



Department of Biotechnology  
School of Life Sciences  
**ASSAM UNIVERSITY, SILCHAR**  
(A Central University constituted under Act XIII of 1989)  
Silchar-788011, Assam, India

## **DECLARATION**

I, **Tarikul Huda Mazumder** bearing Registration No- PhD/2514/14, Dated 12/03/2014, hereby declare that the subject matter of the thesis entitled “***In silico* Study of Codon Usage Bias in Human Proto-oncogenes, Oncogenes and Tumour Suppressor Genes**” is the record of work done by me and that the contents of this thesis did not form the basis for award of any degree to me or to anybody else to the best of my knowledge. The thesis has not been submitted to any other University/Institute.

**Place:** Assam University, Silchar

*Tarikul Huda Mazumder*

**Date:** 22-03-2016



## ACKNOWLEDGEMENTS

At the outset, I express my profound gratitude to my supervisor Professor Supriyo Chakraborty, Head, Department of Biotechnology, Assam University Silchar, Assam, India, for his constant support and experienced professional guidance throughout the course work that made possible to complete this study.

I would like to express my sincere gratitude to all the teachers in the Department of Biotechnology for their valuable suggestions and constructive criticism during the course work of this study.

I am especially thankful to all my lab members of the Computational Biology Laboratory, Department of Biotechnology, Assam University Silchar, for their help and support in compilation of the whole report of this study.

Finally, I am most indebted to my family members, brothers and sister for their blessings and encouragement and especially my parents for their love and support throughout my life.

*(Tarikul Huda Mazumder)*



# CONTENTS

<u>INDEX</u>	<u>Page No.</u>
List of Tables .....	I
List of Figures .....	II-III
<b>CHAPTER - 1 Introduction .....</b>	<b>1-16</b>
1.1. Standard genetic code and synonymous codons.....	1
1.2. Codon usage bias (CUB).....	2-4
1.3. Factors affecting codon usage bias.....	5-10
<i>1.3.1. Selection for optimized translation.....</i>	<i>5</i>
<i>1.3.2. Gene Expression.....</i>	<i>5</i>
<i>1.3.3. Location within genes.....</i>	<i>6</i>
<i>1.3.4. Rate of evolution.....</i>	<i>6</i>
<i>1.3.5. Secondary structure.....</i>	<i>7</i>
<i>1.3.6. Nucleotide composition.....</i>	<i>8</i>
<i>1.3.7. Protein Length.....</i>	<i>8</i>
<i>1.3.8. Environment.....</i>	<i>9</i>
<i>1.3.9. Time.....</i>	<i>10</i>
<i>1.3.10. Neutral alternatives.....</i>	<i>10</i>

1.4.	Cancer: proto-oncogenes/oncogenes.....	11-12
1.5.	Tumour suppressor genes.....	13-14
1.6.	The importance of codon usage bias study.....	15
1.7.	<b>Rationale of the study</b> .....	15
	<i>1.7.1. Objectives</i> .....	16
	<b>CHAPTER - 2 Review of literature</b> .....	17-32
	<b>CHAPTER - 3 Materials and methods</b> .....	33-44
3.1.	Sequence Data and Resources.....	35-37
3.2.	Analysis of synonymous codon usage bias .....	38-44
	3.2.1. <i>Nucleotide Composition Analysis</i> .....	38
	3.2.2. <i>Analysis of nucleotide skewness</i> .....	38
	3.2.3. <i>Effective Number of Codons (ENC) Analysis</i> .....	39
	3.2.4. <i>Frequency of Optimal Codon (Fop) Analysis</i> .....	40
	3.2.5. <i>Relative Synonymous Codon Usage (RSCU) Analysis</i> .....	41
	3.2.6. <i>Relative Dinucleotide Abundance</i> .....	41-42
	3.2.7. <i>Correspondence Analysis</i> .....	42
	3.2.8. <i>Codon Adaptation Index (CAI) Analysis</i> .....	41-42
	3.2.9. <i>tRNA Adaptation Index (tAI)</i> .....	43
	3.3.10. <i>Correlation analysis</i> .....	44
	3.3.12. <i>Software used</i> .....	44

