

Research article

Open Access

## A systems approach to model natural variation in reactive properties of bacterial ribosomes

Julius H Jackson\*<sup>1,2</sup>, Thomas M Schmidt<sup>1</sup> and Patricia A Herring<sup>1,2</sup>

Address: <sup>1</sup>Department of Microbiology & Molecular Genetics, Michigan State University, East Lansing, MI 48824, USA and <sup>2</sup>Theoretical and Computational Biology Group, Department of Microbiology & Molecular Genetics, Michigan State University, East Lansing, MI 48824, USA

Email: Julius H Jackson\* - [jhjacksn@msu.edu](mailto:jhjacksn@msu.edu); Thomas M Schmidt - [tschmidt@msu.edu](mailto:tschmidt@msu.edu); Patricia A Herring - [pherjack@msu.edu](mailto:pherjack@msu.edu)

\* Corresponding author

Published: 13 July 2008

Received: 10 March 2008

BMC Systems Biology 2008, 2:62 doi:10.1186/1752-0509-2-62

Accepted: 13 July 2008

This article is available from: <http://www.biomedcentral.com/1752-0509/2/62>

© 2008 Jackson et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

**Background:** Natural variation in protein output from translation in bacteria and archaea may be an organism-specific property of the ribosome. This paper adopts a systems approach to model the protein output as a measure of specific ribosome reactive properties in a ribosome-mediated translation apparatus. We use the steady-state assumption to define a transition state complex for the ribosome, coupled with mRNA, tRNA, amino acids and reaction factors, as a subsystem that allows a focus on the completed translational output as a measure of specific properties of the ribosome.

**Results:** In analogy to the steady-state reaction of an enzyme complex, we propose a steady-state translation complex for mRNA from any gene, and derive a maximum specific translation activity,  $T_{a(\max)}$ , as a property of the ribosomal reaction complex.  $T_{a(\max)}$  has units of  $a$ -protein output per time per  $a$ -specific mRNA. A related property of the ribosome,  $\tilde{T}_{a(\max)}$ , has units of  $a$ -protein per time per total RNA with the relationship  $\tilde{T}_{a(\max)} = \rho_a T_{a(\max)}$ , where  $\rho_a$  represents the fraction of total RNA committed to translation output of  $P_a$  from gene  $a$  message.  $T_{a(\max)}$  as a ribosome property is analogous to  $k_{\text{cat}}$  for a purified enzyme, and  $\tilde{T}_{a(\max)}$  is analogous to enzyme specific activity in a crude extract.

**Conclusion:** Analogy to an enzyme reaction complex led us to a ribosome reaction model for measuring specific translation activity of a bacterial ribosome. We propose to use this model to design experimental tests of our hypothesis that specific translation activity is a ribosomal property that is subject to natural variation and natural selection much like  $V_{\max}$  and  $K_m$  for any specific enzyme.

### Background

Measures of primary sequence variation in ribosomal RNA, ribosomal protein, and non-ribosomal enzyme protein show that the corresponding gene sequences provide

a record of natural variation associated with phylogeny in bacteria and archaea [1,2]. Recent analysis of a select subset of 32 protein-coding genes, mostly associated with transcription and translation machinery, established phy-

lognetic relationships among the host genomes of 69 bacterial and archaeal organisms, and the basis for a molecular tree of variation calibrated to geological time [1]. The phylogenetic tree constructed in that study largely resembles trees based upon analyses of 16S rDNA sequences [3].

In addition to the phylogenetic perspective based on sequence variation in ribosomal RNA and ribosomal protein encoding genes, we suggest that this sequence variation underlies differences in the performance of the translational machinery. Modeling variation in the kinetics of translation, the primary energy-requiring process of microbes, should lead to an enhanced ecological perspective of microbial life, much like enzyme kinetics has provided the basis for physiological models of metabolism.

Variation in kinetic properties of enzymes has been associated with phylogenetic variation in closely related bacteria but data are less readily available for the variation in enzyme families. In some few studies,  $K_m$  and  $V_{max}$  have been used as functional parameters of enzymes to infer neutral or adaptive properties associated with allelic variation in natural populations [4,5]. Hereafter, we refer to the kinetic parameters of enzymes as specific reactive properties. These studies suggest that we should expect phylogenetic variation in measures of specific reactive properties, e.g.  $K_m$ ,  $V_{max}$ , and  $k_{cat}$  for a family of functionally conserved enzymes. Though a recent, systematic study of the natural variation in specific reactive properties of enzyme families does not appear to be available, a brief survey of literature supports the expectation for natural variation of  $K_m$ ,  $V_{max}$  in an enzyme family. For example, the published values for  $K_m$  of biosynthetic threonine deaminase (EC 4.2.1.16) vary from 0.25 mM in *Thiobacillus acidophilus* and 3.2 mM in *Proteus morgani*, to 8.0 mM in *Escherichia coli*, 13.9 mM in *Geobacillus stearothermophilus*, and 14.0 mM in *Thermus* sp. [6-10]. Similarly, the  $V_{max}$  value for purified threonine deaminase varies over a range from 2.1  $\mu\text{mol}/\text{min}/\text{mg}$  in *Pseudomonas putida* to 210  $\mu\text{mol}/\text{min}/\text{mg}$  in *E. coli* and even higher in other organisms. Overall, the reactive properties of the biosynthetic threonine deaminase (4.2.1.16) family of enzymes appear to reflect a natural variation over an approximate range greater than two orders of magnitude. Simply stated, enzymes and enzyme complexes that evolve to carry out the same reactions in a set of organisms, should be expected to catalyze these reactions with sometimes substantial differences in steady-state rates and reaction efficiencies, owing to natural variation in sequence associated with reaction properties. Variation in reaction properties may confer a fitness advantage on organisms and be fixed rapidly in populations, or be selectively neutral and accumulate in populations as a result of genetic drift.

