mechanisms. *Nucleic Acids Res.* 36 (14), 4680–4688. doi:10.1093/nar/gkn438
- Lodi, R., Cooper, J. M., Bradley, J. L., Manners, D., Styles, P., Taylor, D. J., et al. (1999). Deficit of *in vivo* mitochondrial ATP production in patients with Friedreich ataxia. *Proc. Natl. Acad. Sci. U. S. A.* 96 (20), 11492–11495. doi:10.1073/pnas.96.20.11492
- Lohani, N., and Rajeswari, M. R. (2020). Antigenic and antiproliferative effects of triplex-forming oligonucleotide (TFO) targeted on hmgbl gene in human hepatoma cells. *Anticancer. Agents Med. Chem.* 20 (16), 1943–1955. doi:10.2174/1871520620666200619170438
- London, T. B., Barber, L. J., Mosedale, G., Kelly, G. P., Balasubramanian, S., Hickson, I. D., et al. (2008). FANCD1 is a structure-specific DNA helicase associated with the maintenance of genomic G/C tracts. *J. Biol. Chem.* 283 (52), 36132–36139. doi:10.1074/jbc.M808152200
- Lu, S., Wang, G., Bacolla, A., Zhao, J., Spitzer, S., and Vasquez, K. M. (2015). Short inverted repeats are hotspots for genetic instability: Relevance to cancer genomes. *Cell Rep.* 10 (10), 1674–1680. doi:10.1016/j.celrep.2015.02.039
- Macaya, R. F., Gilbert, D. E., Malek, S., Sinsheimer, J. S., and Feigon, J. (1991). Structure and stability of XGC mismatches in the third strand of intramolecular triplexes. *Science* 254 (5029), 270–274. doi:10.1126/science.254.5029.270
- Maizels, N. (2015). G4-associated human diseases. *EMBO Rep.* 16 (8), 910–922. doi:10.15252/embr.201540607
- Manzini, G., Yathindra, N., and Xodo, L. E. (1994). Evidence for intramolecularly folded i-DNA structures in biologically relevant CCC-repeat sequences. *Nucleic Acids Res.* 22 (22), 4634–4640. doi:10.1093/nar/22.22.4634
- Mayfield, C., Ebbinghaus, S., Gee, J., Jones, D., Rodu, B., Squibb, M., et al. (1994). Triplex formation by the human Ha-ras promoter inhibits Sp1 binding and *in vitro* transcription. *J. Biol. Chem.* 269 (27), 18232–18238. doi:10.1016/s0021-9258(17)32439-0
- McShan, W. M., Rossen, R. D., Laughter, A. H., Trial, J., Kessler, D. J., Zendegei, J. G., et al. (1992). Inhibition of transcription of HIV-1 in infected human cells by oligodeoxynucleotides designed to form DNA triple helices. *J. Biol. Chem.* 267 (8), 5712–5721. doi:10.1016/s0021-9258(18)42824-4
- Meng, H. M., Liu, H., Kuai, H., Peng, R., Mo, L., and Zhang, X. B. (2016). Aptamer-integrated DNA nanostructures for biosensing, bioimaging and cancer therapy. *Chem. Soc. Rev.* 45 (9), 2583–2602. doi:10.1039/c5cs00645g
- Mergny, J. L., Duval-Valentin, G., Nguyen, C. H., Perrouault, L., Faucon, B., Rougé, M., et al. (1992). Triple helix-specific ligands. *Science* 256 (5064), 1681–1684. doi:10.1126/science.256.5064.1681
- Mikami-Terao, Y., Akiyama, M., Yuza, Y., Yanagisawa, T., Yamada, O., Kawano, T., et al. (2009). Antitumor activity of TMPyP4 interacting G-quadruplex in retinoblastoma cell lines. *Exp. Eye Res.* 89 (2), 200–208. doi:10.1016/j.exer.2009.03.008
- Mirkin, S. M., Lyamichev, V. I., Drushlyak, K. N., Dobrynin, V. N., Filippov, S. A., and Frank-Kamenetskii, M. D. (1987). DNA H form requires a homopurine-homopyrimidine mirror repeat. *Nature* 330 (6147), 495–497. doi:10.1038/330495a0
- Mitas, M. (1997). Trinucleotide repeats associated with human disease. *Nucleic Acids Res.* 25 (12), 2245–2254. doi:10.1093/nar/25.12.2245
- Mitrasinovic, P. M. (2018). Structural insights into the binding of small ligand molecules to a G-quadruplex DNA located in the HIV-1 promoter. *J. Biomol. Struct. Dyn.* 36 (9), 2292–2302. doi:10.1080/07391102.2017.1358670
- Mitsui, Y., Langridge, R., Shortle, B. E., Cantor, C. R., Grant, R. C., Kodama, M., et al. (1970). Physical and enzymatic studies on poly d (I-C). Poly d (I-C), an unusual double-helical DNA. *Nature* 228 (5277), 1166–1169. doi:10.1038/2281166a0
- Mochizuki, S., and Okada, Y. (2007). ADAMs in cancer cell proliferation and progression. *Cancer Sci.* 98 (5), 621–628. doi:10.1111/j.1349-7006.2007.00434.x
- Mukherjee, A., and Vasquez, K. M. (2011). Triplex technology in studies of DNA damage, DNA repair, and mutagenesis. *Biochimie* 93 (8), 1197–1208. doi:10.1016/j.biochi.2011.04.001