<b>CHAPTER – 4 Results .....</b>	<b>45-86</b>
4.1. Nucleotide composition in proto-oncogenes/oncogenes.....	45-46
4.2. Nucleotide composition in tumour suppressor gene.....	48-49
4.3. Nucleotide skewness among proto-oncogenes/oncogenes.....	51-53
4.4. Nucleotide skewness among tumour suppressor genes.....	54-56
4.5. Codon usage patterns in the coding sequences of proto-oncogenes/oncogenes.....	56-58
4.6. Codon usage patterns in the coding sequences of tumour suppressor genes.....	58-60
4.7. Relationship between codon usage bias and compositional constraints in the coding sequences of proto-oncogenes/oncogenes.....	60-62
4.8. Relationship between codon usage bias and compositional constraints in the coding sequences of tumour suppressor genes.....	62-64
4.9. Effective number of codons ( <i>ENC</i> ) and its relationship with $GC_{3s}$ values in the coding sequences of proto-oncogenes/oncogenes.....	64-65
4.10. Effective number of codons ( <i>ENC</i> ) and its relationship with $GC_{3s}$ values in the coding sequences of tumour suppressor genes.....	65-66
4.11. Frequency of optimal codons for the corresponding amino acids among proto-oncogenes/oncogenes.....	67-68
4.12. Frequency of optimal codons for the corresponding amino acids among tumour suppressor genes.....	67-69
4.13. Relative synonymous codon usage in the coding sequences of proto-oncogenes/oncogenes.....	70,72
4.14. Relative abundance of dinucleotide in the coding sequences of proto-oncogenes/oncogenes.....	70-73
4.15. Trends of codon usage variation among proto-oncogenes/oncogenes....	73-74
4.16. Relative synonymous codon usage in the coding sequences of tumour suppressor genes.....	74,76
4.17. Relative abundance of dinucleotide in the coding sequences of tumour suppressor genes.....	75,77
4.18. Trends of codon usage variation among tumour suppressor genes.....	77-78

4.19. Prediction of proto-oncogenes/oncogenes expression level and codon usage bias.....	78-79
4.20. Relationship between codon usage and <i>CAI</i> values of proto-oncogenes/oncogenes.....	80-81
4.21. Prediction of translational efficiency as measured by <i>tAI</i> and its relationship with codon usage in proto-oncogenes/oncogenes pool.....	81-82
4.22. Prediction of tumour suppressor gene expression level and codon usage bias.....	83-84
4.23. Relationship between codon usage and <i>CAI</i> values of tumour suppressor genes.....	84-85
4.24. Prediction of translational efficiency as measured by <i>tAI</i> and its relationship with codon usage in tumour suppressor gene pool.....	85-86
<b>CHAPTER – 5 Discussion</b> .....	<b>87-97</b>
<b>SUMMARY</b> .....	<b>98-101</b>
<b>BIBLIOGRAPHY</b> .....	<b>102-112</b>
<b>APPENDIX</b>	
<b>A.</b> Lists of Abbreviations.....	<b>IV</b>
<b>B.</b> Lists of Publications - Published Research Articles.....	<b>V-VI</b>
<b>C.</b> Lists of Workshop & Conference Attended .....	<b>VII-VIII</b>



