First-principles model of optimal translation factors stoichiometry

Jean-Benoît Lalanne[1,2,†] **and Gene-Wei Li**[1*]

***For correspondence:** gwli@mit.edu (GWL)

Present address: †Department of Genome Sciences, University of Washington, Seattle, WA 98105, USA.

[1]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA; [2]Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

Abstract Enzymatic pathways have evolved uniquely preferred protein expression stoichiometry in living cells, but our ability to predict the optimal abundances from basic properties remains underdeveloped. Here we report a biophysical, first-principles model of growth optimization for core mRNA translation, a multi-enzyme system that involves proteins with a broadly conserved stoichiometry spanning two orders of magnitude. We show that predictions from maximization of ribosome usage in a parsimonious flux model constrained by proteome allocation agree with the conserved ratios of translation factors. The analytical solutions, without free parameters, provide an interpretable framework for the observed hierarchy of expression levels based on simple biophysical properties, such as diffusion constants and protein sizes. Our results provide an intuitive and quantitative understanding for the construction of a central process of life, as well as a path toward rational design of pathway-specific enzyme expression stoichiometry.

Introduction

A universal challenge faced by both evolution and synthetic pathway creation is to optimize the cellular abundance of proteins. This abundance optimization problem is not only multidimensional – often involving several proteins participating in the same pathway – but also under systems-wide constraints, such as limited physical space (***Klumpp et al., 2013***) and finite nutrient inputs (***You et al., 2013***). The complexity of this problem has prevented rational design of protein expression for pathway engineering (***Jeschek et al., 2017***). Fundamentally, being able to predict the optimal and observed cellular protein abundances from their individual properties would reflect an ultimate understanding of molecular and systems biology.

Evolutionary comparison of gene expression across microorganisms suggests that basic principles governing the optimization problem may exist. We recently reported broad conservation of relative protein synthesis rates within individual pathways, even under circumstances in which the relative transcription and translation rates for the homologous enzymes have dramatically diverged across species (***Lalanne et al., 2018***). Moreover, distinct proteins that evolved convergently towards the same biological function also displayed the same stoichiometry of protein synthesis in their respective species. These results suggest that the determinants of optimal in-pathway protein stoichiometry are likely modular and independent of detailed biochemical or physiological properties that differ across clades. However, the precise nature of such determinants remains unknown.

Translation of mRNA into proteins is a central pathway required for cell growth and therefore serves as an entry point for establishing a quantitative model of growth-optimized in-pathway stoichiometry. As a group, the total amount of translation-related proteins per cell mass linearly increases with growth rate in most conditions (***Scott et al., 2010***; ***Dai et al., 2016***; ***Schaechter***

et al., 1958), a relationship considered a bacterial 'growth law'. In addition to ribosomes which have well-coordinated synthesis of subunits (*Nomura et al., 1984*), the translation pathway is comprised of nearly 100 protein factors involved in facilitating ribosome assembly, translation initiation, elongation, and termination (*Marintchev and Wagner, 2004*; *Dever and Green, 2012*; *Rodnina, 2018*). The intracellular abundances of these factors vary over 100-fold (*Pedersen et al., 1978*; *Li et al., 2014*), and their ratios are often maintained in different growth conditions and across different species (*Lalanne et al., 2018*). What dictates the observed stoichiometry among translation factors is less understood. Early studies predicted expression of the highly expressed elongation factor Tu (EF-Tu) relative to the ribosome (*Klumpp et al., 2013*; *Ehrenberg and Kurland, 1984*) by maximizing translational flux per unit proteome. More recently, expression of several other components involved in the elongation step (ribosomes, tRNA, mRNA, EF-Tu, and EF-Ts) was predicted by minimizing the total mass of the components at a fixed translational flux (*Hu et al., 2020*). The selective pressure on expression levels remains to be determined for most members of the translation machinery, including initiation and termination factors that are much more lowly expressed and often assumed to be non-limiting.

Here we sought to derive an intuitive model to understand the quantitative abundance hierarchy (Figure 1B) among the core translation factors (tIFs), which have well-characterized functions (Table 1, schematic in Figure 1A). Our goal is not to exhaustively model the heterogeneous movement of ribosomes on the transcriptome (*Shaw et al., 2003*; *Reuveni et al., 2011*; *Subramaniam et al., 2014*; *Dykeman, 2020*) or to include as many details of the underlying molecular steps as possible (*Hu et al., 2020*; *Vieira et al., 2016*). Instead, we coarse-grained global translation into a cycle that consists of sequential steps with interconnected fluxes that depend on core tIFs concentrations. At steady-state cell growth, all individual fluxes are matched and the overall rate of ribosomes completing the full translation cycle is proportional to cell growth. By solving for the maximum flux under proteome allocation constraints, we obtained analytical solutions for the optimal factor concentrations, which agree well with the observed values. The ratios of optimal concentrations depend only on simple biophysical parameters that are broadly conserved across species. For instance, elongation factor EF-G is predicted to be more abundant than initiation and termination tIFs by a multiplicative factor of $\approx \sqrt{\text{average number of codons per protein}} \approx 14$, whereas EF-Tu is predicted to be more abundant than EF-G by a factor of $\approx \sqrt{\text{number of different amino acids}} \approx 4$. These results, arising from the optimization procedure and generic properties of the translation cycle, provide rationales for the order-of-magnitude expression of these important enzymes.

Results

Problem statement and model formulation

Our overall goal is to determine the growth-optimizing proteome allocation for the core translation factors. Conceptually, varying tIF concentrations has two opposing effects on cell proliferation. At the biochemical level, high tIF expression can facilitate growth by allowing more efficient usage of ribosomes. At the systems level, increased tIF expression can nonetheless limit growth by reducing the number of ribosomes and other proteins that can be produced. The tradeoffs between various tIFs and ribosomes create a multidimensional optimization problem.

We solve this multidimensional problem by treating translation as a dynamical system, in which ribosomes cycle through initiation, elongation, and termination. The resulting flux drives cell growth. During steady-state growth, every interlocked step of the translation cycle must have the same ribosome flux that is specified by the growth rate. We show that at the growth optimum, concentrations for distinct tIFs can be solved independently. The resulting analytical solutions can be expressed in terms of the growth rate and simple biophysical parameters.

Table 1. Brief description of the function of core translation factors considered. For reviews of mRNA translation, see (***Rodnina, 2018***; ***Chen et al., 2016***).

Step	Factor	Function
Initiation	IF1	Initiation factor 1: binds to 30S ribosome subunits to facilitate initiator tRNA binding (***Laursen and Sørensen, 2005***; ***Gualerzi and Pon, 2015***).
Initiation	IF2	Initiation factor 2: ribosome-dependent GTPase interacting with 30 ribosome subunits, ensures correct binding of initiator tRNAs (***Laursen and Sørensen, 2005***; ***Gualerzi and Pon, 2015***).
Initiation	IF3	Initiation factor 3: prevents premature docking of 50S ribosomal subunits (***Laursen and Sørensen, 2005***; ***Gualerzi and Pon, 2015***).
Elongation	EF-Tu	Elongation factor Tu: binds to charged tRNAs to form ternary complexes, brings charged tRNAs to empty ribosome A sites. (***Weijland et al., 1992***; ***Agirrezabala and Frank, 2009***; ***Andersen et al., 2003***)
Elongation	aaRS	tRNA synthetases: charge tRNAs with cognate amino acids (***Ibba and Dieter, 2000***; ***Pang et al., 2014***).
Elongation	EF-G	Elongation factor G: catalyzes translocation steps of the ribosome after peptide bond formation (***Andersen et al., 2003***; ***Agirrezabala and Frank, 2009***).
Elongation	EF-Ts	Elongation factor Ts: nucleotide exchange factor for EF-Tu (***Agirrezabala and Frank, 2009***; ***Andersen et al., 2003***).
Termination	RF1/RF2	Peptide chain release factors 1 and 2: recognize stop codon and hydrolyze the completed protein. RF1 recognizes UAA, UAG, and RF2 UAA, UGA (***Bertram et al., 2001***).
Termination	RF4	Ribosome recycling factor: catalyzes the dissociation of ribosome subunits following peptide chain release in translation termination (***Bertram et al., 2001***).

Cell growth driven by tIF-dependent ribosome flux

To describe the biochemical effects of tIF concentrations on cell growth, we first introduce a coarse-grained translation cycle time τ_{tl}, or the time it takes for a ribosome to complete a typical cycle of protein synthesis (Figure 1A), which consists of three sequential steps: initiation ("ini"), elongation ("el"), and termination ("ter"). Each of these steps is catalyzed by multiple tIFs. The full translation cycle time is then sum of ribosome transit times at the three steps ($\tau_{tl} = \tau_{ini} + \tau_{el} + \tau_{ter}$), whose dependence on individual tIF concentrations can be quantitatively described through mass action kinetic schemes (schematically depicted in Figure 1A, see Appendices 2, 3, and 4 for details and examples below). We express tIF concentrations in units of proteome fractions (dry mass fraction of a specified protein to the full proteome), denoted by ϕ (***Scott et al., 2010***) (Materials and Methods, section Conversion between concentration and proteome fraction). Using this notation, the translation cycle time τ_{tl} is a decreasing function of various tIFs concentrations ($\{\phi_{tlF,i}\}$).

In addition to its dependency on tIF concentrations, the translation cycle time provides a bridge between the cell growth rate and ribosome concentration. In steady-state growth (***Monod, 1949***; ***Scott et al., 2010***; ***Dai et al., 2016***), the growth rate of cells and of their protein content (total number of proteins) must be identical, denoted here as λ, as a result of the constant average cellular composition. The protein content grows at a rate determined by the flux of active ribosomes completing the translation cycle, that is N_{ribo}^{act}/τ_{tl}, where N_{ribo}^{act} is the number of active ribosomes per cell, divided by the total number of proteins N_P per cell: $\lambda = N_{ribo}^{act}/\tau_{tl}N_P$. Active ribosomes are defined as those functionally engaged in, and cycling through, the initiation, elongation, and termination reactions of peptide synthesis. Rescaling to the total mass fraction (Materials and Methods, section Conversion between concentration and proteome fraction) of proteome for active ribosomes (ϕ_{ribo}^{act}) yields

$$\lambda = \frac{\phi_{ribo}^{act}}{\tau_{tl}} \frac{\langle \ell \rangle}{\ell_{ribo}}, \tag{1}$$

where ℓ_{ribo} is the number of amino acids in ribosomal proteins and $\langle \ell \rangle$ is the average number of codons per protein, weighted by expression levels (Materials and Methods, section Average number

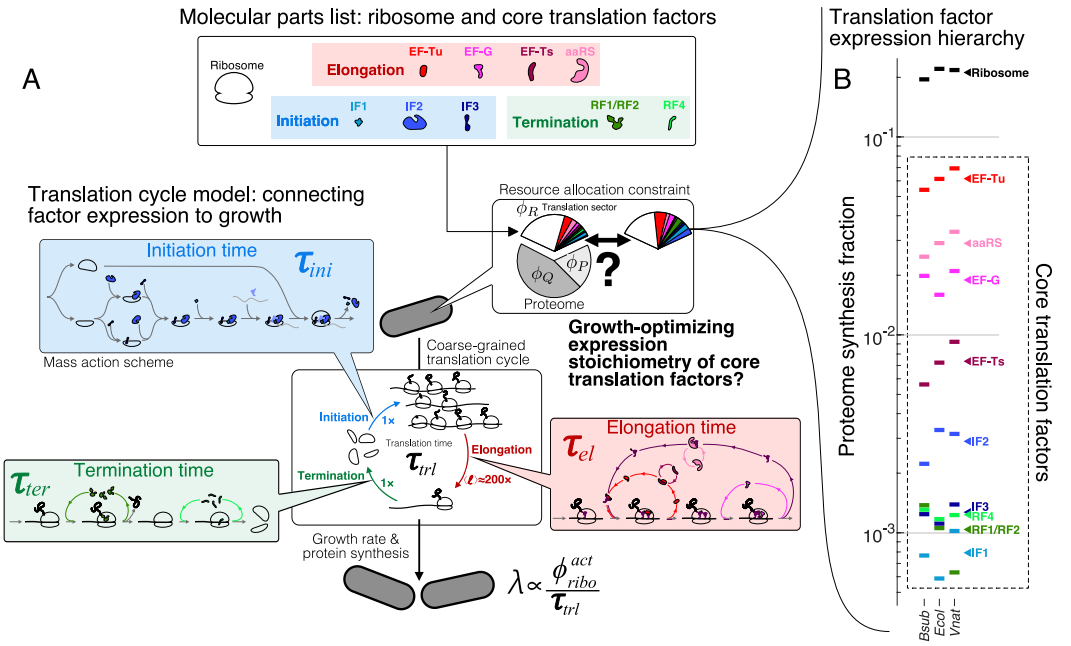


Figure 1. The hierarchy of mRNA translation factor expression stoichiometry. (A) Multiscale model relating translation factor expression to growth rate. The growth rate λ is directly proportional to the active ribosome content (ϕ_{ribo}^{act}) in the cell and inversely proportional to the average time to complete the translation cycle τ_{tl}, consisting of the sum of the initiation (τ_{ini}), elongation (τ_{el}), and termination (τ_{ter}) times. Each of these reaction times are determined by the translation factor abundances. On average, the elongation step is repeated around $\langle \ell \rangle \approx 200\times$ to complete a full protein, compared to $1\times$ for initiation and termination. Our framework of flux optimization under proteome allocation constraint addresses what ribosome and translation factor abundances maximize growth rate. (B) Measured expression hierarchy of bacterial mRNA translation factors, conserved across evolution. Horizontal bars mark the proteome synthesis fractions as measured by ribosome profiling (***Lalanne et al., 2018***) (equal to the proteome fraction by weight for a stable proteome) for key mRNA translation factors in *B. subtilis* (*Bsub*), *E. coli* (*Ecol*), and *V. natriegens* (*Vnat*) and are color-coded according to the protein (or group of proteins) specified. Triangles (◀) on the right indicate the mean synthesis fraction of the protein in the three species. See Table 1 for a short description of the translation factors considered.

Figure 1–source data 1. Synthesis fractions in (B) can be found in Supplementary File 1.

of codons per protein: $\langle \ell \rangle$). The rescaling factor ($\ell_{ribo}/\langle \ell \rangle \approx 7300/200 = 36.5$) is approximately constant across growth conditions (Matrials and Methods, section Average number of codons per protein: $\langle \ell \rangle$). This equation establishes how tIF concentrations affect the growth rate biochemically via τ_{tl}.

We note that equation 1 is a generalized form of the bacterial growth law that relates the mass fraction of elongating ribosomes to growth rate ($\lambda = \frac{\phi_{ribo}^{el}}{\tau_{el}} \frac{\langle \ell \rangle}{\ell_{ribo}} = \gamma \phi_{ribo}^{el}$, where γ is a rescaled translation elongation rate and ϕ_{ribo}^{el} is the proteome fraction of actively translating ribosomes (***Scott et al., 2010***; ***Dai et al., 2016***; ***Scott et al., 2014***)). This classic growth law was derived by considering the steady-state flux of peptide bond formation by elongating ribosomes, whereas our model focuses on the flux of ribosomes that traverse the entire translation cycle, thereby allowing us to consider the effects of translation factors and ribosomes engaged in additional steps (initiation, elongation, and termination). For each step, equation 1 can be extended to show that the growth rate is similarly proportional to the mass fraction of the corresponding ribosomes divided by the transit time at that step (Materials and Methods, section Equality of ribosome flux in steady-state).

Steady-state growth thus imposes the requirement that the growth rate be inversely proportional to the translation cycle time and proportional to the number of active ribosomes engaged in the translation cycle (equation 1). Inactive ribosomes, comprised of assembly intermediates, hibernating ribosomes, or otherwise non-functional ribosomes, have been found to constitute

a small fraction (≈5%) of the total ribosome pool for fast growth (***Lindahl, 1975***; ***Dai et al., 2016***). Based on equation 1, both increasing ribosome concentration and increasing tIF concentrations (which decreases τ_{tl}) can accelerate growth. However, production of ribosomes and tIFs is subject to competition under a limited proteomic space, which we consider next.

Optimization under proteome allocation constraint

To model the production cost tradeoff between tIFs and ribosomes, we integrate the flux-based formulation above with a proteomic constraint. Assuming that components of the translation machinery together accounts for a fixed fraction of proteome, i.e., the 'translation sector' ϕ_{tl} (denoted ϕ_R in the context of growth laws (***Scott et al., 2010***)), the proteome fraction for active ribosomes is related to the proteome fraction for translation factors via

$$\phi_{ribo}^{act} = \phi_{tl} - \phi_{ribo}^{inact} - \sum_i \phi_{tlF,i}. \tag{2}$$

Equations 1 and 2, together with to the kinetic schemes for each step of the translation cycle, constitute the core of our model. Combining the biochemical effects (equation 1) and the systems-level constraints (equation 2) on tIFs, we arrive at a self-contained relationship between growth and tIF concentrations:

$$\lambda = \frac{\phi_{tl} - \phi_{ribo}^{inact} - \sum_i \phi_{tlF,i}}{\tau_{tl}(\{\phi_{tlF,i}\})} \frac{\langle \ell \rangle}{\ell_{ribo}}, \tag{3}$$

where we explicitly express τ_{tl} as a function of $\phi_{tlF,i}$ to reflect the dependence of ribosome transit times on translation factor abundances. The above relationship (equation 3) allows us to ask: what is the stoichiometry of tIFs, or partitioning of the translation sector, that maximizes the growth rate (Figure 1A)?

The condition for the optimal TF abundances, i.e., the set of $\phi_{tlF,i}$ that satisfies $\left(\partial\lambda/\partial\phi_{tlF,i}\right)^* = 0$, can be obtained by considering the $\phi_{tlF,i}$ as independent variables and taking the derivative of equation 3 with respect to a specified tIF abundance. Under the assumptions that the translation sector (ϕ_{tl}) and the proteome fraction for inactive ribosomes (ϕ_{ribo}^{inact}) are both fixed in a given external nutrient condition, this yields

$$\left(\frac{\partial\tau_{tl}}{\partial\phi_{tlF,i}}\right)^* = -\frac{\langle \ell \rangle}{\ell_{ribo}} \frac{1}{\lambda^*}, \tag{4}$$

where the asterisk refers to the growth optimum within our model, i.e., $\left(\partial\lambda/\partial\phi_{tlF,i}\right)^* = 0$. Hence, under this framework, the tIF abundances are growth-optimized when the sensitivity of the translation cycle time to changing the considered tIF abundance ($\partial\tau_{tl}/\partial\phi_{tlF,i}$) reaches a value determined solely by the growth rate and protein size factors. We emphasize that the derivative above corresponds to a perturbation scenario in which the tIF abundance is changed while maintaining fixed the total proteomic resources to the translation sector, as prescribed by our optimization procedure. As such, it does not correspond an actual perturbation easily realizable experimentally.