- Murchie, A. I., and Lilley, D. M. (1992). Retinoblastoma susceptibility genes contain 5' sequences with a high propensity to form guanine-tetrad structures. *Nucleic Acids Res.* 20 (1), 49–53. doi:10.1093/nar/20.1.49
- Neidle, S. (2009). The structures of quadruplex nucleic acids and their drug complexes. *Curr. Opin. Struct. Biol.* 19 (3), 239–250. doi:10.1016/j.sbi.2009.04.001
- Nejedlý, K., Klysiak, J., and Paleček, E. (1989). Supercoil-stabilized left-handed DNA in the plasmid (dA-dT) 16 insert formed in the presence of Ni²⁺. *FEBS Lett.* 243 (2), 313–317. doi:10.1016/0014-5793(89)80152-8
- Nguyen, L., Papenhausen, P., and Shao, H. (2017). The role of c-MYC in B-cell lymphomas: Diagnostic and molecular aspects. *Genes* 8 (4), 116. doi:10.3390/genes8040116
- Niu, K., Zhang, X., Deng, H., Wu, F., Ren, Y., Xiang, H., et al. (2018). BmILF and i-motif structure are involved in transcriptional regulation of BmPOUM2 in *Bombyx mori*. *Nucleic Acids Res.* 46 (4), 1710–1723. doi:10.1093/nar/gkx1207
- Nogueira, C. P., McGuire, M. C., Graeser, C., Bartels, C. F., Arpagaus, M., Van der Spek, A. F., et al. (1990). Identification of a frameshift mutation responsible for the silent phenotype of human serum cholinesterase, Gly 117 (GGT---GGAG). *Am. J. Hum. Genet.* 46 (5), 934–942.
- Oh, J. M., Venters, C. C., Di, C., Pinto, A. M., Wan, L., Younis, I., et al. (2020). U1 snRNP regulates cancer cell migration and invasion *in vitro*. *Nat. Commun.* 11 (1), 1–8. doi:10.1038/s41467-019-13993-7
- Ohshima, K., Montermini, L., Wells, R. D., and Pandolfo, M. (1998). Inhibitory effects of expanded GAA·TTC triplet repeats from intron I of the Friedreich ataxia gene on transcription and replication *in vivo*. *J. Biol. Chem.* 273 (23), 14588–14595. doi:10.1074/jbc.273.23.14588
- Ou, T. M., Lu, Y. J., Zhang, C., Huang, Z. S., Wang, X. D., Tan, J. H., et al. (2007). Stabilization of G-quadruplex DNA and down-regulation of oncogene c-myc by quindoline derivatives. *J. Med. Chem.* 50 (7), 1465–1474. doi:10.1021/jm0610088
- Ozturk, M., Nilsen-Hamilton, M., and Ilgu, M. (2021). Aptamer applications in neuroscience. *Pharmaceuticals* 14 (12), 1260. doi:10.3390/ph14121260
- Pagano, A., Iaccarino, N., Abdelhamid, M. A., Brancaccio, D., Garzarella, E. U., Di Porzio, A., et al. (2018). Common G-quadruplex binding agents found to interact with i-motif-forming DNA: Unexpected multi-target-directed compounds. *Front. Chem.* 6, 281. doi:10.3389/fchem.2018.00281
- Pandya, N., Bhagwat, S. R., and Kumar, A. (2021). Regulatory role of non-canonical DNA Polymorphisms in human genome and their relevance in Cancer. *Biochim. Biophys. Acta. Rev. Cancer* 1876 (2), 188594. doi:10.1016/j.bbcan.2021.188594
- Park, J. H., Lee, H. S., Jang, M. D., Han, S. W., Kim, S. K., and Lee, Y. A. (2018). Enantioselective light switch effect of Δ - and Λ -[Ru(phenanthroline)₂dipyrido [3, 2-a: 2', 3'-c] phenazine] 2+ bound to G-quadruplex DNA. *J. Biomol. Struct. Dyn.* 36 (8), 1948–1957. doi:10.1080/07391102.2017.1345324
- Patel, D. J., Phan, A. T., and Kuryavii, V. (2007). Human telomere, oncogenic promoter and 5'-UTR G-quadruplexes: Diverse higher order DNA and RNA targets for cancer therapeutics. *Nucleic Acids Res.* 35 (22), 7429–7455. doi:10.1093/nar/gkm711
- Patel, H. P., Lu, L., Blaszkak, R. T., and Bissler, J. J. (2004). PKD1 intron 21: Triplex DNA formation and effect on replication. *Nucleic Acids Res.* 32 (4), 1460–1468. doi:10.1093/nar/gkh312
- Patel, P. K., and Hosur, R. V. (1999). NMR observation of T-tetrads in a parallel stranded DNA quadruplex formed by *Saccharomyces cerevisiae* telomere repeats. *Nucleic Acids Res.* 27 (12), 2457–2464. doi:10.1093/nar/27.12.2457
- Patel, P. K., Koti, A. S. R., and Hosur, R. V. (1999). NMR studies on truncated sequences of human telomeric DNA: Observation of a novel A-tetrad. *Nucleic Acids Res.* 27 (19), 3836–3843. doi:10.1093/nar/27.19.3836
- Pérez-Rentero, S., Gargallo, R., González, C., and Eritja, R. (2015). Modulation of the stability of i-motif structures using an acyclic threoinol cytidine derivative. *RSC Adv.* 5 (78), 63278–63281. doi:10.1039/c5ra10096h
- Phan, A. T. (2010). Human telomeric G-quadruplex: Structures of DNA and RNA sequences. *FEBS J.* 277 (5), 1107–1117. doi:10.1111/j.1742-4658.2009.07464.x