In analogy to the natural variation of conserved reactive properties of proteins, ribosomal RNA sequences undergo natural variation that forms the basis for phylogenetic grouping of all organisms and reflects the physiological adaptations associated with the sequence variation [9]. Regions of rRNA that are most highly conserved are speculated to be associated with the ribosomal reaction properties more so than the regions which are more free to vary. Ribosomal proteins also reflect sufficiently conserved variation that allows for their use in phylogenetic grouping [1]. Since phylogenetic variation in enzyme protein sequence is associated with variation in measurable, reactive properties of the enzymes, the suggestion seems reasonable that measurable, reactive properties of ribosomes are associated with regions of rRNA and ribosomal protein sequences that undergo natural variation. Thus, reactive properties of ribosomes should also be expected to vary in ways similar to the natural variation observed for enzymes [11].

A small, phylogenetically varied subset of bacteria, grouped by fast or slow growth rates in response to nutrient availability, demonstrated variation in a measurable property of ribosomal activity termed translational power by Dethlefsen and Schmidt [12]. This translational power displayed natural variation over approximately a five-fold range [11]. By inference from those studies, we postulate that specific reactive properties of ribosomes are measurably constant for any specific bacterial phylotype or species in a constant environment, and that each reactive property is subject to natural variation under selective constraints to maintain function.

In the present study we consider the ribosome and associated translational machinery as a co-evolving complex of functionally conserved molecules with measurably varied reactive properties in distinct bacterial populations. Consequently, we expect to find that ribosomes from different bacteria vary in their steady-state output of proteins as a result of selective pressures that have fine-tuned the translational machinery for optimal activity in a given environment. In analogy to a multi-enzyme complex, we model ribosomal catalyzed translation activity as a steady-state process involving the entire translational complex as a transition-state intermediate such that the measurably important output is a functional protein. The translational apparatus complex is, then, a subsystem of the organism and produces the mass ratios of proteins required for maintenance and growth. Treatment of the translational complex as a subsystem operating in a steady-state is the key factor that makes measurement of protein output sufficient to characterize subsystem properties associated with a ribosome. Our model suggests that variation in the reactive properties of ribosomes can be detected by measuring the steady-state output of specific,

functional proteins. This model substitutes simpler, systems measures to replace the more complex, step-wise, mechanistic measures used elsewhere, [e.g. [13,14]]. Details of the design strategy to develop this measure follow.

**Methods**

**Strategy to measure reactive properties of ribosomes**

In order to measure specific reactive properties of ribosomes, the design of a model of the translational apparatus as a bacterial subsystem must take into account all factors that can influence the rate of protein output that establishes gene expression levels and the mass ratio of proteins for steady-state growth. When all such factors can be held constant, then measurement of translational output records a ribosome-specific property. Under any steady state growth condition the properties of transcription and translation machinery are constant. A few definitions, here, are useful to build a model. For this model we define expression level as the steady state mass ratio of protein,  $P_{a'}$ , from a specific gene compared to the total cell protein,  $P$ , i.e.  $P_a/P$ .

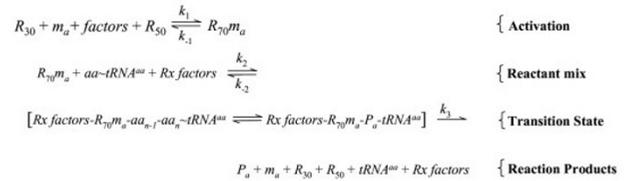
The instantaneous mass ratio or expression level of a protein from any gene in a bacterium under steady state growth conditions is determined by probabilities affecting the following steps: 1, transcription initiation (promoter strength, promoter configuration, transcription factors; and metabolites); 2, transcription termination; 3, mRNA stability; 4, ribosome binding site (e.g. Shine-Dalgarno sequence) affinity for ribosome; 5, ribosome loading at the Shine-Dalgarno (SD) or other sequences; 6, relative codon frequency pattern, i.e. codon usage; 6, translation initiation (start codon; initiation factors); 7, protein elongation (GTP, aminoacylated tRNA, elongation factors), and termination; 8, frequency of polypeptide completion; 9, protein folding; 10, protein stability; and 11, metabolite balance regulating enzyme activity and synthesis [15-18].

For this study, transcription processes in steps 1 and 2 are held constant, and steps 3–11 form the basis to model the steady-state output activity of the translation apparatus. Fig. 1 and the accompanying description refer to these steps 3–11.

**Rationale for the model**

In the exponential state of bacterial growth in a closed system, the rate of production of cell mass,  $dx/dt$ , is proportional to the initial cell mass,  $x$ , with the proportionality represented by the instantaneous growth rate constant,  $k$ , as

$$dx/dt = kx. \tag{1}$$



**Figure 1**  
**Ribosome reaction model, analogy to an enzyme complex.** Abbreviations:  $R_{30}$ , 30s ribosomal subunit;  $m_a$ , mRNA from gene  $a$ ;  $R_{50}$ , 50s ribosomal subunit;  $aa$ , amino acid;  $tRNA^{aa}$ , tRNA isoaccepting species for amino acid  $aa$ ;  $aa-tRNA^{aa}$ , aminoacylated tRNA;  $Rx$ , reaction; and  $P_a$ , protein from translation of mRNA from gene  $a$ .

