Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Microbiology

How do bacteria tune translation efficiency?

Gene-Wei Li

Bacterial proteins are translated with precisely determined rates to meet cellular demand. In contrast, efforts to express recombinant proteins in bacteria are often met with large unpredictability in their levels of translation. The disconnect between translation of natural and synthetic mRNA stems from the lack of understanding of the strategy used by bacteria to tune translation efficiency (TE). The development of array-based oligonucleotide synthesis and ribosome profiling provides new approaches to address this issue. Although the major determinant for TE is still unknown, these high-throughput studies point out a statistically significant but mild contribution from the mRNA secondary structure around the start codon. Here I summarize those findings and provide a theoretical framework for measuring TE.

Addresses
Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, United States

Corresponding author: Li, Gene-Wei (gwli@mit.edu)

Current Opinion in Microbiology 2015, **24**:66–71

This review comes from a themed issue on **Cell regulation**

Edited by **Carol Gross** and **Angelika Gründling**

For a complete overview see the Issue and the Editorial

Available online 28th January 2015

http://dx.doi.org/10.1016/j.mib.2015.01.001

Introduction

Soon after Jacob and Monod proposed the existence of polycistronic mRNA [1], it was noticed that different proteins originating from the same mRNA are translated at very different rates [2]. This observation was made for the RNA genome of a bacteriophage that was translated upon entrance to the host bacterium. The difference in translation rates was deemed necessary to synthesize a large excess of bacteriophage coat proteins relative to RNA polymerases for viral particle production. Although the initial studies suggested the use of mRNA secondary structure to modulate translation efficiency (TE) [3], the later discovery of the Shine–Dalgarno (SD) sequence pointed to the potential for tuning TE by changing the affinity of mRNA to the anti-Shine–Dalgarno sequence on the 16S rRNA [4,5]. Detailed studies on individual genes and operons then revealed a plethora of means to modulate TE [6–9]. It remains unclear whether there exists a general principle for setting the TE for the 4000+ genes in *Escherichia coli*.

Recently, the promise of synthetic biology — the design of biological devices from genetic and protein components — increased the demand for better understanding and control of TE. To address this issue, several research groups created large-scale libraries of synthetic mRNAs to probe the sequence features that influence TE [13••,10••,11••,12••]. Meanwhile, with the development of ribosome profiling (deep sequencing of ribosome protected mRNA fragments) it became possible to monitor the TE of endogenous genes at genome-scale [14]. Here, I will summarize the conclusions from these recent studies and discuss the missing pieces of the puzzle.

Defining TE

In this review, 'TE' is referred to as the rate of protein production per mRNA [14–16]. In other contexts, the same phrase has been defined as the rate of translation elongation, which affects the efficiency with which ribosomes are used [13••,17,18]. Both definitions are widely used in the literature, and this can lead to profound confusion when the exact definition is not specified. For example, factors that influence the efficiency of elongation should not be confused with the determinant of protein production per mRNA, or TE as defined here [19–21]. In cells, these two processes are sometimes connected because they both concern the cellular pool of ribosomes [22–24], but they are not the same. The possible connection between elongation and production per mRNA (or lack thereof) has been reviewed in several recent studies [21,25–27]. Here I focus on understanding the meaning and utility of defining TE as the rate of protein production per mRNA.

It is clear that protein abundance is not equal to TE. Protein abundance is a product of mRNA level, TE, and protein lifetime (Box 1). Changes to the mRNA sequence can often affect some or all of these factors, making it difficult to attribute the resulting difference in protein level to changes in TE alone. Moreover, TE itself can also directly influence mRNA levels. If an mRNA is more stable when TE is high, the amount of proteins produced scales nonlinearly with TE (Box 1). It is therefore important to normalize protein abundance by differences in protein lifetime and, in particular, mRNA levels. As described later, the combination of ribosome profiling and RNA-seq enables accurate determination of protein production rate per mRNA.

Using synthetic DNA libraries to interrogate determinant of TE

Examining the effect of mutations on protein production is a common approach to dissect the determinants of TE.

Box 1 Definition of translation efficiency.