- Phan, A. T., Kuryavii, V., Gaw, H. Y., and Patel, D. J. (2005). Small-molecule interaction with a five-guanine-tract G-quadruplex structure from the human MYC promoter. *Nat. Chem. Biol.* 1 (3), 167–173. doi:10.1038/nchembio723
- Phan, A. T., Modi, Y. S., and Patel, D. J. (2004). Propeller-type parallel-stranded G-quadruplexes in the human c-myc promoter. *J. Am. Chem. Soc.* 126 (28), 8710–8716. doi:10.1021/ja048805k
- Pilch, D. S., Levenson, C., and Shafer, R. H. (1991). Structure, stability, and thermodynamics of a short intermolecular purine-purine-pyrimidine triple helix. *Biochemistry* 30 (25), 6081–6088. doi:10.1021/bi00239a001
- Platt, J. R. (1955). Possible separation of intertwined nucleic acid chains by transfer-twist. *Proc. Natl. Acad. Sci. U. S. A.* 41 (3), 181–183. doi:10.1073/pnas.41.3.181
- Plum, E. G., and Breslauer, K. J. (1995). Thermodynamics of an intramolecular DNA triple helix: A calorimetric and spectroscopic study of the pH and salt dependence of thermally induced structural transitions. *J. Mol. Biol.* 248 (3), 679–695. doi:10.1006/jmbi.1995.0251
- Pollard, L. M., Sharma, R., Gomez, M., Shah, S., Delatycki, M. B., Pianese, L., et al. (2004). Replication-mediated instability of the GAA triplet repeat mutation in Friedreich ataxia. *Nucleic Acids Res.* 32 (19), 5962–5971. doi:10.1093/nar/gkh933
- Pomponio, R. J., Narasimhan, V., Reynolds, T. R., Buck, G. A., Povirk, L. F., and Wolf, B. (1996). Deletion/insertion mutation that causes biotinidase deficiency may result from the formation of a quasispalindromic structure. *Hum. Mol. Genet.* 5 (10), 1657–1661. doi:10.1093/hmg/5.10.1657
- PoshtehShirani, M., Rezaei, B., Khayamian, T., Dinari, M., Karami, K., Mehri-Lighvan, Z., et al. (2018). Folate receptor-targeted multimodal fluorescence mesosilica nanoparticles for imaging, delivery palladium complex and *in vitro* G-quadruplex DNA interaction. *J. Biomol. Struct. Dyn.* 36 (16), 4156–4169. doi:10.1080/07391102.2017.1411294
- Postel, E. H., Flint, S. J., Kessler, D. J., and Hogan, M. E. (1991). Evidence that a triplex-forming oligodeoxyribonucleotide binds to the c-myc promoter in HeLa cells, thereby reducing c-myc mRNA levels. *Proc. Natl. Acad. Sci. U. S. A.* 88 (18), 8227–8231. doi:10.1073/pnas.88.18.8227
- Praseuth, D., Guieysse, A. L., and Helene, C. (1999). Triple helix formation and the antigene strategy for sequence-specific control of gene expression. *Biochim. Biophys. Acta* 1489 (1), 181–206. doi:10.1016/s0167-4781(99)00149-9
- Rachwal, P. A., Brown, T., and Fox, K. R. (2007). Effect of G-tract length on the topology and stability of intramolecular DNA quadruplexes. *Biochemistry* 46 (11), 3036–3044. doi:10.1021/bi062118j
- Rademakers, R. (2012). C9orf72 repeat expansions in patients with ALS and FTD. *Lancet. Neurol.* 11 (4), 297–298. doi:10.1016/S1474-4422(12)70046-7
- Raghavan, S. C., Chastain, P., Lee, J. S., Hegde, B. G., Houston, S., Langen, R., et al. (2005). Evidence for a triplex DNA conformation at the bcl-2 major breakpoint region of the t (14; 18) translocation. *J. Biol. Chem.* 280 (24), 22749–22760. doi:10.1074/jbc.M502952200
- Rajeswari, M. R. (2012). DNA triplex structures in neurodegenerative disorder, Friedreich's ataxia. *J. Biosci.* 37 (3), 519–532. doi:10.1007/s12038-012-9219-1
- Rapozzi, V., Zorzet, S., Zacchigna, M., Della Pietra, E., Cogoi, S., and Xodo, L. E. (2014). Anticancer activity of cationic porphyrins in melanoma tumour-bearing mice and mechanistic *in vitro* studies. *Mol. Cancer* 13 (1), 75–17. doi:10.1186/1476-4598-13-75
- Ravichandran, S., Subramani, V. K., and Kim, K. K. (2019). Z-DNA in the genome: From structure to disease. *Biophys. Rev.* 11 (3), 383–387. doi:10.1007/s12551-019-00534-1
- Ray, B. K., Dhar, S., Shakya, A., and Ray, A. (2011). Z-DNA-forming silencer in the first exon regulates human ADAM-12 gene expression. *Proc. Natl. Acad. Sci. U. S. A.* 108 (1), 103–108. doi:10.1073/pnas.1008831108
- Remes, A. M., Peuhkurinen, K. J., Herva, R., Majamaa, K., and Hassinen, I. E. (1993). Kearns-Sayre syndrome case presenting a mitochondrial DNA deletion with unusual direct repeats and a rudimentary RNAase mitochondrial ribonucleotide processing target sequence. *Genomics* 16 (1), 256–258. doi:10.1006/geno.1993.1170