Although equation 3 and the resulting optimization conditions (equation 4, one for every tIF) corresponds to a coupled nonlinear system of multiple $\phi_{tlF,i}$, substantial decoupling occurs at the optimal growth rate. In this situation, most $\phi_{tlF,i}$ are only connected through the resulting growth rate. The optimization problem is then further simplified by the fact that the translation cycle consists of sequential and largely independent steps. The translation cycle time τ_{tl} corresponds to the sum of the coarse-grained initiation, elongation, and termination times, i.e., $\tau_{tl} = \tau_{ini} + \tau_{el} + \tau_{ter}$. Given that each tIF is involved in a specific molecular step, the sensitivity matrix of these times to tIF concentration is sparse: $\left(\partial\tau_j/\partial\phi_{tlF,i}\right)^* = 0$ for most combinations of τ_j and $\phi_{tlF,i}$. This lack of 'cross-reactivity' expresses that, for example, the initiation time τ_{ini} is unaffected by the tRNA synthetase concentration. This sparsity only occurs at the optimal expression levels, as the transit times typically depend on the growth rate (see an example in section Non binding-limited regime (one stop codon)) and $\partial\lambda/\partial\phi_{tlF,i} \neq 0$ away from the optimum. The optimum condition for factor i

then simplifies to:

$$\left(\frac{\partial \tau_j}{\partial \phi_{tlF,i}}\right)^* = -\frac{\langle \ell \rangle}{\ell_{ribo}} \frac{1}{\lambda^*}, \tag{5}$$

where j above denotes the translation step(s) that tlF_i participates in. This leads to simplifications that allow the system to be solved analytically in most cases: instead of solving the full system at once, individual reactions within the translation cycle can be considered in isolation. The resulting optimal concentrations are connected via the growth rate λ^*. Interestingly, the optimal stoichiometry among most tlFs is independent of λ^* if the reactions are in the binding-limited regime, as we show below.

Case study: Translation termination

We first illustrate the process of solving for the optimal tlF concentration for the relatively simple case of translation termination. The principles used here and the form of solutions provide conceptual guideposts for solving other steps of the translation cycle.

In bacteria, translation termination (***Bertram et al., 2001***) consists of two distinct, sequential steps: (1) stop codon recognition and peptidyl-tRNA hydrolysis catalyzed by class I peptide chain release factors RF1 and RF2, followed by (2) dissociation of ribosomal subunits from the mRNA, i.e., ribosome recycling, catalyzed by RF4. We do not explicitly consider the additional factors (e.g., RF3 and EF-G) due to their lack of conservation or because they are non-limiting for this specific step (Appendix 2, section Omitted molecular details). RF1 and RF2 have the same molecular functions but recognize different stop codons (***Scolnick et al., 1968***): RF1 recognizes stops UAA and UAG, whereas RF2 recognizes UAA and UGA. For simplicity, we describe here a scenario where RF1 and RF2 have no specificity towards the three stop codons, which allows us to combine them in a single factor (denoted RFI). The model is readily generalized, with similar results, to the case of the two RFs with their specificity towards the three stop codons (Appendix 2, section Full three stop codons model).

Under a coarse-grained description, the total ribosome transit time at termination τ_{ter} can be decomposed into a sum of peptide release time and ribosome recycling time. In the treatment below, we consider a regime of binding-limited reactions for simplicity (rapid catalytic rate). A full model with catalytic components can also be solved analytically (Appendix 2, section Non binding-limited regime (one stop codon), Figure 2A). In the binding-limited regime ($k_{cat} \to \infty$), the peptide release time and ribosome recycling time are inversely proportional to the corresponding tlF concentrations:

$$\tau_{ter} = \frac{1}{k_{on}^{RFI} \phi_{RFI}} + \frac{1}{k_{on}^{RF4} \phi_{RF4}}, \tag{6}$$

where the association rate constants k_{on}^i are rescaled by the factor's sizes in proteome fraction units (Materials and Methods, section Conversion between concentration and proteome fraction). The above expression constitutes the solution of the mass action scheme for termination, connecting factor abundances to termination time.

The termination time (equation 6) can then be directly substituted into the optimality condition (equation 5) and solved in terms of λ^*:

$$\phi_{RFI}^* = \sqrt{\frac{\ell_{ribo} \lambda^*}{\langle \ell \rangle k_{on}^{RFI}}}, \quad \phi_{RF4}^* = \sqrt{\frac{\ell_{ribo} \lambda^*}{\langle \ell \rangle k_{on}^{RF4}}}. \tag{7}$$

If the reactions are not binding-limited, an additional catalytic term $\propto \lambda^*/k_{cat}$ is added to the minimally required levels above (Appendix 2, section Non binding-limited regime (one stop codon)). The square-root dependence in the optimal RF concentrations emerges from the ϕ_i^{-1} dependence of τ_i, e.g., for ribosome recycling $\tau_{recyc} \propto \phi_{RF4}^{-1}$, which becomes $(\phi_i^*)^{-2}$ upon taking the derivative in the optimality condition (equation 5). The square root is then obtained by solving for ϕ_i^*. A similar square-root dependence has been noted in optimization of the ternary complex and tRNA

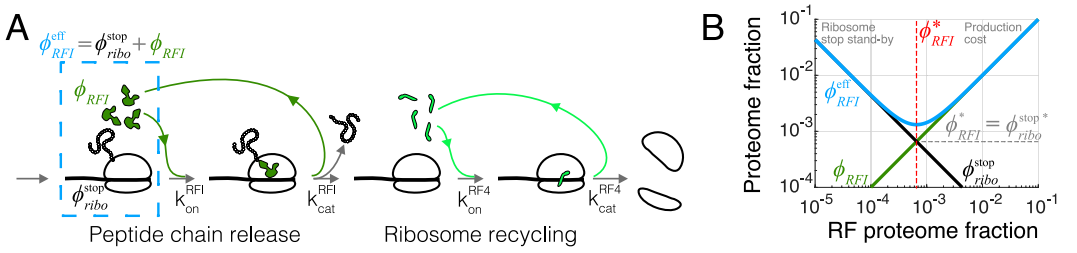


Figure 2. Case study with translation termination (A) Coarse-grained translation termination scheme. (B) Illustration of the minimization of effective proteome fraction corresponding to peptide chain release factors, leading to the equipartition principle.

abundances (***Ehrenberg and Kurland, 1984***; ***Berg and Kurland, 1997***). Analysis of tIF expression across slower growth conditions supports the derived square root dependence (Figure 4-Figure supplement 2). As a result of the square-root, the optimal RF concentrations are weakly affected by biophysical properties such as the association rate constants and protein sizes. In the binding-limited regime above, the ratio of the optimal concentrations between RFI and RF4 is independent of the growth rate and only depends on the kinetics of binding.

As a side note, the expression for termination time τ_{ter} in equation 6 must be modified in a regime where ribosomes are frequently queued upstream of stop codons. This would occur if the termination rate were slow and approached initiation rates on mRNAs (***Bergmann and Lodish, 1979***; ***Lalanne et al., 2021***). In this regime, queues of ribosomes at stop codons would incur an additional time to terminate. In a general description, the resulting additional termination time can be absorbed in a queuing factor $\mathcal{Q}$: $\tau_{ter}^{full} := \tau_{ter}\ \mathcal{Q}(\tau_{ter})$ (Appendix 1 for derivation and discussion). The resulting nonlinearity would forbid the decoupling in the optimization procedure between RFI and RF4. Although absolute rates of termination are difficult to measure *in vivo*, translation on mRNAs is generally thought to be limited at the initiation step (***Laursen and Sørensen, 2005***), and consistently, ribosome queuing at stop codons in bacteria is not usually observed (except under severe perturbations, e.g., (***Kavčič et al., 2020***; ***Baggett et al., 2017***; ***Mangano et al., 2020***; ***Saito et al., 2020***; ***Lalanne et al., 2021***)). In the physiological regime of fast termination, the queuing factor converges to 1, yielding simple solutions that depend only on biophysical parameters (equations 7).

Equipartition between tIF and corresponding ribosomes

The optimal tIF concentrations (e.g., equation 7) can also be intuitively derived from another viewpoint. For each reaction in the translation cycle, we can define an effective proteome fraction allocated to that process, combining the proteome fractions of the corresponding tIF and the ribosomes waiting at that specific step. As an example, for the case of peptide chain release factor (RFI) just treated, the effective proteome fraction includes the release factors and ribosomes with completed peptides waiting at stop codons (dashed box in Figure 2A), i.e., $\phi_{RFI}^{eff} := \phi_{RFI} + \phi_{ribo}^{stop}$. This effective proteome fraction corresponds to the total proteomic space associated to a tIF in the context of the translation cycle.

During steady-state growth, the concentration of ribosomes waiting at any specific step of the translation cycle is equal to the total active ribosome concentration multiplied by the ratio of the transit time of that step to the full cycle: e.g., here $\phi_{ribo}^{stop} = \frac{\tau_{stop}}{\tau_{tl}}\phi_{ribo}^{act}$, where $\tau_{stop} = 1/(k_{on}^{RFI}\phi_{RFI})$ is the time to arrival of RFI. Using equation 1 for ϕ_{ribo}^{act}, the effective proteome fraction satisfies:

$$\begin{aligned}\phi_{RFI}^{eff} := \phi_{RFI} + \phi_{ribo}^{stop} &= \phi_{RFI} + \frac{1}{\phi_{RFI}}\frac{\lambda}{k_{on}^{RFI}}\frac{\ell_{ribo}}{\langle \ell \rangle} \\ &\geq 2\sqrt{\frac{\lambda}{k_{on}^{RFI}}\frac{\ell_{ribo}}{\langle \ell \rangle}}.\end{aligned} \tag{8}$$

In the last line, we used the inequality of arithmetic and geometric means ($a + b \geq 2\sqrt{ab}$) to obtain

the minimum of the effective proteome fraction. The equality holds when the two proteome fractions are equal ($\phi_{RFI} = \phi_{ribo}^{stop}$), which provides the solution for optimal ϕ_{RFI}:

$$\phi_{RFI}^{*} = \sqrt{\frac{\ell_{ribo}\lambda^{*}}{\langle \ell \rangle k_{on}^{RFI}}}, \tag{9}$$

Hence, we recover equation 7 by minimizing the effective proteome fraction allocated to a given process in the translation cycle (the above argument applies to the optimal free concentration in the non-binding limited regime, see Appendix 2, section Non binding-limited regime (one stop codon) for an example). From this perspective, optimization of the translation apparatus balances the production cost of the enzyme of interest with the improved efficiency of a having less ribosomes idle at that step, Figure 2B. The optimal abundance in our model corresponds to a point of equipartition: the proteome fraction of free cognate factors equals the proteome fraction of ribosomes waiting at the corresponding step (Figure 2B).

Case study: Ternary complex and tRNA cycle (EF-Tu and aaRS)

We next consider a more complex step of the translation cycle – elongation – and demonstrate that the optimality criterion (equation 5) can similarly provide simple analytical solutions in the physiologically relevant regime. Translation elongation involves multiple interlocked cycles (one for each chemical species) and enzymes (EF-Tu, EF-G, EF-Ts, aminoacyl-tRNA synthetases (aaRS), and more). Our simplified kinetic scheme for translation elongation is shown in Figure 3A: charged tRNAs are brought to ribosomes through a ternary complex (TC), corresponding to a bound tRNA and EF-Tu. Following tRNA delivery and GTP hydrolysis, EF-Tu is released from the ribosome, and nucleotide exchange factor EF-Ts recycles EF-Tu back into the active pool, after which EF-Tu can bind a charged tRNA again and form another TC. At the ribosome, translocation to the next codon is catalyzed by EF-G, followed by release of uncharged tRNAs. Aminoacyl-tRNA synthetases then charge tRNAs to complete the elongation cycle.

To reduce the complexity due to different tRNA isoacceptors and aaRSs, we self-consistently coarse-grained the translation elongation cycle to have a single codon (derived in Appendix 3, section Coarse-grained one-codon model). The resulting model harbors a single effective species for tRNA, aaRSs, and TCs, respectively. A rescaling factor ($1/n_{aa} \approx 1/20$, estimated in section Estimation of coarse-grained rates) arises in the procedure to decrease the rates of codon specific reactions and can be attached to either the respective rate constants or chemical species concentrations. In our formulation, we choose to rescale the association rate constants such that the coarse-grained abundance for each effective species corresponds to the sum over all individual codon-specific components. For example, ϕ_{aaRS} in our coarse-grained model corresponds to the summed proteome fraction of all aaRSs in the cell, and its association rate constant with the total tRNAs is rescaled by a factor of $1/n_{aa}$.

As a result of this choice of rescaling within our coarse-grained model, there are two classes of reactions in the elongation cycle that are distinguished by different kinetics: those that were codon specific (scaled by $1/n_{aa}$) and those that are not. Codon-specific reactions, e.g., aaRS binding to cognate tRNAs and TC binding to cognate codons, are coarse-grained into one-codon reactions with reduced association rate constants (marked by # in Figure 3A). By contrast, codon-agnostic reactions do not incur such a rescaling and are thus much faster. We refer to this as a separation of timescale between the two classes of reactions (codon-specific vs. codon-agnostic), and note that this is not a reflection of slower underlying microscopic bimolecular reaction rates, but rather a result of our choice of variable in the coarse-graining.

Similar to translation termination, the factor-dependent ribosome transit time through a single codon (τ_{aa}) is comprised of two steps, corresponding to binding of the TC and EF-G, respectively (formal derivation and non binding-limited regime in Appendix 3, section Coarse-grained translation

elongation time):

$$\tau_{aa} = \frac{1}{\frac{k_{on}^{TC}}{n_{aa}}\phi_{TC}} + \frac{1}{k_{on}^{G}\phi_{G}}. \tag{10}$$

The coarse-grained factor-dependent portion of the total translation elongation time in our model is then given by the single codon time above multiplied by the average number of codons per protein, i.e., $\langle \ell \rangle \tau_{aa}$. As discussed above, the rescaling of the TC association rate constant by n_{aa}^{-1} arises as a result of our coarse-graining to a one-codon model (Appendix 3, section Coarse-grained one-codon model). Note that the ternary complex concentration, ϕ_{TC}, is a nonlinear function of the concentrations of all elongation factors (including ϕ_G).

Despite the complexity of τ_{aa} as a function of the $\phi_{tIF,i}$, the fact that all fluxes are equal in steady-state allows several steps to be isolated and solved separately (EF-Ts and EF-G, greyed out in Figure 3A, respectively solved in Appendix 3, sections Optimal EF-Ts abundance and Optimal EF-G abundance). For example, the approximate binding-limited solution for optimal EF-G concentration parallels that for termination factors:

$$\phi_G^* \approx \sqrt{\frac{\ell_{ribo}\lambda^*}{k_{on}^{G}}}. \tag{11}$$

Importantly, the optimum for EF-G is larger than the optimum for RFs by a factor $\sqrt{\langle \ell \rangle}$, reflecting that the typical translation cycle to produce a protein requires $\langle \ell \rangle$ steps catalyzed by EF-G and only one step for RFs (i.e, $\langle \ell \rangle \tau_{aa}$ enters the optimality condition, equation 5, in contrast to τ_{ter} which is not multiplied by a scaling factor). The square root dependence arises here for the same reason as in the case of translation termination (derivative of ϕ^{-1}).

In contrast to EF-G and EF-Ts, EF-Tu and aaRS cannot *a priori* be treated in isolation because the TC is composed of both EF-Tu and charged tRNAs. Still, the separation of timescales within our coarse-grained model (see Appendix 3, section Interpretation of the sharp separation between aaRS and EF-Tu limited regimes) simplifies the solution considerably. Indeed, rapid binding of charged tRNAs to EF-Tu leads to either component being limiting for ternary complex concentration in most of the aaRS/EF-Tu expression space, leading to two clearly delineated regimes (Figure 3B). In one regime, charged tRNAs are limiting (low aaRS), whereas EF-Tu is limiting in the other (low EF-Tu). These regimes are separated by a narrow transition region, whose sharpness is a reflection of the smallness of the rate rescaling parameter n_{aa}^{-1} (see Appendix 3, section Interpretation of the sharp separation between aaRS and EF-Tu limited regimes). We term the focal region separating the two regimes in the aaRS/EF-Tu expression space the "transition line" (see Box 1 for derivation and additional details).

The transition line corresponds to conditions in which EF-Tu and aaRS are co-limiting for TC concentration. In the EF-Tu limited region, increasing aaRS abundance does not increase ternary complex concentration: since all EF-Tu proteins are already bound to charged tRNAs, increasing tRNA charging cannot further increase TC concentration. Conversely, in the aatRNA limited region, increasing EF-Tu abundance does not increase TC concentration: since all charged tRNAs are already bound by EF-Tu, increasing EF-Tu concentration does not alleviate the requirement for more charged tRNAs. Given that the optimality condition requires non-zero increase in ternary complex concentration with increasing factor abundance (equation 5 using τ_{aa} from equation 10), the optimal EF-Tu and aaRS abundances must be on the transition line.

