Sequence Specific Structural Polymorphism of DNA

Shikha Kaushik
Department of Chemistry,
Rajdhani College, University of Delhi
New Delhi -110015.

Received June 10, 2017 Accepted July 18, 2017

ABSTRACT *DNA double helical structure exhibits structural polymorphism and numerous studies have revealed that this ambiguous biomolecule adopts various polymorphic forms. A galaxy of non-canonical structure of DNA have been explored and reported across the world. The physiological form of DNA is "B-form", where two antiparallel strands are joined by Watson-Crick hydrogen bonding. DNA is a versatile molecule which can adoptstructuresother than Watson-Crick duplex, depending on the sequence and solution conditions like triplexes, G-quadruplexes, i-motif (C-tetraplexes), parallel stranded structure, tri-G-quadruplexes and many more. Extensive studies have discussed the biological importance of such structures and have also shown their occurrencein-vivo. This review is an attempt to summarize the different non-canonical structures adopted by DNA oligonucleotide and their biological relevance. The discovery of multi-stranded DNA structures provide new openings towards treating genetic disorders and has paved the way for the treatment of many lethal diseases.*

***Key words*:** DNA Double Helix, Triplex, Quadruplex, DNA polymorphism.

1. INTRODUCTION

DNA double helical structure was discovered by James Watson and Francis Crick in 1953 and this discovery proved to be a milestone in the history of science (Watson and Crick, 1953). Since its discovery, this inscrutable molecule has unravel many opportunities in the field of molecular biology, forensics, gene therapy andstructural biology. DNA acts as a carrier of genetic information in nearly all living organisms except few viruses, viral genomes generally used RNA as genetic material. During the essential biological processes such as replication and transcription, unwinding of DNA occurs with the aid of enzymes.So, for the normal functioning of the cells, the integrity of this delicate and multifariousbiomolecule is important.

For many years, DNA was considered as a rigid molecule having a sole purpose of storing genetic information, but numerous studies has revealed the dynamic stature of this biomolecule. The double helical structure of DNA is capable of forming myriad structures consisting of one, two, three, or four strands having different folding pattern. Such structure plays significant role in various cellular processes, at genomic and molecular level(Sharma, 2011). Single-stranded hairpins, triplexes, G-Quadruplexes, i-motif and parallel-stranded DNA are some examples of alternative forms of DNA. These structures arisesbecauseDNA bases have numerous hydrogen bonding sites (apart from Watson-Crick). Structural Polymorphism of DNA is governed by several factors like base sequence, concentration of the cation, pH of the solution, temperature, and other solvent conditions, in which transition from one conformation to another can take place.Some of these unusual DNA structures are shown in figure 1.