- Renčíuk, D., Kypř, J., and Vorlíčková, M. (2011). CGG repeats associated with fragile X chromosome form left-handed Z-DNA structure. *Biopolymers* 95 (3), 174–181. doi:10.1002/bip.21555
- Renton, A. E., Majounie, E., Waite, A., Simón-Sánchez, J., Rollinson, S., Gibbs, J. R., et al. (2011). A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron* 72 (2), 257–268. doi:10.1016/j.neuron.2011.09.010
- Rhodes, D., and Lipps, H. J. (2015). G-quadruplexes and their regulatory roles in biology. *Nucleic Acids Res.* 43 (18), 8627–8637. doi:10.1093/nar/gkv862
- Riccardi, C., Napolitano, E., Platella, C., Musumeci, D., and Montesarchio, D. (2021). G-quadruplex-based aptamers targeting human thrombin: Discovery, chemical modifications and antithrombotic effects. *Pharmacol. Ther.* 217, 107649. doi:10.1016/j.pharmthera.2020.107649
- Rimokh, R., Rouault, J. P., Wahbi, K., Gadoux, M., Lafage, M., Archimbaud, E., et al. (1991). A chromosome 12 coding region is juxtaposed to the MYC protooncogene locus in at (8; 12)(q24; q22) translocation in a case of B-cell chronic lymphocytic leukemia. *Genes Chromosom. Cancer* 3 (1), 24–36. doi:10.1002/gcc.2870030106
- Robinson, D. O., Bunyan, D. J., Gabb, H. A., Temple, I. K., and Yau, S. C. (1997). A small intraexon deletion within the dystrophin gene suggests a possible

- mechanism of mutagenesis. *Hum. Genet.* 99 (5), 658–662. doi:10.1007/s004390050424
- Rooks, H., Clark, B., Best, S., Rushton, P., Oakley, M., Thein, O. S., et al. (2012). A novel 506 kb deletion causing $\epsilon\gamma\delta\beta$ thalassemia. *Blood Cells Mol. Dis.* 49 (3–4), 121–127. doi:10.1016/j.bcmd.2012.05.010
- Rossetti, S., Hopp, K., Sikkink, R. A., Sundsbak, J. L., Lee, Y. K., Kubly, V., et al. (2012). Identification of gene mutations in autosomal dominant polycystic kidney disease through targeted resequencing. *J. Am. Soc. Nephrol.* 23 (5), 915–933. doi:10.1681/ASN.2011101032
- Saada, A., Costa, A. B., Sheng, Z., Guo, W., Haber, J. E., and Lobachev, K. S. (2021). Structural parameters of palindromic repeats determine the specificity of nuclease attack of secondary structures. *Nucleic Acids Res.* 49 (7), 3932–3947. doi:10.1093/nar/gkab168
- Sacca, B., Lacroix, L., and Mergny, J. L. (2005). The effect of chemical modifications on the thermal stability of different G-quadruplex-forming oligonucleotides. *Nucleic Acids Res.* 33 (4), 1182–1192. doi:10.1093/nar/gki257
- Saglio, G., Borrello, M. G., Guerrasio, A., Sozzi, G., Serra, A., Celle, P. F. D., et al. (1993). Preferential clustering of chromosomal breakpoints in Burkitt's lymphomas and L3 type acute lymphoblastic leukemias with at (8; 14) translocation. *Genes Chromosom. Cancer* 8 (1), 1–7. doi:10.1002/gcc.2870080102
- Saha, U., Yasmeen Khan, A., Bhuiya, S., Das, S., Fiorillo, G., Lombardi, P., et al. (2019). Targeting human telomeric DNA quadruplex with novel berberubine derivatives: Insights from spectroscopic and docking studies. *J. Biomol. Struct. Dyn.* 37 (6), 1375–1389. doi:10.1080/07391102.2018.1459319
- Sanchez-Martin, V., Soriano, M., and Garcia-Salcedo, J. A. (2021). Quadruplex ligands in cancer therapy. *Cancers* 13 (13), 3156. doi:10.3390/cancers13133156
- SanthanaMariappan, S. V., Catasti, P., Chen, X., Ratliff, R., Moyzis, R. K., Morton Bradbury, E., et al. (1996). Solution structures of the individual single strands of the fragile X DNA triplets (GCC) $_n$ (GGC) $_n$. *Nucleic acids Res.* 24 (4), 784–792. doi:10.1093/nar/24.4.784
- Santoro, M. R., Bray, S. M., and Warren, S. T. (2012). Molecular mechanisms of fragile X syndrome: A twenty-year perspective. *Annu. Rev. Pathol.* 7, 219–245. doi:10.1146/annurev-pathol-011811-132457
- Sarkies, P., Reams, C., Simpson, L. J., and Sale, J. E. (2010). Epigenetic instability due to defective replication of structured DNA. *Mol. Cell* 40 (5), 703–713. doi:10.1016/j.molcel.2010.11.009
- Sarmiento, B. F. C. C., and Neves, J. D. (Editors) (2018). *Biomedical applications of functionalized nanomaterials: Concepts, development and clinical translation* (Norwich: William Andrew).