Which point on the transition line corresponds to the optimum? Note that inside the EF-Tu limited region, the ternary complex concentration is entirely set by the total EF-Tu concentration: $\phi_{TC} \approx \phi_{Tu}$ (since most EF-Tu proteins are bound by charged tRNAs, Figure 3-Figure supplement 1). As an approximation resulting from the narrow range of transition region (Figure 3 and Figure 3-Figure supplement 1), we assume that the EF-Tu limited regime solution $\phi_{TC} \approx \phi_{Tu}$ holds up to very close to the transition line. Replacing ϕ_{TC} by ϕ_{Tu} in the elongation time equation 10 and substituting in the optimality condition (equation 5), the approximate optimal abundance for EF-Tu (the full solution

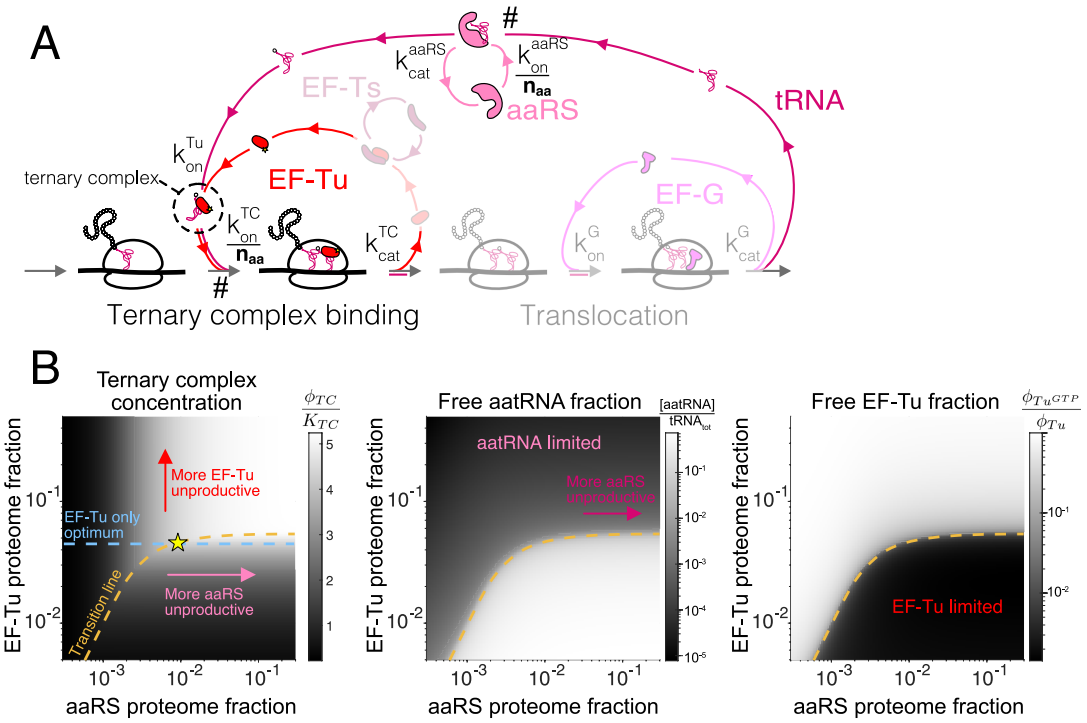


Figure 3. Case study with elongation factors (EF-Tu/aaRS) (A) Schematic of the translation elongation scheme, with the tRNA cycle, involving aminoacyl-tRNA synthetases (aaRS) and EF-Tu. Reactions with a # have their association rate constants rescaled by a factor of $n_{aa}^{-1} \approx 1/20$ through our coarse-graining to a single codon model. Greyed out cycles (EF-Ts and EF-G) can be solved in isolation (Appendix 3, sections Optimal EF-Ts abundance and Optimal EF-G abundance). (B) Exploration of the aaRS/EF-Tu expression space from numerical solution of the elongation model (Appendix 3, section Optimal EF-Tu and aaRS abundances). The transition line (orange) marks the boundary between the EF-Tu limited and aaRS limited regimes. Left panel shows the ternary complex concentration (which is closely related to the elongation rate, equation 10). The ternary complex concentration is scaled by the dissociation constant K_{TC} to the ribosome A site (see equation 39). Middle panel shows the free charged tRNA fraction. Right panel shows the free EF-Tu fraction ($\phi_{Tu^{GTP}}$ denotes the proteome fraction of EF-Tu GTP that can bind to charged tRNAs to form the ternary complex). The star marks the optimal solution, as described in the text.

Figure 3–Figure supplement 1. Geometrical interpretation of the sharpness of the separation of the aaRS limited and EF-Tu limited regimes.

Figure 3–source code 1. Source code to obtain panel (B) can be found in the associated scripts submitted with this work.

includes additional terms from the EF-Ts cycle, section Optimal EF-Tu and aaRS abundances) can then be obtained in the same way as for translation termination factors:

$$\phi_{Tu}^{*} \approx \sqrt{\frac{\ell_{ribo} n_{aa} \lambda^{*}}{k_{on}^{TC}}}. \tag{12}$$

Importantly, compared to the solution for EF-G, the above is multiplied by an additional factor of $\sqrt{n_{aa}}$. This contribution arises from the rescaling of the association rate for the ternary complex to the ribosome in our coarse-grained one-codon model, increasing the requirement on EF-Tu abundance.

From the necessity for the combined EF-Tu and aaRS solution to fall on the transition line, the approximate solution for the optimal aminoacyl-tRNA synthetase abundance is then the intersection (yellow star in Figure 3B) of the transition line with the EF-Tu-only solution described above (dashed blue line in Figure 3B, derivation of solution in Box 1).

For the above derivation to be valid, the total number of tRNAs in the cell must be sufficient to accommodate all ribosomes (about 2 per ribosome, A- and P-sites) and binding to all EF-Tu (about

> 4 per ribosome based on endogenous expression stoichiometry (***Li et al., 2014***; ***Lalanne et al., 2018***)). The number of tRNAs per ribosomes in the cell should thus be at least $6\times$. Remarkably, estimates of this ratio in the cell suggest that this is barely the case (between 6-7 tRNAs/ribosome at fast growth (***Dong et al., 1996***)). Although our model treats the total tRNA abundance as a measured parameter and omits its selective pressure (see (***Hu et al., 2020***) which includes RNA mass in their optimization procedure), the abundance of three core components of the tRNA cycle appear to be at the special point where the transition line plateau, that is set by total tRNA abundance, just crosses the EF-Tu-only optimum (blue line in Figure 3B). At this point, all three components are co-limiting.

Optimal stoichiometry of mRNA translation factors

Analogous to the case studies above, optimal concentrations for all core translation factors can be solved using the optimality condition (equation 5) and their respective kinetics schemes (the case of translation initiation is solved in Appendix 4). The analytical forms of the optimal solutions are shown in Table 2. In the binding-limited regime, the ratios of growth-optimized tIF concentrations are independent of the growth rate (except for aaRS), and are dependent only on basic biophysical parameters, such as protein sizes and diffusion constants.

To obtain the numerical values of association rates needed for calculate the optimal tIF stoichiometry (Table 2), we rely on a biophysically motivated scaling of the measured association between TC and ribosomes *in vivo*, $\hat{k}_{on}^{TC} = 6.4\ \mu\text{M}^{-1}\text{s}^{-1}$ (***Dai et al., 2016***) ($\hat{k}$ denotes the raw association rate constant in units $\mu\text{M}^{-1}\text{s}^{-1}$, which is different from the rescaled k, see section Conversion between concentration and proteome fraction). To our knowledge this is the only measurement of a tIF's association constant in a physiological context. We estimate the association rates for reactions involving other tIFs by scaling $\hat{k}_{on}^{TC}$ by the respective diffusion coefficients of the chemical species, that is for reaction involving species A and B: $\hat{k}_{on}^{AB}/\hat{k}_{on}^{TC} = (D_A + D_B)/(D_{TC} + D_{ribo})$, where D_i is the diffusion constant for the molecular species i (see Appendix 5 Table 2). Diffusion constants for several tIFs have been measured experimentally (***Bakshi et al., 2012***; ***Sanamrad et al., 2014***; ***Plochowietz et al., 2017***; ***Volkov et al., 2018***), and uncharacterized ones can be estimated using the cubic-root scaling with number of codons per protein from the Stokes-Einstein relation (***Nenninger et al., 2010***) (see Appendix 5 Table 1). This approach to arrive at plausible numerical estimates of $\hat{k}_{on}$'s assumes in particular that reactive radii and orientational constraints are similar for the different reactions (see Discussion for additional assumptions). These are strong assumptions which are necessary given the lack of *in vivo* biochemical parameter measurements, and can be relaxed as refined empirical measurements for more physiological association rates become available. Nonetheless, we note that the square-root dependence on these parameters (Table 2) for our predictions makes the numerical values less sensitive to possible tIF-specific effects. For our estimates in fast growth, we take the growth rate λ^* to be the average of the fast-growing species considered, corresponding to a doubling time of 21 ± 1 min (*E. coli*: 21.5 ± 1 min, *B. subtilis*: 21 ± 1 min, *V. natriegens*: 19 ± 1 min).

The estimated optimal tIF concentrations show concordance with the observed ones, both in terms of the absolute levels and the stoichiometry among tIFs (Figure 4 for fast growth, see Supplementary File 1 for data and Figure 4-Figure supplement 1 for additional growth conditions). A hierarchy of expression levels emerges such that the factors involved in elongation are more abundant compared to initiation and termination factors. The separation of these two classes is driven by the scaling factor $\sqrt{\langle \ell \rangle} \approx 14$ in our analytical solutions, which reflects the fact that the flux for elongation factors is $\langle \ell \rangle \approx 200$ times higher than that for initiation and termination factors. Within each class, the finer hierarchy of expression levels can also be further explained by simple parameters. For example, EF-Tu is predicted to be more abundant than EF-G by a factor of $\sqrt{n_{aa}\ell_{Tu}/\ell_G} \approx 3.3$ (observed ϕ_{Tu}/ϕ_G: *E. coli* 3.9, *B. subtilis* 2.7, *V. natriegens* 3.3). A higher abundance is required for EF-Tu because it is bound to the different tRNAs, which effectively decreases the concentration by a factor of $n_{aa} \approx 20$ (see section Estimation of coarse-grained rates for derivation

Box 1. The EF-Tu and aaRS transition line

Within our framework, optimality of translation factors is dictated by how coarse-grained ribosome transit times depend on factors' abundances (equation 4). For elongation factors aaRS and EF-Tu, contribution to the ribosome elongation time ($\tau_{el} = \langle \ell \rangle \tau_{aa}$) is through the concentration of the ternary complex (equation 10). Obtaining the optimal EF-Tu and aaRS abundance therefore requires solving for the ternary complex concentration as a function of these two variables.

The steady-state solution for the ternary complex concentration in the aaRS/EF-Tu expression displays two sharply separated regime (Figure 3B), separated by a narrow transition region (the 'transition line'). As described in the main text, the transition line plays a critical role for identifying the optimal EF-Tu and aaRS abundances within our model. Away from the line, there is an unproductive excess of either factors, viz. either $\partial\phi_{TC}/\partial\phi_{Tu} \approx 0$ or $\partial\phi_{TC}/\partial\phi_{aaRS} \approx 0$. Here, we derive the equation for the transition line. First, we leverage the constraint imposed by the conservation of tRNAs, which in our model is:

$$\text{tRNA}_{tot} = [\text{R}_\emptyset] + \underbrace{2[\text{R}_{TC}] + 2[\text{R}_{tRNA}] + 2[\text{R}_G]}_{\propto\, \lambda/k_{el}^{max}} + [\text{tRNA}] + [\text{tRNA:aaRS}] + [\text{aatRNA}] + [\text{TC}].$$

Above, tRNA_{tot} corresponds to the total tRNA concentration in the cell. In addition: $\text{R}_\emptyset$: elongating ribosomes with empty A-site, R_{TC}: ribosomes with bound TC, R_{tRNA}: ribosomes with filled A-site and no bound factor, R_G: ribosomes with bound EF-G, tRNA: free uncharged tRNAs, tRNA:aaRS: tRNA and aaRS complex, aatRNA: free charged tRNAs, and TC: ternary complex. Here, we assume that the elongating ribosomes always have a tRNA in the P-site, and a negligible occupancy in the E-site.

Using the system of equations from the mass action scheme at steady-state (section Translation elongation: optimal solutions), variables in the tRNA conservation equation above can be solved for in terms of the total abundance of EF-Tu and aaRS, the growth rate, and the steady-state ternary complex concentration. We note that the three ribosome species with a filled A site (R_{TC}, R_{tRNA}, and R_G) do not depend on EF-Tu concentration, and can be coarse-grained to a term proportional to λ/k_{el}^{max}, where k_{el}^{max} is the maximal translation elongation rate (not including the TC diffusion contribution) (***Dai et al., 2016***). In the binding-limited regime, converting to proteome fraction units, and leaving out the EF-Ts contribution without loss of generality (see section Optimal EF-Tu and aaRS abundances for a full treatment), we have:

$$\psi_{tRNA} = \underbrace{\frac{\lambda(\phi_{TC})}{\frac{k_{on}^{TC}}{n_{aa}}\phi_{TC}}}_{\text{R}_\emptyset} + \frac{2\lambda(\phi_{TC})}{k_{el}^{max}} + \underbrace{\frac{\lambda(\phi_{TC})}{\frac{k_{on}^{aaRS}}{n_{aa}}\phi_{aaRS}}}_{\text{free uncharged tRNA}} + \underbrace{\frac{\lambda(\phi_{TC})}{k_{on}^{Tu}\phi_{Tu^{\text{GTP}}}}}_{\text{free aatRNA}} + \frac{\phi_{TC}}{\ell_{Tu}}, \text{ where } \phi_{Tu^{\text{GTP}}} := \phi_{Tu} - \phi_{TC}. \tag{13}$$

Above, ψ_{tRNA} is a normalized tRNA concentration (see equation 28). We have explicitly highlighted that the growth rate is dependent on EF-Tu and aaRS only through the ternary complex concentration ϕ_{TC}. From the definition of of the elongation time (equation 10), we have $\lambda(\phi_{TC}) \propto \phi_{TC}/(K_{TC} + \phi_{TC})$ (***Klumpp et al., 2013***; ***Dai et al., 2016***) (definition of K_{TC} in terms of model parameters: supplement, equation 39). Equation 13 is closed and can be solved for ϕ_{TC} at given abundances of EF-Tu (ϕ_{Tu}) and aaRS (ϕ_{aaRS}).

Although equation 13 is non-linear and cannot be solved exactly in general, the separation of timescales in our coarse-grained description simplifies the problem considerably. Indeed, numerical solutions of equation 13 (Figure 3B, section Optimal EF-Tu and aaRS abundances) show that the behavior of TC concentration in the two-dimensional EF-Tu/aaRS expression space is split into two distinct regimes, sharply delineated by a transition line (orange line in Figure 3B, a geometric heuristic explaining the sharp separation between the regimes is presented in Appendix 3, section Interpretation of the sharp separation between aaRS and EF-Tu limited regimes, Figure 3-Figure supplement 1). Since TC concentration only increases as a function of both aaRS and EF-Tu on the transition line, the optimal solutions for the two factors must fall on it.

An expression for the transition line can be derived. Conceptually, the region of transition between the two regimes has both a low concentration of free EF-Tu molecules ($\phi_{Tu^{GTP}}/\phi_{Tu} \approx 0$) and a low concentration of free charged tRNAs ([aatRNAs]/$\text{tRNA}_{tot} \approx 0$). Although no values in the aaRS/EF-Tu expression plane can formally satisfy these two conditions simultaneously, the transition line is specified by setting the free charged tRNA term to 0 and replacing ϕ_{TC} by ϕ_{Tu} (no free EF-Tu) in equation 13. We denote by $(\bar{\phi}_{Tu}, \bar{\phi}_{aaRS})$ points satisfying the resulting requirement, namely (see equation 40 for non binding-limited case):

$$\text{Transition line: } \psi_{tRNA} - \frac{\lambda\left(\bar{\phi}_{Tu}\right) n_{aa}}{k_{on}^{TC}\bar{\phi}_{Tu}} - \frac{2\lambda\left(\bar{\phi}_{Tu}\right)}{k_{el}^{max}} - \frac{\bar{\phi}_{Tu}}{\ell_{Tu}} := \Delta_{tRNA}(\bar{\phi}_{Tu}) = \frac{n_{aa}\lambda\left(\bar{\phi}_{Tu}\right)}{k_{on}^{aaRS}\bar{\phi}_{aaRS}}, \tag{14}$$

where we have defined the excess tRNA (Δ_{tRNA}) above. In words, Δ_{tRNA} corresponds to the available tRNAs after the tRNAs sequestered on ribosomes and EF-Tu in the TC are subtracted from the total tRNA budget. At large aaRS concentrations, the transition line plateaus as a result of the finite total tRNA budget within the cell (Figure 3B, middle panel). The plateau is reached once all tRNAs aaRS are charged: the system is then no longer limited by aaRSs, but by the amount of tRNAs. Using the requirement that the optimum must fall on the transition line and the approximate solution for the EF-Tu optimum, the approximate optimal solution for aaRS is, from equation 14 (section Optimal EF-Tu and aaRS abundances for non binding-limited solution):

$$\phi_{aaRS}^* \approx \frac{n_{aa}\lambda^*}{k_{on}^{aaRS}\Delta_{tRNA}^*}, \text{ where: } \Delta_{tRNA}^* = \psi_{tRNA} - \frac{n_{aa}\lambda^*}{k_{on}^{TC}\phi_{Tu}^*} - \frac{2\lambda^*}{k_{el}^{max}} - \frac{\phi_{Tu}^*}{\ell_{Tu}} \tag{15}$$

Within our model, the optimal aaRS concentration is thus set by the excess tRNAs at the EF-Tu optimum (Δ_{tRNA}^*).

and discussion of why the factor is not equal to the number of different tRNAs). Taken together, our model offers straightforward explanations for the observed tIF stoichiometry.

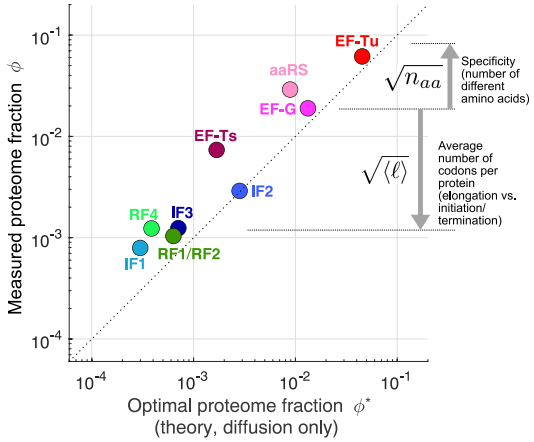


Figure 4. Predicted optimal abundance (no catalytic contribution, $k_{cat} \to \infty$) versus observed abundance. Measured proteome fractions are the average of *E. coli, B. subtilis, V. natriegens* (***Lalanne et al., 2018***). We note that given the sensitivity of the optimal aaRS abundance on the total tRNA/ribosome ratio (visually: yellow star's position in Figure 3B moves rapidly along x-axis upon changes in plateau of transition line), the prediction for aaRS should be interpreted with caution.

Figure 4–Figure supplement 1. Measured and predicted proteome fraction for core translation factors in individual conditions.

Figure 4–Figure supplement 2. Expression stoichiometry of core translation factors in different species and at different growth rates.

Figure 4–source data 1. Data and predicted values can be found in Supplementary File 1 and 2.

For a few tIFs, the observed concentrations are 2- to 5-fold higher than the predicted optimal levels (e.g., EF-Ts, RF4, and IF1 in Fig. 4). A potential explanation is that the corresponding reactions may not be binding or diffusion-limited, which would lead to a non-negligible fraction of tIFs sequestered at the catalytic step and thereby require higher total concentrations. Indeed, recent detailed modeling of the EF-Ts (***Hu et al., 2020***) cycle estimated only a small fraction (6 to 48%) of its abundance was in the free form in the cell, consistent with the large deviation we observe for this factor from our diffusion only prediction. Our optimization model can also be solved analytically in the non-binding-limited regime (Table 2), with the finite catalytic rate leading to an additional contribution of the form $\propto \ell\lambda^*/k_{cat}$. However, the numerical values for these solutions are in general difficult to obtain because the estimates for catalytic rates are sparse and often inconsistent with estimates of kinetics in live cells. As an example, median estimated aaRS catalytic rates (***Jeske et al., 2019***) measured *in vitro* is $\approx 3\ s^{-1}$, well below the *minimal* value of $15\ s^{-1}$, required to sustain translation flux at the measured value (Appendix 5), suggesting substantial deviation between *in vitro* and *in vivo* kinetics. While technically demanding, the fraction of free vs. bound factors can in principle be determined through live cell microscopy of tagged factors by partitioning the diffusive states of the tagged enzyme. Using that approach, (***Volkov et al., 2018***) estimated that EF-Tu was in its bound state <10% of the time (consistent with our diffusion-limited prediction closed to the observed value for this factor).