In this steady state, the rate of production of any component of cell mass is commonly and reasonably assumed to be a constant fraction of the total cell mass, and  $k$  would be the expected proportionality constant for production of any component of the cell. This is a reasonable assumption and subject to a simple proof that need not be shown here. From this inference, as suggested by Neidhardt et al. [19],  $k$  defines the steady-state rate of protein production in Eq. (2).

$$k = \frac{1}{x} \frac{dx}{dt} = \frac{1}{P} \frac{dP}{dt}, \tag{2}$$

In the steady state translation output of a genome represented by Eq. (2), the protein output per unit total RNA is

$$T = \frac{kP}{R}, \tag{3}$$

where  $T$  = the protein translation output of a genome, measured in units of total protein,  $P$ , per unit total RNA,  $R$ , in a bacterium in steady state growth with instantaneous growth rate,  $k$  [12,20]. This relationship approximates a property of average translation activity per unit total RNA for an organism in a specific state of growth. We might also refer to  $T$  as the translation completion rate in recognition that not all starts of polypeptide synthesis reach completion [18]. Considering the protein product from a single gene as a regulated, fractional component of cell mass, Eq. (2) implies that the change in the mass of protein,  $P_{a'}$ , from any gene  $a$  in genome  $A$ , i.e.  $a \in A$ , is

$$dP_{a'}/dt = kP_{a'}. \tag{4}$$

It follows that a translation output  $T_{a'}$ , from any gene  $a$  per unit total RNA results from translation activity by the subset of the ribosomes translating the message from gene  $a$ . Ribosomal RNA constitutes approximately 85% of total RNA in bacteria [21], so total RNA is mostly ribosomal.

Let  $\rho_a$  represent the fraction of total RNA committed to translation output of  $P_a$  from gene  $a$  message, then the amount of  $R$  committed to producing  $P_a$  is  $\rho_a R$ . Thus  $\rho_a R$  approximates the fraction of ribosomes translating the mRNA from gene  $a$ . Since Eq. (3) defines the general relationship of the overall translational output to total RNA in a growing organism or culture, i.e. the translational power, we propose Eq. (5) as a special case of Eq. (3) to represent the specific translational output from any one specific gene,  $a$ .

$$T_a = \frac{kP_a}{\rho_a R} \quad (5)$$

## Results

### Ribosome analogy to an enzyme complex

Through simplifying assumptions, treat the 70S ribosome as a multi-subunit enzyme that catalyzes reactions to polymerize amino acids to make a functional protein product,  $P_a$ . A summary representation of this process is diagrammed as a chemical reaction sequence (Fig. 1). For any gene, treat the initiation process and assembly of the 70S ribosome as analogous to activation of the ribosome as an enzyme. The formation of the initiation complex involves binding of 30S ribosome ( $R_{30}$ ) with mRNA from gene  $a$ , ( $m_a$ ), at a Shine-Dalgarno (SD) or other ribosome binding site; f-met-tRNA; the initiation factors IF1, IF2 and IF3; GTP; and subsequent assembly with 50S ribosome ( $R_{50}$ ) to form the activated ribosome,  $R_{70}m_a$ . In this analogy, reaction cofactors are not rate limiting,  $m_a$  is an activator, and aminoacylated tRNAs ( $aa\sim tRNA^{aa}$ ) are substrate molecules. The amino acid polymerization reactions that follow utilize the pools of  $aa\sim tRNA^{aa}$  to produce the polypeptides for the functional  $P_a$  protein. Treat all steps of polypeptide chain elongation reactions as a single transition state intermediate. The overall reaction product is  $P_a$  formed in a step involving release from the ribosome-mRNA complex and assembly of the active protein. We consider the last step as rate-determining. Measurement of the rate of increase of  $P_a$  reactivity is the measure of the rate of production of functional proteins from translation of gene  $a$  message.

Continuing the analogy, the active ribosome complex,  $R_{70}m_a$ , would be expected to vary the rate of polypeptide output in response to the variables of input, especially the pool of  $aa\sim tRNA^{aa}$ , and the codon composition of the gene transcript,  $m_a$ . The rate of the translation output would be expected to vary as a function of the degree of match of the  $aa\sim tRNA^{aa}$  composition and abundance with the relative codon frequency of the transcript. We posit that a major source of the variation of specific translation activity is the degree of mismatch between the relative codon frequency and the tRNA abundance and composition. The greater the degree of mismatch, the higher the

frequency of codon-specific pauses and the greater is the probability of premature termination of polypeptide synthesis, i.e. a higher drop-off frequency. Thus, within a genome,  $A$ , the translation activity of its ribosomes would be expected to vary over a definable range up to a maximum translation activity that is a property of the ribosome.

The rate of protein output in this Ribosome Reaction Model is a function of the catalytic capacity of the activated ribosome complex acting upon the pool of  $aa\sim tRNA^{aa}$  as shown by Eq. (6).

$$dP_a/dt = k_3[R_{70}m_a] \quad (6)$$

We can see that the rate of production of  $P_a$  is the initial reaction velocity of output from the gene  $a$  message, thus

$$dP_a/dt = v_a \quad (7)$$

From Eqs. (6-7) we see that

$$k_3 = v_a/[R_{70}m_a]. \quad (8)$$

Eqs. (4-6) imply that  $T_a = \frac{k_3[R_{70}m_a]}{\rho_a R}$ , and from our definition of  $\rho_a R$  as the amount of  $R$  involved with translation of mRNA from gene  $a$ , we infer that  $\rho_a R \approx [R_{70}m_a]$  and thus  $k_3 \approx T_a$ , the specific translation activity. The units for  $k_3$  are units of  $P_a$  formed per unit time per unit  $R$  involved in translating message  $m_a$  from gene  $a$ . These units are the same as the units for  $T_a$ , the specific translation activity in Eq. (5). Under conditions where  $[R_{70}m_a]$  is maximum, all mRNA from gene  $a$  is loaded onto ribosomes, and the maximum output rate is achieved, then  $v_a$  becomes  $V_{a(max)}$ . For conditions producing  $V_{a(max)}$ ,  $k_3$  becomes analogous to  $k_{cat}$  for an enzyme reaction, as in Eq. (9).