- Sasco, A. J., Secretan, M. B., and Straif, K. (2004). Tobacco smoking and cancer: A brief review of recent epidemiological evidence. *Lung cancer* 45, S3–S9. doi:10.1016/j.lungcan.2004.07.998
- Scaggiante, B., Morassutti, C., Tolazzi, G., Michelutti, A., Baccarani, M., and Quadrioglio, F. (1994). Effect of unmodified triple helix-forming oligodeoxyribonucleotide targeted to human multidrug-resistance gene mdr1 in MDR cancer cells. *FEBS Lett.* 352 (3), 380–384. doi:10.1016/0014-5793(94)00995-3
- Schaffitzel, C., Berger, I., Postberg, J., Hanes, J., Lipps, H. J., and Plückthun, A. (2001). In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with *Stylomychia lemnae* macronuclei. *Proc. Natl. Acad. Sci. U S A* 98 (15), 8572–8577. doi:10.1073/pnas.141229498
- Schwartz, T., Rould, M. A., Lowenhaupt, K., Herbert, A., and Rich, A. (1999). Crystal structure of the Z-DNA domain of the human editing enzyme ADAR1 bound to left-handed Z-DNA. *Science* 284 (5421), 1841–1845. doi:10.1126/science.284.5421.1841
- Seité, P., Leroux, D., Hillion, J., Monteil, M., Berger, R., Mathieu-Mahul, D., et al. (1993). Molecular analysis of a variant 18; 22 translocation in a case of lymphocytic lymphoma. *Genes Chromosom. Cancer* 6 (1), 39–44. doi:10.1002/gcc.2870060108
- Sen, D., and Gilbert, W. (1988). Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *nature* 334 (6180), 364–366. doi:10.1038/334364a0
- Serrano-Chacón, I., Mir, B., Escaja, N., and González, C. (2021). Structure of i-motif/duplex junctions at neutral pH. *J. Am. Chem. Soc.* 143 (33), 12919–12923. doi:10.1021/jacs.1c04679
- Shammas, M. A., Reis, R. J. S., Li, C., Koley, H., Hurley, L. H., Anderson, K. C., et al. (2004). Telomerase inhibition and cell growth arrest after telomestatin treatment in multiple myeloma. *Clin. Cancer Res.* 10 (2), 770–776. doi:10.1158/1078-0432.ccr-0793-03
- Shay, J. W., Reddel, R. R., and Wright, W. E. (2012). Cancer. Cancer and telomeres—an ALTernative to telomerase. *Science* 336 (6087), 1388–1390. doi:10.1126/science.1222394
- Sheng, Q., Neaverson, J. C., Mahmoud, T., Stevenson, C. E., Matthews, S. E., and Waller, Z. A. (2017). Identification of new DNA i-motif binding ligands through a fluorescent intercalator displacement assay. *Org. Biomol. Chem.* 15 (27), 5669–5673. doi:10.1039/c7ob00710h
- Shi, H., He, X., Wang, K., Wu, X., Ye, X., Guo, Q., et al. (2011). Activatable aptamer probe for contrast-enhanced *in vivo* cancer imaging based on cell membrane protein-triggered conformation alteration. *Proc. Natl. Acad. Sci. U. S. A.* 108 (10), 3900–3905. doi:10.1073/pnas.1016197108
- Siddiqui-Jain, A., Grand, C. L., Bearss, D. J., and Hurley, L. H. (2002). Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proc. Natl. Acad. Sci. U. S. A.* 99 (18), 11593–11598. doi:10.1073/pnas.182256799
- Simón-Sánchez, J., Dopfer, E. G., Cohn-Hokke, P. E., Hukema, R. K., Nicolaou, N., Seelaar, H., et al. (2012). The clinical and pathological phenotype of C9ORF72 hexanucleotide repeat expansions. *Brain.* 135 (3), 723–735. doi:10.1093/brain/awr353
- Simonsson, T. (2001). G-quadruplex DNA structures variations on a theme. *Biol. Chem.* 382, 621–628. doi:10.1515/BC.2001.073
- Simonsson, T., Kubista, M., and Pecinka, P. (1998). DNA tetraplex formation in the control region of c-myc. *Nucleic Acids Res.* 26 (5), 1167–1172. doi:10.1093/nar/26.5.1167
- Sinden, R. R. (1994). *DNA structure and function*. Houston: Gulf Professional Publishing.
