Contents

Contents iv

List of Figures vi

List of Tables vii

1 Introduction 1

1.1 Central Dogma of Molecular Biology 4

1.2 Genetic Code Table 5

1.3 Codon Usage Bias (CUB) 5

1.3.1 Overview 5

1.3.2 Causes of CUB 6

2 Related Research Studies 9

2.1 Methods of Quantifying Codon Usage Bias 9

2.2 Measuring Effective No of Codon(ENc) 9

2.2.1 Finding Nc Value 9

2.2.1.1 Derivation 9

2.2.1.2 Analysis 10

2.2.1.3 Limitation of Wrightś formula with example . . 11

2.2.1.4 Suggested solution by Wright: 12

2.2.2 Modified Wrightś N_c value by Sun 13

2.2.2.1 Description of Sun formula 13

2.2.2.2 Analysis 15

2.2.2.3 Pitfalls 16

2.2.3 Proposed Idea: 17

3 Improved Implementation of Nc value **18**

3.1 Derivation 18

3.2 Analysis 18

4 Result and Analysis **20**

4.1 ROC Curve 20

4.2 Analysis of Results by plotting ROC curve 21

4.3 Conclusion and Future Work 22

Bibliography **24**

List of Figures

1.1 Central Dogma of Molecular Biology 1
1.2 Double helical structure of DNA 1
1.3 single helical structure of RNA 2
1.4 mRNA 2
1.5 rRNA 2
1.6 tRNA 2
1.7 Deoxyribose and Ribose 3
1.8 purine and pyrimidine 3
1.9 Pairing Bases 3
1.10 Protein Structure 3
1.11 Transcription and Translation 4
1.12 DNA Replication 4
1.13 Protein Synthesis 5
1.14 Universal genetic code table 6
1.15 Codon usage in E.coli 6
1.16 Codon usage in Yeast 6
1.17 Leading and lagging strand 7

2.1 Synonymous Family 10

4.1 AUC for ROC curve 21
4.2 ROC Curve by three versions of N_c for E.coli 22
4.3 web portal[http://agnigarh.tezu.ernet.in/ ssankar/cub.php] 22

List of Tables

2.1 Uniform CUB 12
2.2 Uniform CUB with equal no of codon 13
2.3 High CUB 14
2.4 Highly Uniform 4-fold amino acid 15

3.1 Correlation study in E. coli 19