$$k_{cat} = V_{a(max)}/[R_{70}m_a]. \quad (9)$$

Activated ribosomes within an organism will therefore be expected to display a property of maximum specific translation activity,  $T_{a(max)}$ , under optimal conditions of reactant availability. By inference, it may be possible to measure the catalytic efficiency for use of any specific  $aa\sim tRNA^{aa}$  in translation of  $m_a$  by measure of  $k_{cat}/K_m$ , i.e.  $T_{a(max)}/K_m$ . We might expect, then, that the translation properties of ribosomes in a  $R_{70}m_a$  complex are constant, are organism specific, and therefore are genome-specific.  $k_{cat}$  for the complex would be expected to vary according to  $m_a$  properties that influence the expression level.

**Single gene translation activity, factors and determinants**

Modifying Eq. (5) to introduce a term for the expression level of gene *a*, define the specific translation activity for gene *a* as

$$T_a = \frac{kP_a}{\rho_a R} = \frac{kP\lambda_a}{\rho_a R} \tag{10}$$

where  $\lambda_a = P_a/P$  is the expression level. The  $\lambda_a$  term shows the fractional contribution of functional protein from any gene, *a*, to the total protein content of the organism in any steady state of a growth cycle. Thus,  $\lambda_a$  is the mass ratio of protein  $P_a$  compared to the total cellular protein. The  $dP_a/dt$  varies among the genes of a genome such that a gene or gene set that is highly expressed has a higher production rate for  $P_a$  than genes expressed less frequently. Therefore, multiple expression states must be expected to exist according to the functional role of a gene or set of genes in the life cycle of an organism. The consequence of these multiple expression states is that the steady state translation of transcripts from any low expression gene contributes fewer proteins than a high expression gene to the total cellular protein.

The half-life of proteins that comprise  $P$  would be expected to vary over some range and that variation will constitute some frequency distribution. Considering the variation in stability of any protein  $P_{a'}$ , a more accurate representation of  $\lambda_a = P_a/P$  may be

$$\lambda_a = \frac{P_a}{\sum P_a} \tag{11}$$

In theory, the measurable total protein mass,  $P$ , will be experimentally indistinguishable from  $\sum P_a$  even if the  $t_{1/2}$  of one or several  $P_a$  is extremely short. Therefore, we expect that  $P \approx \sum P_a$ . However, in the case of an extremely short half-life for any  $P_{a'}$ , the expression of  $P_a$  would appear to be low and lead to an erroneous conclusion that the mass ratio  $\lambda_a$  is low. Though a condition of high instability of  $P_a$  may be rare, calculation of the initial mass ratio from Eq. (11) can be accomplished by substituting the initial protein output  $P_{a'}$  for  $P_a$

$$P_{a'} = P_a e^{c_a t}, \tag{12}$$

where  $c_a$  is a protein stability constant, and  $t$  is the time increment between completion of protein synthesis and sampling. These relationships derive from treating the decay of the functional protein  $P_a$  as  $P_a = P_{a'} \exp(-c_a t)$ . Note that initially, i.e. when  $t = 0$ , the measured value and initial value for  $P_a$  are equal, and when  $c_a$  is small, i.e. stable protein,  $P_a \approx P_{a'}$ . Thus  $P_{a'}$  is an estimate of the original protein mass when  $t$ ,  $c_{a'}$  and  $P_{a_s}$  are known.

Since not all translation starts,  $P_{a_s}$ , reach completion for a gene message owing to premature termination,  $\psi_a$  in Eq. (13) represents a translation completion efficiency for each gene.

$$\psi_a = \frac{P_a}{P_{a_s}} \tag{13}$$

Then  $\delta$ , a so-called "drop-off frequency" sometimes used to describe translation efficiency, may be defined as

$$\delta_a = 1 - \psi_a = 1 - \frac{P_a}{P_{a_s}} \tag{14}$$

For unstable  $P_{a'}$ , i.e. when the protein stability constant  $c_a$  is large, then it would be appropriate to substitute  $P_{a'}$  for  $P_a$  from Eq. (13) and to introduce an initial  $P_{a_s'}$  to substitute for  $P_{a_s}$  to get

$$\psi_{a'} = \frac{P_{a'}}{P_{a_s'}} \tag{15}$$

where

$$P_{a_s'} = P_{a_s} e^{c_a t} \tag{16}$$

The rationale for Eq. (16) is analogous to that for Eq. (12), and  $P_{a_s'}$  is an estimate of the original protein mass from completion of all translation starts when  $t$ ,  $c_{a'}$  and  $P_{a_s}$  are known. Generally, it would be appropriate to use the efficiency factor of Eq.(13) and reserve Eq. (15) for probably rare instances of a highly unstable  $P_a$ .

We may now also define specific translation activity for a single gene *a* as

$$T_a = \frac{kP_{a_s} \psi_a}{\rho_a R}, \tag{17}$$

and when considering the instability of proteins from any gene *a*,

$$T_a = \frac{kP_{a_s} \psi_{a'}}{\rho_a R} \tag{18}$$

**Ribosome loading and transcript stability**

We will next treat, specifically, the initiation step, also called the ribosome loading step. Thus far we have considered factors affecting the specific translation activity,  $T_a$ . The overall translation rate, however, also depends upon

the frequency of ribosome loading onto mRNA, the stability of that mRNA, and the probability and frequency of the gene transcription. The first step of the reaction sequence (Fig. 1) represents ribosomal loading onto transcript  $m_a$  from gene  $a \in A$ , described by rate equation Eq. (19).

