Since the discovery of DNA as a double helix molecule, many other non B-DNA forms of DNA are being uncovered. The investigators are interested in other forms of DNA which could potentially be formed in certain specialized regions of DNA. Various regions of DNA which are rich in purines (guanines) or pyrimidines (cytosine) can adopt folded four stranded structures called as G-quadruplex and i-motif respectively. Such sequences are known to be present at the telomeric ends of chromosomes, gene promoter regions, recombination sites, RNA dimerization domains and RNA packaging sites. Among all those, four stranded cytosine rich (tetraplex or i-motif) or guanine rich motif (Quadruplex or G4 DNA) have gained a lot of attention. The four stranded structure of these motifs is stabilized by a special kind of hydrogen bonding called hoogsteen bonding between two hemi-protonated cytosines, in case of tetraplex and between two guanines, in case of quadruplex structures. Tetraplex motif existence, inside the cells, has not been shown but there are in vitro evidences which support the possibility of their formation. Biological importance of tetraplex motifs has been shown in case of progressive myoclonus epilepsy, where multiple dodecamer cytosine repeats were shown to be involved in the disease. Also tetraplex formation is most favored in acidic conditions thus their existence inside cells is questionable. But certainly the complementary strand of potential i-motif forming sequence has the possibility to form a four stranded guanine rich quadruplex motifs. Various studies have implicated the role of quadruplex motifs in several biological phenomenons like transcription and recombination. Some of the recent evidences indicate that several oncogenes like c-MYC, c-KIT, KRAS and VEGF have G4 DNA motifs as potential regulatory signals in their promoters. For a long time the relevance of G4 DNA as a biologically important structure was doubted because its existence inside the cells was not shown. But later many studies have been published which support the fact that these structures indeed have a role in several biological processes. The discovery of various proteins, Mre11, MyoD, and nucleolin which binds specifically to quadruplex structures indicates the relevance of existence of quadruplex inside the cells. Though there are vast numbers of studies being done but the direct evidence of G-quadruplex as a ‘transcription regulatory signal’ is still lacking. The present work is done to understand the role of DNA-tetraplex in the transcriptional regulation of genes at genome wide level. The work is primarily
Abstract

divided into two parts: whole genome expression studies and 2) molecular level studies. In whole genome expression studies, the regulatory significance of quadruplex motifs were analyzed at whole genome scale employing microarrays and *in silico* approaches while in case of molecular level studies, certain genes were selected and their promoters were further analyzed using *in silico* tools and molecular techniques like cloning and expression assays. This thesis presents the results of the experiments performed, the inferences drawn along with the pertinent review of literature in the order described below:

➢ **GENOME WIDE EXPRESSION STUDIES**

Based on already known facts about quadruplex elements being involved in transcription regulation of many oncogenes, the present work was done to study this mechanism at a global level. We tested the association of quadruplex motifs with transcription process at the whole genome level. A widely accepted quadruplex binding molecule, a cationic porphyrin - TMPyP4, was employed to probe these motifs. Herein, we demonstrate the widespread role of quadruplex motifs in gene regulation. Presence of potential quadruplex motifs (PG4) was observed in the promoters of most of the differentially expressed genes, after TMPyP4 treatment. Expression studies done in HeLa S3 cells indicate that significantly differentially expressed genes (67/93 at 24 hrs and 46/69 at 48 hrs) harbor one or more quadruplex motifs near transcription start site (TSS) while expression profile of A549 and HeLa S3 cells was compared together, it was found that 711/1161 genes have PG4 motifs in their promoters. TMPyP4 besides as a telomerase inhibitor is also a well studied anti-cancer drug. To test this, the expression of down regulated genes in microarrays was compared to gene expression profiles in different cancer tissues, which showed that most of the down regulated genes in microarrays were up regulated in several cancer tissues. Gene ontology (GO) analysis indicated the important biological classes to which these genes belong, which are: transcription activator activity, intracellular transport, cellular localization and protein binding.

➢ **MOLECULAR LEVEL STUDIES**

To test the above finding at single gene level several genes were cloned in luciferase reporter plasmids. Few of the selected genes which harbor quadruplex motifs in their promoter regions are: Thymidine kinase1 (*TK1*), Recombination activating gene1
Abstract

(RAG1) and c-MYC, whose expression was potentially checked for the existence of PG4 motifs in physiological conditions. It was observed that in presence of TMPyP4 the luciferase expression, driven by the promoters of these genes, decreases (30% and 80% reduction at 10μM and 100μM TMPyP4 respectively) indicating the suppressive effect of potential quadruplex motif. To study the molecular role of these motifs in transcription, interaction of a well known tumor suppressor protein, nucleoside diphosphate kinase 2 (NM23-H2), with the promoters of these genes was studied. The results indicated transactivation of genes in presence of NM23-H2 followed by abrogation of transactivation potential in presence of TMPyP4, indicating interaction of protein with a folded DNA structure. Previous studies indicated that the expression of many oncogenes like c-MYC, c-KIT and VEGF, were regulated by quadruplex motifs. To further explore the potential of quadruplex motifs in cancer related genes, we selected 62 cancer related genes for this part of study. Using bioinformatics tools, the proximal promoters (-1kb to +100 bp) of 62 genes were searched for potential quadruplex motifs. Interestingly, it was found that most of the genes have such motifs, maximum density being near the transcription start site of the genes. 23 genes out of 62 genes also showed orthologous conservation of quadruplex motifs across human, mouse and rat. Transcription regulatory potential of these motifs was also showed by significant association of certain transcription factors with the potential quadruplex forming sequences. However, this was not specific to cancer related genes because a similar association was observed in case of non-cancer related genes also.

In addition to quadruplex structures, cytosine rich sequences can adopt tetraplex structures or i-motif which are relatively less studied owing to the acidic conditions required for their formation. This questions the existence of i-motif inside the cells. It was envisaged that the complementary strand (which may fold into a tetraplex motif) to guanine rich strand can potentially help in remodeling of a quadruplex structure on the other strand. To study the thermodynamic parameters of tetraplex formation this part of the study was designed. We have taken a 31 mer cytosine rich sequence of c-MYC promoter to carry out this study as this sequence was already shown to form a tetraplex structure previously but in acidic conditions. The results of this study indicate the presence of tetraplex structure at near physiological conditions (pH 7.0) in addition to favorable kinetic parameters for the formation of tetraplex motifs. In another study it was proposed that a bacterial homolog of human NM23-H2, mNDK
in *Mycobacterium Tuberculosis* (*Mtb*), shows nuclear localization. In earlier part of this study, it was demonstrated that human NM23-H2 may principally bind to quadruplex structures to transactivate the expression of several genes. To study the biological importance of these unusual structures mNDK was employed. Interaction of mNDK with 31 mer cytosine rich sequence showed that this protein binds to i-motif structures and cleaves it while the three mutants (which disrupts the tetraplex structure) did not showed similar cleavage profiles at pH 7.0. This supports the previous notion of nuclear localization of mNDK as this protein enters nucleus and may chop DNA at certain specific sites (potential tetraplex/quadruplex forming sequences) and thus contributes to *Mtb* infection. Thus this study presents results, which together, indicates the role of quadruplex/tetraplex motifs as cis-regulatory elements in transcription process.