Various definitions of translation efficiency have been a source of confusion. To make a clear definition of the term used in this review, consider the four basic kinetic parameters in the central dogma (Figure 1). mRNAs are produced at a rate k_1, and proteins are produced from mRNA with a first-order rate constant k_2. λ_1 and λ_2 are the first-order decay rate constants for mRNA and protein, respectively. The master equations for the scheme in Figure 1 are

$$\frac{d}{dt}M = k_1 - \lambda_1 M \quad (1)$$

$$\frac{d}{dt}P = k_2 M - \lambda_2 P \quad (2)$$

where M and P are the concentration of mRNA and protein, respectively. In this review, translation efficiency is defined as the rate of protein production per mRNA, which is equal to k_2.

$$TE \equiv k_2 \quad (3)$$

Operationally, most studies report changes in P or P/M as a measure for changes in TE. At steady state, the master equations yield

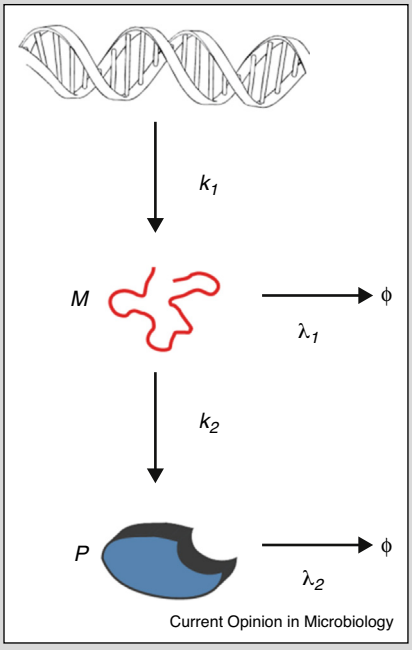
$$M = \frac{k_1}{\lambda_1} \quad (4)$$

$$P = \frac{k_2}{\lambda_2} M = \frac{k_2}{\lambda_2}\frac{k_1}{\lambda_1} \quad (5)$$

Therefore, both P and P/M are sensitive to changes in the degradation rate. In contrast, ribosome profiling in combination with RNA-seq reports k_2, a direct measure for TE (Box 2).

Consider a simple case in which mRNA degradation depends on the level of translation, which is common in bacteria [37]. If the mRNA decay rate is inversely proportional to TE ($\lambda_1 = A/k_2$), the amount of protein produced then scales quadratically with TE ($P = ((k_1 k_2^2)/(A\lambda_2)) \propto k_2^2$), whereas $P/M = k_2/\lambda_2$ still scales linearly with TE ($P/M = \propto k_2$). This example illustrates the importance of normalizing protein levels by mRNA levels when reporting TE. In ribosome profiling, the ratio between ribosome density and mRNA level is not affected by changes in mRNA decay.

Figure 1



Defining translation efficiency and other kinetic constants in gene expression. The mRNA (M) is transcribed from DNA at a rate k_1 and degraded (ϕ) with a rate constant λ_1. The corresponding protein (P) is translated from mRNA with a rate constant k_2 and degraded (ϕ) with a rate constant λ_2. Translation efficiency is defined as k_2 in this review. See Box 1.

An important limitation is that it is not feasible to explore every possible combination of mutations, as a 200-aa protein can have 10^{120} possible synonymous coding variants multiplied by additional variations for the untranslated regions of mRNA. Therefore, even with high-throughput oligonucleotide synthesis technology, a library of mutations can only cover a very small subset of the sequence space. Therefore, instead of aiming for unbiased coverage, synthetic libraries are often designed to test specific hypotheses. For example, a library of ~100 mutations in the 5′ un-translated region (UTR) was constructed to systematically examine the effects of sequences surrounding the ribosome binding site using RFP fluorescence as a readout [10••]. Note that this approach assumes that the potential impact of 5′ UTR on mRNA stability is negligible. The results of this study suggested that TE was influenced by multiple factors including the SD sequence, the thermostability of RNA secondary structure, and other features of the ribosome binding site.

In parallel to the 5′ UTR study, several groups constructed fluorescent reporter libraries of similar size (~10^2) to test the effect of different synonymous mutations within the open reading frame (ORF) [11••,13••]. These studies found that the usage of rare codons has little or no effect on protein abundance. Instead, the lack of mRNA secondary structure at the start site has the most significant, albeit weak, correlation with protein abundance [11••,13••]. As was true of the previous study, the parameter measured was the final amount of protein produced. Thus, it was unclear whether the effects of RNA folding were on mRNA decay or TE. In fact, a later study reanalyzed the data and found that when protein abundance is normalized by mRNA levels, the correlation with RNA folding vanishes even though TE remains variable over two orders of magnitude [28]. How the observed TE is modulated is still unknown.

Bacterial mRNAs tend to have a lower amount of secondary structure around the translational start site, both in