Another potential explanation for the observed deviations from our predictions is that the selective pressure for these tIFs may be lower compared to the more highly expressed tIFs. This explanation is unlikely both because their stoichiometry are observed to be conserved (Figure 1B, Figure 4-Figure supplement 2) and given that the expression of other lowly expressed tIFs (e.g., RF1, RF2, and individual aaRSs) has been shown to acutely affect cell growth (***Lalanne et al., 2021***; ***Parker et al., 2020***). Nevertheless, the deviations from the predicted optimal levels suggest that a

more refined model may be required than our first-principles derivation.

Discussion

Despite the comprehensive characterization of their molecular mechanisms, the 'mixology' for the protein synthesis machineries inside living cells has remained elusive. Here we establish a first-principles framework to provide analytical solutions for the growth-optimizing concentrations of translation factors. We find reasonable agreements between our parameter-free parsimonious predictions and the observed tIF stoichiometry (Figure 4). These results provide simple rationales for the hierarchy of expression levels, as well as insights into several construction principles for biological pathways.

An important implication from the agreement between observed stoichiometries and our predictions is that most tIFs are co-limiting for growth. Previous models have focused on expression optimization for the full translation sector, ribosomes (***Scott et al., 2010***, ***2014***; ***Belliveau et al., 2021***), and the abundant elongation factors EF-Tu (***Ehrenberg and Kurland, 1984***; ***Klumpp et al., 2013***). In a recent study, Hu and colleagues considered additional RNA components and EF-Ts in their optimization procedure (***Hu et al., 2020***). In line with the conclusions of these previous studies, our results demonstrate that multiple components of the translation machinery, regardless of their observed expression level, are simultaneously co-limiting for cell growth. By virtue of the interlocked translation cycles at steady state, the flux through every cycle must be matched. In our model, the optimality occurs when there are just enough tIFs to support the required flux in every cycle, such that the proteome fraction of free factors equals that of waiting ribosomes at that step (equipartition). If the concentration of any one tIF falls below the optimal point, it becomes the limiting factor for protein synthesis and growth. This result is supported by experimental evidence that slight knockdowns of individual RFs and aaRSs are detrimental to growth (***Parker et al., 2020***; ***Lalanne et al., 2021***). Figuratively, the translation apparatus is analogous to a vulnerable supply chain, in which slowdown in any of the steps affects the full output.

In the binding-limited regime, the optimal tIF stoichiometry is independent of the specific growth rate (except for aaRS). This is consistent with the observation that relative tIF expression remains unchanged in *E. coli* in conditions with growth rates ranging from 20-min to 2-hr doubling time (***Lalanne et al., 2018***; ***Li et al., 2014***) (Figure 4-Figure supplement 2A).

Our results are also consistent with the maintenance of the relative tIF expression across large phylogenetic distances even though the underlying regulation and cellular physiology has diverged (***Lalanne et al., 2018***) (Figure 1B, and additional comparison to slow growing *C. crescentus* in Figure 4-Figure supplement 2A). Under the assumption of diffusion-limited association to estimate parameters, the optimal tIF stoichiometry depends only on simple biophysical parameters, including protein sizes and diffusion constants, that are likely conserved in distant species. It remains to be determined if similar biophysical principles apply to the other pathways that also exhibit conserved enzyme expression stoichiometry.

In principle, our model can also make predictions on the growth defects at suboptimal tIF concentrations. However, experimentally testing these predictions will be difficult due to secondary effects of gene regulation that are not considered in our model near optimality. For example, we have recently shown that small changes in RF levels lead to idiosyncratic induction of the general stress response in *B. subtilis* due to a single ultrasensitive stop codon (***Lalanne et al., 2021***). As a result, the growth defect not only arises from reduced translation flux, but is in fact dictated by spurious regulatory connections that are normally not activated when tIF expression is at the optimum. We propose that tIF expression may be set at the optimal levels as our first-principles model suggests but entrenched by connections in the regulatory network. To predict the full expression-to-fitness landscape away from the optimum, a more comprehensive model may be required to take into account all the molecular interactions in the cell (***Karr et al., 2012***; ***Macklin et al., 2020***).

Our coarse-graining approach has several limitations in its connection to detailed biochemical parameters. Foremost, coarse-grained association rate constants remain difficult to numerically estimate, and possibly neglect important features. In particular, given the sparsity of available *in vivo* rate constants, we scaled the measured TC association rate ($\hat{k}_{on}^{TC}$) by the respective diffusion coefficients to estimate $\hat{k}_{on}$ for all tIFs reactions. This approach generates more plausible values than the unrealistic overestimate from Smoluchowski theory (diffusion-limited rate for perfectly absorbing spheres, see Appendix 5), but however assumes that certain molecular properties of other modeled reactions are similar. These include the size of the reactive surfaces, orientational constraints of the bimolecular interaction, and possible non-cognate binding events. We also do not explicitly consider off-rates in our model. Instead, our parameters correspond to effective rate constants that account for possible sequential binding and unbinding events, i.e., $\tilde{k}_{on} = k_{on}/n_{bind}$, with $n_{bind} = k_{cat}/(k_{cat}+k_{off})$. The effective association rate constants in our model thus contain information about catalytic and possible proofreading steps, which could be tIF-specific and are challenging to estimate. All of these effects may contribute to the discrepancy between our predicted and observed tIF concentrations. As more physiological and molecular data become available, these tIF-specific features could be used to individually refine our estimate for the association rates constants and our predictions. For example, elaborate calculations from structural data could account for rotational constraints (***Schlosshauer and Baker, 2004***), but are beyond the scope of the present work. Overall, we expect these tIF-specific corrections to be of limited influence on the final predictions due to the square-root dependence of the optimal expression (Table 2). We further note that a number of conclusions from our model, such as the factor of $\sqrt{\langle \ell \rangle}$ separating the optimal abundance of elongation from initiation/termination tIFs, are generic and do not depend on the specific association rates.

Taken together, our model provides the biophysical basis for the stoichiometry of translation factors in living cells. The first-principles approach complements more comprehensive models that include many biochemical parameters (***Hu et al., 2020***; ***Vieira et al., 2016***), while providing intuitive rationales for the expression hierarchy. We anticipate that our approach will be generalizable to elucidate or design enzyme stoichiometry of other biological pathways, especially those whose activities are required for cell growth.

Materials and Methods

Average number of codons per protein: $\langle \ell \rangle$

We calculate the average number of codons per protein, weighted by expression, as

$$\langle \ell \rangle := \frac{\sum_i e_i \ell_i}{\sum_i e_i}, \tag{16}$$

where ℓ_i is the number of codon for the protein product of gene i, and e_i is the protein synthesis rate (as estimated from ribosome profiling (***Li et al., 2014***; ***Lalanne et al., 2018***)) for gene i. For a stable proteome (in fast growing bacteria, the cell doubling time is shorter than the active degradation of most proteins (***Larrabee et al., 1980***)), the protein synthesis rate equals to the proteome mass fraction (***Li et al., 2014***). Changes in the expression of genes across growth conditions do not lead to substantial changes in $\langle \ell \rangle$. In *E. coli*, across growth conditions spanning ≈ 20 min doubling time to ≈ 120 min, $\langle \ell \rangle$ changes by about 20%. Specifically, we find $\langle \ell \rangle$ =196, 210, and 240 in respectively MOPS complete (≈ 20 min doubling time (***Li et al., 2014***)), MOPS minimal (≈ 56 min doubling time (***Li et al., 2014***)), and NQ1390 forced glucose limitation (≈ 120 min doubling time (***Mori et al., 2021***)), based on ribosome profiling data. Here for simplicity, we take $\langle \ell \rangle \approx 200$ throughout.

Conversion between concentration and proteome fraction

Throughout, we use both units of concentration (molar), denoted as e.g., $[A]$ for protein A, and proteome fraction, denoted by ϕ_A (***Scott et al., 2010***). The correspondence between the two is $\phi_A = [A]\ell_A/P$, where ℓ_A is the number of amino acid in protein A, and P is the in-protein amino

acid concentration in the cell. $P \approx 2.6 \times 10^6$ μM, and has a value approximately independent of growth rate (***Klumpp et al., 2013***; ***Bremer and Dennis, 2008***). This change in units also relates to how association constants are defined in units of proteome fraction: $\hat{k}_{on}[A] := k_{on}\phi_A$, where the hat $\hat{}$ refers to the association constant in usual units of $\mu M^{-1}\ s^{-1}$ (used to connect to empirical data). Hence, $k_{on} := \hat{k}_{on} P \ell^{-1}$ is the rescaled association rate in units of proteome fraction.

Equality of ribosome flux in steady-state

In steady-state exponential growth, the ribosome flux in and out of each intermediate state is equal to the total flux. This results from the fact that no ribosome can accumulate in any intermediate state. Since the flux out of state i is given by ϕ^i_{ribo}/τ_i, we must have:

$$\frac{\lambda \ell_{ribo}}{\langle \ell \rangle} = \frac{\phi^{act}_{ribo}}{\tau_{trl}} = \frac{\phi^{ini}_{ribo}}{\tau_{ini}} = \frac{\phi^{el}_{ribo}}{\tau_{el}} = \frac{\phi^{ter}_{ribo}}{\tau_{ter}}. \tag{17}$$

As a consequence, the proportion of ribosome in each state is equal to the proportion of time spent at that given step, for example for translation initiation:

$$\frac{\phi^{ini}_{ribo}}{\phi^{act}_{ribo}} = \frac{\tau_{ini}}{\tau_{ini} + \tau_{el} + \tau_{ter}}.$$

Protein production flux and growth rate

In order to write the mass action kinetic scheme for more complex models, it is useful to recast our framework in terms of the protein number production flux J, defined as the number of full length proteins produced per cell volume per unit time. The production of each protein requires a ribosome to go through the full synthesis cycle, and as such J provides a convenient quantity in mass action schemes formulated in molar units.

In steady-state of exponential growth (***Monod, 1949***; ***Scott et al., 2010***; ***Dai et al., 2016***), there is a direct relationship between the growth rate λ (defined through $\mathrm{d}N/\mathrm{d}t = \lambda N$, where N is the number of cells per unit volume of culture) and the protein production flux J. Explicitly, the protein mass accumulation rate is λM, where M is the total protein mass per unit volume of culture. If V is the mean cell volume, then $\lambda M/V = N m_{aa} \langle \ell \rangle J$, where m_{aa} is the mean amino acid mass. Defining $P := M/(m_{aa} N V)$, the in-protein amino acid concentration per cell (Materials and Methods, section Conversion between concentration and proteome fraction), the connection between protein production flux J and growth rate λ is then $J = \frac{P\lambda}{\langle \ell \rangle}$. This relationship will be used to convert between molar and proteome fraction in some equations below.

Summary of optimal solutions

Solutions for the factor predicted optimal abundances as a function of effective biochemical parameters and the growth rate at the optimum, are presented in Table 2. The table breaks down terms in each solution by categories: direct diffusion term (arising from diffusive search time), catalytic sequestration, and delay incurred by the diffusion of other proteins in part of the cycle of the factor of interest. Solutions are listed in terms of on-rate $\hat{k}_{on}$ (units of $\mu M^{-1} s^{-1}$). The aaRS solution follows a different form:

$$\phi^*_{aaRS} = \frac{n_{aa}\ell_{aaRS}\lambda^*}{\hat{k}^{aaRS}_{on} P \Delta^*_{tRNA}} + \frac{\ell_{aaRS}\lambda^*}{k^{aaRS}_{cat}}, \tag{18}$$

$$\text{with } \Delta^*_{tRNA} := \frac{\mathrm{tRNA}_{tot}}{P} - \frac{\lambda^*}{k^{TC}_{on}\phi^*_{TC}} - \frac{2\lambda^*}{k^{max}_{el}} - \frac{\phi^*_{TC}}{\ell_{Tu}} - \frac{\lambda^*}{k^{aaRS}_{cat}}, \text{ and } \phi^*_{TC} := \sqrt{\frac{n_{aa}\ell_{ribo}\ell_{Tu}\lambda^*}{\hat{k}^{TC}_{on} P}}.$$

Factor	Diffusion (direct) $\propto \sqrt{\frac{\lambda^*}{P}}$	Diffusion (other) $\propto \sqrt{\frac{\lambda^*}{P}}$	Catalytic sequestration $\propto \lambda^*$
IF1	$\sqrt{\frac{\ell_{ribo}\ell_{IF1}}{\langle\ell\rangle\hat{k}_{on}^{IF1}}\left[1+\frac{\ell_{IF2}+\ell_{IF3}}{\ell_{ribo}}\right]}$	$\frac{\ell_{IF1}}{\langle\ell\rangle}\sqrt{\frac{\langle\ell\rangle}{\hat{k}_{on}^{50S}}}$	$\frac{\ell_{IF1}}{\langle\ell\rangle}\left(\frac{1}{k^{RNA}}+\frac{1}{k_{cat}^{ini}}\right)$
IF2	$\sqrt{\frac{3}{4}}\sqrt{\frac{\ell_{ribo}\ell_{IF2}}{\langle\ell\rangle\hat{k}_{on}^{IF2}}}$	$\frac{\ell_{IF2}}{\langle\ell\rangle}\left(\sqrt{\frac{\ell_{ribo}\ell_{IF1}}{\langle\ell\rangle\hat{k}_{on}^{IF1}}}+\sqrt{\frac{\langle\ell\rangle}{\hat{k}_{on}^{50S}}}\right)$	$\frac{\ell_{IF2}}{\langle\ell\rangle}\left(\frac{1}{k^{RNA}}+\frac{1}{k_{cat}^{ini}}\right)$
IF3	$\sqrt{\frac{3}{4}}\sqrt{\frac{\ell_{ribo}\ell_{IF3}}{\langle\ell\rangle\hat{k}_{on}^{IF3}}}$	$\frac{\ell_{IF3}}{\langle\ell\rangle}\left(\sqrt{\frac{\ell_{ribo}\ell_{IF1}}{\langle\ell\rangle\hat{k}_{on}^{IF1}}}+\sqrt{\frac{\langle\ell\rangle}{\hat{k}_{on}^{50S}}}\right)$	$\frac{\ell_{IF3}}{\langle\ell\rangle}\left(\frac{1}{k^{RNA}}+\frac{1}{k_{cat}^{ini}}\right)$
EF-G	$\sqrt{\frac{\ell_{ribo}\ell_{G}}{\hat{k}_{on}^{G}}}$		$\frac{\ell_{G}}{k_{cat}^{G}}$
EF-Ts	$\sqrt{\frac{\ell_{Tu}\ell_{Ts}}{\hat{k}_{on}^{Ts}}}$		$\frac{\ell_{Ts}}{k_{cat}^{Ts}}$
EF-Tu	$\sqrt{\frac{\ell_{ribo}\ell_{Tu}n_{aa}}{\hat{k}_{on}^{TC}}}$	$\sqrt{\frac{\ell_{Tu}\ell_{Ts}}{\hat{k}_{on}^{Ts}}}$	$\ell_{Tu}\left(\frac{1}{k_{cat}^{TC}}+\frac{1}{k_{cat}^{Ts}}\right)$
RF1+RF2	$\sqrt{\frac{\ell_{ribo}\ell_{RFI}(1+\delta)}{\langle\ell\rangle\hat{k}_{on}^{RFI}}}$		$\frac{\ell_{RFI}}{\langle\ell\rangle k_{cat}^{RFI}}$
RF4	$\sqrt{\frac{\ell_{ribo}\ell_{RF4}}{\langle\ell\rangle\hat{k}_{on}^{RF4}}}$		$\frac{\ell_{RF4}}{\langle\ell\rangle k_{cat}^{RF4}}$

Table 2. Compilation of predicted optimal abundances for translation factors. The optimal abundance is the sum of the terms in each row. Columns correspond to contributions of different nature (diffusion of factor itself, diffusion of other factors involved in the factor's cycle, catalytic term). Terms must be multiplied by the common factors indicated in each column's header ($\propto$). For RF1+RF2, $\delta := 2\sqrt{f_{UAG}f_{UGA}}$ (see section Optimal abundances for RF1/RF2).

Acknowledgments

We thank R. Battaglia, J. Cascino, M. Gill, M. Parker, D. Parker, and G. Schmidt for critical reading of the manuscript, and all members of the Li lab for discussion. This research was supported by NIH grant R35GM124732, the NSF CAREER Award, the Smith Odyssey Award, the Pew Biomedical Scholars Program, a Sloan Research Fellowship, the Searle Scholars Program, the Smith Family Award for Excellence in Biomedical Research; NSERC doctoral Fellowship and HHMI International Student Research Fellowship (to J.-B.L.).

Supplementary Files

Supplementary File 1: Proteome synthesis fraction (in %) of core mRNA translation factors for species and growth conditions with fast growth estimated from ribosome profiling data (***Li et al., 2014***; ***Lalanne et al., 2018***).

Supplementary File 2: Diffusion-limited optima predicted for translation factors for fast-growth conditions.

Supplementary File 3: Proteome synthesis fraction (in %) of core mRNA translation factors for species/conditions with slower growth estimated from ribosome profiling. Ribosome profiling data: *E. coli* (MOPS minimal (***Li et al., 2014***), M9 glucose (***Mori et al., 2021***), *C. crescentus* ((***Schrader et al., 2014***), with synthesis rates estimated in (***Lalanne et al., 2018***)).

Supplementary File 4: Diffusion-limited optima predicted for translation factors for slower growth conditions.

References

Agirrezabala X, Frank J. Elongation in translation as a dynamic interaction among the ribosome, tRNA, and elongation factors EF-G and EF-Tu. Quarterly Reviews of Biophysics. 2009; 42(3):159–200. doi: 10.1017/S0033583509990060.

Andersen GR, Nissen P, Nyborg J. Elongation factors in protein biosynthesis. Trends in Biochemical Sciences. 2003; 28(8):434–441. doi: 10.1016/S0968-0004(03)00162-2.

Baggett NE, Zhang Y, Gross CA. Global analysis of translation termination in E. coli. PLoS Genetics. 2017; 13(3):1–27. http://journals.plos.org/plosgenetics/article/file?id=10.1371/journal.pgen.1006676{&}type=printable, doi: 10.1371/journal.pgen.1006676.

Bakshi S, Siryaporn A, Goulian M, Weisshaar JC. Superresolution imaging of ribosomes and RNA polymerase in live Escherichia coli cells. Molecular Microbiology. 2012; 85(1):21–38. doi: 10.1111/j.1365-2958.2012.08081.x.

