Length-dependent translation of messenger RNA by ribosomes

Angelo Valleriani,[1,*] Gong Zhang,[2] Apoorva Nagar,[1] Zoya Ignatova,[2] and Reinhard Lipowsky[1]
[1]*Max Planck Institute of Colloids and Interfaces, Department of Theory and Bio-Systems, D-14424 Potsdam, Germany*
[2]*University of Potsdam, Institute of Biochemistry and Biology, D-14467 Potsdam, Germany*
(Received 15 May 2010; published 21 April 2011)

A simple measure for the efficiency of protein synthesis by ribosomes is provided by the steady state amount of protein per messenger RNA (mRNA), the so-called translational ratio, which is proportional to the translation rate. Taking the degradation of mRNA into account, we show theoretically that both the translation rate and the translational ratio decrease with increasing mRNA length, in agreement with available experimental data for the prokaryote Escherichia coli. We also show that, compared to prokaryotes, mRNA degradation in eukaryotes leads to a less rapid decrease of the translational ratio. This finding is consistent with the fact that, compared to prokaryotes, eukaryotes tend to have longer proteins.

DOI: 10.1103/PhysRevE.83.042903 PACS number(s): 87.10.Mn, 87.14.gn, 87.15.A−

Understanding and controlling the different mechanisms that determine the rate of protein synthesis is a fundamental issue both in cell biology and biomedicine [1–4]. On the one hand, the predicted stationary or homeostatic level of proteins under different environmental conditions allows one to estimate the growth rate of the cells and their sensitivity to changes in their environment [5]. On the other hand, an accurate description of the underlying molecular processes allows one to determine the corresponding response time of cells [6] and is therefore a central issue in cell biology.

Degradation of messenger RNA (mRNA) plays an important role in shaping the balance between mRNA and protein molecules. This process follows different pathways in prokaryotic and eukaryotic cells, as illustrated in Fig. 1 and in Ref. [7]. In prokaryotic cells, the mRNA can be degraded cotranslationally, all ribosomes that are on the mRNA at the moment of degradation are lost and the mRNA chains are degraded immediately [8–10]. In eukaryotic cells, the degradation of mRNA proceeds in two steps. First, the initiation region is decapped to prevent the assembly of new ribosomes. Second, those ribosomes that are present on the mRNA at the moment of decapping are allowed to complete the translation [8,11].

We assume here that the process of mRNA degradation is governed by a certain rate ω_r, or equivalently, that any intact mRNA is characterized by a random lifetime U, which is the time until fast degradation in prokaryotes or decapping in eukaryotes. Thus, the time U has a probability density that is given by

$$\phi_U(t) = \omega_r \exp(-\omega_r t)\,. \tag{1}$$

As we shall see, in both prokaryotic and eukaryotic cells, the degradation rate ω_r of mRNA will enter into the expression that determines the rate by which proteins are synthesized.

One relatively simple and successful theory that relates mRNA and protein abundance is based on flux balance relations that include both transcription and translation [1]. For fixed external conditions, an active gene is steadily transcribed with a certain transcription rate $\omega_{\rm ts}$. This process generates several copies of mRNA. Each mRNA is then steadily translated into a protein with a certain synthesis or translation rate $\omega_{\rm tl}$, defined here as the mean number of protein molecules synthesized per mRNA per unit time [1]. Furthermore, both the proteins and the mRNAs are degraded with rates ω_p and ω_r, respectively [12]. Since all these processes are stochastic in nature, the number of mRNA chains and the number of proteins in the cell will, in general, fluctuate [13–15]. Nevertheless, the average numbers N_r and N_p of mRNAs and proteins that arise from a certain gene are governed by the simple equations

$$\frac{dN_r}{dt} = \omega_{\rm ts} - \omega_r N_r\,, \quad \frac{dN_p}{dt} = \omega_{\rm tl} N_r - \omega_p N_p\,, \tag{2}$$

which determine N_r and N_p as a function of time t. Starting from any initial condition, the average numbers N_r and N_p eventually attain their steady-state values

$$N_r^{\rm st} = \frac{\omega_{\rm ts}}{\omega_r} \quad \text{and} \quad N_p^{\rm st} = \frac{\omega_{\rm ts}\omega_{\rm tl}}{\omega_r \omega_p}. \tag{3}$$

