ABSTRACT

Genetic code is degenerate and a set of sixty-one sense codons encode only twenty standard amino acids found in protein sequences. Most of the amino acids are encoded by more than one codon with the exception of amino acid methionine and tryptophan. The number of alternative or synonymous codons varies from two to six, which mostly differs in their third position. The synonymous codons are distributed unevenly among genes within the genome and most genes from different organisms exhibit bias towards codon usage. Different organisms have different codon usage profiles, and codon usage often varies significantly within the genes of the same organism. The pattern of synonymous codon usage must reflect the combined influence of mutation, natural selection and random genetic drift. Investigations on codon usage have provided interesting insights into the basic aspects of cell biochemistry, genetics and evolutionary biology.

In this study, we have analyzed the nucleotide compositional constraints and estimated the expression level of the coding sequences of human proto-oncogenes/oncogenes and tumour suppressor genes in the cell using several genetic indices namely, the codon adaptation index (*CAI*), tRNA adaptation index (*tAI*), frequency of optimal codon (*Fop*), relative synonymous codon usage (*RSCU*), effective number of codons (*ENC*) and compositional dynamics of the background nucleotide constraints.

Our results showed that nearly all G/C-ending codons were positively correlated (significant p value<0.01) with GC3s indicating that codon usage in human proto-oncogenes/oncogenes and tumour suppressor genes has been influenced by GC bias, mainly due to GC3s. The majority of the frequently used codons were G/C ending in which C –ending codons were mostly favored compared to G –ending codons for the corresponding amino acid. In the neutrality plot analysis we compared the values of GC12 (average of GC1 and GC2) and GC3s, and observed a significant positive correlation (p<0.01) in the coding sequences of proto-oncogenes/oncogenes as well as tumour suppressor genes, which revealed that intragenomic GC mutation bias influences the codon usage patterns in the coding sequence of human proto-oncogenes/oncogenes as well as tumour suppressor gene, since its effects are present at all codon positions. However, the linear

regression coefficient of GC12 on GC3s in these genes suggested that natural selection played a major role while mutation pressure played a minor role in the codon usage patterns of proto-oncogenes/oncogenes as well as tumour suppressor genes in human. Relatively weak codon bias was observed in these genes as reflected by high *ENC* values which ranged from 35 to 60. Further, correspondence analysis (COA) based on *RSCU* values showed that the principal axis was significantly correlated with the four major indices of codon bias namely *ENC*, GC, GC3s and *CAI* which revealed that nucleotide composition and mutation bias might play a pivotal role in shaping the codon usage patterns of human proto-oncogenes/oncogenes and tumour suppressor genes. In addition, we observed that nature might have preferred the over-representation (highest *RSCU* value) of the codon CTG encoding leucine amino acid in the coding sequences of human proto-oncogenes/oncogenes and tumour suppressor genes. The frequently used optimal codon (*Fop*) values in the coding sequences of proto-oncogenes/oncogenes were similar to that of tumour suppressor genes, except the codons TTT, CCT and ACA encoding amino acids phenylalanine, proline and threonine respectively. It was evident that CpG dinucleotides were under-represented, whereas GpC dinucleotides were over-represented in both the cases of proto-oncogenes/oncogenes and tumour suppressor genes. However, nearly all the codons containing dinucleotide CpA and TpG were over-represented and most of them were also used as preferred codons for their corresponding amino acid based on *RSCU* analysis of the selected genes under study. Moreover, significant correlation was observed between compositional constraints (codon usage) and gene expression (measured by *CAI*) suggesting that expression level might play a pivotal role in the codon usage of human proto-oncogenes/oncogenes and tumour suppressor genes. In addition, a significant positive correlation (Pearson, $r=0.676$, $p<0.01$) was observed between tRNA adaptation index (*tAI*) and codon adaptation index (*CAI*) suggesting that the expression of proto-oncogenes/oncogenes and tumour suppressor genes (*CAI*) was remarkably influenced by the genomic tRNA pool. Our present findings certainly report a novel insight into the codon usage patterns in gaining the clues for the functional conservation of gene expression, translational studies and codon optimization for desired expression and the significance of the nucleotide composition in the coding sequences of human proto-oncogenes/oncogenes and tumour suppressor genes.