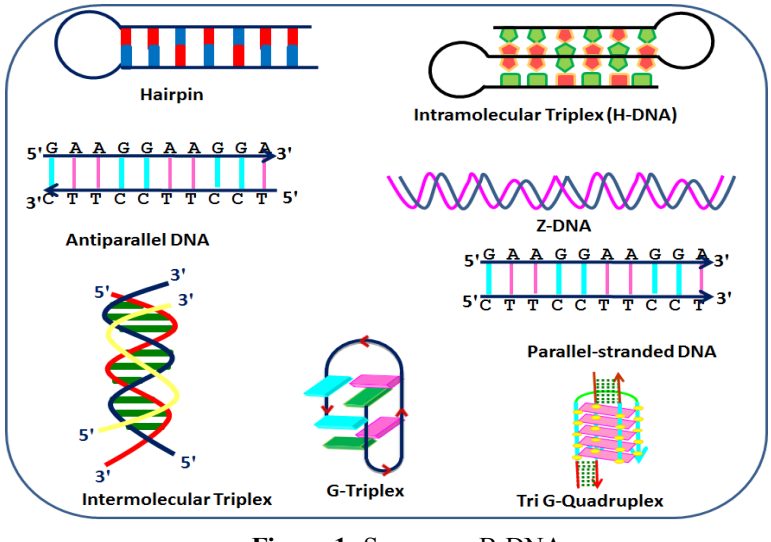

Figure 1: Some non-B-DNA

2. DNA DOUBLE HELIX

The majority of DNA within the cell is considered to be present in the "B-form", it also has the potential to adopt numerous other conformations, known as non-B-DNA forms. In "B-form" of DNA, the two strands of the helix are held together by the complementary base pairing- adenine (A) is hydrogen bonded to thymine (T) via two hydrogen bonds whereas guanine (G) pairs with cytosine (C) via three hydrogen bonds. The sugar phosphate backbones forms the outside of the helix and the purine and pyrimidine bases lie inside the helix (Figure 2).DNA is anextremely important biomolecule as it serves two main biological purposes. First, it allow the biosynthesis of new and identical DNA strand through the process of replication. In addition, the specific DNA strand, act as a template and direct the synthesis of mRNA (by the process of transcription) and, finally mRNA strand translated into proteins. The message stored in DNA bases is crucial for the biological processes like transcription, replication and recombination.

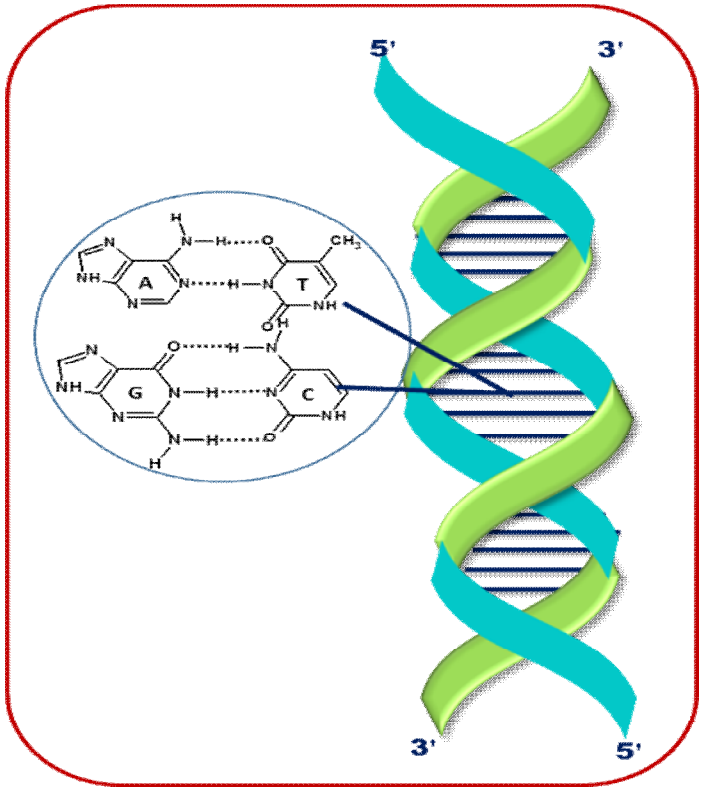

Figure 2: Watson-Crick Double Helix

3. PARALLEL-STRANDED DNA

Conformational polymorphism of DNA has now extended beyond the Watson-Crick double helix, and one such example is parallel-stranded (ps) DNA structure. Parallel-stranded DNA is formed by purine rich sequences(dG•dA)n which are present in the eukaryotic genome. The hydrogen bonding scheme present in the ps-DNA is reverse Watson-Crick.Alternatively, ps duplexes can also be formed using Hoogsteen A•T base pairing (Raghunathan et al, 1994). The base pairing pattern involved in Watson-Crick, reverse Watson-Crick and Hoogsteenbonding is shown in figure. 3. The biological role of such structures is not well understood but it may possible that two single stranded regions of DNA could interact in a parallel fashion and single strand regions do occurs at telomeres and such sequences can form parallel stranded double helix.