The ratio

$$r \equiv \frac{N_p^{\rm st}}{N_r^{\rm st}} = \frac{\omega_{\rm tl}}{\omega_p}, \tag{4}$$

called the translational ratio, is a simple measurable quantity that depends only on the translation rate $\omega_{\rm tl}$ and on the protein degradation rate ω_p. In the theories based on Eq. (2), it has been convenient to consider rates that are independent of each other. In this Brief Report, we will show that the rate of translation $\omega_{\rm tl}$ is determined by the mRNA degradation process and by the time scale of translation, which is, in turn, proportional to the length of the mRNA.

Schematically, the process of translation can be summarized as follows: Ribosomes enter the coding region with a certain rate $\omega_{\rm on}$, will pass this region with a mean velocity v and will leave the end of the coding region instantaneously. During their walk through the coding region, the ribosomes may prematurely drop off from the chain with a rate $\omega_{\rm off}$. The number of ribosomes on a single mRNA molecule is therefore a stochastic variable. Assuming a constant density of ribosomes in the cell and a low density of ribosomes on each mRNA, the average ribosome number $M(t)$ on a given mRNA obeys the deterministic equation [16]

$$\frac{dM}{dt} = \omega_{\rm on} - \omega_{\rm off} M \quad \text{for} \quad t < t_L = L/v, \tag{5}$$

*angelo.valleriani@mpikg.mpg.de

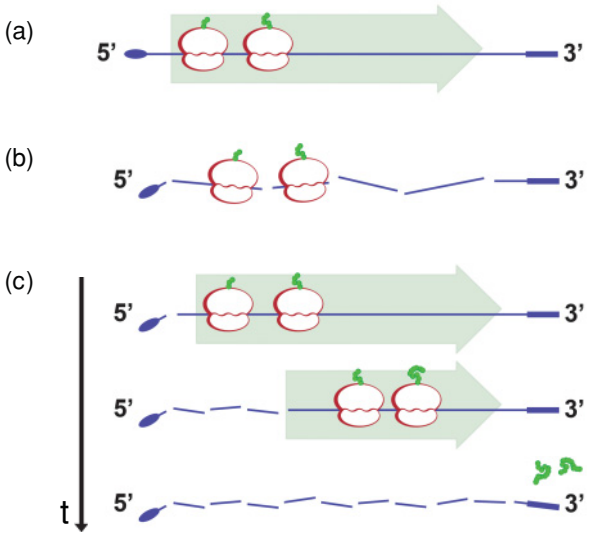

FIG. 1. (Color online) Degradation pathways for mRNA in prokaryotic and eukaryotic cells. (a) Single mRNA strand with two loaded ribosomes that move from the 5′ to the 3′ end of the mRNA; at a later time more ribosomes may initiate translation at the 5′ end. (b) In prokaryotes, mRNA degradation aborts translation and protein synthesis for all loaded ribosomes. (c) In eukaryotes, decapping prevents the loading of new ribosomes but the mRNA is not degraded until the previously loaded ribosomes have completed their protein synthesis and are released from the 3′ end of the mRNA. The arrow indicates the evolution with time t.

where the run time t_L is the time needed by the first ribosome to reach the end of the chain. For $t \geqslant t_L$, the number M of ribosomes attains the constant value M_L, which is the average number of ribosomes in the steady state [16]. Since the drop-off rate $\omega_{\rm off}$ is very small [17], we will henceforth consider only the limit of zero $\omega_{\rm off}$. The solution of Eq. (5), with initial condition $M(0) = 0$ is then given by

$$M(t) = \begin{cases} \omega_{\rm on} t & \text{for} \quad t < t_L, \\ M_L = \omega_{\rm on} t_L & \text{for} \quad t \geqslant t_L. \end{cases} \tag{6}$$

The relations (6) describe the mean number of ribosomes on a single mRNA starting from the initial time $t = 0$, at which ribosomes can initiate translation.