$$\frac{d[R_{70}m_a]}{dt} = k_R[R_{30}][m_a][R_{50}] - k_{-1}[R_{70}m_a] \quad (19)$$

In Eq. (19),  $k_R$  is a third-order rate constant with units of  $(t^{-1})(\text{conc}^{-1})(\text{conc}^{-1})$ . The conditions of growth result in  $[m_a] \ll [R_{30}]$  and  $[m_a] \ll [R_{50}]$ . These conditions suggest that the association reaction may display first-order behavior. Let  $k_{+1}$  be a pseudo first-order rate constant, where  $k_{+1} = k_R [R_{30}] [R_{50}]$ , and substitute into Eq. (19) to obtain

$$\frac{d[R_{70}m_a]}{dt} = k_{+1}[m_a] - k_{-1}[R_{70}m_a]. \quad (19a)$$

We see then that Eq. (19a) shows the expected pseudo first-order behavior of  $m_a$  binding to ribosomes. At any steady state of growth, ribosomal activation can be represented by an association constant,  $K_a$ , for the message  $m_a$  from gene  $a$ , as

$$K_a = \frac{k_{+1}}{k_{-1}} = \frac{[R_{70}m_a]}{[m_a][R_{30}][R_{70}]} \quad (20)$$

This association constant can be thought of as a loading factor that is a ratio of bound to free mRNA from gene  $a$ , and can be related to the probability of message binding to ribosome. The amount of  $m_a$  undergoing translation compared to the total  $m_a$  is the relative frequency of loading and approximates the probability of ribosomal loading, i.e.

$$\pi_a \approx f_1 = \frac{[R_{70}m_a]}{[m_a] + [R_{70}m_a]}, \quad (20a)$$

where  $f_1$  is the relative loading frequency and  $\pi_a$  is the probability of message from gene  $a$  binding to ribosomes for translation. The specific translation rate,  $T_a$ , weighted by the fraction of total message loaded,  $\pi_a([m_a] + [R_{70}m_a]) = \pi_a m_{a(\text{total})}$ , gives an overall expectation for translation rate of gene  $a$  message to make the functional protein,  $P_a$  (Eq. (21)).

$$T_{\text{rate}} = \pi_a m_{a(\text{total})} T_a \quad (21)$$

Factors and growth conditions that affect transcription initiation are too numerous and variable to treat in the context of overall translation rate. Therefore, the last step we consider in this treatment of the overall specific trans-

lation rate,  $T_{\text{rate}}$ , is transcript stability. The higher the stability of  $m_a$ , the greater is the output of  $P_a$  per time, and conversely, the lower the stability of  $m_a$ , the lower the overall translation rate. Analogous to treatment of protein stability in Eq. (12),  $m_a$  represents the amount of mRNA from gene  $a$  that is available to bind a ribosome. It is easy to see that instability of  $m_a$  reduces the amount of transcript loaded for translation and reduces the overall translation rate. If the stability of mRNA, i.e. half-life, for gene  $a$  is known, it should be possible to use Eq. (22) to estimate the actual  $m_a$  available for loading,

$$m_a = m_{a'} e^{-\omega_a t}, \quad (22)$$

where  $\omega_a$  is a constant of mRNA stability for gene  $a$ , and  $m_{a'}$  is the initial amount of message.

### Maximum translation rate

Ideal conditions that include optimal mRNA stability and protein stability would be expected to maximize production of proteins from a single, protein-coding gene. Every translation start results in a completed protein under optimal conditions when  $\psi_a = 1.0$ . Thus, an expected maximum, steady state specific translation activity by any ribosome to make any  $P_a$  may be defined as a maximum specific translation activity

$$T_{a(\text{max})} = \frac{kP_{a_s}}{\rho_a R} \quad (23)$$

To approximate a  $T_{\text{max}}$  for any gene  $a$  under non-ideal laboratory or natural conditions we may use Eq. (18) to correct for protein instabilities and translation efficiency to establish that

$$T_{\text{max}} = \lim_{\psi_{a'} \rightarrow 1.0} T_a = \lim_{\psi_{a'} \rightarrow 1.0} \left( \frac{kP_{a_s} \psi_{a'}}{\rho_a R} \right) = \frac{kP_{a_s'}}{\rho_a R} \quad (24)$$

is the maximum speed at which ribosomes can translate any gene in one organism at a steady state of growth. Such optimal conditions imply existence of a  $K_{a'}$ , a maximum efficiency of ribosome binding to mRNA such that

$$K_{a'} = \frac{[R_{70}m_{a'}]}{[m_{a'}][R_{30}][R_{50}]}, \quad (25)$$

where  $m_{a'} = m_a e^{\omega_a t}$  and this also implies a corresponding  $\pi_{a'}$  for the  $K_{a'}$ . It follows, then, from Eqs. (24) and (25) that a maximum overall translation output rate is given by

$$T_{\text{rate}(\text{max})} = \pi_{a'} m_{a(\text{total})} T_{\text{max}} \quad (26)$$

It is now possible to account for the multiple effect(s) of growth conditions on  $T_{a'}$ , including factors that affect translation efficiency, e.g. the drop-off rate and completion rate.

