

ABSTRACT

In the translation process, mRNA is broken down into nucleotides group of three called codons. 61 codons correspond to 20 amino acids. Due to this degeneracy, more than one codon code for the same amino acid. Different codons encoding the same amino acids are known as synonymous codons. The unequal usage of synonymous codons is called "Codon Usage Bias".

Effective number of Codons (N_c) is one of the important methods that have been proposed to estimate the degree of synonymous codon usage bias. N_c provides an intuitively meaningful measure of the extent of codon preference in a gene. To calculate N_c used in a gene, the number of effective number of codons is used for each amino acid need to be calculated. Then, summing the 'effective number of codons' used by each of the 20 amino acids will then yield an N_c , used in a gene. N_c value should be 61 when codon usage is highly uniform and it should be 20 for extreme codon bias. At the very beginning, multiple problems were recognized in the N_c formula. One of these problems is N_c per codon family of a particular amino acid having values much greater than maximum expected for that codon family. Another problem found is N_c having garbage values for gene with smallest total number of codons. These problems are analyzed; resolved and previous N_c formula is modified. Modified N_c is showing accuracy in previous cases.