e ISSN 2348 –1269, Print ISSN 2349-5138
Cosmos Impact Factor 4.236

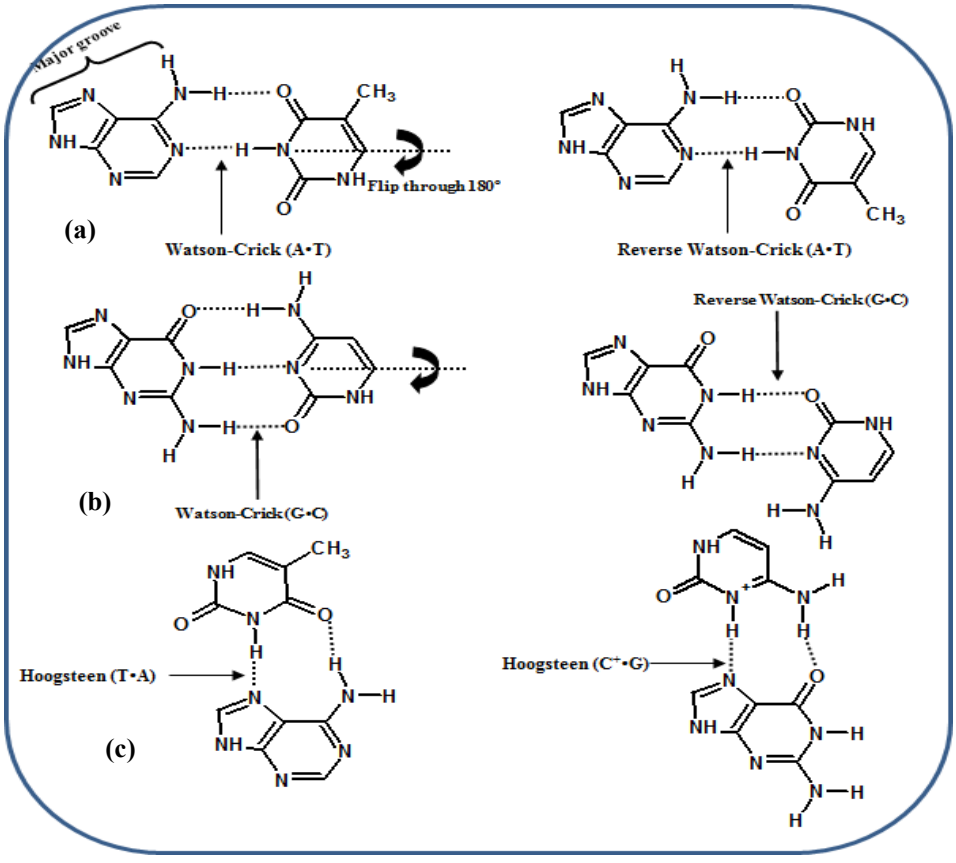

Figure 3: Base Pairing schemes in DNA Duplexes (a) Watson-Crick and reverse Watson-Crick (A•T) (b) Watson-Crick and reverse Watson-Crick (G•C) and Hoogsteen (A•T) and ($G•C^+$)

4. TRIPLE-STRANDED DNA (DNA TRIPLEXES)

Polypurine•polypyrimidine (Pu•Py) regions are abundant in eukaryotic DNA. These regions of the genome exhibits structural polymorphism and therefore can assume several non B- DNA conformations.The formation of triple-stranded DNA structures was reported soon after the discovery of DNA double helix (Felsenfeld et al, 1957) but their role in biological and therapeutic intervention were understood in the late 80s (Moser and Dervan, 1987). A DNA triple helix is formed either by the association of three strands (intermolecular triplex) or by the folding of a single DNA strand (intramolecular triplex). Third strand that binds to the target duplex (Watson-Crick) can be purine rich (A, G), pyrimidine rich (C,T) or mixed (G,T). So, DNA triplexes can be classified into three categories on the basis of the orientation of third strand, (i) purine motif, (ii) pyrimidine motif and (iii) mixed motif (Figures 4). In the pyrimidine motif, third strand align itself parallel to the central purine rich strand forming T*A•T and C^+*G•C triplets (Moser and Dervan, 1987). Such type of triplexes require acidic conditions [low pH] so as to protonate the cytosines of the third strand. In the purine motif, the orientation of third strand is antiparallel to the central purine rich strand and form A*A•T and G*G•C triplets (Beal and Dervan, 1991). Intermolecular purine motif triplexes require the presence of divalent cations for their formation. However, presence of divalent cations is not necessary for the formation of intramolecular purine motif triplexes (Kaushik et al, 2011).Mixed motifwhich form T*A•T and G*G•C can have parallel/antiparallel orientation, depending on the bases in the third strand (Durland et al, 1991).

