

7. Highly Expressed Proteins Contain Increased Frequency of Alanine in the Second Amino Acid Position

Age Tats¹§, Mairo Remm¹, Tanel Tenson²

¹Department of Bioinformatics, Institute of Molecular and Cell Biology, University of Tartu, Estonia

²Institute of Technology, University of Tartu, Estonia

§Presenting author

We analysed the mRNA regions downstream of initiation codons in nine bacteria, three archaea and three unicellular eukaryotes comparing the dataset of highly expressed genes (HEG) to the dataset of all genes.

Strong alanine over-representation at second amino acid position of highly expressed proteins was observed in all three domains of life.

Keywords: highly expressed genes, codon, nucleotide, amino acid, alanine, initiation of translation

The mRNA sequence immediately following the initiation codon influences the efficiency of translation. Although the sequence requirements for translation initiation regions have been frequently analysed, usually the highly expressed genes have not been treated as a separate dataset. Therefore we analysed the mRNA regions downstream of initiation codons in nine bacteria, three archaea and three unicellular eukaryotes comparing the dataset of highly expressed genes (HEG) to the dataset of all genes.

The HEG datasets for the studied organisms were compiled based on assumption that functional conservation implies conservation of relative gene expression level. Therefore we used orthologues to HEG of *Escherichia coli* and *Saccharomyces cerevisiae*, organisms for which correlation between synonymous codon usage and expression level is widely accepted. In addition to the detailed analysis of the nucleotide and codon frequencies we compared the N-termini of highly expressed proteins to the N-termini of all proteins coded in the genome. For comparisons of nucleotide, codon and amino acid frequencies in two datasets of each organism we used two-tailed Fisher's exact test.

The most conserved pattern was observed on amino acid level: strong alanine over-representation was observed at second amino acid position of highly expressed proteins. As the bacteria and eukaryotes use different mechanisms for initiation of translation, it has been anticipated that the initiation context effects are different. Our studies revealed a well conserved pattern in all three domains of life.