- Singhal, G., Akhter, M. Z., Stern, D. F., Gupta, S. D., Ahuja, A., Sharma, U., et al. (2011). DNA triplex-mediated inhibition of MET leads to cell death and tumor regression in hepatoma. *Cancer Gene Ther.* 18 (7), 520–530. doi:10.1038/cgt.2011.21
- Solomon, O., Di Segni, A., Cesarkas, K., Porath, H. T., Marcu-Malina, V., Mizrahi, O., et al. (2017). RNA editing by ADAR1 leads to context-dependent transcriptome-wide changes in RNA secondary structure. *Nat. Commun.* 8 (1), 1–14. doi:10.1038/s41467-017-01458-8
- Subramani, V. K., Ravichandran, S., Bansal, V., and Kim, K. K. (2019). Chemical-induced formation of BZ-junction with base extrusion. *Biochem. Biophys. Res. Commun.* 508 (4), 1215–1220. doi:10.1016/j.bbrc.2018.12.045
- Suganthi, S., Sivaraj, R., Selvakumar, P. M., and Enoch, I. V. (2018). Supramolecular complex binding to G-quadruplex DNA: Berberine encapsulated by a planar side arm-tethered β -cyclodextrin. *J. Biomol. Struct. Dyn.* 37, 3305–3313. doi:10.1080/07391102.2018.1512420
- Sun, H., Karow, J. K., Hickson, I. D., and Maizels, N. (1998). The Bloom's syndrome helicase unwinds G4 DNA. *J. Biol. Chem.* 273 (42), 27587–27592. doi:10.1074/jbc.273.42.27587
- Sun, J. S., and Hélène, C. (1993). Oligonucleotide-directed triple-helix formation. *Curr. Opin. Struct. Biol.* 3 (3), 345–356. doi:10.1016/s0959-440x(05)80105-8
- Suram, A., Rao, J. K., Latha, K. S., and Viswamitra, M. A. (2002). First evidence to show the topological change of DNA from B-DNA to Z-DNA conformation in the hippocampus of Alzheimer's brain. *Neuromolecular Med.* 2 (3), 289–297. doi:10.1385/nmm:2:3:289
- SvetećMiklenić, M., and Sveteć, I. K. (2021). Palindromes in DNA—a risk for genome stability and implications in cancer. *Int. J. Mol. Sci.* 22 (6), 2840. doi:10.3390/ijms22062840
- Tateishi-Karimata, H., and Sugimoto, N. (2020). Chemical biology of non-canonical structures of nucleic acids for therapeutic applications. *Chem. Commun.* 56 (16), 2379–2390. doi:10.1039/c9cc09771f
- Tateishi-Karimata, H., and Sugimoto, N. (2021). Roles of non-canonical structures of nucleic acids in cancer and neurodegenerative diseases. *Nucleic Acids Res.* 49 (14), 7839–7855. doi:10.1093/nar/gkab580
- Teng, F. Y., Jiang, Z. Z., Guo, M., Tan, X. Z., Chen, F., Xi, Xu-G., et al. (2021). G-quadruplex DNA: A novel target for drug design. *Cell. Mol. Life Sci.* 78, 6557–6583. doi:10.1007/s00018-021-03921-8
- Thakur, R. K., Kumar, P., Halder, K., Verma, A., Kar, A., Parent, J. L., et al. (2009). Metastases suppressor NM23-H2 interaction with G-quadruplex DNA within c-MYC promoter nuclease hypersensitive element induces c-MYC expression. *Nucleic Acids Res.* 37 (1), 172–183. doi:10.1093/nar/gkn919
- the IgH locus up to several hundred kb. *Hum. Mol. Genet.* 1 (8), 625–632.
- Thuong, N. T., and Hélène, C. (1993). Sequence-specific recognition and modification of double-helical DNA by oligonucleotides. *Angew. Chem. Int. Ed. Engl.* 32 (5), 666–690. doi:10.1002/anie.199306661
- Tian, B. L., Wen, J. M., Zhang, M., Xie, D., Xu, R. B., and Luo, C. J. (2002). The expression of ADAM12 (meltrin α) in human giant cell tumours of bone. *Mol. Pathol.* 55 (6), 394–397. doi:10.1136/mp.55.6.394
- Tiner, W. J., Sr, Potaman, V. N., Sinden, R. R., and Lyubchenko, Y. L. (2001). The structure of intramolecular triplex DNA: Atomic force microscopy study. *J. Mol. Biol.* 314 (3), 353–357. doi:10.1006/jmbi.2001.5174

- Torres, V. E., Harris, P. C., and Pirson, Y. (2007). Autosomal dominant polycystic kidney disease. *Lancet* 369 (9569), 1287–1301. doi:10.1016/S0140-6736(07)60601-1
- Umek, T., Sollander, K., Bergquist, H., Wengel, J., Lundin, K. E., Smith, C. L., et al. (2019). Oligonucleotide binding to non-B-DNA in MYC. *Molecules* 24 (5), 1000. doi:10.3390/molecules24051000
- University of Washington (1993). Seattle, Seattle (WA): University of Washington. GeneReviews®
- Van Raay, T. J., Burn, T. C., Connors, T. D., Petry, L. R., Germino, G. G., Klinger, K. W., et al. (1996). A 2.5 kb polypyrimidine tract in the PKD1 gene contains at least 23 H-DNA-forming sequences. *Microb. Comp. Genomics* 1 (4), 317–327. doi:10.1089/mcg.1996.1.317