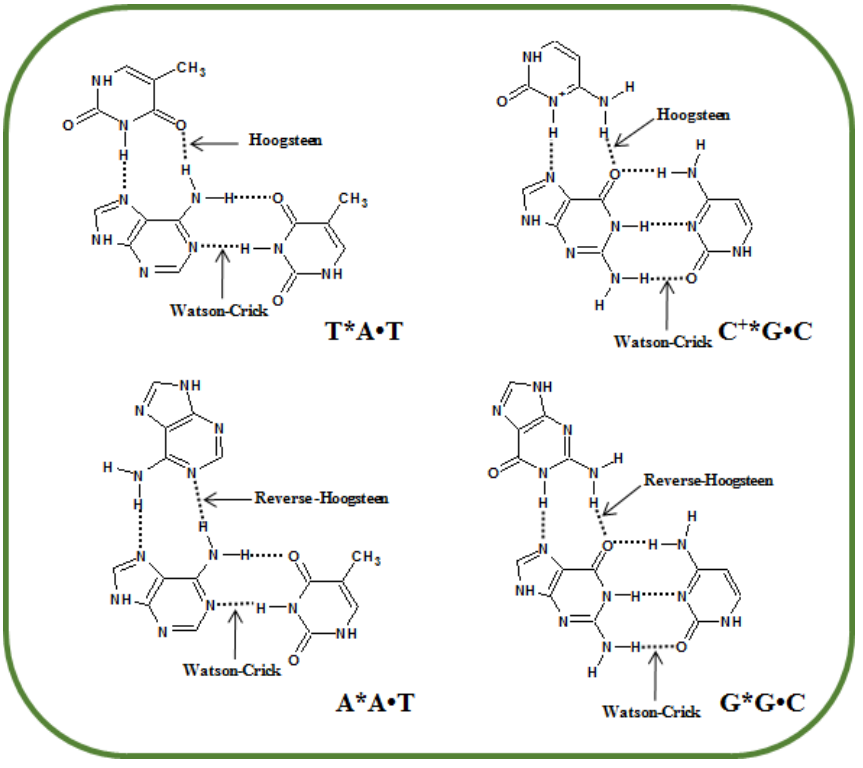

Figure 4: Hydrogen bonding pattern involved in DNA Triplexes (a) Pyrimidine Triplexes and (b) Purine Triplexes

5. GUANINE-QUADRUPLEXES

G-Quadruplexes are formed by DNA sequences that are rich in guanines and these structures are stabilized by Hoogsteen hydrogen bonding between four guanines (Figure 5).The formation of G-Quadruplex structure was discovered by Gellert and co-workers in 1962 (Gellert et al, 1962). The first G-quadruplex formation was observed in telomeric DNA (Yang and Okamoto, 2010). G-Quadruplexes exhibit extensive polymorphism and can be categorized into different groups on the basis of strand stoichiometry, strand polarity, glycosidic torsion angle variation and other parameters. These structures can be formed either by the intermolecular association of two/four DNA G-rich sequences or by the folding of a single strand (intramolecular association) and may have parallel/antiparallel orientation.

Sodium and potassium ions are among the most effective at stabilizing quadruplex structures. Because these cations are present in the cellular milieu, *in vivo* conditions are generally favorable for the existence of these structures. With the advancement of technology, various new G-rich DNA structureslike tri-G-quadruplex, (3+1) G-quadruplex and G-triplex are added to the pre-existing cluster of G-quadruplexes (Cerofolini et al, 2014).

6. i-MOTIF

It was proposed in 1960s that stretches of cytosines can form parallel duplexes under acidic conditions(Langridge and Rich, 1963). The formation of novel i-motif structure was first reportedin 1994 for the a C-rich sequence d(CCCC)(Chen et al, 1994). i-motif (intercalated motif)is another class of multistranded DNA structures which forms at low pH andconsisted of two parallel-stranded DNA duplexes held together in an antiparallel orientation(Day et al, 2014).These sequences can be found in the regulatory regions of eukaryotic genome and i-motif structures are involved in cellular processes like replication, recombination, transcription and chromatin organization.Depending on the sequence length, strand orientation, number of cytosine base pairs and pH of the solution, these sequences can form uni, bi- or tetramolecular structures. Several studies have been reported in the recent time, confirming the existence of such structures *in-vivo*by using different biophysical andbiochemical techniques (Amato et al, 2014). Figure 5 displays the different types of G-Quadruplex and i-motif DNA structures.