Belliveau NM, Chure G, Hueschen CL, Fisher DS, Theriot JA, Phillips R. Fundamental limits on the rate of bacterial growth and their influence on proteomic composition. Cell Systems. 2021; p. 1–21. doi: 10.1016/j.cels.2021.06.002.

Berg OG, Kurland CG. Growth rate-optimised tRNA abundance and codon usage. Journal of molecular biology. 1997; 270(4):544–550. doi: 10.1006/jmbi.1997.1142.

Bergmann JE, Lodish HF. A kinetic model of protein synthesis. Application to hemoglobin synthesis and translational control. Journal of Biological Chemistry. 1979; 254(23):11927–11937. http://www.jbc.org/content/254/23/11927.full.pdf, doi: 10.1016/s0021-9258(19)86406-2.

Bertram G, Innes S, Minella O, Richardson JP, Stansfield I. Endless possibilities: Translation termination and stop codon recognition. Microbiology. 2001; 147(2):255–269. doi: 10.1099/00221287-147-2-255.

Björk GR, Hagervall TG. Transfer RNA Modification: Presence, Synthesis, and Function. EcoSal Plus. 2014; 6(1). doi: 10.1128/ecosalplus.esp-0007-2013.

664 **Borg A**, Pavlov M, Ehrenberg M. Complete kinetic mechanism for recycling of the bacterial ribosome. RNA.
665 2016; 22(1):10–21. doi: 10.1261/rna.053157.115.

666 **Bremer H**, Dennis PP. Modulation of Chemical Composition and Other Parameters of the Cell at Different
667 Exponential Growth Rates. EcoSal Plus. 2008; http://ctbp.ucsd.edu/qbio/beemer96.pdf, doi: 10.1016/0022-
668 2836(72)90190-8.

669 **Chen J**, Choi J, O'Leary SE, Prabhakar A, Petrov A, Grosely R, Puglisi EV, Puglisi JD. The molecular choreography of
670 protein synthesis: translational control, regulation, and pathways. Quarterly Reviews of Biophysics. 2016;
671 49:e11. doi: 10.1017/s0033583516000056.

672 **Dai X**, Zhu M, Warren M, Balakrishnan R, Patsalo V, Okano H, Williamson JR, Fredrick K, Wang YP, Hwa T. Re-
673 duction of translating ribosomes enables Escherichia coli to maintain elongation rates during slow growth.
674 Nature Microbiology. 2016; 2(2):1–9. http://dx.doi.org/10.1038/nmicrobiol.2016.231, doi: 10.1038/nmicro-
675 biol.2016.231.

676 **Dever TE**, Green R. The elongation, termination, and recycling phases of translation in eukaryotes. Cold Spring
677 Harbor Perspectives in Biology. 2012; 4(7):1–16. doi: 10.1101/cshperspect.a013706.

678 **Dong H**, Nilsson L, Kurland CG. Co-variation of tRNA Abundance and Codon Usage in Escherichia coli at Different
679 Growth Rates. Journal of molecular biology. 1996; 260(5):649–663. doi: 10.1006/jmbi.1996.0428.

680 **Dykeman EC**. A stochastic model for simulating ribosome kinetics in vivo. PLoS computational biology. 2020;
681 16(2):e1007618. http://dx.doi.org/10.1371/journal.pcbi.1007618, doi: 10.1371/journal.pcbi.1007618.

682 **Ehrenberg M**, Kurland CG. Costs of accuracy determined by a maximal growth rate constraint. Quarterly
683 reviews of biophysics. 1984; 17:45–80. doi: 10.1080/10643389.2012.728825.

684 **Elowitz MB**, Surette MG, Wolf PE, Stock JB, Leibler S. Protein mobility in the cytoplasm of Escherichia coli. Journal
685 of Bacteriology. 1999; 181(1):197–203. doi: 10.1128/jb.181.1.197-203.1999.

686 **Gorochowski TE**, Chelysheva I, Eriksen M, Nair P, Pedersen S, Ignatova Z. Absolute quantification of translational
687 regulation and burden using combined sequencing approaches. Molecular Systems Biology. 2019; 15(5):e8719.
688 doi: 10.15252/msb.20188719.

689 **Gualerzi CO**, Pon CL. Initiation of mRNA translation in bacteria: Structural and dynamic aspects. Cellular and
690 Molecular Life Sciences. 2015; 72(22):4341–4367. doi: 10.1007/s00018-015-2010-3.

691 **Hu XP**, Dourado H, Schubert P, Lercher MJ. The protein translation machinery is expressed for maximal efficiency
692 in Escherichia coli. Nature Communications. 2020; 11(1):1–10. http://dx.doi.org/10.1038/s41467-020-18948-x,
693 doi: 10.1038/s41467-020-18948-x.

694 **Ibba M**, Dieter S. Aminoacyl-tRNA synthesis. Annu Rev Biochem. 2000; 69:617–650.

695 **Jeschek M**, Gerngross D, Panke S. Combinatorial pathway optimization for streamlined metabolic engineering.
696 Current Opinion in Biotechnology. 2017; 47:142–151. http://dx.doi.org/10.1016/j.copbio.2017.06.014, doi:
697 10.1016/j.copbio.2017.06.014.

698 **Jeske L**, Placzek S, Schomburg I, Chang A, Schomburg D. BRENDA in 2019: A European ELIXIR core data resource.
699 Nucleic Acids Research. 2019; 47(D1):D542–D549. doi: 10.1093/nar/gky1048.

700 **Johnson GE**, Lalanne JB, Peters ML, Li GW. Functionally uncoupled transcription-translation in Bacillus subtilis.
701 Nature. 2020; 585(7823):124–128. doi: 10.1038/s41586-020-2638-5.

702 **Karr JR**, Sanghvi JC, MacKlin DN, Gutschow MV, Jacobs JM, Bolival B, Assad-Garcia N, Glass JI, Covert MW.
703 A whole-cell computational model predicts phenotype from genotype. Cell. 2012; 150(2):389–401. doi:
704 10.1016/j.cell.2012.05.044.

705 **Kavčič B**, Tkačik G, Bollenbach T. Mechanisms of drug interactions between translation-inhibiting antibiotics.
706 Nature Communications. 2020; 11(1):4013. doi: 10.1038/s41467-020-17734-z.

707 **Kennell D**, Riezman H. Transcription and translation initiation frequencies of the Escherichia coli lac operon.
708 Journal of Molecular Biology. 1977; 114(1):1–21. doi: 10.1016/0022-2836(77)90279-0.

Keseler IM, Mackie A, Santos-Zavaleta A, Billington R, Bonavides-Martínez C, Caspi R, Fulcher C, Gama-Castro S, Kothari A, Krummenacker M, Latendresse M, Muñiz-Rascado L, Ong Q, Paley S, Peralta-Gil M, Subhraveti P, Velázquez-Ramírez DA, Weaver D, Collado-Vides J, Paulsen I, et al. The EcoCyc database: reflecting new knowledge about Escherichia coli K-12. Nucleic acids research. 2016; 45(November 2016):gkw1003. doi: 10.1093/nar/gkw1003.

Klumpp S, Scott M, Pedersen S, Hwa T. Molecular crowding limits translation and cell growth. Proceedings of the National Academy of Sciences of the United States of America. 2013; 110(42):16754–9. doi: 10.1073/pnas.1310377110.

Kumar M, Mommer MS, Sourjik V. Mobility of cytoplasmic, membrane, and DNA-binding proteins in Escherichia coli. Biophysical Journal. 2010; 98(4):552–559. doi: 10.1016/j.bpj.2009.11.002.

Lalanne JB, Parker DJ, Li GW. Spurious regulatory connections dictate the expression-fitness landscape of translation factors. Molecular Systems Biology. 2021; 17(4):1–23. doi: 10.15252/msb.202110302.

Lalanne JB, Taggart JC, Guo MS, Herzel L, Schieler A, Li GW. Evolutionary Convergence of Pathway-Specific Enzyme Expression Stoichiometry. Cell. 2018; p. 749–761. doi: 10.1016/j.cell.2018.03.007.

Larrabee KL, Phillips JO, Williams GJ, Larrabee AR. The relative rates of protein synthesis and degradation in a growing culture of Escherichia coli. Journal of Biological Chemistry. 1980; 255(9):4125–4130.

Laursen BS, Sørensen HP. Initiation of protein synthesis in bacteria. Microbiology and Molecular Biology Reviews. 2005; 69(1):101–123. doi: 10.1128/MMBR.69.1.101.

Li GW. How do bacteria tune translation efficiency? Current Opinion in Microbiology. 2015; 24:66–71. http://dx.doi.org/10.1016/j.mib.2015.01.001, doi: 10.1016/j.mib.2015.01.001.

Li GW, Burkhardt D, Gross C, Weissman JS. Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. Cell. 2014; 157(3):624–635. doi: 10.1016/j.cell.2014.02.033.

Lindahl L. Intermediates and time kinetics of the in vivo assembly of Escherichia coli ribosomes. Journal of Molecular Biology. 1975; 92(1):15–37. doi: 10.1016/0022-2836(75)90089-3.

Macklin DN, Ahn-Horst TA, Choi H, Ruggero NA, Carrera J, Mason JC, Sun G, Agmon E, DeFelice MM, Maayan I, Lane K, Spangler RK, Gillies TE, Paull ML, Akhter S, Bray SR, Weaver DS, Keseler IM, Karp PD, Morrison JH, et al. Simultaneous cross-evaluation of heterogeneous E. coli datasets via mechanistic simulation. Science. 2020; 369(6502). doi: 10.1126/science.aav3751.

Mangano K, Florin T, Shao X, Klepacki D, Chelysheva I, Ignatova Z, Gao Y, Mankin AS, Vázquez-Laslop N. Genome-wide effects of the antimicrobial peptide apidaecin on translation termination in bacteria. eLife. 2020; 9:1–24. doi: 10.7554/eLife.62655.

Margus T, Remm M, Tenson T. Phylogenetic distribution of translational GTPases in bacteria. BMC Genomics. 2007; 8:1–18. doi: 10.1186/1471-2164-8-15.

Marintchev A, Wagner G. Translation initiation: structures, mechanisms and evolution. Quarterly reviews of biophysics. 2004; 37(3-4):197–284. doi: 10.1017/S0033583505004026.

Milón P, Maracci C, Filonava L, Gualerzi CO, Rodnina MV. Real-time assembly landscape of bacterial 30S translation initiation complex. Nature Structural & Molecular Biology. 2012; 19(6):609–615. doi: 10.1038/nsmb.2285.

Mohammad F, Green R, Buskirk AR. A systematically-revised ribosome profiling method for bacteria reveals pauses at single-codon resolution. eLife. 2019; 8:1–25. doi: 10.7554/eLife.42591.

Monod J. The Growth of Bacterial Cultures. Annual Review of Microbiology. 1949; 3(1):371–394. doi: 10.1146/annurev.mi.03.100149.002103.

Mora L, Heurgué-Hamard V, De Zamaroczy M, Kervestin S, Buckingham RH. Methylation of bacterial release factors RF1 and RF2 is required for normal translation termination in vivo. Journal of Biological Chemistry. 2007; 282(49):35638–35645. doi: 10.1074/jbc.M706076200.

Mori M, Zhang Z, Banaei-Esfahani A, Lalanne J, Okano H, Collins BC, Schmidt A, Schubert OT, Lee D, Li G, Aebersold R, Hwa T, Ludwig C. From coarse to fine: the absolute Escherichia coli proteome under diverse growth conditions . Molecular Systems Biology. 2021; 17(5). doi: 10.15252/msb.20209536.

Nenninger A, Mastroianni G, Mullineaux CW. Size dependence of protein diffusion in the cytoplasm of Escherichia coli. Journal of Bacteriology. 2010; 192(18):4535–4540. doi: 10.1128/JB.00284-10.

Nomura M, Gourse R, Baughman G, Regulation of the synthesis of ribosomes and ribosomal components.; 1984. doi: 10.1146/annurev.bi.53.070184.000451.

Pang YLJ, Poruri K, Martinis SA. tRNA synthetase: TRNA aminoacylation and beyond. Wiley Interdisciplinary Reviews: RNA. 2014; 5(4):461–480. doi: 10.1002/wrna.1224.

Parker DJ, Lalanne JB, Kimura S, Johnson GE, Waldor MK, Li GW. Growth-Optimized Aminoacyl-tRNA Synthetase Levels Prevent Maximal tRNA Charging. Cell Systems. 2020; 11:1–10. doi: 10.1016/j.cels.2020.07.005.

Pavlov MY, Freistroffer DV, Heurgué-Hamard V, Buckingham RH, Ehrenberg M. Release factor RF3 abolishes competition between release factor RF1 and ribosome recycling factor (RRF) for a ribosome binding site. Journal of Molecular Biology. 1997; 273(2):389–401. doi: 10.1006/jmbi.1997.1324.

Pedersen S, Bloch PL, Reeh S, Neidhardt FC. Patterns of protein synthesis in E. coli: a catalog of the amount of 140 individual proteins at different growth rates. Cell. 1978; 14(1):179–190. doi: 10.1016/0092-8674(78)90312-4.

Plochowietz A, Farrell I, Smilansky Z, Cooperman BS, Kapanidis AN. In vivo single-RNA tracking shows that most tRNA diffuses freely in live bacteria. Nucleic Acids Research. 2017; 45(2):926–937. doi: 10.1093/nar/gkw787.

Reuveni S, Meilijson I, Kupiec M, Ruppin E, Tuller T. Genome-scale analysis of translation elongation with a ribosome flow model. PLoS Computational Biology. 2011; 7(9). doi: 10.1371/journal.pcbi.1002127.

Rodnina MV. Translation in prokaryotes. Cold Spring Harbor Perspectives in Biology. 2018; 10(9):1–22. doi: 10.1101/cshperspect.a032664.

Saito K, Green R, Buskirk AR. Ribosome recycling is not critical for translational coupling in E. Coli. eLife. 2020; 9:1–37. doi: 10.7554/ELIFE.59974.

Sanamrad A, Persson F, Lundius EG, Fange D, Gynnå AH, Elf J. Single-particle tracking reveals that free ribosomal subunits are not excluded from the Escherichia coli nucleoid. Proceedings of the National Academy of Sciences of the United States of America. 2014; 111(31):11413–11418. doi: 10.1073/pnas.1411558111.

Schaechter M, MaalOe O, Kjeldgaard NO. Dependency on Medium and Temperature of Cell Size and Chemical Composition during Balanced Growth of Salmonella typhimurium. Journal of General Microbiology. 1958; 19(3):592–606. http://mic.microbiologyresearch.org/content/journal/micro/10.1099/00221287-19-3-592, doi: 10.1099/00221287-19-3-592.

Schlosshauer M, Baker D. Realistic protein-protein association rates from a simple diffusional model neglecting long-range interactions, free energy barriers, and landscape ruggedness. Protein Science. 2004; 13(6):1660–1669. doi: 10.1110/ps.03517304.

Schrader JM, Zhou B, Li GW, Lasker K, Childers WS, Williams B, Long T, Crosson S, McAdams HH, Weissman JS, Shapiro L. The Coding and Noncoding Architecture of the Caulobacter crescentus Genome. PLoS Genetics. 2014; 10(7). doi: 10.1371/journal.pgen.1004463.

Scolnick E, Tompkins R, Caskey T, Nirenberg M. Release factors differing in specificity for terminator codons. Proceedings of the National Academy of Sciences of the United States of America. 1968; 61(2):768–774. doi: 10.1073/pnas.61.2.768.

Scott M, Gunderson CW, Mateescu EM, Zhang Z, Hwa T, Gunderson CW, Mateescu EM, Zhang Z, Hwa T. Interdependence of Cell Growth Origins and Consequences. Science. 2010; 330:1099–1102. doi: 10.1126/science.1192588.

Scott M, Klumpp S, Mateescu EM, Hwa T. Emergence of robust growth laws from optimal regulation of ribosome synthesis. Molecular Systems Biology. 2014; 10(8):747. doi: 10.15252/msb.20145379.

Shaw LB, Zia RKP, Lee KH. Totally asymmetric exclusion process with extended objects: a model for protein synthesis. Physical review E, Statistical, nonlinear, and soft matter physics. 2003; 68(2 Pt 1):021910. doi: 10.1103/PhysRevE.68.021910.

Subramaniam AR, Zid BM, O'Shea EK. An integrated approach reveals regulatory controls on bacterial translation elongation. Cell. 2014; 159(5):1200–1211. http://dx.doi.org/10.1016/j.cell.2014.10.043, doi: 10.1016/j.cell.2014.10.043.

Vieira JP, Racle J, Hatzimanikatis V. Analysis of Translation Elongation Dynamics in the Context of an Escherichia coli Cell. Biophysical Journal. 2016; 110(9):2120–2131. http://dx.doi.org/10.1016/j.bpj.2016.04.004, doi: 10.1016/j.bpj.2016.04.004.

Volkov IL, Lindén M, Aguirre Rivera J, Ieong KW, Metelev M, Elf J, Johansson M. tRNA tracking for direct measurements of protein synthesis kinetics in live cells. Nature Chemical Biology. 2018; 14(6):618–626. http://dx.doi.org/10.1038/s41589-018-0063-y, doi: 10.1038/s41589-018-0063-y.

Weijland A, Harmark K, Cool RH, Anborgh PH, Parmeggiani A. Elongation factor Tu: a molecular switch in protein biosynthesis. Molecular Microbiology. 1992; 6(6):683–688. doi: 10.1111/j.1365-2958.1992.tb01516.x.

Wittmann HG. Components of Bacterial Ribosomes. Annual Review of Biochemistry. 1982; 51(1):155–183. doi: 10.1146/annurev.bi.51.070182.001103.

You C, Okano H, Hui S, Zhang Z, Kim M, Gunderson CW, Wang YP, Lenz P, Yan D, Hwa T. Coordination of bacterial proteome with metabolism by cyclic AMP signalling. Nature. 2013; 500(7462):301–306. doi: 10.1038/nature12446.

Zavialov AV, Hauryliuk VV, Ehrenberg M. Splitting of the posttermination ribosome into subunits by the concerted action of RRF and EF-G. Molecular Cell. 2005; 18(6):675–686. doi: 10.1016/j.molcel.2005.05.016.

Appendix 1

Coarse-grained transition times: models of ribosome traffic

Our coarse-grained model of ribosome transitions between categories of initiation, elongation, and termination need to be distinguished from the individual molecular times of the respective steps in one important regard: ribosome traffic on mRNAs can lead to effective delays arising from transient queuing. For example, if translation termination is slow and ribosomes start to pile up and form queues upstream of stop codons on mRNAs, the molecular time of termination (time between ribosome arrival to the stop codon and its recycling to the free ribosome pool) will not be a correct reflection of the actual termination time of a ribosome, because of the additional wait time in the queue. A similar argument can be made for transient queuing forming in the body of genes for elongating ribosomes.