Translational ratio in prokaryotic cells. Inspection of Eq. (6) shows that two mutually exclusive events can occur: Either the mRNA is degraded before the first ribosome reaches the end of the coding region or it is degraded after steady state with $M(t) = M_L$ has been established. In the first case, there are no proteins produced from the mRNA. In the second case, the number of produced proteins depends on the excess life time $t - t_L^{\rm pro}$ of the mRNA, where $t_L^{\rm pro} = L/v^{\rm pro}$ and $v^{\rm pro}$ is the average velocity of ribosomes in prokaryotic cells [7].

If the ribosomal traffic on the mRNA has reached the stationary state, the number of proteins produced per unit time is equal to the number of ribosomes leaving the chain per unit of time, which is given by the ribosomal current or flux $M_L/t_L^{\rm pro}$. Thus, to obtain the number of ribosomes that are able to successfully complete the translation, we need to multiply the ribosomal flux by the random excess lifetime $t - t_L^{\rm pro}$. The translational yield η, defined as the mean number of proteins per transcript, is thus given by

$$\eta^{\rm pro} = \frac{M_L}{t_L^{\rm pro}} \int_{t_L^{\rm pro}}^{\infty} dt \, (t - t_L^{\rm pro}) \phi_U(t), \tag{7}$$

which, after substituting Eq. (1) and taking the limit of zero $\omega_{\rm off}$, is described by the simple expression

$$\eta^{\rm pro} = \frac{\omega_{\rm on}}{\omega_r} \exp\left(-\omega_r t_L^{\rm pro}\right), \tag{8}$$

which shows that longer mRNAs tend to produce fewer proteins. To compute the translation rate, therefore, one has to compute the average of the number of ribosomes that complete translation divided by the lifetime of the transcript. Taking Eq. (1) into account in the limit of zero $\omega_{\rm off}$, we obtain the translation rate

$$\omega_{\rm tl}^{\rm pro} = \omega_{\rm on} \int_{t_L^{\rm pro}}^{\infty} dt \, \frac{t - t_L^{\rm pro}}{t} \omega_r \exp(-\omega_r t), \tag{9}$$

which is a decreasing function of L and contains an explicit dependence on the degradation rate ω_r. Finally, after inserting Eq. (9) into Eq. (4) we can determine the dependence of the translational ratio on the values of the rates, including the mRNA degradation rate. Thus, for a prokaryotic cell the translational ratio (4) is given by

$$r^{\rm pro} = \frac{\omega_{\rm on}}{\omega_p} \int_{t_L^{\rm pro}}^{\infty} dt \, \frac{t - t_L^{\rm pro}}{t} \omega_r \exp(-\omega_r t), \tag{10}$$

which has the property to decrease linearly for small L and exponentially for large L. This means that the larger L, the larger the pool of mRNA molecules that are needed to maintain a certain level of protein abundance.

We have compared Eq. (10) with experimental data from Ref. [18]. Using these data, we have been able to extract information about 423 genes of the bacterium *E. coli*, for which both mRNA and protein abundances have been reported. Figure 2 presents the ratio between protein and mRNA abundance from the original data (small red dots) as a function of mRNA length in number of codons. The data show a very large scatter, which is due to the large variation in the protein half-life. Nevertheless, our correlation analysis reveals a significant negative correlation between translational ratio $r^{\rm pro}$ and mRNA length L. To reveal this pattern, we have binned the data: since the number of genes in the pool decreases exponentially with the length of the gene [18], the width of the bins is taken to increase exponentially with the length. The binned data are shown as large (blue) circles in Fig. 2. Finally, together with these data, we have plotted the theoretical expression (10) with the parameter values $\omega_{\rm on}/\omega_p = 708.2$ and $\omega_r = 0.2$ min^{-1} [20–22]. Taking into account that $\omega_{\rm on}$ can be about 0.167 s^{-1} [23,24], ω_p^{-1} corresponds to an average protein life time of about one hour, which is a commonly accepted value [25]. Thus, in spite of several simplifications, we obtain qualitative agreement between our theory and the data.