As defined in Eq. (24), the maximum specific translation output activity for a ribosome,  $T_{\max}$ , defines a genome-specific property as suggested by Dethlefsen and Schmidt [11]. As a general property,  $T_{\max}$  defines a capacity for production of protein, and that capacity will variably express in relation to the transcript composition from a particular gene. A convenient and practical way to look at specific translation output capacity for a single gene tied to the steady state growth rate of the organism is to combine the factors that cause variability from gene to gene within a genome into the expression for maximum specific translation output activity per unit total RNA in translation of  $m_a$ . From Eq. (23) for expression of protein from a specific gene,  $a$  in genome  $A$ , we can see that

$$T_{a(\max)} = \frac{kP_a s}{\rho_a R} = \frac{kP\lambda_a}{\rho_a R\psi_a}, \quad (27)$$

and in this form it is not necessary to know or measure protein starts from gene  $a$ , i.e.  $P_a$ , as long as the expression level,  $\lambda_a$ , and the translation completion efficiency,  $\psi_a$ , are known. In the section on Discussion and Conclusions we suggest a way to estimate a value for  $\psi_a$ . When  $\rho_a$  is known, i.e. the approximate fraction of ribosomes translating  $m_a$ , the value of  $T_{a(\max)}$  will be highest because it is a measure of the protein expression per unit of total RNA translating  $m_a$ . When  $\rho_a$  is not known, a fractional approximation of  $T_{a(\max)}$  is

$$\tilde{T}_{a(\max)} = \frac{kP\lambda_a}{R\psi_a} \quad (28)$$

which has units of units  $a$ -protein per time per unit total RNA. It is clear that  $\tilde{T}_{a(\max)} < T_{a(\max)}$  because the approximation of Eq. (28) compares the gene  $a$ -protein output rate to the total translational capacity as opposed to the fraction of capacity,  $\rho_a$ , committed to  $m_a$  translation in Eq. (27). Thus,  $\rho_a$  relates  $\tilde{T}_{a(\max)}$  to  $T_{a(\max)}$ , by  $\tilde{T}_{a(\max)} = \rho_a T_{a(\max)}$ .

The maximum specific translation output rate of functional, gene  $a$  protein can be measured in multiple organisms as a means to compare this ribosome-specific property among the set of organisms by using Eq. (27). As

long as the functional proteins measured derive from an orthologous set of genes, the actual measurement of  $T_{a(\max)}$  should reveal the variation in the specific translation rate property of ribosomes from one organism with respect to that property in every other organism. Recognizing that the approximation  $\tilde{T}_{a(\max)}$  is always less than but fractionally related to  $T_{a(\max)}$  by  $\rho_a$  to a specific gene  $a$ , we suggest that this approximation will provide a basis to detect, quantify and compare variation in the specific translation rate property of a ribosome. We suggest that  $\tilde{T}_{a(\max)}$  measures the specific translation activity of ribosomes analogous to enzyme specific activity measurements in crude extracts, e.g.  $V_{\max}/P$ , and that  $T_{a(\max)}$  is analogous to the  $k_{\text{cat}}$  of a purified enzyme, e.g.  $V_{\max}/P_a$ . We propose that Eq. (27) is an idealized mathematical model of the organism-specific translation output rate as a specific, functional property of the ribosome and that Eq. (28) is a readily measurable, fractional approximation.

### Discussion and conclusion

Analogy to the reaction of an enzyme complex formed the basis to derive Eq. (27) as a simple model for specific translation activity as a specific reaction property of a ribosome. The measure of  $T_{a(\max)}$  is made by the measure of  $kP/\rho_a R$  and  $\lambda_a/\psi_a$  for the species-specific reaction property of a ribosome. Factors that make the measure species-specific are the expression level of the protein measured, captured in  $\lambda_a$  as the relative ratio of  $P_a$  to the total protein  $P$ , and the translation completion efficiency,  $\psi_a$ . This model of translation activity of a ribosome provides a way to compare ribosome function in one bacterium to another, in bacteria that differ in their growth responses to changing environments or growth conditions. A key postulate in deriving this model is the steady-state assumption that for the time interval of constant growth conditions, the level of biosynthetic intermediates is constant and sufficient to support growth at a constant rate. The steady-state assumption, in turn, permits treatment of the ribosomal reaction complex (Fig. 1) as a cellular reaction subsystem, and to measure its output as an indicator of the subsystem function and therefore as a property of the subsystem. These two simplifications make it possible to treat ribosomal protein synthesis as a "black box" subsystem comprised by many component activities, and to use only the output of functional protein to measure this subsystem property. Any gene product can be used to measure a functional property of the protein synthesis subsystem. For example, the specific activity of a biosynthetic pathway enzyme can be used to measure the reaction property of specific protein output per unit of translating RNA associated with a single gene in a crude

extract of a pure culture of a bacterium. If the  $k_{cat}$  of the enzyme is known, i.e. reaction units per mg enzyme, the expression level,  $\lambda_a$ , is the mass ratio of the amount of enzyme protein, measured by activity, to the total extract protein. The translation efficiency correlated to  $\psi_a$  can be calibrated to a scale of codon usage relative to the highest expression set of genes. The definition of  $\psi_a$  requires values for the codon usage measure to vary over the range  $0 \leq \psi_a \leq 1.0$ , and several possible calculations of bias in codon usage previously reported could be used. As an example, the codon adaptation index of Sharp and Li [21] has been shown to correlate directly with the expression-level of protein-coding genes in *Escherichia coli* and may be used as a value for  $\psi_a$  in this context [16]. We also suggest that a Pearson cross correlation measure of similarity between the relative codon frequency vector of a test gene,  $\mathbf{a}$ , and a reference vector,  $\mathbf{b}$ , of a high-expression gene set, i.e.  $r_s(\mathbf{a}|\mathbf{b})$ , may be well suited and sufficient for this measure.

Measurement of the activity of the functional protein output from a gene permits use of any protein-coding gene or set of genes to calibrate specific translation activity as the measurable reactive property or parameter of ribosomes, using  $T_{a(\max)}$  when  $\rho_a$  is known or using  $\tilde{T}_{a(\max)}$  when  $\rho_a$  is unknown. It is reasonable to expect that measurement of  $\tilde{T}_{a(\max)}$  as the specific translation activity for  $m_a$  is fractionally related to  $T_{a(\max)}$  just as the specific activity of an enzyme measured in a crude cell extract,  $V_{\max}/P$ , is fractionally related to  $k_{cat} = V_{\max}/P_a$  for the purified enzyme. So, it is appropriate to treat the measurement of the specific translation activity,  $\tilde{T}_{a(\max)}$ , for bacterial ribosomes under defined growth conditions as one would treat measuring the specific activity,  $V_{\max}/P$ , for an enzyme under defined reaction conditions in a crude cell extract.