We connect these two (molecular and coarse-grained) levels of description by noting that our mass action schemes relating the translation factor abundance to the times of the specific steps can be used as input parameters in traffic models of ribosome movement along mRNAs taking into account possible many-body interactions (e.g., totally asymmetric exclusion processes (***Shaw et al., 2003***; ***Kavčič et al., 2020***)). Solving these traffic models can then be used to obtain transition times in our coarse-grained translation cycle model. As we show below, corrections arising from transient queuing are small (for endogenous translation factor abundances) based on current estimates the absolute rates of initiation, elongation, and termination, on individual mRNAs, such that stochastic queuing does not play a dominant role in determining optimal translation factor expression levels.

As a first example, we relate the on-stop codon molecular termination time τ_{ter}, which we obtain from solving our mass action scheme (see equation 6), to the termination time in presence of queuing: τ_{ter}^{full}. The difference between the two, as described above, being related to possible queues upstream of stop codons leading to further delays in the process of translation termination, and thus to a longer termination time than that of the molecular on-stop codon termination. The delay factor will be denoted $\mathcal{Q}\left(\tau_{ter}\right)$, defined through:

$$\tau_{ter}^{full} := \tau_{ter}\, \mathcal{Q}\left(\tau_{ter}\right).$$

To derive the expression for the $\mathcal{Q}$ factor, note that in steady-state, ribosome numbers in a given state is directly proportional to the time to transition out of that state. Let m_i be the mRNA concentration for gene i in the cell, $n_{ter}(\alpha_i, \tau_{ter})$ the number of terminating ribosomes (including queues if present) on a transcript with per mRNA translation initiation rate (i.e., translation efficiency (***Li, 2015***)) α_i, then:

$$\tau_{ter}^{full} \propto \sum_i m_i\, n_{ter}(\alpha_i, \tau_{ter}),$$

whereas

$$\tau_{ter} \propto \sum_i m_i\, n_{ter}^{\emptyset\mathcal{Q}}(\alpha_i, \tau_{ter}),$$

with $n_{ter}^{\emptyset\mathcal{Q}}(\alpha_i, \tau_{ter})$ the average number of terminating ribosomes on a transcript with translation efficiency α_i, assuming no queue upstream of the stop codon. Note that $n_{ter}(\alpha_i, \tau_{ter}) \geq n_{ter}^{\emptyset\mathcal{Q}}(\alpha_i, \tau_{ter})$ (the differences being queued ribosomes). Hence, the queuing factor $\mathcal{Q}$ is:

$$\mathcal{Q}\left(\tau_{ter}\right) := \frac{\tau_{ter}^{full}}{\tau_{ter}} = \frac{\sum_i m_i\, n_{ter}(\alpha_i, \tau_{ter})}{\sum_i m_i\, n_{ter}^{\emptyset\mathcal{Q}}(\alpha_i, \tau_{ter})}.$$

Formally, n_{ter} can be obtained by solving a TASEP model (***Shaw et al., 2003***), but a simplified queue model (***Bergmann and Lodish, 1979***; ***Lalanne et al., 2021***) disregarding spatial information recapitulates the statistics of queue formation (as verified by full stochastic simulations, data not shown). The state space of the queue model is the number of ribosomes N in the queue. Ribosomes arrive at a rate α (initiation rate on the transcript), and leave at the molecular termination rate τ_{ter}^{-1}. The ribosome arrival rate at the queue is rigorously correct in steady-state, unless the queue becomes large enough to affect the initiation process (fully jammed transcript), or RNA degradation. The stochastic process (away from the jammed state) is then described by: $N \to N+1$ at rate α, and $N \to N-1$ at rate τ_{ter}^{-1} for $N > 0$. The probability for the queue to have N ribosomes, $P(N)$, can be obtained as the steady-state from the resulting master equation, leading to a geometric series: $P(N) = \left(\alpha\tau_{ter}\right)^N \left(1 - \alpha\tau_{ter}\right)$. Hence, the prevalence of higher order queues scales as the ratio of the initiation to termination rate on the transcript. The average queue size, corresponding to $n_{ter}(\alpha_i, \tau_{ter})$, is:

$$n_{ter}(\alpha_i, \tau_{ter}) \approx \begin{cases} \dfrac{\tau_{ter}\alpha_i}{1 - \tau_{ter}\alpha_i}, & \tau_{ter}^{-1} \geq \alpha_i(1 + \ell_{footprint}\ell_i^{-1}), \\ \dfrac{\ell_i}{\ell_{footprint}}, & \tau_{ter}^{-1} < \alpha_i(1 + \ell_{footprint}\ell_i^{-1}). \end{cases}$$

Above, the solution of the simple model is truncated at the value where the transcript becomes fully jammed with $\ell_i/\ell_{footprint}$ ribosomes (ℓ_i and $\ell_{footprint}$ being the size of gene i and the size occupied by a ribosome respectively). The no queue ribosome number is simply equal to a model where queues with $N > 1$ do not arise, hence $n_{ter}^{\emptyset Q}(\alpha_i, \tau_{ter}) = \alpha_i\tau_{ter}$. Therefore, the queuing factor, under the stated assumptions (and assuming no transcript is in the jammed state), is

$$Q\left(\tau_{ter}\right) \approx \frac{\sum_i m_i \frac{\alpha_i}{1-\tau_{ter}\alpha_i}}{\sum_i m_i \alpha_i}.$$

Expanding for fast termination gives $Q - 1 = \frac{\tau_{ter}\langle\alpha^2\rangle}{\langle\alpha\rangle}$ as the leading order correction, where the averages are weighted by mRNA levels. The above was derived assuming exponentially distributed initiation and termination times, but could be modified to account for more complex dynamics of the initiation and initiation steps.

The queuing factor can be estimated based on absolute measurements of the initiation and termination rates in cells. Kennell and Riezman (***Kennell and Riezman, 1977***) estimate 3.2 s between initiation events on the *lacZ* mRNA (at 48 min per cell doubling). Bremer and Dennis (***Bremer and Dennis, 2008***) estimate 1 s per ribosome initiation events at 20 min doubling time. Recent calibrated high-throughput measurements report a genome-wide median of 5.6 s per initiation events (***Gorochowski et al., 2019***). To our knowledge, estimation of absolute *in vivo* termination rates have not been performed, but we can estimate bounds. Indirect assessment based on steady-state protein production measurements place the fraction of actively elongating ribosome at about 95% (***Dai et al., 2016***). Assuming (upper bound) that the 5% of non elongating ribosomes are in the process of termination would give a termination time of $5\% \times 11.1s \approx 0.6\ s$ (fraction of ribosomes in a given state equal to the ratio of transition times), where we have used that the elongation time of an average protein is about 11.1 s ($200/18\ s^{-1}$) at fast growth (***Dai et al., 2016***). This upper bound is still much smaller than the reported median initiation time, suggesting that the queuing factor for termination is small. As additional support to the view that translation is far from being termination limited, small that queues at stop codons are only globally observed

in ribosome profiling upon severe perturbations (***Kavčič et al., 2020***; ***Baggett et al., 2017***; ***Mangano et al., 2020***; ***Saito et al., 2020***; ***Lalanne et al., 2021***).

With regards to translation elongation, transient queuing in the body of gene can also lead to a difference between molecular and coarse-grained transition times in our model. However, the fraction of ribosomes transiently stalled due to this queuing scales as $\alpha\tau_{aa}$ in the low density phase (defined by requirements $\alpha\tau_{ter} < 1$ and $\alpha\tau_{aa} < (1 + \sqrt{\ell_{footprint}})^{-1} \approx 0.25$) of the TASEP model (***Shaw et al., 2003***). Since measured estimates place $\alpha\tau_{aa} \sim 0.01$ (***Dai et al., 2016***; ***Gorochowski et al., 2019***), we do not consider the queuing effect for elongating ribosomes within our optimization framework for elongation factor abundances.

Appendix 2

Translation termination

Omitted molecular details

The kinetic scheme presented in Figure 2A does not include some known molecular details of translation termination. For example, GTPase RF3 has been shown to catalyze the release of RF1/RF2 post peptide hydrolysis and to effectively prevent rebinding to empty A site ribosome without peptide (***Pavlov et al., 1997***). RF3 is not included in our model given our desire for a parsimonious description and due to the absence of identifiable homologs in multiple bacteria (e.g., *B. subtilis*) (***Margus et al., 2007***). Our scheme aggregates the RF1/RF2 recycling rate with the catalytic rate, and further assume a unidirectional reaction without rebinding (consistent with a lower bound), effectively taking into account the action of RF3. In addition, translocation factor EF-G is known to be implicated in ribosome recycling via translocation post RF4 binding (***Zavialov et al., 2005***). We assume EF-G's abundance requirement towards the function of termination to be a minor fraction of its total requirement (non-sense to sense codons ≈0.5%) and to be non-limiting for this step. We thus coarse-grain EF-G's role in ribosome recycling through an effective catalytic rate for RF4, see (***Borg et al., 2016***) for details of EF-G's involvement in ribosome recycling. As another example of simplification in our coarse-graining, we also do not explicitly model RF1/RF2's post-translational modification by methyltransferase PrmC (***Mora et al., 2007***). Thus, the activity of the RFs within our description to correspond to the average within a possibly heterogeneous pool of modified and unmodified factors in the cell.

Non binding-limited regime (one stop codon)

If translation termination is not diffusion limited, terms corresponding to the finite catalytic times must be included in addition to the diffusive contributions in the termination time (equation 6). Under our simplified scheme (Figure 2A) and with a single stop codons (grouping RF1 and RF2), the molecular termination time is then sum of the four separate times corresponding to distinct events:

$$\tau_{ter} = \frac{1}{k_{on}^{RF1}\phi_{RF1}^{free}} + \frac{1}{k_{cat}^{RF1}} + \frac{1}{k_{on}^{RF4}\phi_{RF4}^{free}} + \frac{1}{k_{cat}^{RF4}}$$

The two novelties compared to the diffusion-limited regime (equation 6) are: (1) addition of the catalytic times k_{cat}^{-1} for the two steps, and importantly (2) the mass action diffusion terms now involve the free concentration of release factors. Generally, the free concentration of the tIFs can be obtained by solving the steady-state solutions of kinetic schemes under constraints imposed by conservation equations. The examples in e.g., sections 2, 2, and 1 below provide the mathematical details associated with the procedure.

Here, the difference between the total and free concentration of release factor arises from the finite catalytic turnover of the enzymes, and corresponds to the concentration of ribosome bound release factors. Given the flux J through the system in steady-state of growth, the concentration of ribosome bound release factor (e.g., for RF4) is J/k_{cat}^{RF4}, which becomes $\frac{\ell_{RF4}\lambda}{\langle\ell\rangle k_{cat}^{RF4}}$ upon converting to proteome fraction. This quantity sets the absolute minimum for the release factor abundance necessary to sustain growth λ for a given k_{cat}. The free concentrations for the release factors are then:

$$\phi_{RF1}^{free} = \phi_{RF1} - \frac{\ell_{RF1}\lambda}{\langle\ell\rangle k_{cat}^{RF1}}, \quad \phi_{RF4}^{free} = \phi_{RF4} - \frac{\ell_{RF4}\lambda}{\langle\ell\rangle k_{cat}^{RF4}}. \tag{19}$$

Hence, the final solution for the steady-state termination time as a function of the total abundance of the release factors and growth rate is:

$$\tau_{ter} = \frac{1}{k_{on}^{RFI}\left(\phi_{RFI} - \frac{\ell_{RFI}\lambda}{\langle\ell\rangle k_{cat}^{RFI}}\right)} + \frac{1}{k_{cat}^{RFI}} + \frac{1}{k_{on}^{RF4}\left(\phi_{RF4} - \frac{\ell_{RF4}\lambda}{\langle\ell\rangle k_{cat}^{RF4}}\right)} + \frac{1}{k_{cat}^{RF4}}.$$

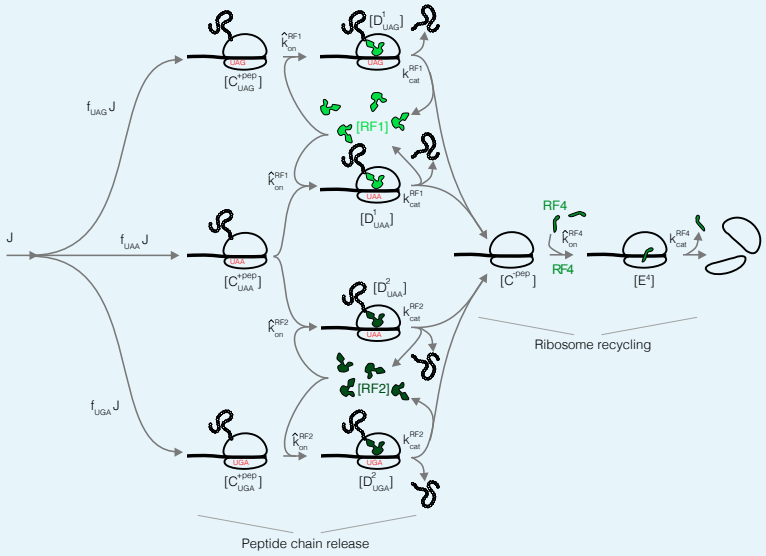
The relationship above, between termination time, total tIF abundance, and growth rate λ closes the solution of the kinetic scheme. Substituting the above in the optimality condition (equation 5) leads to the solution:

$$\phi_{RFI}^{*} = \sqrt{\frac{\ell_{ribo}\lambda^{*}}{\langle\ell\rangle k_{on}^{RFI}}} + \frac{\ell_{RFI}\lambda^{*}}{\langle\ell\rangle k_{cat}^{RFI}}, \quad \phi_{RF4}^{*} = \sqrt{\frac{\ell_{ribo}\lambda^{*}}{\langle\ell\rangle k_{on}^{RF4}}} + \frac{\ell_{RF4}\lambda^{*}}{\langle\ell\rangle k_{cat}^{RF4}}. \tag{20}$$

The additional terms $\propto \lambda^{*}$ correspond to the contribution to the optimal abundance arising from the finite catalytic rates, no present in the diffusion limited regime (equation 7).

Full three stop codons model

The full model with three different stop codons (UAA, UGA, UAG) and RF1/RF2 with different specificities (RF1: UAA, UAG; RF2: UAA, UGA) can also be solved exactly, leading to a small correction on the summed optimal abundance for RF1 and RF2 of $\sqrt{1+2\sqrt{f_{UAG}f_{UGA}}} < 1.05$ (fast growing species considered, where f_{UAG} and f_{UGA} are the fractional fluxes through the RF1 and RF2 stop codons respectively) compared to the single stop codon optimum derived above (ϕ_{RFI}^{*}, equation 20). We provide details below. With three stop codons, the coarse-grained reaction scheme is shown in Appendix 2 Figure 1. The relevant chemical species and parameters are listed in Appendix 2 Table 1.



Appendix 2 Figure 1. Coarse-grained translation termination scheme with three stop codons and RF1/RF2.

Variable	Description
$[C^{+pep}_{UAA}]$	Ribosomes at UAA with peptide chain [μM]
$[C^{+pep}_{UAG}]$	Ribosomes at UAG with peptide chain [μM]
$[C^{+pep}_{UGA}]$	Ribosomes at UGA with peptide chain [μM]
$[D^{1}_{UAA}]$	Ribosomes at UAA with peptide chain and RF1 bound [μM]
$[D^{1}_{UAG}]$	Ribosomes at UAG with peptide chain and RF1 bound [μM]
$[D^{2}_{UAA}]$	Ribosomes at UAA with peptide chain and RF2 bound [μM]
$[D^{2}_{UGA}]$	Ribosomes at UGA with peptide chain and RF2 bound [μM]
$[C^{-pep}]$	Ribosomes at all stops without peptide chain [μM]
$[E^{4}]$	Ribosomes at all stops without peptide chain and RF4 bound [μM]
$[RF1]$	Free RF1 [μM]
$[RF2]$	Free RF2 [μM]
$[RF4]$	Free RF4 [μM]
$J^{\mathrm{UAA}} = f_{\mathrm{UAA}} J$	Ribosome flux through UAA [μM s^{-1}]
$J^{\mathrm{UAG}} = f_{\mathrm{UAG}} J$	Ribosome flux through UAG [μM s^{-1}]
$J^{\mathrm{UGA}} = f_{\mathrm{UGA}} J$	Ribosome flux through UGA [μM s^{-1}]
$\hat{k}^{RF1}_{on}$	On-rate for RF1 [μM^{-1} s^{-1}]
$\hat{k}^{RF2}_{on}$	On-rate for RF2 [μM^{-1} s^{-1}]
$\hat{k}^{RF4}_{on}$	On-rate for RF4 [μM^{-1} s^{-1}]
k^{RF1}_{cat}	Catalytic rate for RF1 [s^{-1}]
k^{RF2}_{cat}	Catalytic rate for RF2 [s^{-1}]
k^{RF4}_{cat}	Catalytic rate for RF4 [s^{-1}]
$RF1_{tot}$	Total RF1 [μM]
$RF2_{tot}$	Total RF2 [μM]
$RF4_{tot}$	Total RF4 [μM]

Appendix 2 Table 1. Chemical species and parameters in three stop codons termination model.