Translational ratio in eukaryotic cells. Let us first observe that after decapping it takes a time $t_L^{\rm eu} = L/v^{\rm eu}$ for the last ribosome to reach the end of the mRNA [16]. It has been shown [13,27] that the translational yield $\eta^{\rm eu}$ for the eukaryotic

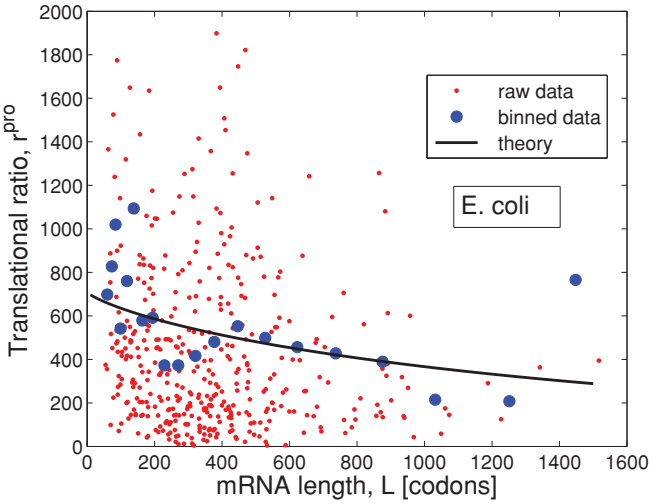

FIG. 2. (Color online) Translational ratio $r^{\rm pro}$ of *E. coli* as a function of mRNA length L. The red points correspond to the experimental data for 409 genes of E. coli as reported in Ref. [18]. Fourteen data points are out of the range of the figure, but have been included in the data analysis. The blue circles represent average values over bins of unequal length, see text. Using Eq. (10) a least-square fit fixes the multiplicative constant $\omega_{\rm on}/\omega_p = 708.2$. The analytical expression (10) leads to the black line and provides a reasonable fit to the binned data. The other parameters have the values $v^{\rm pro} = 800$ codons/min (cd/min) [19] and $\omega_r = 0.2\,{\rm min}^{-1}$.

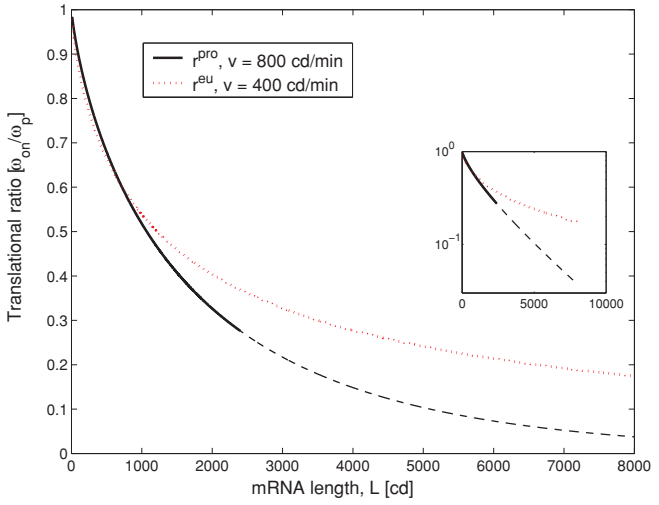

FIG. 3. (Color online) The translational ratio $r^{\rm pro}$ and $r^{\rm eu}$ for prokaryotes as in Eq. (10) (full line) and eukaryotes as in Eq. (13) (dotted line) as a function of mRNA length L. In the inset, the same quantities are plotted on a semilogarithmic scale. To allow a comparison, the values on the y axis are in units of $\omega_{\rm on}/\omega_p$. The other parameters are $v^{\rm pro} = 800$ codons/min (cd/min) for *E. coli* [19], $v^{\rm eu} = 400$ cd/min for budding yeast *S. cerevisiae* [26] and $\omega_r = 0.2$ min^{-1}. The full line extends from 10 to 2400 codons, which is the range of length for *E. coli* mRNA. The dotted line extends up to 8000 codons, which is the maximal length of yeast mRNA. The dashed line indicates the values of $r^{\rm pro}$ if the maximum mRNA length of bacteria would be the same as in yeast. In the inset, the semilog plot shows that $r^{\rm pro}$ decreases exponentially whereas $r^{\rm eu}$ decreases as $1/L$.

degradation shown in Fig. 1, in the limit of zero $\omega_{\rm off}$ is independent of L and is given by

$$\eta^{\rm eu} = \frac{\omega_{\rm on}}{\omega_r}. \tag{11}$$

A comparison with Eq. (8) shows how mRNA degradation contributes to the remarkable differences between prokaryotic and eukaryotic cells.

