CHAPTER 1

INTRODUCTION
1.1 G-quadruplex (G4 DNA) Motif

Apart from the right-handed B-DNA double helix, DNA is known to form other structural motifs like Z-DNA (reviewed in (1,2)), triple- helix (H-DNA) (reviewed in (3,4)), Holliday-junction (reviewed in (5)), hairpin (reviewed in (6,7)), etc. Similarly, repetitive guanine-tract oligonucleotide sequences can foldback to form a non B-DNA conformation known as ‘G-quadruplex’ or ‘G4-DNA' motif (reviewed in (8,9,10)). These potential structure forming sequences are known to form at telomeres (reviewed in (11)) and in the promoters of many important genes (12,13,14). The presence of these motifs in the regions of biological significance has opened new avenues for its targeting and modulation as an anti-telomerase and oncogene-specific drug development. Additionally, recent advances in nanotechnology as G-quadruplex mediated nano-motors and nano-switches (reviewed in (15)) is also gaining more attention than ever.

1.1.1 G-quadruplex DNA Conformation(s)

G-quadruplex motifs are a degenerate set of structural conformations formed by two or more than two ‘guanine-tetrad’, also known as ‘G-quartet’ motifs stacked one over the other consecutively and further stabilized by in-between sandwiched monovalent cations. One of the most intriguing aspects of G-quadruplex conformations is their extensive polymorphism, which is due to the matrix combinations of the two contributing factors, (i) the loop-connectivity, viz., tetramolecular, bimolecular or unimolecular and (ii) the orientation of the strands, viz., parallel, anti-parallel or mixed.

a. ‘G-quartet’ Motif

The cyclo- tetrameric spatial arrangement of four guanine nucleotides was first noticed in 1962 using X-ray diffraction data (16) which could satisfactorily explain the ‘viscous’ nature for millimolar guanosine- 5’-monophosphate solutions prepared in water. The tetrameric structure is a square co-planar array of four guanine bases in which each base acts both as the donor and acceptor of two
hydrogen bonds simultaneously with its two neighboring guanine bases, was termed as the ‘guanine-tetrad’ or simply ‘G-quartet’ (Figure 1.1). Additional biophysical studies with more than 60 different guanine derivatives with either ribose or deoxyribose sugar moieties have confirmed the tetrameric coplanar cyclic reverse Hoogsteen hydrogen bonding of the guanine tetrad arrangement (17).

![Figure 1.1 A ‘G-quartet’ motif and its hydrogen bonding scheme in spacefill (A), skeletal (B) with atom details (C)](image)

Similarly, solutions of guanine- homopolymers in millimolar range prepared in water were also found to form a highly viscous gel and interactions with the complementary cytosine- homopolymers were poor (18). Later on, using circular dichroism (19), infrared spectroscopy (20) and X-ray fiber diffraction studies (21,22,23) of guanine-homopolymers, the four-stranded G-quadruplex conformation with stacked ‘G-tetrads’ was confirmed.

1.1.2 G-quadruplex DNA Structural Polymorphism

a. Strand Stoichiometry

G-quadruplex motif structural polymorphism can arise due to multiple factors. One of the attributes is the strand stoichiometry which can lead to unimolecular (24,25,26) (Figure 1.2A), bimolecular [which can be further differentiated into two
subclasses—two sequences contribute equally two strands in the G-quadruplex motif (27,28) (Figure 1.2B) or one sequence contributing three strands while the other sequence contributing the one left (29) (Figure 1.2C)], or tetramolecular (30,31) (Figure 1.2D). It would be interesting to mention that the strands having the ability to form an unimolecular G-quadruplex, theoretically can also form bimolecular or tetramolecular structures (32,33) and is mostly governed by the strand concentration in solution.

![Figure 1.2](image-url)

**Figure 1.2** Various G-quadruplex topologies: (A) unimolecular – all four strands from a single oligonucleotide sequence, (B) bimolecular – two strands each from two different oligonucleotide sequences, (C) bimolecular – three strands from one oligonucleotide sequence and fourth one from another oligonucleotide sequence and (D) tetramolecular – all four strand from four independent oligonucleotide sequences.