- Vasquez, K. M., Marburger, K., Intody, Z., and Wilson, J. H. (2001). Manipulating the mammalian genome by homologous recombination. *Proc. Natl. Acad. Sci. U. S. A.* 98 (15), 8403–8410. doi:10.1073/pnas.111009698
- Vasquez, K. M., Narayanan, L., and Glazer, P. M. (2000). Specific mutations induced by triplex-forming oligonucleotides in mice. *Science* 290 (5491), 530–533. doi:10.1126/science.290.5491.530
- Vasquez, K. M., Wang, G., Havre, P. A., and Glazer, P. M. (1999). Chromosomal mutations induced by triplex-forming oligonucleotides in mammalian cells. *Nucleic Acids Res.* 27 (4), 1176–1181. doi:10.1093/nar/27.4.1176
- Venczel, E. A., and Sen, D. (1993). Parallel and antiparallel G-DNA structures from a complex telomeric sequence. *Biochemistry* 32 (24), 6220–6228. doi:10.1021/bi00075a015
- Verma, A., Yadav, V. K., Basundra, R., Kumar, A., and Chowdhury, S. (2009). Evidence of genome-wide G4 DNA-mediated gene expression in human cancer cells. *Nucleic Acids Res.* 37 (13), 4194–4204. doi:10.1093/nar/gkn1076
- Vinnarasi, S., Radhika, R., Vijayakumar, S., and Shankar, R. (2019). Structural insights into the anti-cancer activity of quercetin on G-tetrad, mixed G-tetrad, and G-quadruplex DNA using quantum chemical and molecular dynamics simulations. *J. Biomol. Struct. Dyn.* 38, 317–339. doi:10.1080/07391102.2019.1574239
- Vorlicková, M., Kejniovská, I., Tumová, M., and Kyrp, J. (2001). Conformational properties of DNA fragments containing GAC trinucleotide repeats associated with skeletal displasias. *Eur. Biophys. J.* 30 (3), 179–185. doi:10.1007/s002490000121
- Wang, A. H. J., Quigley, G. J., Kolpak, F. J., Crawford, J. L., Van Boom, J. H., van der Marel, G., et al. (1979). Molecular structure of a left-handed double helical DNA fragment at atomic resolution. *Nature* 282 (5740), 680–686. doi:10.1038/282680a0
- Wang, G., Carbajal, S., Vijg, J., DiGiovanni, J., and Vasquez, K. M. (2008). DNA structure-induced genomic instability *in vivo*. *J. Natl. Cancer Inst.* 100 (24), 1815–1817. doi:10.1093/jnci/djn385
- Wang, G., Christensen, L. A., and Vasquez, K. M. (2006). Z-DNA-forming sequences generate large-scale deletions in mammalian cells. *Proc. Natl. Acad. Sci. U. S. A.* 103 (8), 2677–2682. doi:10.1073/pnas.0511084103
- Wang, G., and Vasquez, K. M. (2014). Impact of alternative DNA structures on DNA damage, DNA repair, and genetic instability. *DNA repair* 19, 143–151. doi:10.1016/j.dnarep.2014.03.017
- Wang, G., and Vasquez, K. M. (2009). Models for chromosomal replication-independent non-B DNA structure-induced genetic instability. *Mol. Carcinog.* 48 (4), 286–298. Published in cooperation with the University of Texas MD Anderson Cancer Center.
- Wang, G., and Vasquez, K. M. (2004). Naturally occurring H-DNA-forming sequences are mutagenic in mammalian cells. *Proc. Natl. Acad. Sci. U. S. A.* 101 (37), 13448–13453. doi:10.1073/pnas.0405116101
- Wang, G., and Vasquez, K. M. (2006). Non-B DNA structure-induced genetic instability. *Mutat. Res.* 598 (1–2), 103–119. doi:10.1016/j.mrfmmm.2006.01.019
- Wang, G., and Vasquez, K. M. (2007). Z-DNA, an active element in the genome. *Front. Biosci.* 12 (4424), 4424–4438. doi:10.2741/2399
- Wang, Y., Cortez, D., Yazdi, P., Neff, N., Elledge, S. J., and Qin, J. (2000). BASC, a super complex of BRCA1-associated proteins involved in the recognition and repair of aberrant DNA structures. *Genes Dev.* 14 (8), 927–939. doi:10.1101/gad.14.8.927
- Wells, R. D., Collier, D. A., Hanvey, J. C., Shimizu, M., and Wohlrab, F. (1988). The chemistry and biology of unusual DNA structures adopted by oligopurine-oligopyrimidine sequences. *FASEB J.* 2 (14), 2939–2949. doi:10.1096/fasebj.2.14.3053307
- Wells, R. D. (2008). DNA triplexes and Friedreich ataxia. *FASEB J.* 22 (6), 1625–1634. doi:10.1096/fj.07-097857
- Wells, R. D. (2007). Non-B DNA conformations, mutagenesis and disease. *Trends Biochem. Sci.* 32 (6), 271–278. doi:10.1016/j.tibs.2007.04.003
- Wells, R. D., Warren, S. T., and Sarmiento, M. (Editors) (1998). *Genetic instabilities and hereditary neurological diseases* (Cambridge: Academic Press).