The ribosome reaction models for this study expect values for  $T_{a(\max)}$  and  $\tilde{T}_{a(\max)}$  to vary according to the gene-specific mRNA undergoing translation and the specific reactive properties of the ribosomes within an organism. Owing to the  $\rho_a$  relationship between  $T_{a(\max)}$  and  $\tilde{T}_{a(\max)}$ , what is true for  $\tilde{T}_{a(\max)}$  is also true for  $T_{a(\max)}$ . Since it is likely that a value for  $\rho_a$  will frequently be unknown, this discussion will treat  $\tilde{T}_{a(\max)}$  as a proxy for  $T_{a(\max)}$ . A basic assumption of our model is that the reactive properties of ribosomes are constant within an organism under a defined set of growth conditions. The model pioneered by

Kurland [20,18] and uniquely applied by Dethlefsen and Schmidt [11] is based upon treating the generalized protein production per unit time per total RNA ( $kP/R$ ) as constant for a specific growth condition of an organism. If the apparent difference in translational power between ribosomes of a bacterium  $A$  compared to  $B$  can be attributed to differences in the ribosomes [11], we suggest that  $\tilde{T}_{a(\max)}$  may improve the accuracy of measurement of the ribosomal differences and offer a way to study the natural variation in the properties associated with ribosomal reactivity. Our proposed improvement in accuracy is based upon the measurement of translational output, or power, from the mRNA from an individual gene in our ribosome reaction model. This model predicts that the natural variation in specific translation activity for mRNA from any genes within a genome should show a central tendency that is characteristic for that organism, such that

$$\tilde{T}_{a_1(\max)} \approx \tilde{T}_{a_2(\max)} \approx \tilde{T}_{a_3(\max)} \dots \approx \tilde{T}_{a_n(\max)} \quad (29)$$

for specific genes  $a_1$  through  $a_n$  in the genome of an organism  $A$ . This means, of course, that while the specific translation activity for different genes will be expected to vary within a genome, an average specific translation activity will represent the ribosome property for the host organism. According to Eq. (28), the steady state condition under which the relationships of Eq. (29) would hold requires that  $\lambda_a/\psi_a$  be approximately constant for all  $a_i$ . This means, then, that if specific translation activity is an approximately constant property of the ribosomes of an organism, the ratio of the expression level to codon usage bias must also be approximately constant. This inference forms the basis of postulates that are experimentally testable and should reveal the natural variation in  $\tilde{T}_{a(\max)}$  within an organism.

Measurement of  $\tilde{T}_{a(\max)}$  should also provide the basis to compare and assess the natural variation in ribosome reactivity using the same gene or set of genes in all organisms. For that purpose, crude extracts of steady-state growth cultures will yield  $\tilde{T}_{a(\max)}$  values for each defined growth condition to compare with  $\tilde{T}_{a(\max)}$  for a different organism under a different growth condition. Expect to observe, measuring the same protein for  $n$  different bacterial genera, that  $T_{a(\max)}^{(1)}, T_{a(\max)}^{(2)}, T_{a(\max)}^{(3)}, \dots, T_{a(\max)}^{(\dots)}, T_{a(\max)}^{(n)}$  comprise a skewed frequency distribution that character-

izes the natural variation in this ribosomal reactive property.

In addition to using the reaction rate measurement  $T_{a(\max)}$  and  $\tilde{T}_{a(\max)}$  to characterize ribosomes, it should be possible to measure a  $K_m$  for the ribosome reaction model by using in vitro translation with rate-limiting concentrations of an aminoacylated tRNA as conceptually depicted in Fig. 1. Measurement of a  $K_m$  for a charged tRNA would then allow characterization of a ribosome reaction efficiency as  $T_{a(\max)}/K_m$  or  $\tilde{T}_{a(\max)}/K_m$ . A further advantage of this model is that it corrects for various specific errors, e.g. translational frameshift and other output errors, by basing the ribosome reactivity upon production of functional enzymes and measuring those functions directly.

A useful model should reflect the behavior of a system under naturally occurring as well as experimental conditions. One observation is that ribosomes are rarely saturated under a specific growth condition [11]. Achievement of a steady-state proposed by this model requires only a constant balance of reactants to achieve a stable output over time. This steady-state model applies for all degrees of saturation. This Ribosome Reaction model,  $T_{a(\max)} = \frac{kP\lambda_a}{\rho_a R \psi_a}$ , behaves such that high speed of translation is associated with high values for the growth rate,  $k$ , and expression level  $\lambda_a$ , and low values for the translation efficiency  $\psi_a$ . Conversely, slower translation speed is associated with slower growth,  $k$ , lower expression levels,  $\lambda_a$ , and higher translation efficiency,  $\psi_a$ . In other words, organisms that are rapid responders to change of environmental conditions will tend to have sloppier translation than slow responders [12], and our Ribosome Reaction model supports these translation relationships.

Development of this model provides a potential boost to successful inference of physiology from genomic sequence. If the relationships of Eq. (27) and the proportionate measurement of Eq. (28) hold in analysis of codon usage and gene expression level, it may become possible to associate specific ribosomal characteristics to bacterial lifestyle and ecological niche, i.e. to establish that ribosomes are customized for their host environment. For example, using a sufficient number of measurements of threonine deaminase specific activity in steady-state growth of a phylogenetically varied set of bacteria, along with codon usage patterns of these genes, the calcu-

lated value of  $T_{a(\max)}$  may distribute into expected ranges that are consistent with specific lifestyles from mostly dormant to rapid responders. This model provides an accessible tool and potential to measure a physiological bias of ribosomes to phylogeny and the environment in which they operate.