**b. Strand Polarity**

The relative 5’-3’ strand polarity arrangement of the four strands in the G-quadruplex ‘stem’ can retrospectively generate structural variations and depending upon the different strand polarities, the resultant G-quadruplex can be all-strand parallel (Figure 1.3A) as evidenced by NMR (34,35,26,36) and crystallography (31,37), three-strand parallel and one antiparallel (Figure 1.3B) as found in *Tetrahymena* (38) and *Oxytricha* (26) telomeric sequences using NMR, adjacent parallel (39,40,41) (Figure 1.3C), or alternating antiparallel (25,42,43,44,45) (Figure 1.3D).
c. Loop Connectivity

G-quadruplex motif ‘loops’ (usually non-guanosine bases) can show polymorphism in their connectivity. In a bimolecular G-quadruplex motif, the loops that would connect guanine strands can be connected either diagonally or edgewise. Although bimolecular diagonal loop connectivity cris-crossing on the same side is theoretically possible, their formation is highly unlikely due to both steric hindrance and electrostatic repulsion between the two negatively charged backbones. Hence, diagonal loops on opposite sides of the G-quadruplex motif is usually evidenced (45) (Figure 1.4A).

Figure 1.3 Various G-quadruplex strand polarities (arrowheads indicating 5’-3’ strand polarity): (A) All-parallel (B) Three parallel - one anti-parallel (C) two adjacent parallel and (D) two diagonally parallel strand combinations.

Figure 1.4 Various bimolecular loop connectivities (arrowheads indicating 5’-3’ strand polarity): (A) diagonally anti-parallel cross-over, (B) edgewise parallel, (C) diagonally parallel, (D) diagonally anti-parallel and (E) edgewise anti-parallel.
Alternatively, the edgewise loop connectivity taken together with loop polarity can be either parallel (Figure 1.4B) or antiparallel (Figure 1.4C), in case the loops are on the same side of the motif, otherwise, two more configurations are possible (Figures 1.4D and 1.4E).

Moreover, bimolecular G-quadruplex can also arise due to one guanine-rich sequence contributing three strands of the ‘stem’ and the fourth strand coming from another guanine-rich tract (29). This would lead to two loops generated from the three strand-contributing guanine-rich sequence in both edgewise or edge-and-diagonal-wise loop connectivity and depending upon the strand orientation of the fourth stand, a total of four distinct G-quadruplex conformations can result (Figures 1.5A, 1.5B, 1.5C and 1.5D).

Unimolecular G-quadruplexes are expected to have relatively fewer loop-connectivity possibilities in order to avoid the steric and negative charge-charge repulsion of two diagonal loops on the same side. The three loops can adopt the order edge-edge-edge (Figure 1.6A) or edge-diagonal-edge (Figure 1.6B).
d. Loop Size

The DNA sequence required for the formation of intrastrand fold-back G-quadruplex motifs can be argued as $G_nN_{L1}G_nN_{L2}G_nN_{L3}G_n$ where G is guanine and N is any nucleotide including G. The number of guanines constituting the stem of G-quadruplex motif is given by n, which can be two or more than two, while L1, L2 and L3 indicate the number of nucleotides within loops which can vary from 1-7 each. If all the combinations are allowed, there would be 349 (7x7x7) combinations possible for each stem size and would get further complicated with the nucleotide composition, connectivity and orientation. Though longer loop lengths are possible, entropically it would not be favoured.

e. Role of Cations

One of the intriguing characteristics of G-quadruplex motifs is their selective interaction with certain cations that sandwich well in between the stacked G-tetrads. The cavity between two planes of guanine tetrads generates eight carbonyl oxygen-6 atoms which can participate to fulfill the cationic coordination number, given that the cation has appropriate size to fit in between. The ability of potassium ion ($K^+$, ionic radii of approximately 1.3 Å) to fit well in the cavities between stacked G-tetrads was found first (46), which later found to be relatively higher than other cations in the order $K^+ >> Na^+ > Rb^+ > NH_4^+ > Cs^+ >> Li^+$ (17).
1.1.2 Structure Determination of G-quadruplex Motif

a. Nuclear Magnetic Resonance

The imino proton chemical shifts (10.5–12 ppm) are characteristic Hoogsteen hydrogen bonding signals and indicate for G-quartet arrangement, while selective $^{13}$C and/or $^{15}$N labeling is usually done to further correlate intranucleotide imino protons and base protons providing finer structure coordinate information. First NMR resolved G-quadruplex conformation was reported as early as 1987 (24), followed by first RNA quadruplex in solution state in 1992 (47), first solution structures of human, Tetrahymena and Oxytricha telomeric repeat sequences in 1990 (48), 1994 (38) and 1995 (26) respectively, while first $^{23}$Na, $^{39}$K and $^{17}$O solid state NMR in 2002 (49), 2003 (50) and 2007 (51) respectively.

