Summary

Genomic DNA is usually present as a canonical right-handed double helix, but it also has potential to form other structures. The ability for DNA to form structures containing more than two strands has been renewed for the potential use of nucleic acids as therapeutics and the specific functional roles played by these alternative structures in vivo. The ability of DNA to adopt a variety of conformations and the sequence specificity arising from complementary base pairing make it a potentially useful starting material for constructing two- and three-dimensional structures of controlled geometry such as DNA combs, knots and oligonucleotide dendrimers. Structural polymorphism and dynamism in a DNA is of paramount importance for genetic expression and regulation. Usually DNA must be a duplex with complementary base pairing, multiple folding and protein interactions can be conceived to cause separation of the strands either transiently during processes of replication, transcription and translation or permanently in specific regions and then these strands may associate in a variety of ways to form different type of structures such as G-quadruplex and i-motif.

Self-assembling multistranded complexes of DNA oligonucleotides constitute a special class of DNA superstructures. Nucleic acid sequences rich in guanines (G) in particular are predisposed to form higher-order structures because of the capacity of guanine bases to self-associate via Hoogsteen hydrogen bonds to form planar G-quartets. Several G-rich DNA sequences identified in biologically important regions of the genome (telomeric and non-telomeric locations). G-quadruplex forming sequences have been identified in the gene promoters. Evidences of G-quadruplex formation in the regulatory sequences of muscle specific genes has also been reported. Genome-wide bioinformatic searches have revealed a significant number (~376000) of putative quadruplex-forming sequences in the genome, with > 40% of human gene promoters enriched in G-rich sequences. Potential for DNA quadruplex formation is correlated with gene function. Recent, interest in G-quadruplex and i-motif DNA structures is accelerating because of their prevalent occurrence in the human genome and their possible biological relevance. Molecular recognition of these structures
adopted by DNA by small molecules and protein is a fundamental problem in structural biology and drug design. This study shows the formation of G-quadruplexes by G5-tracts at genomic location (promoter region of Myosin Heavy Chain β Gene). For a detailed analysis of G-quadruplexes and the associated polymorphism, sequence of varied length containing G-tracts were designed. The C-counterparts of such G-rich sequences were studied for i-motif (C-tetraplexes). Their structural status and stabilities under varied solution conditions were investigated by various biochemical and biophysical techniques like Gel electrophoresis, UV-Visible absorption spectroscopy, UV-thermal denaturation and circular dichroism spectroscopy, Fluorescence Spectroscopy. For any biological relevance efforts were made towards stabilization of G-quadruplexes in solution using ligands and a designed peptide. The thesis comprises of seven chapters.

Chapter 1 includes a brief account of the structure of G-quadruplex and i-motif adopted by various G-rich and C-rich sequences and their existence at various genomic locations. This chapter also includes the techniques used to study the G-quadruplex structures and their application in biological processes. A brief account for their existence and biological significance is also discussed. A brief account of Myosin gene (MYH7) harbour G5-tract and importance of gene has also been discussed. An uptodate literature on G-quadruplex and their types and ligand interactions is included here.

Chapter 2 includes the details of oligonucleotide sequences, ligands, peptide and chemicals used for the study. This chapter also deals with the methodology adopted for the study. Bioinformatics tools were used to access and analyze the G-rich sequence at genomic locations. General principles of the techniques like UV-thermal denaturation, Circular Dichroism, gel electrophoresis and Fluorescence Spectroscopy have been described alongwith the details of the instruments used.

Chapter 3 deals with Bioinformatic studies on the DNA sequences of MYH7 gene. BLAST is one of the most widely used bioinformatic tool used for the searching sequence in genome. ClustalW2 software was used for the multiple sequence alignment of MYH7 gene. Multiple Sequence Alignment (MSA) is often used in
identifying conserved sequence regions across a group of sequences hypothesized to be evolutionary related. *In silico* studies showed that sequences in promoter region of MYH7 gene from different organisms (Human, Chimpanzee, Rabbit, Mouse and Bos) was a conserved region. It also includes the construction of phylogenetic tree which tells that evolutionarily which organism is close to another organism.

Chapter 4 deals with the biochemical and biophysical characterization of the 23-mer and 34-mer G-rich DNA segments named as HM23, HM34, RM23 RM34 containing unique G-tracts belonging to the promoter of Myosin Heavy Chain β Gene of Human and Rabbit. The Rabbit myosin gene (RM23, RM34) differs from the Human counterpart at nucleotide position number 8, 10 and 23. The techniques used for the study of these sequences are Gel electrophoresis, UV-Thermal Denaturation and Circular Dichroism. It was found that HM23 and RM23 adopt novel trimeric (three stranded) G-quadruplex structures. Mutations were introduced in the HM23 to investigate the involvement of G5 in this unique G-quadruplex formation. The induction of mutation proved that the G5 tract is involved in the three stranded structure formation. The hypothetical models have been proposed to explain the experimental results.

Sub-section of Chapter 4 includes the studies on the formation of G-quadruplex by the various designed sequences having G5-tract with the intervening Thymine bases. Studies were carried out in presence of Na\(^+\) as well as K\(^+\). It was concluded that all the G-quadruplexes formed by short as well as longer verison are more thermally stable in K\(^+\) than in Na\(^+\). This study shows that G-rich sequences are highly polymorphic and adopt structure of varied molecularity.

Chapter 5 presents the studies conducted to investigate the formation of C-tetraplexes (i-motif) by the counterparts of G-rich sequences of MYH7 gene as well as G-tracts with intervening bases thymine. This study shows that all the C-rich counterparts adopt i-motif structures at acidic pH. Interestingly, though acidic pH condition is a prerequisite for i-motif formation, the sequence RM34C was shown to exist as i-motif as well as in homoduplex state at neutral pH 7.0. It was shown that at low temperature (4-15 °C) it adopts i-motif structure whereas at 25 °C it transforms to
homoduplex. C5-tract with intervening Adenine bases were investigated to see the i-motif formation. Thermal denaturation study revealed that thermal stabilization increases with A1, A2 and A3 in loops but presence of A4 destabilizes the i-motif structure.

Chapter 6 Stabilization of G-quadruplex is particularly important aspect for any biological application. This chapter covers the interaction studies of G-quadruplex binding ligands with parallel, mixed and antiparallel G-quadruplex having G5-tract. The studied ligands include TMPYP4 and Sanguinarine Chloride. Thermal denaturation, CD and Fluorescence studies showed that TMPYP4 specifically stabilizes parallel G-quadruplex having G5-tract. Sanguinarine recognizes and stabilizes the antiparallel G-quadruplexes and also facilitate the formation of antiparallel G-quadruplex. Interestingly, this study also showed that Sanguinarine specifically bind strongly to the intermolecular antiparallel than intramolecular antiparallel G-quadruplexes. Recognition of G-quadruplexes was also studied by a11-mer designed peptide and it was observed that the peptide specifically destabilizes the parallel G-quadruplexes.

Chapter 7 concludes the overall findings presented in the study. This study is an attempt to understand and gaining information on the formation of G-quadruplexes and C-tetraplexes formed at genomic locations as well as some designed sequences. It also covered the mode of interactions of some ligand sand peptide, binding specifically to these structures. This study adds on the knowledge and understanding the G-quadruplex polymorphism and interest in designing and development of new molecules with improved binding affinity and selectivity for G-quadruplexes over current therapeutics.