STATEMENT OF THE PROBLEM
IV. STATEMENT OF THE PROBLEM

Mammalian genome contains a large amount of repetitive DNA. In the post-genomic era, although the genes and their regulatory sequences are known to large extents, information about the cellular ‘RNA world’ is continuously expanding. In addition to the known protein-coding and RNA-coding genes/sequences, many other chromosomal regions are actively transcribing into RNAs. This also includes a variety of repetitive DNAs. Repeat sequences at RNA level can attain greater complexity of secondary structures for intermolecular interactions. Therefore, repeat sequences containing transcripts may provide new information about cellular functions. Earlier in our laboratory, a 227 bp novel rat genomic simple repeat DNA (X97459) was isolated, which contains a (GA)7A9AG)7 mirror repeat and CAG, GAAA, CA, GGG, CATACA, AGAGAA simple repeat sequences. It showed the property of forming a triplex-like (H-DNA) structure in vitro and homology with a variety of eukaryotic mRNAs mostly due to presence of (GA)n, (CU)n repeats in the 5'-untranslated (5'-UTR), 3'-UTR and amino acid coding regions of the mRNAs. The simple repeat DNA detected large cellular RNAs of 9.1 and 7.5 kb in the rat brain indicating transcription of large RNAs from the genomic repeats of rat. The simple repeat DNA was used to isolate several cDNAs from a rat and human testis cDNA library to study candidate cDNAs representing repeat sequence containing cellular RNAs.

With this background, the present study entitled: “Expression and functional characterization of candidate-cDNAs isolated by a simple repeat DNA probe” was designed to study RNA-expression and functional characterization of a few cDNAs. For this study, five cDNAs were selected: two human cDNAs, the TPIP-C2 cDNA and the H3.SRY cDNA, which represent protein-coding mRNAs; three rat cDNAs, 4.1-Diphospho inositol polyphosphatase phosphohydrolase (4.1-DIPP) cDNA, which represents protein-coding mRNA and the 5.5 cDNA, the 11.4 cDNA, both represent repeat-rich noncoding RNAs generating small regulatory RNAs. Experiments involving Bioinformatics, RNA expression, in vitro characterization of the recombinant protein, transfection of expression plasmids into human cells and analysis of small regulatory RNAs were carried out in mouse, rat tissues and human cells to characterize the cDNAs. The following questions were investigated:

1. What are the sequence-properties of the five candidate-cDNAs?
2. If the cDNAs code for any ORPs (Open reading frames) and how the repeat sequences are related to the ORPs?
3. If the cDNAs represents any noncoding RNAs and how the repeat sequences are related to the noncoding RNAs?
4. What is the RNA expression pattern of these repeat sequences-associated cDNAs in mammalian tissues and cells?
5. If the human cDNAs: TPIP-C \textsubscript{2} and H\textsubscript{3}-SRY are also expressed in mouse tissues and what is their function?

To address the above questions following experiments were carried out.
1. Computational analysis of the five candidate-cDNAs were carried out for presence of repeat sequences, genomic localization, Open Reading Frames (ORFs), motifs, post-translational modifications and structures to predict possible function of respective cDNAs, the possibility of the noncoding cDNAs to make extensive RNA secondary structures, which can act as precursor for generation of small regulatory RNAs.

2. The RNA expression pattern was studied in different rat tissues by RT-PCR and northern blot hybridization for the rat 4.1, 5.5 and 1104 cDNAs. The 4.1 RNA was localized in rat tissues by \textit{in situ} hybridization. The 5.5 and 11.4 RNAs were shown to be precursors for generation of a number of miRNAs and piRNAs by northern blot hybridiation.

3. Expression of the human TPIP-C\textsubscript{2} RNA was studied by RT-PCR in different mouse tissues and transfected (human embryonic kidney) HEK-293 cells. The RT-PCR products from mouse tissues were cloned and sequenced to confirm its expression in mouse tissues. Mouse TPIP-C\textsubscript{2} RNA expression was also checked by northern blot hybridization.

4. The human TPIP-C\textsubscript{2} cDNA was over-expressed in HEK-293 cells. The expression was studied by RT-PCR and effect of over-expression of TPIP-C\textsubscript{2} RNA was studied by cell counting and staining of the co-transfected cells.

5. The human H\textsubscript{3}-SRY cDNA was characterized by computational prediction. The 241 aa ORF was expressed as a GST-fusion protein in \textit{E. coli} by GST-pulldown and electrophoretic gel mobility shift assay (EMSA) were carried out to find out interaction between GST-H\textsubscript{3}-SRY and His-SRY protein and its effect on DNA binding activity of His-SRY.