b. X-ray Crystallography

G-quadruplex structure determination using X-ray crystallography is more quantitative than other biophysical methods. In most cases, the crystals were prepared by vapor diffusion using hanging- drops method. Some recent reports for G-quadruplex structure determination using X-ray crystallography used molecular replacements in guanine stretches (37,42). As a note of caution the crystal-resolved structure may not be the most stable or dominant species in solution or at times the conformation may get distorted during crystallization. For example, two different structural conformations were found using solution NMR and crystal structure for the same human telomeric repeat sequence (52).

c. Circular Dichroism Spectroscopy

Unlike NMR spectroscopy and X-ray crystallography, Circular Dichroism (CD) spectroscopy does not provide finer structural details, however it is a fast and qualitative method to test for G-quadruplex motif formation. Moreover, low sample requirement and easy preparation are added advantage. The characteristic CD signature for G-quadruplex motif can distinguish between parallel (positive peak at 260 nm and a negative trough at 240 nm) and anti-parallel (positive peak at 295 nm
and a negative trough at 260 nm) motifs (33,53). Although the overall sensitivity is low, CD spectroscopy helps in initial screening and hence is one of the widely used methods to characterize G-quadruplex motifs.

1.1.3 Kinetics of G-quadruplex formation/deformation

The functionality of G-quadruplex motifs depends much on their existence which in turn is a function of its inherent stability. To elucidate this aspect, the folding and unfolding kinetics of G-quadruplexes have been studied for quite some time (54,55), but only recently more quantitative data have been obtained for various sequences using different techniques, viz., Fluorescence Resonance Energy Transfer (FRET, (56)), Surface Plasmon Resonance (57), UV spectroscopy (58), CD spectroscopy (59) and NMR spectroscopy (60).

Notably, regardless of the techniques and oligonucleotide sequences, the general conclusion of the above studies indicated very slow folding-unfolding rate constants (ranging from few seconds to hours), which implied that a G-quadruplex motif would be stable and exist for significant time, once allowed to form. However, multiple other factors like the stem size of the motif (number of stackes of G-tetrads), cation concentration and loop sizes (in case of intramolecular G-quadruplexes) can also influence the folding-unfolding equilibrium.

1.2 Biological Existence and Significance

1.2.1 Genomic Distribution

Recently, the mosaic distribution of G-quadruplex motifs in human (61,62), chicken (63) and other higher eukaryotic (64) genomes has been determined computationally. But initially, it was the individual gene promoter G-quadruplex studies which helped in understanding the functional role of these motifs. In-vitro and indirect studies implicated mostly a transcriptional-suppressive role for these motifs. Apart from the gene promoters, the G-quadruplex motif with telomeric repeats has been shown in-vitro using crystal structures, CD spectra, FRET probes, etc
(discussed above), while the recent in-vivo G-quadruplex specific antibody staining (65) and small-molecule staining (66,67) have added to its intracellular existence.

1.2.2 G-quadruplex Interacting Proteins

Of all the possible evidence in support of in-vivo G-quadruplex subsistence, the most convincing impression comes from the G-quadruplex specific interacting proteins. The ever-increasing list of G-quadruplex interacting proteins range from novel proteins (68,69) to transcription factors (70,71), helicases (72), nucleases (73,74) and ribonucleases (75) (reviewed in (76)). Apart from the above set of proteins, additional alluring evidence came from the in-vitro generated G-quadruplex specific antibodies stained to the telomeric regions of Stylonychia lamnae macronuclei (77) and later shown that the telomere-end-binding proteins (TEBPα and TEBPβ) cooperate in formation of G-quadruplex motifs, in-vivo (65). A detailed list of G-quadruplex interacting proteins and their sub-classification can be found elsewhere (76).

1.2.3 G-quadruplex Interacting Small Molecules

Presence of G-quadruplex motif in telomere and promoter regions of certain proto-oncogenes implied for certain functional role in aging and cancer, and hence, G-quadruplex specific molecules evolved as an interesting research field. Nevertheless these molecules would further ascertain the existence of G-quadruplex motif in-vitro and may be extended to in-vivo studies in future. The first crystal structure solved G-quadruplex-daunomycin complex indicated groove-binding with hydrogen-bonding and/or van-der-waals interactions (78), while a NMR solution structure showed telomeric G-quadruplex as end-stacked by fluorinated polycyclic acridinium salt molecules (79). Noteworthy to mention, the c-MYC promoter G-quadruplex motif complexed with a cationic porphyrin (TmPyP4) indicated a 5’-end stacking and electrostatic interactions contributing in the complex stability (80). In future, selective intervention and/or promotion of G-quadruplex motif in regions of interest (promoter, telomere, etc.) using conformation specific ligands would have therapeutic values.