Gene Architectures that Minimize Cost of Gene Expression

Idan Frumkin*1, Aviv Rotman*1, Dvir Schirman*1, Fangfei Li2,3, Liron Zahavi1, Omer Asraf2, Gilad Shaham4, Ernest Mordret1, Tamir Tuller4,5, Song Wu3, Sasha F Levy2,6 & Yitzhak Pilpel1

1 - Department of Molecular Genetics, Weizmann Institute of Science, Rehovot, Israel.
2 - Lauder Center for Physical and Quantitative Biology, Stony Brook University, Stony Brook, New York 11794, USA.
3 - Department of Applied Mathematics and Statistics, Stony Brook University, Stony Brook, New York 11794, USA.
4 - Department of Biomedical Engineering, the Engineering Faculty, Tel Aviv University, Tel-Aviv, Israel.
5 - The Sagol School of Neuroscience, Tel-Aviv University, Tel-Aviv, Israel.
6 - Department of Biochemistry and Cell Biology, Stony Brook University, Stony Brook, New York 11794, USA.

Abstract

Gene expression burdens cellular growth by consumption of resources, raw materials and energy. Here we ask how cells minimize production costs while maintaining a desired expression level, and if there are sequence parameters that can optimize this process. To this end, we utilized a synthetic library with ~14K Escherichia coli variants, each expressing a GFP gene with a unique variable upstream region. We measured the fitness of all strains and found that it declines linearly with expression level. Yet, for many variants fitness was significantly either above or below its expected value given the expression level of the unneeded GFP protein. Four factors appear to modulate expression costs: The first is the production of more proteins per mRNA, a potential means to minimize transcription costs. The second appears to attenuate ribosomes at early elongation by three independent mechanisms: ribosomal codon-decoding speed, tight mRNA secondary structure and high affinity to the anti-Shine Dalgarno motif of the ribosome. The third mechanism relates to energetic cost of raw materials: cells that express peptides with expensive-to-synthesize amino acids have low relative fitness. The fourth factor is amino acid hydrophobicity, which is negatively associated with relative fitness, potentially reflecting toxicity of aggregation-prone proteins. By projecting these results onto the natural E. coli genome, we found that the higher the cellular expression level for a given gene, the more likely it is to follow the architectures that minimize expression cost. In conclusion, our study elucidates basic cellular mechanisms that minimize costs of gene expression and are utilized by natural selection to shape the genome.