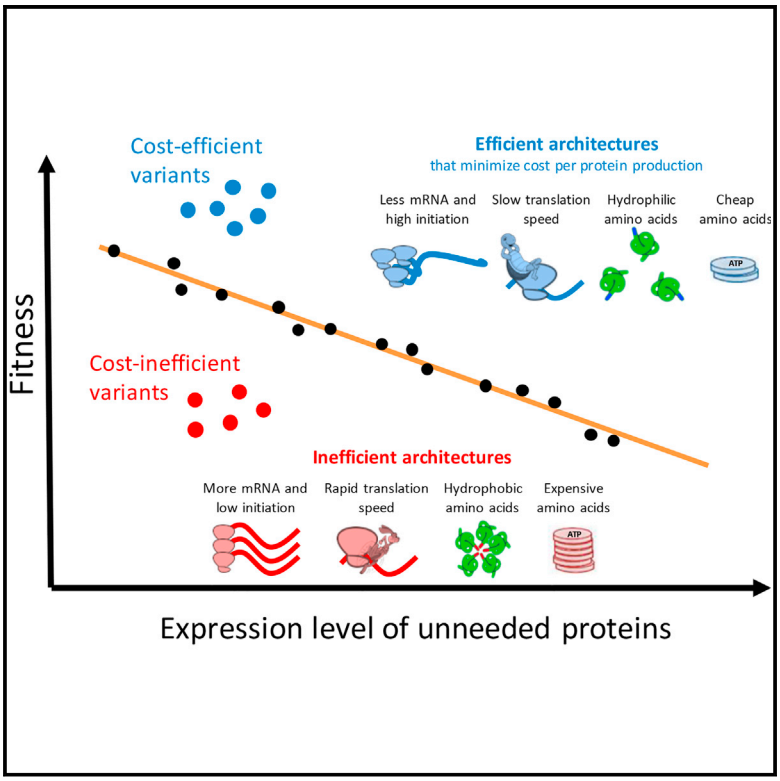
Article

Molecular Cell

Gene Architectures that Minimize Cost of Gene Expression

Graphical Abstract



Authors

Idan Frumkin, Dvir Schirman, Aviv Rotman, ..., Song Wu, Sasha F. Levy, Yitzhak Pilpel

Correspondence

pilpel@weizmann.ac.il

In Brief

While numerous studies have investigated regulation of expression level, Frumkin et al. study gene design elements that govern expression costs and allow cells to minimize such costs while maintaining a given protein expression level.

Highlights

- Microorganisms can minimize expression cost with diverse molecular means
- Some design elements can produce more unneeded proteins but maintain high fitness
- Such elements optimize use of production machineries and utilize cheap materials
- Natural highly expressed genes evolved more forcefully to lower expression costs

Frumkin et al., 2017, Molecular Cell 65, 1–12
January 5, 2017
http://dx.doi.org/10.1016/j.molcel.2016.11.007

CellPress

Please cite this article in press as: Frumkin et al., Gene Architectures that Minimize Cost of Gene Expression, Molecular Cell (2016), http://dx.doi.org/10.1016/j.molcel.2016.11.007

Molecular Cell
Article

CellPress

Gene Architectures that Minimize Cost of Gene Expression

Idan Frumkin,[1,5] Dvir Schirman,[1,5] Aviv Rotman,[1,5] Fangfei Li,[2,3] Liron Zahavi,[1] Ernest Mordret,[1] Omer Asraf,[1] Song Wu,[3] Sasha F. Levy,[2,4] and Yitzhak Pilpel[1,6,*]
[1]Department of Molecular Genetics, Weizmann Institute of Science, 7610001 Rehovot, Israel
[2]Laufer Center for Physical and Quantitative Biology
[3]Department of Applied Mathematics and Statistics
[4]Department of Biochemistry and Cell Biology
Stony Brook University, Stony Brook, NY 11794, USA
[5]Co-first author
[6]Lead Contact
*Correspondence: pilpel@weizmann.ac.il
http://dx.doi.org/10.1016/j.molcel.2016.11.007

SUMMARY

Gene expression burdens cells by consuming resources and energy. While numerous studies have investigated regulation of expression level, little is known about gene design elements that govern expression costs. Here, we ask how cells minimize production costs while maintaining a given protein expression level and whether there are gene architectures that optimize this process. We measured fitness of ~14,000 *E. coli* strains, each expressing a reporter gene with a unique 5′ architecture. By comparing cost-effective and ineffective architectures, we found that cost per protein molecule could be minimized by lowering transcription levels, regulating translation speeds, and utilizing amino acids that are cheap to synthesize and that are less hydrophobic. We then examined natural *E. coli* genes and found that highly expressed genes have evolved more forcefully to minimize costs associated with their expression. Our study thus elucidates gene design elements that improve the economy of protein expression in natural and heterologous systems.

INTRODUCTION

In nature, cells must express different genes in a regulated manner. On one hand, genes must be expressed at levels that maximize their benefit, and on the other, cells need to minimize the genes' production costs (Dekel and Alon, 2005; Wagner, 2005). Costs of expression originate from spending cellular resources, such as building blocks (amino acids and nucleotides), from allocation of cellular machineries (RNA polymerase and ribosome), and from energy and reducing power consumption (Bienick et al., 2014; Glick, 1995; Ibarra et al., 2002; Rang et al., 2003). Even after their production, proteins might still impose costs when degraded or by exerting toxicity, e.g., due to aggregation (Geiler-Samerotte et al., 2011). Understanding what molecular processes determine expression cost, its relation to cellular growth and gene regulation, and how costs evolutionarily shape the genome are key aspects of cell biology that remain largely elusive. While numerous studies investigated molecular mechanisms and gene sequence architectures that regulate expression level (Gingold and Pilpel, 2011; Kudla et al., 2009; Qian et al., 2012; Sharp et al., 1986; Subramaniam et al., 2013), very little is known about design elements that govern expression costs.

Different works have studied expression costs in unicellular organisms by imposing the expression of an unneeded protein (Bentley et al., 1990; Dekel and Alon, 2005; Dong et al., 1995; Kafri et al., 2016; Rang et al., 2003; Scott et al., 2010). The production of such unneeded proteins diverts resources from synthesis of the cell's own proteins, thus decreasing cellular fitness (Emilsson and Kurland, 1990; Marr, 1991; Vind et al., 1993). Central to these studies is the characterization of the correlation between the imposed expression levels of the unneeded proteins to the cost. Yet, ultimately natural selection dictates the expression level of natural genes according to the required concentration of each protein. Thus, a fundamental question, which has not been addressed before, is how cells can achieve a specific expression level of a gene while minimizing its expression costs.

Addressing this question is challenging because changes in sequence could affect both expression level and expression costs. To disentangle expression level and expression costs and reveal mechanisms that affect cost per protein molecule, we utilized a synthetic reporter library of ~14,000 different sequence variants, each fused upstream to a GFP gene (Goodman et al., 2013). We then combined competition assays and deep sequencing to measure the fitness of all variants in parallel. This procedure allowed us to elucidate gene architectures that minimize expression cost at a given protein expression level. We show that various molecular mechanisms, such as protein/mRNA ratios, ribosome early elongation pauses, amino acid synthesis costs, and peptide hydrophobicity, determine the cost per protein molecule. We then generated a model that predicts the cost effectiveness of gene architectures and applied it to natural *E. coli* genes. We found that highly expressed genes have