e ISSN 2348 –1269, Print ISSN 2349-5138
Cosmos Impact Factor 4.236

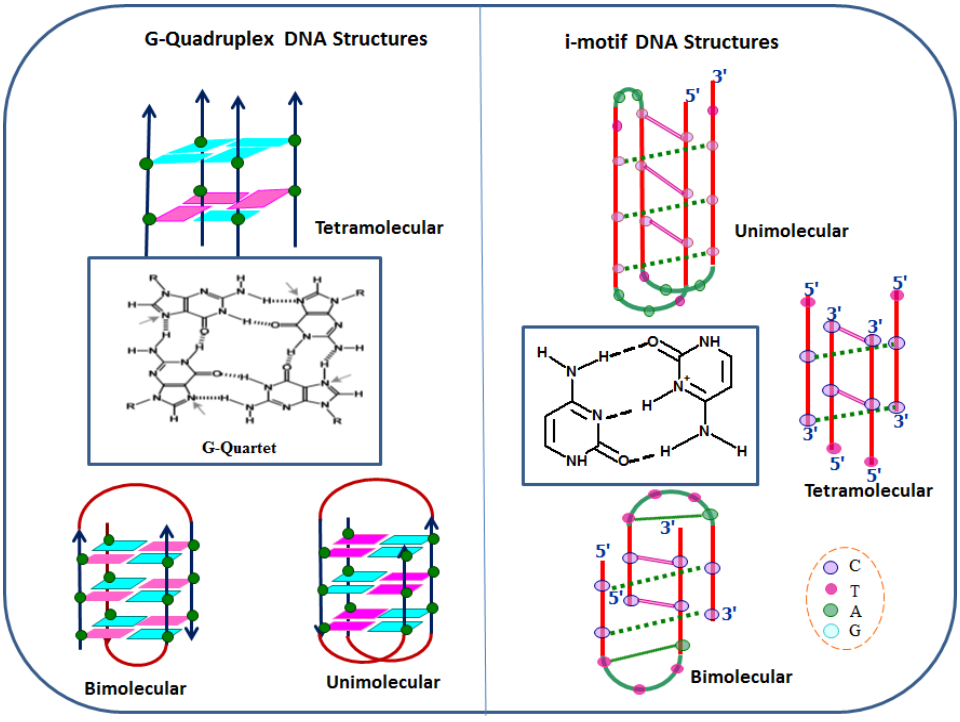

Figure 5: G-Quadruplex and i-motif DNA structures

7.BIOLOGICAL SIGNIFICANCE

Recent research has revealed that 98% of DNA present in the human genome consisted of non-coding regions and can be transcribed to regulatory non-coding RNAs, introns, centromeres, repeat sequences and telomeres. Non-coding regions are generally considered as "junk DNA" but the sequences present in this region do have biological significance. These non-coding regions may consists repetitive sequences andthese repetitive DNA sequences have the tendency to form non-B-DNA structures under certain experimental conditions**.** These alternative forms of DNA play an important role in regulation of replication and transcription, RNA splicing**,** chromosome folding, genetic recombination, mutational processes (Rhodes and Hans, 2015).

Sequences capable of forming triplexes are located in non-coding region i.e. introns and are involved in processes like recombination, regulation of gene-expression and many other DNA metabolism events in different ways (Jain et al., 2008). The intramolecular DNA triplexes (H-DNA) are significantly important because many sequences present in the human genome are capable of forming such structures. Triplex forming oligonucleotides act as promising drugs which inhibit the process of transcription and this process is known as anti-gene strategy.

The occurrence of G-rich and C-rich sequences *in-vivo* has always been the subject of debate, but some recent studies have addressed their existence and possible biological roles. It has been reported that 43% of genes are capable of forming least one G-quadruplex structure (Kendrick and Hurley, 2010).Sequences forming G-quadruplex structures are significantly present at some specific locations in the human genome including chromosome 19, centromeres and gene promoters and introns (Paulo and Francisco, 2016). Presence of such repeats at different locations suggests its crucial biological role in the treatment of diseases other than cancer (Simone et al, 2015).The various topologies adopted by G-rich sequences fascinated the galaxy of scientists in almost all fields like structural biology, biotechnology, molecular biology andnanotechnology.

The formation of i-motif structure is pH-dependent, so cytosine-rich DNA sequences have been used extensively in the field of

nanotechnology. This pH dependent property can be exploited to fabricate nanoswitches, nanomachines, electrodes and nanomaterials. Recent studies have suggested the use of cytosine-rich sequences for analytical and biomedical purposes which includes development of biosensors and drug-delivery biomaterials (Dembska, 2016).