## Abbreviations

Except in Fig. 1 and noted in the Fig. 1 legend, only standard abbreviations were used in this paper.

## Authors' contributions

JHJ, TMS and PAH fully shared in all aspects of this study to develop a quantitative model of bacterial translation to measure natural variation in specific protein output by ribosomes. JHJ led in the model development, TMS contributed specific expertise in ribosomal properties, and PAH brought perspectives on enzyme reaction models into the study. All authors shared in all phases of evolution of the concepts, approaches, and the final construct of the model, as well as preparation of the manuscript.

## Acknowledgements

The authors are indebted to L. Dethlefsen for stimulating discussions and extensive encouragement in early stages of this work. The website databases BRENDA <http://www.brenda-enzymes.info/> and BioCyc <http://www.biocyc.org/> were used as initial information sources for comparative enzyme properties. This study was supported by NSF grants MCB-0135880 to TMS and JHJ, IBN-0421900 to TMS, and DBI-0139577 to JHJ.

## References

- Battistuzzi FU, Feijao A, Hedges SB: **A genomic timescale of prokaryote evolution: insights into the origin of methanogenesis, phototrophy, and the colonization of land.** *BMC Evol Biol* 2004, **4**:44.
- Hansmann S, Martin W: **Phylogeny of 33 ribosomal and six other proteins encoded in an ancient gene cluster that is conserved across prokaryotic genomes: influence of excluding poorly alignable sites from analysis.** *Int J Syst Evol Microbiol* 2000, **50 Pt 4**:1655-1663.
- Woese CR: **Interpreting the universal phylogenetic tree.** *Proc Natl Acad Sci USA* 2000, **97**:8392-8396.
- Demetrius L: **Selective neutrality and enzyme kinetics.** *J Mol Evol* 1997, **45**:370-377.
- Wang I-N, Dykhuizen DE: **Variation of enzyme activities at a branched pathway involved in the utilization of gluconate in *Escherichia coli*.** *Evolution* 2001, **55**:897-908.
- Eisenstein E: **Cloning, expression, purification, and characterization of biosynthetic threonine deaminase from *Escherichia coli*.** *J Biol Chem* 1991, **266**:5801-5807.
- Higa EH, Ramaley RF: **Purification and properties of threonine deaminase from the X-1 isolate of the genus *Thermus*.** *J Bacteriol* 1973, **114**:556-562.
- Muramatsu N, Nosoh Y: **Some catalytic and molecular properties of threonine deaminase from *Bacillus stearothermophilus*.** *J Biochem* 1996, **80(3)**:485-490.
- Proteau G, Silver M: **Effects of amino acids on *Thiobacillus acidophilus*. II. Threonine deaminase.** *Can J Microbiol* 1980, **26**:385-388.
- Yoshida H, Hanada K, Ohsawa H, Kumagai H, Yamada H: **Biosynthetic threonine deaminase from *Proteus morganii*.** *Agric Biol Chem* 1982, **46**:1035-1042.
- Dethlefsen L, Schmidt TM: **Performance of the translational apparatus varies with the ecological strategies of bacteria.** *J Bacteriol* 2007, **189**:3237-3245.

12. Dethlefsen L, Schmidt TM: **Differences in codon bias cannot explain differences in translational power among microbes.** *BMC Bioinformatics* 2005, **6**:3.
13. Drew DA: **A mathematical model for prokaryotic protein synthesis.** *Bull Math Biol* 2001, **63**:329-351.
14. Heyd A, Drew DA: **A mathematical model for elongation of a peptide chain.** *Bull Math Biol* 2003, **65**:1095-1109.
15. Dong H, Nilsson L, Kurland CG: **Co-variation of tRNA abundance and codon usage in *Escherichia coli* at different growth rates.** *J Mol Biol* 1996, **260**:649-663.
16. Karlin S, Mrazek J, Campbell AM: **Codon usages in different gene classes of the *Escherichia coli* genome.** *Mol Microbiol* 1998, **29**:1341-1355.
17. Karlin S, Mrazek J: **Predicted highly expressed genes of diverse prokaryotic genomes.** *J Bacteriol* 2000, **182**:5238-5250.
18. Kurland CG, Hughes D, Ehrenberg M: **Limitations of translational accuracy.** In *Escherichia coli and Salmonella Cellular and Molecular Biology Volume 1*. Edited by: Neidhardt F, Curtiss III R, Ingraham JL, Lin ECC, Low KB, Magasanik B, Reznikoff VS, Riley M, Schaechter M, Umberger HE. Washington, D. C.: ASM Press; 1996:979-1004.
19. Neidhardt FC, Ingraham JL, Schaechter M: *Physiology of the Bacterial Cell: A Molecular Approach* Sinauer Associates; 1990.
20. Kjeldgaard NO, Kurland CG: **The distribution of soluble and ribosomal RNA as a function of growth rate.** *J Molec Biol* 1963, **6**:341-348.
21. Karpinets TV, Greenwood DJ, Sams CE, Ammons JT: **RNA:protein ratio of the unicellular organism as a characteristic of phosphorus and nitrogen stoichiometry and of the cellular requirement of ribosomes for protein synthesis.** *BMC Biology* 2006, **4**:30.
22. Sharp PM, Li W-H: **The codon adaptation index – a measure of directional synonymous codon usage bias and its potential applications.** *Nucleic Acids Res* 1987, **15**:1281-1295.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)