The translation rate $\omega_{\rm tl}^{\rm eu}$ for eukaryotes is given by the sum of two contributions. Namely, the one arising from chains that have been decapped before reaching steady state (i.e., before time $t_L^{\rm eu}$) and the other deriving from chains that have been decapped after having reached steady-state ribosomal loading [7]. The computation of these two contributions leads to

$$\omega_{\rm tl}^{\rm eu} = \omega_{\rm on} \int_0^\infty dt\, \frac{t}{t + t_L^{\rm eu}} \omega_r \exp(-\omega_r t), \tag{12}$$

in the limit of zero $\omega_{\rm off}$ and taking Eq. (1) into account. If we turn now our attention back to Eq. (4), we find that the translational ratio for eukaryotes is given by

$$r^{\rm eu} = \frac{\omega_{\rm on}}{\omega_p} \int_0^\infty dt\, \frac{t}{t + t_L^{\rm eu}} \omega_r \exp(-\omega_r t), \tag{13}$$

which predicts a decrease of the translational ratio as a function of the transcript length L for eukaryotes. Our analysis of the data in Ref. [18] for the yeast *S. cerevisiae* shows that there is a significant negative correlation between translational ratio $r^{\rm eu}$ and mRNA length L also for this organism (see Ref. [7]). If the velocities of ribosomes in prokaryotes and eukaryotes were identical (i.e., for $v^{\rm pro} = v^{\rm eu} = v$) the expression (13) would imply

$$r^{\rm pro} = \exp(-\omega_r t_L)\, r^{\rm eu}. \tag{14}$$

Since $v^{\rm pro} > v^{\rm eu}$ [19,26] the prokaryotic translational ratio exceeds the eukayotic one at small lengths L, while the exponential decrease of $r^{\rm pro}$ leads to the opposite ordering at large L. The differences between the prokaryotic and the eukaryotic translational ratio $r^{\rm pro}$ and $r^{\rm eu}$ are illustrated in Fig. 3, where we consider a class of genes for which the rates ω_p and ω_r are the same. The translational ratio for the prokaryotic degradation is similar to the ratio for the eukaryotic degradation up to a mRNA length of about 1000 codons, if we take into account that the velocity of translation in eukaryotes is about one half of that in prokaryotes [19,26]. On the other hand, for larger mRNA length, prokaryotic translation becomes much less efficient (Fig. 3). Thus, the different degradation mechanism allows eukaryotes to considerably decrease the translation velocity while keeping the same translational ratio as prokaryotes for short mRNAs and have a better ratio for long mRNAs. This allows eukaryotic cells to perform a more accurate translation process [28] for long proteins [29,30].

In summary, our theory predicts (i) that the process of mRNA degradation leads to a strong dependence of the translational ratio on the length L of the mRNA and (ii) that this L dependence shows marked differences between prokaryotic

and eukaryotic cells, see Fig. 3. The latter differences reflect the distinct degradation mechanisms for these two types of cells as illustrated in Fig. 1. We also compared the predictions of our theory for prokaryotes with experimental data for *E. coli*, see Fig. 2. In addition, we find that our prediction for eukaryotes is in qualitative agreement with the data for *S. cerevisiae* [7,18] and for fission yeast in Ref. [31].

As shown in Fig. 3, the decrease of the translational ratio with the mRNA length is found to be faster in prokaryotic cells compared to eukaryotic cells. This difference is consistent with the observation that proteins in eukaryotic cells are on average longer than those in prokaryotic ones. Thus, our theory implies that this larger size of the eukaryotic proteins is facilitated by a more parsimonious use of the transcriptional resources.

Acknowledgment. A.N. was financially supported by GoFORSYS; Z.I. was supported by the DFG (IG 73/10-1). The illustrations were prepared by Susann Valleriani.