The corresponding mass action system of equations for peptide release:

$$\begin{aligned}
\frac{\mathrm{d}[C^{+pep}_{UAA}]}{\mathrm{dt}} &= f_{UAA}J - [C^{+pep}_{UAA}]\left(\hat{k}^{RF1}_{on}[RF1] + \hat{k}^{RF2}_{on}[RF1]\right),\\
\frac{\mathrm{d}[C^{+pep}_{UAG}]}{\mathrm{dt}} &= f_{UAG}J - \hat{k}^{RF1}_{on}[C^{+pep}_{UAG}][RF1],\\
\frac{\mathrm{d}[C^{+pep}_{UGA}]}{\mathrm{dt}} &= f_{UGA}J - \hat{k}^{RF2}_{on}[C^{+pep}_{UGA}][RF1],\\
\frac{\mathrm{d}[D^{1}_{UAA}]}{\mathrm{dt}} &= \hat{k}^{RF1}_{on}[RF1][C^{+pep}_{UAA}] - k^{RF1}_{cat}[D^{1}_{UAA}],\\
\frac{\mathrm{d}[D^{1}_{UAG}]}{\mathrm{dt}} &= \hat{k}^{RF1}_{on}[RF1][C^{+pep}_{UAG}] - k^{RF1}_{cat}[D^{1}_{UAG}],\\
\frac{\mathrm{d}[D^{2}_{UAA}]}{\mathrm{dt}} &= \hat{k}^{RF2}_{on}[RF2][C^{+pep}_{UAA}] - k^{RF1}_{cat}[D^{2}_{UAA}],\\
\frac{\mathrm{d}[D^{2}_{UGA}]}{\mathrm{dt}} &= \hat{k}^{RF2}_{on}[RF2][C^{+pep}_{UGA}] - k^{RF1}_{cat}[D^{2}_{UGA}],\\
\frac{\mathrm{d}[RF1]}{\mathrm{dt}} &= -\hat{k}^{RF1}_{on}[RF1]\left([C^{+pep}_{UAA}] + [C^{+pep}_{UAG}]\right) + k^{RF1}_{cat}\left([D^{1}_{UAA}] + [D^{1}_{UAG}]\right),\\
\frac{\mathrm{d}[RF2]}{\mathrm{dt}} &= -\hat{k}^{RF2}_{on}[RF2]\left([C^{+pep}_{UAA}] + [C^{+pep}_{UGA}]\right) + k^{RF2}_{cat}\left([D^{2}_{UAA}] + [D^{2}_{UGA}]\right).
\end{aligned}$$

And for ribosome recycling:

$$\begin{aligned}
\frac{\mathrm{d}[C^{-pep}]}{\mathrm{dt}} &= k^{RF1}_{cat}\left([D^{1}_{UAA}] + [D^{1}_{UAG}]\right) + k^{RF2}_{cat}\left([D^{2}_{UAA}] + [D^{2}_{UGA}]\right) - \hat{k}^{RF4}_{on}[C^{-pep}][RF4],\\
\frac{\mathrm{d}[E^{4}]}{\mathrm{dt}} &= \hat{k}^{RF4}_{on}[C^{-pep}][RF4] - k^{RF4}_{cat}[E^{4}],\\
\frac{\mathrm{d}[RF4]}{\mathrm{dt}} &= -\hat{k}^{RF4}_{on}[C^{-pep}][RF4] + k^{RF4}_{cat}[E^{4}].
\end{aligned}$$

The conservation equations for RF1, RF2 and RF4 are:

$$\begin{aligned}
RF1_{tot} &= [RF1] + [D^{1}_{UAA}] + [D^{1}_{UAG}],\\
RF2_{tot} &= [RF2] + [D^{2}_{UAA}] + [D^{2}_{UGA}],\\
RF4_{tot} &= [RF4] + [E^{4}].
\end{aligned}$$

With a more complex scheme such as the one above, the optimization problem can be solved in three steps. First, we obtain the steady-state concentration of the chemical species. Second, we determine the effective coarse-grained termination time. Finally,

the optimal abundance is found by substituting the termination time in the optimality condition (equation 5), and solving the resulting system of equation.

Steady-state concentrations for RFs

Note that the RF1/RF2 and RF4 completely decouple, and that the solution for RF4 is identical to the one stop codon case solved above (section Non binding-limited regime (one stop codon)). For peptide chain release, the steady-state of the system can be solved by expressing the all chemical species in terms of $[RF1]$, and $[RF2]$:

$$
\begin{aligned}
[C^{+pep}_{UAA}] &= \frac{f_{UAA} J}{\hat{k}^{RF1}_{on}[RF1] + \hat{k}^{RF2}_{on}[RF2]} \\
[D^{1}_{UAA}] &= f_{UAA} \frac{J}{k^{RF1}_{cat}} \left(\frac{\hat{k}^{RF1}_{on}[RF1]}{\hat{k}^{RF1}_{on}[RF1] + \hat{k}^{RF2}_{on}[RF2]} \right), \\
[D^{2}_{UAA}] &= f_{UAA} \frac{J}{k^{RF2}_{cat}} \left(\frac{\hat{k}^{RF2}_{on}[RF2]}{\hat{k}^{RF1}_{on}[RF1] + \hat{k}^{RF2}_{on}[RF2]} \right), \\
[C^{+pep}_{UAG}] &= \frac{f_{UAG} J}{\hat{k}^{RF1}_{on}[RF1]}, \; [C^{+pep}_{UGA}] = \frac{f_{UGA} J}{\hat{k}^{RF2}_{on}[RF2]}, \; [D^{1}_{UAG}] = f_{UAG} \frac{J}{k^{RF1}_{cat}}, \; [D^{2}_{UGA}] = f_{UGA} \frac{J}{k^{RF2}_{cat}}.
\end{aligned}
\tag{21}
$$

Substituting these in the conservation equations for RF1 and RF2 leads to a closed system in terms of $[RF1]$ and $[RF2]$:

$$
\begin{aligned}
RF1_{tot} &= [RF1] \left[1 + f_{UAA} \frac{J}{k^{RF1}_{cat}} \left(\frac{\hat{k}^{RF1}_{on}}{\hat{k}^{RF1}_{on}[RF1] + \hat{k}^{RF2}_{on}[RF2]} \right) \right] + f_{UAG} \frac{J}{k^{RF1}_{cat}}, \\
RF2_{tot} &= [RF2] \left[1 + f_{UAA} \frac{J}{k^{RF2}_{cat}} \left(\frac{\hat{k}^{RF2}_{on}}{\hat{k}^{RF1}_{on}[RF1] + \hat{k}^{RF2}_{on}[RF2]} \right) \right] + f_{UGA} \frac{J}{k^{RF2}_{cat}}.
\end{aligned}
$$

Under the assumption of identical biochemical properties for RF1 and RF2, namely $k^{RF1}_{cat} = k^{RF2}_{cat} := k^{RFI}_{cat}$ and $\hat{k}^{RF1}_{on} = \hat{k}^{RF2}_{on} := \hat{k}^{RFI}_{on}$, the total free concentration of RF1 and RF2 simplifies to: $[RF1]+[RF2] = RF1_{tot} + RF2_{tot} - \frac{J}{k^{RFI}_{cat}}$, where we used $f_{UAA} + f_{UAG} + f_{UGA} = 1$ (by definition). Using this relation to eliminate $[RF2]$ from the $[RF1]$ equation (and vice-versa), we obtain, upon conversion to proteome fraction:

$$
\begin{aligned}
\phi^{free}_{RF,tot} &:= \phi_{RF1} + \phi_{RF2} - \frac{\ell_{RFI} \lambda}{\langle \ell \rangle k^{RFI}_{cat}}, \\
\phi^{free}_{RF1} &= \chi_{RF1} \, \phi^{free}_{RF,tot}, \;\; \phi^{free}_{RF2} = \chi_{RF2} \, \phi^{free}_{RF,tot},
\end{aligned}
\tag{22}
$$

where

$$
\begin{aligned}
\chi_{RF1} &:= \frac{\phi_{RF1} - \frac{\ell_{RFI} \lambda}{\langle \ell \rangle k^{RFI}_{cat}} f_{\mathrm{UAG}}}{(\phi_{RF1} - \frac{\ell_{RFI} \lambda}{\langle \ell \rangle k^{RFI}_{cat}} f_{\mathrm{UAG}}) + (\phi_{RF2} - \frac{\ell_{RFI} \lambda}{\langle \ell \rangle k^{RFI}_{cat}} f_{\mathrm{UGA}})}, \\
\chi_{RF2} &:= \frac{\phi_{RF2} - \frac{\ell_{RFI} \lambda}{\langle \ell \rangle k^{RFI}_{cat}} f_{\mathrm{UGA}}}{(\phi_{RF1} - \frac{\ell_{RFI} \lambda}{\langle \ell \rangle k^{RFI}_{cat}} f_{\mathrm{UAG}}) + (\phi_{RF2} - \frac{\ell_{RFI} \lambda}{\langle \ell \rangle k^{RFI}_{cat}} f_{\mathrm{UGA}})}.
\end{aligned}
$$

These constitute the steady-state solutions of the system of equation.

Coarse-grained translation termination time

In order to obtain an expression for the termination time (peptide release portion), needed to determine the optimal RF abundance (i.e., to substitute in equation 5), the peptide chain release contribution arises from the ribosome containing species listed in equation 21, which

sum to (under the assumption of identical biochemical properties for RF1/RF2):

$$[R^{pep}_{ter}] = [C^{+pep}_{UAA}] + [C^{+pep}_{UAG}] + [C^{+pep}_{UGA}] + [D^{1}_{UAA}] + [D^{1}_{UAG}] + [D^{2}_{UAA}] + [D^{2}_{UGA}],$$

$$[R^{pep}_{ter}] = J\left(\frac{f_{UAG}}{\hat{k}^{RFI}_{on}[RF1]} + \frac{f_{UGA}}{\hat{k}^{RFI}_{on}[RF2]} + \frac{f_{UAA}}{\hat{k}^{RFI}_{on}([RF1]+[RF2])} + \frac{1}{k^{RFI}_{cat}}\right).$$

Upon conversion to proteome fraction, the above becomes:

$$\phi^{pep}_{ribo} = \frac{\ell_{ribo}}{\langle \ell \rangle}\lambda\left(\frac{f_{UAG}}{k^{RFI}_{on}\phi^{free}_{RF1}} + \frac{f_{UGA}}{k^{RFI}_{on}\phi^{free}_{RF2}} + \frac{f_{UAA}}{k^{RFI}_{on}\left(\phi^{free}_{RF1}+\phi^{free}_{RF2}\right)} + \frac{1}{k^{RFI}_{cat}}\right) := \frac{\ell_{ribo}}{\langle \ell \rangle}\lambda\tau_{pep}.$$

The bracketed term corresponds to the coarse-grained time associated with peptide chain release τ_{pep}, and the free concentrations are given by equations 22.

Optimal abundances for RF1/RF2

The solved concentrations in steady-state (as a function of proteome fractions) and coarse-grained times allow us to determine the optimal RF1 and RF2 solutions (within our model). The optimality condition (equation 5) is now:

$$\left(\frac{\partial \tau_{pep}}{\partial \phi_{\mathrm{RF1}}}\right)^{*} = -\frac{\langle \ell \rangle}{\ell_{ribo}\lambda^{*}}, \quad \left(\frac{\partial \tau_{pep}}{\partial \phi_{\mathrm{RF2}}}\right)^{*} = -\frac{\langle \ell \rangle}{\ell_{ribo}\lambda^{*}}.$$

Solving the above system leads to optima ϕ^{*}_{RF1} and ϕ^{*}_{RF2}:

$$\phi^{*}_{RF1} + \phi^{*}_{RF2} = \sqrt{\frac{\ell_{ribo}\lambda^{*}(1+\delta)}{\langle \ell \rangle k^{RFI}_{on}}} + \frac{\ell_{RFI}\lambda^{*}}{\langle \ell \rangle k^{RFI}_{cat}}, \tag{23}$$

$$\frac{\phi^{*}_{RF1} - \frac{f_{UAG}\ell_{RFI}\lambda^{*}}{\langle \ell \rangle k^{RFI}_{cat}}}{\phi^{*}_{RF2} - \frac{f_{UGA}\ell_{RFI}\lambda^{*}}{\langle \ell \rangle k^{RFI}_{cat}}} = \sqrt{\frac{f_{UAG}}{f_{UGA}}}. \tag{24}$$

where the new factor $\delta := 2\sqrt{f_{UAG}f_{UGA}}$.

The relative flux through each stop codon ($f_{UAA}, f_{UAG}, f_{UGA}$) can be estimated in a variety of bacteria from ribosome profiling data (***Lalanne et al., 2018***) as the total synthesis fraction of genes with the respective stop codon. For fast growing species considered in the current study, $f_{UAA} \approx 0.9$, and the correction term to the optimal solution for the summed abundance of RF1 and RF2 ($\sqrt{1+\delta}$) is consequently small (*E. coli*: $f_{UAA} = 0.888, f_{UAG} = 0.015, f_{UGA} = 0.097, \sqrt{1+\delta} = 1.04$; *B. subtilis*: $f_{UAA} = 0.888, f_{UAG} = 0.064, f_{UGA} = 0.049, \sqrt{1+\delta} = 1.05$, *V. natriegens*: $f_{UAA} = 0.929, f_{UAG} = 0.041, f_{UGA} = 0.031, \sqrt{1+\delta} = 1.04$)

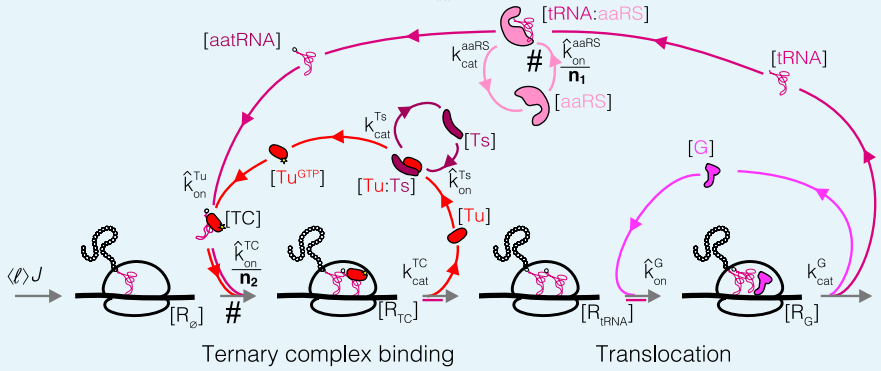
Appendix 3

Translation elongation

Coarse-grained one-codon model

Translation elongation is a more complicated process than termination, involving multiple factors to bring the charged tRNA to the ribosome (EF-Tu), charge the tRNAs (aaRS), translocate the ribosome (EF-G), and perform nucleotide exchange on EF-Tu to drive the process (EF-Ts), in addition to others not included here. Our simplified kinetic scheme is illustrated in Appendix 3 Figure 1. In anticipation coarse-graining procedure detailed below, rates rescaled in the conversion to a one-codon model are marked by $*$.

To simplify our model, we coarse-grain the elongation cycle by considering a single codon type (section Estimation of coarse-grained rates below or details of the coarse-graining procedure), effectively grouping the tRNA's, tRNA synthetases, and different ternary complexes to single entities. Importantly, as a result, the on-rates associated with these processes are rescaled by a factor close to n_{aa}^{-1}, where $n_{aa} = 20$.



Appendix 3 Figure 1. Coarse-grained reaction scheme for a single step (amino acid incorporation) of translation elongation. Tu: EF-Tu, Ts: EF-Ts, G: EF-G, aaRS: aminoacyl tRNA synthetases. Steps with slower rates as a result of the coarse-graining to one effective codon are marked by #.

An important distinction for elongation compared to initiation and termination is that multiple elongation steps (average $\langle \ell \rangle \approx 200$) are required to generate a protein. Hence, the flux into the through the elongation cycle is $\langle \ell \rangle$ larger than that through the initiation and termination steps (there is one initiation and termination event for each protein made, but about 200 elongation steps on average).

The mass action reaction scheme for translation elongation:

$$\begin{aligned}
&\xrightarrow{\langle \ell \rangle J} \mathrm{R}_{\emptyset}, \\
\mathrm{tRNA} + \mathrm{aaRS} &\xrightarrow{\hat{k}_{on}^{aaRS}/n_1} \mathrm{tRNA{:}aaRS}, \\
\mathrm{tRNA{:}aaRS} &\xrightarrow{k_{cat}^{aaRS}} \mathrm{aatRNA} + \mathrm{aaRS} \\
\mathrm{Tu} + \mathrm{Ts} &\xrightarrow{\hat{k}_{on}^{Ts}} \mathrm{Tu{:}Ts}, \\
\mathrm{Tu{:}Ts} &\xrightarrow{k_{cat}^{Ts}} \mathrm{Tu^{GTP}} + \mathrm{Ts}, \\
\mathrm{Tu^{GTP}} + \mathrm{aatRNA} &\xrightarrow{\hat{k}_{on}^{Tu}} \mathrm{TC}, \\
\mathrm{TC} + \mathrm{R}_{\emptyset} &\xrightarrow{\hat{k}_{on}^{TC}/n_2} \mathrm{R}_{TC}, \\
\mathrm{R}_{TC} &\xrightarrow{k_{cat}^{TC}} \mathrm{R}_{tRNA}, \\
\mathrm{R}_{tRNA} + \mathrm{G} &\xrightarrow{\hat{k}_{on}^{G}} \mathrm{R}_{G}, \\
\mathrm{R}_{G} &\xrightarrow{k_{cat}^{G}} \mathrm{G} + \mathrm{tRNA}.
\end{aligned} \tag{25}$$

To arrive at the above, we started with a full model of translation (not shown), will all possible codons, tRNA species, and ribosomes with different codons. To coarse-grain the model, we introduced the following effective variables, which correspond to the total concentration of each type of species involved, summed over the of the codon/amino acid specificity:

$$[\text{tRNA}] := \sum_i [\text{tRNA}_i],\ \ [\text{aatRNA}] := \sum_i [\text{aatRNA}_i],\ \ [\text{aaRS}] := \sum_{\hat{i}} [\text{aaRS}_{\hat{i}}],\ \ [\text{TC}] := \sum_i [\text{TC}_i]$$

$$[\text{R}_\emptyset] := \sum_{i,\nu,\mu} [\text{R}^i_{\nu\mu}],\ \ [\text{R}_{TC}] := \sum_{i,j,\nu,\mu} [\text{R}^{i\,TC_j}_{\nu\mu}],\ \ [\text{R}_{tRNA}] := \sum_{i,j,\nu,\mu} [\text{R}^{ij}_{\nu\mu}],\ \ [\text{R}_G] := \sum_{i,j,\nu,\mu} [\text{R}^{ij}_{\nu\mu} :: G].$$

In the above, Greek indices correspond to different codons on mRNAs, and Roman indices to different tRNAs. Roman indices with a hat ($\hat{i}$) correspond to tRNA synthetases recognizing specific tRNAs (multiple amino acids have more than one tRNA isoacceptor). In defining these coarse-grained species (our approach is analogous to that of (***Dai et al., 2016***)), we redefined the two following kinetic parameters:

$$\frac{\hat{k}^{aaRS}_{on}}{n_1} := \hat{k}^{aaRS}_{on} \sum_i \frac{[\text{tRNA}_i][\text{aaRS}_{\hat{i}}]}{[\text{tRNA}][\text{aaRS}]},\ \text{and}\ \ \frac{\hat{k}^{TC}_{on}}{n_2} := \hat{k}^{TC}_{on} \sum_{\mu,\nu,i,j} \frac{[\text{R}^i_{\mu\nu}] S_{\nu,j} [\text{TC}_j]}{[\text{R}_\emptyset][\text{TC}]}. \tag{26}$$

$\hat{k}^{aaRS}_{on}$ and $\hat{k}^{TC}_{on}$ correspond to the microscopic bimolecular rates (assumed equal for the different chemical species). $S_{\nu,j}$ is the tRNA isoacceptor/codon specificity matrix (1 if tRNA i can recognize codon ν, 0 otherwise) (***Björk and Hagervall, 2014***). Rescaling terms n_1 and n_2 are estimated below.

Estimation of coarse-grained rates

The definition of coarse-grained parameters (equations 26) involves sums:

$$\frac{1}{n_1} := \sum_i \frac{[\text{tRNA}_i][\text{aaRS}_{\hat{i}}]}{[\text{tRNA}][\text{aaRS}]}\ \ \text{and}\ \ \frac{1}{n_2} := \sum_{\mu,\nu,i,j} \frac{[\text{R}^i_{\mu\nu}] S_{\nu,j} [\text{TC}_j]}{[\text{R}_\emptyset][\text{TC}]}.$$

These can be estimated from tRNA abundances, codon usage and individual synthetases' levels obtained from ribosome profiling data in *E. coli* (***Li et al., 2014***).