8. FUTURE PERSPECTIVE

For years after the discovery of double helical structure by Watson and Crick, DNA was regarded as a rigid and uniform molecule with the sole purpose to store genetic information. Further advancement in technology and, experimental data has unraveled the full dynamic repertoire of this biomolecule. There are numerous non-B-DNA structures existing in the carnival of DNA secondary structures like parallel-stranded DNA, triplexes, G- quadruplexes, cruciform, i-motifs, hairpin and many more. Understanding the mechanism involved in various biological processes involving such structures and designing the drug which can target the specificstructure could help us to develop proficient strategies to combat lethal disease at the genomic level.

9. REFERENCES

1. Amato, Jussara, et al. "Noncanonical DNA Secondary Structures as Drug Targets: the Prospect of the i-Motif." *ChemMedChem* 9.9 (2014): 2026-2030.
2. Beal, Peter A., and Peter B. Dervan. "Second structural motif for recognition of DNA by oligonucleotide-directed triple-helix formation." *Science* 251.4999 (1991): 1360-1363. *Biochemistry* 28.24 (1989): 9528-9535.
3. Cerofolini, Linda, et al. "G-triplex structure and formation propensity." *Nucleic acids research* 42.21 (2014): 13393-13404.
4. Chen, Liqing, et al. "Crystal structure of a four-stranded intercalated DNA: (C4)." *Biochemistry* 33.46 (1994): 13540-13546.
5. Day, Henry A., Pavlos Pavlou, and Zoë AE Waller. "i-Motif DNA: Structure, stability and targeting with ligands." *Bioorganic & medicinal chemistry* 22.16 (2014): 4407- 4418.
6. Dembska, Anna. "The analytical and biomedical potential of cytosine-richoligonucleotides: A review." *Analytica chimica acta* 930 (2016): 1-12.
7. Durland, Ross H., et al. "Binding of triple helix forming oligonucleotides to sites in gene promoters." *Biochemistry* 30.38 (1991): 9246-9255.
8. Felsenfeld, G., David R. Davies, and Alexander Rich. "Formation of a three-stranded polynucleotide molecule." *Journal of the American Chemical Society* 79.8 (1957): 2023-2024.
9. Gellert, Martin, Marie N. Lipsett, and David R. Davies. "Helix formation by guanylic acid." *Proceedings of the National Academy of Sciences* 48.12 (1962): 2013-2018.
10. Jain, Aklank, Guliang Wang, and Karen M. Vasquez. "DNA triple helices: biological consequences and therapeutic potential." *Biochimie* 90.8 (2008): 1117-1130.
11. Kaushik, Shikha, et al. "Presence of divalent cation is not mandatory for the formation of intramolecular purine-motif triplex containing human c-jun protooncogene target." *Biochemistry* 50.19 (2011): 4132-4142.
12. Kendrick, Samantha, and Laurence H. Hurley. "The role of G-quadruplex/i-motif secondary structures as cis-acting regulatory elements." *Pure and Applied Chemistry* 82.8 (2010): 1609-1621.
13. Langridge, Robert, and Alexander Rich. "Molecular structure of helical polycytidylic acid." *Nature* 198 (1963): 725-728.
14. Moser, Heinz E., and Peter B. Dervan. "Sequence-specific cleavage of double helical DNA by triple helix formation." *Science* 238.4827 (1987): 645-650.
15. Paulo, A., and A. P. Francisco. "Oncogene Expression Modulation in Cancer Cell Lines by DNA G-Quadruplex-Interactive Small Molecules." *Current medicinal chemistry* (2016).
16. Raghunathan, G., H. Todd Miles, and V. Sasisekharan. "Parallel nucleic acid helices with Hoogsteen base pairing: symmetry and structure." *Biopolymers*34.12 (1994): 1573-1581.
17. Sharma, Sudha. "Non-B DNA secondary structures and their resolution by RecQ helicases." *Journal of nucleic acids* 2011 (2011).
18. Simone, Roberto, et al. "G-quadruplexes: Emerging roles in neurodegenerative diseases and the non-coding transcriptome." *FEBS letters* 589.14 (2015): 1653-1668.
19. Watson, J. D. and Crick, F. H. C. "A structure for deoxyribose nucleic acid"*Nature* 171.4356 (1953): 737–738
20. Yang, Danzhou, and Keika Okamoto. "Structural insights into G-quadruplexes: towards new anticancer drugs." *Future medicinal chemistry* 2.4 (2010): 619-646