[1] J. L. Hargrove and F. H. Schmidt, FASEB Journal **3**, 2360 (1989).
[2] J. L. Hargrove, M. G. Hulsey, F. H. Schmidt, and E. G. Beale, Biotechniques **8**, 654 (1990).
[3] G. Hoffmann, R. Ostermeir, H. Mueller, N. Bitterlich, P. Martus, and M. Neumaier, Clin. Lab. **54**, 19 (2008).
[4] J. L. Hargrove, FASEB Journal **7**, 1163 (1993).
[5] M. Piques, W. X. Schulze, M. Hohne, B. Usadel, Y. Gibon, J. Rohwer, and M. Stitt, Mol. Syst. Biol. **5** (2009).
[6] Q. W. Yang, P. A. Lindahl, and J. J. Morgan, J. Theoretical Biology **222**, 407 (2003).
[7] See supplemental material at [http://link.aps.org/supplemental/10.1103/PhysRevE.83.042903] for additional illustrations.
[8] S. Kushner, IUBMB Life **56**, 585 (2004).
[9] V. R. Kaberdin and U. Blaesi, FEMS Microbiol. Rev. **30**, 967 (2006).
[10] C. Condon, Curr. Opin. Microbiol. **10**, 271 (2007).
[11] W. Hu, T. J. Sweet, S. Chamnongpol, K. E. Baker, and J. Coller, Nature (London) **461**, 225 (2009).
[12] For simplicity, the degradation rate ω_r is taken to be constant and independent of length L. An increasing degradation rate with increasing L, as discussed in [20,21,32] would enhance the effects described here but does not significantly change our results.
[13] H. McAdams and A. Arkin, Proc. Natl. Acad. Sci. **94**, 814 (1997).
[14] E. Ozbudak, M. Thattai, I. Kurtser, A. Grossman, and A. van Oudenaarden, Nat. Genet. **31**, 69 (2002).
[15] J. R. S. Newman, S. Ghaemmaghami, J. Ihmels, D. K. Breslow, M. Noble, J. L. DeRisi, and J. S. Weissman, Nature (London) **441**, 840 (2006).
[16] A. Valleriani, Z. Ignatova, A. Nagar, and R. Lipowsky, Europhys. Lett. **89**, 58003 (2010).
[17] C. Kurland, Annu. Rev. Genet. **26**, 29 (1992).
[18] P. Lu, C. Vogel, R. Wang, X. Yao, and E. M. Marcotte, Nat. Biotechnol. **25**, 117 (2007).
[19] H. Bremer and P. Dennis, in *Escherichia Coli and Salmonella Typhimurium: Cellular and Molecular Biologgy*, edited by F. Neidhardt (American Society for Microbiology, Washington, D. C., 1996), Chap. 97, p. 1559; G. Zhang, I. Fedyunin, O. Miekley, A. Valleriani, A. Moura, Z. Ignatova, Nucleic Acids Research **38**, 1 (2010)
[20] J. Bernstein, A. Khodursky, P. Lin, S. Lin-Chao, and S. Cohen, Proc. Natl. Acad. Sci. **99**, 9697 (2002).
[21] A. F. Andersson, M. Lundgren, S. Eriksson, M. Rosenlund, R. Bernander, and P. Nilsson, Gen. Biol. **7** (2006).
[22] J. Bernstein, P. Lin, S. Cohen, and S. Lin-Chao, Proc. Natl. Acad. Sci. **101**, 2758 (2004).
[23] N. Jacques and M. Dreyfus, Mol. Microbiol. **4**, 1063 (1990).
[24] G. Zhang and Z. Ignatova, PLoS ONE **4**, e5036 (2009).
[25] M. Thattai and A. van Oudenaarden, Proc. Natl. Acad. Sci. **98**, 8614 (2000).
[26] B. Bonven and K. Gullov, Mol. Gen. Genetics **170**, 225 (1979).
[27] M. A. Suchard, K. Lange, and J. S. Sinsheimer, J. Stat. Theory Pract. **2**, 173 (2008).
[28] W. Netzer and F. Hartl, Nature (London) **388**, 343 (1997).
[29] A. Hawkins and H. Lamb, Europ. J. Biochem. **232**, 7 (1995).
[30] L. Brocchieri and S. Karlin, Proc. Natl. Acad. Sci. **92**, 12136 (1995).
[31] D. Lackner, T. Beilharz, S. Marguerat, J. Mata, S. Watt, F. Schubert, T. Preiss, and J. Bähler, Mol. Cell **26**, 145 (2007).
[32] T. Santiago, I. Purvis, A. Bettany, and A. Brown, Nucleic Acids Res. **14**, 8347 (1986).