- Wilda, M., Busch, K., Klose, I., Keller, T., Woessmann, W., Kreuder, J., et al. (2004). Level of MYC overexpression in pediatric Burkitt's lymphoma is strongly dependent on genomic breakpoint location within the MYC locus. *Genes Chromosom. Cancer* 41 (2), 178–182. doi:10.1002/gcc.20063
- Wölfel, S., Wittig, B., and Rich, A. (1995). Identification of transcriptionally induced Z-DNA segments in the human c-myc gene. *Biochim. Biophys. Acta* 1264 (3), 294–302. doi:10.1016/0167-4781(95)00155-7
- Wolski, P., Wojton, P., Nieszporek, K., and Panczyk, T. (2019). Interaction of human telomeric i-motif DNA with single-walled carbon nanotubes: Insights from molecular dynamics simulations. *J. Phys. Chem. B* 123 (49), 10343–10353. doi:10.1021/acs.jpcc.9b07292
- Wright, E. P., Huppert, J. L., and Waller, Z. A. (2017). Identification of multiple genomic DNA sequences which form i-motif structures at neutral pH. *Nucleic Acids Res.* 45 (6), 2951–2959. doi:10.1093/nar/gkx090
- Wu, L., Zhou, W., Lin, L., Chen, A., Feng, J., Qu, X., et al. (2022). Delivery of therapeutic oligonucleotides in nanoscale. *Bioact. Mat.* 7, 292–323. doi:10.1016/j.bioactmat.2021.05.038
- Wu, Y., Shin-ya, K., and Brosh, R. M., Jr (2008). FANCD1 helicase defective in Fanconi anemia and breast cancer unwinds G-quadruplex DNA to defend genomic stability. *Mol. Cell. Biol.* 28 (12), 4116–4128. doi:10.1128/MCB.02210-07
- Yamakawa-Kobayashi, K., Kobayashi, T., Yanagi, H., Shimakura, Y., Satoh, J., and Hamaguchi, H. (1994). A novel complex mutation in the LDL receptor gene probably caused by the simultaneous occurrence of deletion and insertion in the same region. *Hum. Genet.* 93 (6), 625–628. doi:10.1007/BF00201560
- Yang, D., and Okamoto, K. (2010). Structural insights into G-quadruplexes: Towards new anticancer drugs. *Future Med. Chem.* 2 (4), 619–646. doi:10.4155/fmc.09.172
- Yang, S., Li, H., Xu, L., Deng, Z., Han, W., Liu, Y., et al. (2018). Oligonucleotide aptamer-mediated precision therapy of hematological malignancies. *Mol. Ther. Nucleic Acids* 13, 164–175. doi:10.1016/j.omtn.2018.08.023
- Yang, Y., Fu, H., Qian, C., Li, H., and Chen, D. D. (2020). Characterization of interaction between Bcl-2 oncogene promoter I-Motif DNA and flavonoids using electrospray ionization mass spectrometry and pressure-assisted capillary electrophoresis frontal analysis. *Talanta* 215, 120885. doi:10.1016/j.talanta.2020.120885
- Yoga, Y. M., Traore, D. A., Sidiqi, M., Szeto, C., Pardini, N. R., Barker, A., et al. (2012). Contribution of the first K-homology domain of poly (C)-binding protein 1 to its affinity and specificity for C-rich oligonucleotides. *Nucleic Acids Res.* 40 (11), 5101–5114. doi:10.1093/nar/gks058
- Yüce, M., Kurt, H., Hussain, B., and Budak, H. (2018). “Systematic evolution of ligands by exponential enrichment for aptamer selection,” in *Biomedical applications of functionalized nanomaterials*. (Elsevier), 211–243.
- Yu-Ru, X., Fang-Yuan, T., Xiao-Zhen, T., Hong-Ting, Z., and Yong, X. (2020). Anti-inflammatory activities of berberine in the treatment of metabolic disorders by regulating the gut microbiota. *Prog. Biochem. BIOPHYSICS* 47 (8), 835–843.
- Zeraati, M., Langley, D. B., Schofield, P., Moye, A. L., Rouet, R., Hughes, W. E., et al. (2018). I-motif DNA structures are formed in the nuclei of human cells. *Nat. Chem.* 10 (6), 631–637. doi:10.1038/s41557-018-0046-3
- Zhou, G., Lin, M., Song, P., Chen, X., Chao, J., Wang, L., et al. (2014). Multivalent capture and detection of cancer cells with DNA nanostructured biosensors and multibranch hybridization chain reaction amplification. *Anal. Chem.* 86 (15), 7843–7848. doi:10.1021/ac502276w
- Zhou, J. L., Lu, Y. J., Ou, T. M., Zhou, J. M., Huang, Z. S., Zhu, X. F., et al. (2005). Synthesis and evaluation of quindoline derivatives as G-quadruplex inducing and stabilizing ligands and potential inhibitors of telomerase. *J. Med. Chem.* 48 (23), 7315–7321. doi:10.1021/jm050041b
- Zhu, H., Shang, D., Sun, M., Choi, S., Liu, Q., Hao, J., et al. (2011). X-linked congenital hypertrichosis syndrome is associated with interchromosomal insertions mediated by a human-specific palindromic near SOX3. *Am. J. Hum. Genet.* 88 (6), 819–826. doi:10.1016/j.ajhg.2011.05.004