## List of Tables

Table No.	Title	Page No.
1	Accession numbers of coding sequences (cds) of eighty-two human protooncogenes/oncogenes with their lengths (bp)	34-35
2	Accession numbers of coding sequences (cds) of sixty-three human tumour suppressor genes with their lengths (bp)	36-37
3	Nucleotide composition analysis in the coding sequences of eighty-two proto-oncogenes/oncogenes	46-47
4	Nucleotide composition analysis in the coding sequences of sixty-three tumour suppressor genes	49-50
5	Skew values of eighty-two proto-oncogenes/oncogenes	52-53
6	Skew values of sixty-three tumour suppressor genes	55-56
7	Correlation analysis between A, T, G, C, GC contents and A <sub>3</sub> , T <sub>3</sub> , G <sub>3</sub> , C <sub>3</sub> , GC <sub>3</sub> contents in all the selected coding sequences of proto-oncogenes/oncogenes	62
8	Correlation analysis between A, T, G, C, GC contents and A <sub>3</sub> , T <sub>3</sub> , G <sub>3</sub> , C <sub>3</sub> , GC <sub>3</sub> contents in all selected coding sequences of tumour suppressor genes	63
9	The average of <i>Fop</i> values showing the highest percentage of codon usage for the corresponding amino acid among proto-oncogenes/oncogenes	68
10	The average of <i>Fop</i> values showing the highest percentage of codon usage for the corresponding amino acid among tumour suppressor genes	69
11	Overall relative synonymous codon usage patterns ( <i>RSCU</i> ) in the coding sequences of proto-oncogenes/oncogenes	72
12	Overall relative synonymous codon usage patterns ( <i>RSCU</i> ) in the coding sequences of tumour suppressor genes	76

<b>List of Figures</b>		
<b>Figure No.</b>	<b>Title</b>	<b>Page No.</b>
1	Standard genetic code along with 20 amino acids and three stop codons	2
2	Codon bias may select for the prevalence of KRas mutations in human cancers	12
3	A schematic representation of tumour suppressor gene <i>p53</i> pathway and its mechanism of action during DNA damage repair and cell cycle progression	14
4	Percentage of GC content at three codon positions in the coding sequences of proto-oncogenes/oncogenes	45
5	Percentage of GC content at three codon positions in the coding sequences of tumour suppressor genes	48
6	GC skew values of eighty-two proto-oncogenes/oncogenes	51
7	GC skew values of sixty-three tumour suppressor genes	54
8	Heat maps of correlation coefficient of codons with GC <sub>3s</sub> (proto-oncogenes/oncogenes)	57
9	Scatter plots of negatively correlated codons with GC <sub>3s</sub> (proto-oncogenes/oncogenes)	58
10	Heat maps of correlation coefficient of codons with GC <sub>3s</sub> (tumour suppressor genes)	59
11	Scatter plots of negatively correlated codons with GC <sub>3s</sub> (tumour suppressor genes)	60
12	Neutrality plot of GC <sub>12</sub> versus GC <sub>3</sub> (proto-oncogenes/oncogenes)	62
13	Neutrality plot of GC <sub>12</sub> versus GC <sub>3</sub> (tumour suppressor genes)	64

Continued



Figure No.	Title	Page No.
14	Distribution of effective number of codons ( <i>ENC</i> ) and GC3s in the coding sequences of proto-oncogenes/oncogenes	65
15	Distribution of effective number of codons ( <i>ENC</i> ) and GC3s in the coding sequences of tumour suppressor genes	66
16	Line diagram showing the relative abundance of sixteen dinucleotides in the coding sequence of proto-oncogenes/oncogenes in human	73
17	Correspondence analysis of <i>RSCU</i> values in human proto-oncogenes/oncogenes	74
18	Line diagram showing the relative abundance of sixteen dinucleotides in the coding sequence of tumour suppressor genes in human	77
19	Correspondence analysis of <i>RSCU</i> values in human tumour suppressor genes	78
20	Trends of the <i>CAI</i> values in the coding sequences of proto-oncogenes/oncogenes	79
21	Correlation analysis between <i>CAI</i> values and <i>ENC</i> , GC, GC <sub>3</sub> values (proto-oncogenes/oncogenes)	79
22	Heat maps of correlation coefficient of codons with <i>CAI</i> (proto-oncogenes/oncogenes)	80-81
23	Relationship between <i>tAI</i> and codon usage bias in proto-oncogene/oncogene pool	82
24	Trends of the <i>CAI</i> values in the coding sequences of tumour suppressor genes	83
25	Correlation analysis between <i>CAI</i> values and <i>ENC</i> , GC, GC <sub>3</sub> values (tumour suppressor genes)	84
26	Heat maps of correlation coefficient of codons with <i>CAI</i> (tumour suppressor genes)	85
27	Relationship between <i>tAI</i> and codon usage bias in tumour suppressor gene pool	86