We first consider n_1. Note that the fraction of free tRNA of type i to the total number of free tRNA (not bound to any protein) is not readily measurable. Assuming similarities between types of tRNA's, we approximate this fraction with the fraction of total tRNA of type i to the total tRNA concentration, or

$$\frac{[\text{tRNA}_i]}{[\text{tRNA}]} \approx \frac{\text{tRNA}^{tot}_i}{\text{tRNA}_{tot}}.$$

The total tRNA concentration has been measured at fast growth for *E. coli* (***Dong et al., 1996***). The relative concentration of each tRNA synthetases (appropriately corrected for stoichiometry for the different classes) can be computed from the ribosome profiling data (***Li et al., 2014***), and we obtain

$$\frac{1}{n_1} := \sum_i \left(\frac{[\text{tRNA}_i]}{[\text{tRNA}]} \frac{[\text{aaRS}_{\hat{i}}]}{[\text{aaRS}]} \right) \approx \sum_i \left(\frac{\text{tRNA}^{tot}_i}{\text{tRNA}_{tot}} \frac{[\text{aaRS}_{\hat{i}}]}{[\text{aaRS}]} \right) \approx 0.056 \ \ \Rightarrow n_1 \approx 17.8$$

This was to be expected since the synthetases in *E. coli* show little variability around their mean, and in the case of equal synthetase concentration, $n_1 = 20$ would strictly hold.

For the second sum (n_2), we use distribution of ribosome footprint reads across the transcriptome to estimate ribosome occupancies at different codons. We first make the following approximation for one of the sub-sum:

$$\sum_{\mu,i} \frac{[\text{R}^i_{\mu\nu}]}{[\text{R}_\emptyset]} \approx \sum_\mu \frac{N^{FP}_{\mu\nu}}{N^{FP}_{tot}},$$

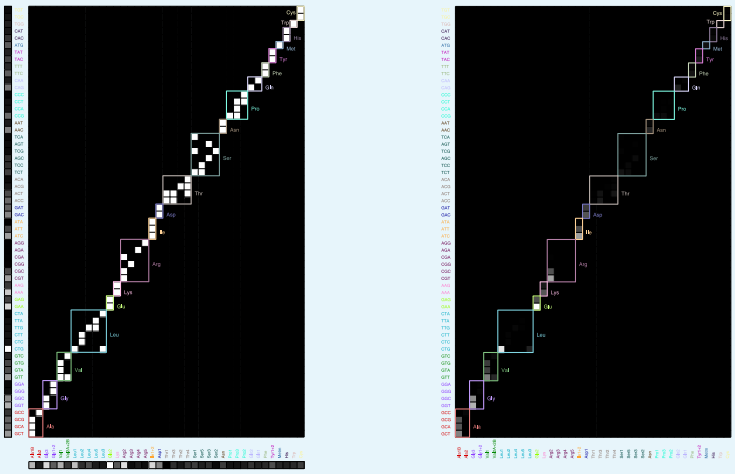
where $N^{FP}_{\mu\nu}$ is the total number of ribosome footprint reads at codon pairs μ, ν and N^{FP}_{tot} is the total number of footprint reads mapping to coding sequences. The nature of the approximation is that we are taking relative fraction of ribosome footprints (representing ribosomes across the elongation cycle at that codon pair) at a given codon pair to be equal to the relative fraction of ribosomes waiting for the ternary complex to derliver a tRNA to the A site. The modest differences in elongation rates at different codons seen in ribosome profiling data (***Mohammad et al., 2019***) justify this approximation.
From our data (not shown), we have that

$$\sum_{\mu} \frac{N^{FP}_{\mu\nu}}{N^{FP}_{tot}} \approx \sum_{\mu} \frac{N^{FP}_{\nu\mu}}{N^{FP}_{tot}} = \frac{N^{FP}_{\nu}}{N^{FP}_{tot}} := f_{\nu}$$

holds to better than 0.5% for each codon. f_{ν} above is the (expression weighted) codon usage. As before with the free tRNA concentrations, we can approximate the relative ternary complexes concentrations by the corresponding total tRNA concentrations:

$$\frac{1}{n_2} := \sum_{\mu,\nu,i,j} \frac{[R^{i}_{\mu\nu}] S_{\nu,j} [TC_j]}{[R][TC]} \approx \sum_{\nu,j} \frac{f_{\nu}\ S_{\nu,j}\ \text{tRNA}^{tot}_{j}}{\text{tRNA}_{tot}} \approx 0.048 \quad \Rightarrow n_2 \approx 20.8 \tag{27}$$

We used the same dataset as before for the total tRNA concentration in *E. coli* (***Dong et al., 1996***). The codon usage was determined directly from ribosome profiling data (***Li et al., 2014***). The sum of these products is graphically represented in Appendix 3 Figure 2. The above sum of product of tRNA fraction and codon usage provides an effective number of different ternary complexes. *A priori*, that might have been expected to equal to the number of tRNAs (≈40). However, as is apparent in Appendix 3 Figure 2, certain tRNA-codon pairs are much more prevalent than others (even for amino acid with multiple codons and/or tRNA isoacceptors), which leads to a decrease in the effective concentration. The exact value depends on the detailed codon usage and tRNA abundance.



Appendix 3 Figure 2. Graphical illustration of the sum (equation 27). Left: codon usage (vertical, from analysis of ribosome profiling data from (***Li et al., 2014***)), tRNA-codon specificity (matrix, from (***Björk and Hagervall, 2014***), with different amino acids outlined with different colors), and tRNA abundance (horizontal, from (***Dong et al., 1996***)) organized by amino acid. Right: product matrix.

Given the results above, we take for simplicity $n_1 = n_2 = n_{aa} = 20$.

Translation elongation: optimal solutions

The mass action reactions corresponding to the one codon elongation cycle model are (equations 25):

$$\begin{aligned}
\frac{\mathrm{d}[\mathrm{R}_\emptyset]}{\mathrm{dt}} &= \langle \ell \rangle J - \frac{\hat{k}_{on}^{TC}}{n_{aa}}[\mathrm{TC}][\mathrm{R}_\emptyset],\\
\frac{\mathrm{d}[\mathrm{R}_{TC}]}{\mathrm{dt}} &= \frac{\hat{k}_{on}^{TC}}{n_{aa}}[\mathrm{TC}][\mathrm{R}_\emptyset] - k_{cat}^{TC}[\mathrm{R}_{TC}],\\
\frac{\mathrm{d}[\mathrm{Tu}]}{\mathrm{dt}} &= k_{cat}^{TC}[\mathrm{R}_{TC}] - \hat{k}_{on}^{Ts}[\mathrm{Tu}][\mathrm{Ts}],\\
\frac{\mathrm{d}[\mathrm{tRNA}]}{\mathrm{dt}} &= -\frac{\hat{k}_{on}^{aaRS}}{n_{aa}}[\mathrm{tRNA}][\mathrm{aaRS}] + k_{cat}^{G}[\mathrm{R}_G],\\
\frac{\mathrm{d}[\mathrm{tRNA::aaRS}]}{\mathrm{dt}} &= \frac{\hat{k}_{on}^{aaRS}}{n_{aa}}[\mathrm{tRNA}][\mathrm{aaRS}] - k_{cat}^{aaRS}[\mathrm{tRNA::aaRS}] = -\frac{\mathrm{d}[\mathrm{aaRS}]}{\mathrm{dt}},\\
\frac{\mathrm{d}[\mathrm{aatRNA}]}{\mathrm{dt}} &= k_{cat}^{aaRS}[\mathrm{tRNA::aaRS}] - \hat{k}_{on}^{Tu}[\mathrm{aatRNA}][\mathrm{Tu}^{\mathrm{GTP}}],\\
\frac{\mathrm{d}[\mathrm{Tu}^{\mathrm{GTP}}]}{\mathrm{dt}} &= k_{cat}^{Ts}[\mathrm{Tu{:}Ts}] - \hat{k}_{on}^{Tu}[\mathrm{aatRNA}][\mathrm{Tu}^{\mathrm{GTP}}],\\
\frac{\mathrm{d}[\mathrm{Tu{:}Ts}]}{\mathrm{dt}} &= -k_{cat}^{Ts}[\mathrm{Tu{:}Ts}] + \hat{k}_{on}^{Ts}[\mathrm{Tu}][\mathrm{Ts}] = -\frac{\mathrm{d}[\mathrm{Ts}]}{\mathrm{dt}},\\
\frac{\mathrm{d}[\mathrm{TC}]}{\mathrm{dt}} &= \hat{k}_{on}^{Tu}[\mathrm{aatRNA}][\mathrm{Tu}^{\mathrm{GTP}}] - \frac{\hat{k}_{on}^{TC}}{n_{aa}}[\mathrm{TC}][\mathrm{R}_\emptyset],\\
\frac{\mathrm{d}[\mathrm{R}_{tRNA}]}{\mathrm{dt}} &= k_{cat}^{TC}[\mathrm{R}_{TC}] - \hat{k}_{on}^{G}[\mathrm{R}_{tRNA}][\mathrm{G}],\\
\frac{\mathrm{d}[\mathrm{R}_G]}{\mathrm{dt}} &= \hat{k}_{on}^{G}[\mathrm{R}_{tRNA}][\mathrm{G}] - k_{cat}^{G}[\mathrm{R}_G] = -\frac{\mathrm{d}[\mathrm{G}]}{\mathrm{dt}}.
\end{aligned}$$

Conservation equations close the system:

$$\begin{aligned}
\mathrm{Ts}_{tot} &= [\mathrm{Ts}] + [\mathrm{Tu{:}Ts}],\\
\mathrm{Tu}_{tot} &= [\mathrm{Tu}] + [\mathrm{Tu}^{\mathrm{GTP}}] + [\mathrm{Tu{:}Ts}] + [\mathrm{TC}] + [\mathrm{R}_{TC}],\\
\mathrm{tRNA}_{tot} &= [\mathrm{R}_\emptyset] + 2[\mathrm{R}_{TC}] + 2[\mathrm{R}_{tRNA}] + 2[\mathrm{R}_G] + [\mathrm{tRNA}] + [\mathrm{tRNA{:}aaRS}] + [\mathrm{aatRNA}] + [\mathrm{TC}],\\
\mathrm{aaRS}_{tot} &= [\mathrm{tRNA{:}aaRS}] + [\mathrm{aaRS}],\\
\mathrm{G}_{tot} &= [\mathrm{G}] + [\mathrm{R}_G].
\end{aligned}$$

The ternary complex concentration and free EF-G concentration enter the translation elongation time (equation 10, which is the diffusion limited and factor dependent contribution to the elongation time) and are required to infer optimal abundances of elongation factors. Both can to be obtained by solving the system of non-linear equations above.
First, catalytic steps must equal to the flux through in the system in steady-state and thus:

$$[\mathrm{R}_G] = \frac{\langle \ell \rangle J}{k_{cat}^{G}},\ [\mathrm{R}_{TC}] = \frac{\langle \ell \rangle J}{k_{cat}^{TC}},\ [\mathrm{tRNA::aaRS}] = \frac{\langle \ell \rangle J}{k_{cat}^{aaRS}},\ [\mathrm{Tu::Ts}] = \frac{\langle \ell \rangle J}{k_{cat}^{Ts}}.$$

Together with the conservation equations, these allow for immediate solutions for the free concentrations [Ts], [aaRS], and [G]:

$$\begin{aligned}
[\mathrm{Ts}] &= \mathrm{Ts}_{tot} - \frac{\langle \ell \rangle J}{k_{cat}^{Ts}},\\
[\mathrm{aaRS}] &= \mathrm{aaRS}_{tot} - \frac{\langle \ell \rangle J}{k_{cat}^{aaRS}},\\
[\mathrm{G}] &= \mathrm{G}_{tot} - \frac{\langle \ell \rangle J}{k_{cat}^{G}}.
\end{aligned}$$

The solution for other species can then also be obtained in terms [Tu^{GTP}], and [TC]:

$$[R_{tRNA}] = \frac{\langle \ell \rangle \boldsymbol{J}}{\hat{k}^{G}_{on}\left(G_{tot} - \frac{\langle \ell \rangle J}{k^{G}_{cat}}\right)},\ [R_{\emptyset}] = \frac{\langle \ell \rangle n_{aa} \boldsymbol{J}}{\hat{k}^{TC}_{on}[TC]}$$

$$[tRNA] = \frac{\langle \ell \rangle n_{aa} \boldsymbol{J}}{\hat{k}^{aaRS}_{on}\left(aaRS_{tot} - \frac{\langle \ell \rangle J}{k^{aaRS}_{cat}}\right)},\ [aatRNA] = \frac{\langle \ell \rangle \boldsymbol{J}}{\hat{k}^{Tu}_{on}[Tu^{GTP}]},$$

$$[Tu] = \frac{\langle \ell \rangle \boldsymbol{J}}{\hat{k}^{Ts}_{on}\left(Ts_{tot} - \frac{\langle \ell \rangle J}{k^{Ts}_{cat}}\right)}.$$

Substituting these in the conservation equations for tRNAs and EF-Tu lead to the final system to solve (converting to proteome fraction):

$$\frac{tRNA_{tot}}{P} := \psi_{tRNA} = \frac{\lambda n_{aa}}{k^{TC}_{on}\phi_{TC}} + \frac{2\lambda}{k^{TC}_{cat}} + \frac{2\lambda}{k^{G}_{on}\left(\phi_G - \frac{\ell_G \lambda}{k^{G}_{cat}}\right)} + \frac{2\lambda}{k^{G}_{cat}} + \ldots \tag{28}$$

$$\frac{\lambda n_{aa}}{k^{aaRS}_{on}\left(\phi_{aaRS} - \frac{\ell_{aaRS}\lambda}{k^{aaRS}_{cat}}\right)} + \frac{\lambda}{k^{aaRS}_{cat}} + \frac{\lambda}{k^{Tu}_{on}\phi_{Tu^{GTP}}} + \frac{\phi_{TC}}{\ell_{Tu}},$$

$$\text{where } \phi_{Tu^{GTP}} := \phi_{Tu} - \frac{\ell_{Tu}\lambda}{k^{Ts}_{on}\left(\phi_{Ts} - \frac{\ell_{Ts}\lambda}{k^{Ts}_{cat}}\right)} - \frac{\ell_{Tu}\lambda}{k^{Ts}_{cat}} - \phi_{TC} - \frac{\ell_{Tu}\lambda}{k^{TC}_{cat}}. \tag{29}$$

where the solution for $\phi_{Tu^{GTP}}$ in terms of the ternary concentration was obtained from the conservation equation for EF-Tu. Equations 28 and 29 are closed, and the only variables to solve for is ϕ_{TC} in terms of the tIF abundances: $\phi_{Tu}, \phi_{Ts}, \phi_G, \phi_{aaRS}$, tRNA abundances, kinetic parameters, and the growth rate λ.

Coarse-grained translation elongation time

In order to obtain the coarse-grained translation elongation time, we proceed as for translation termination (section Coarse-grained translation termination time). The summed concentration of the ribosome containing species for translation elongation in our model is:

$$[R_{el}] = [R_{\emptyset}] + [R_{TC}] + [R_{tRNA}] + [R_G],$$
$$= \frac{\langle \ell \rangle n_{aa} \boldsymbol{J}}{\hat{k}^{TC}_{on}[TC]} + \frac{\langle \ell \rangle \boldsymbol{J}}{k^{TC}_{cat}} + \frac{\langle \ell \rangle \boldsymbol{J}}{\hat{k}^{G}_{on}\left(G_{tot} - \frac{\langle \ell \rangle J}{k^{G}_{cat}}\right)} + \frac{\langle \ell \rangle \boldsymbol{J}}{k^{G}_{cat}}.$$

Converting to proteome fraction:

$$\frac{1}{\ell_{ribo}}\phi^{el}_{ribo} = \lambda\left(\frac{n_{aa}}{k^{TC}_{on}\phi_{TC}} + \frac{1}{k^{TC}_{cat}} + \frac{1}{k^{G}_{on}\left(\phi_G - \frac{\ell_G\lambda}{k^{G}_{cat}}\right)} + \frac{1}{k^{G}_{cat}}\right).$$

From the coarse-grained flux relations through the different categories (equation 17), which defines the coarse-grained transition times, we thus have:

$$\tau_{el} = \langle \ell \rangle \tau_{aa}, \text{ where } \tau_{aa} = \frac{n_{aa}}{k^{TC}_{on}\phi_{TC}} + \frac{1}{k^{TC}_{cat}} + \frac{1}{k^{G}_{on}\left(\phi_G - \frac{\ell_G\lambda}{k^{G}_{cat}}\right)} + \frac{1}{k^{G}_{cat}}. \tag{30}$$

Above, τ_{aa} is the effective time for a single step (by one codon) of translation elongation, and τ_{ind} corresponds to the summed time of factor independent transitions in each elongation step (not explicitly included in the kinetic scheme).

Optimality conditions for translation elongation factors

The optimality condition (equation 5) applied to translation elongation factors leads to:

$$\left(\frac{\partial \tau_{taa}}{\partial \phi_G}\right)^* = \left(\frac{\partial \tau_{taa}}{\partial \phi_{Tu}}\right)^* = \left(\frac{\partial \tau_{taa}}{\partial \phi_{Ts}}\right)^* = \left(\frac{\partial \tau_{taa}}{\partial \phi_{aaRS}}\right)^* = -\frac{1}{\ell_{ribo}\lambda^*}. \tag{31}$$

where equation 30 was used for τ_{aa}. Since the free EF-G concentration does not depend on EF-Tu, EF-Ts, or aaRS concentration, the conditions for EF-Tu, EF-Ts and aaRS simplify to:

$$\frac{\partial}{\partial \phi_{Tu}}\left(\frac{n_{aa}}{k_{on}^{TC}\phi_{TC}}\right)^* = \frac{\partial}{\partial \phi_{Ts}}\left(\frac{n_{aa}}{k_{on}^{TC}\phi_{TC}}\right)^* = \frac{\partial}{\partial \phi_{aaRS}}\left(\frac{n_{aa}}{k_{on}^{TC}\phi_{TC}}\right)^* = -\frac{1}{\ell_{ribo}\lambda^*}. \tag{32}$$

Carrying through the differentiation also leads to conditions on the derivatives of the ternary complex concentration at the optimum:

$$\left(\frac{\partial \phi_{TC}}{\partial \phi_{Tu}}\right)^* = \left(\frac{\partial \phi_{TC}}{\partial \phi_{Ts}}\right)^* = \left(\frac{\partial \phi_{TC}}{\partial \phi_{aaRS}}\right)^* = \frac{k_{on}^{TC}\left(\phi_{TC}^*\right)^2}{\ell_{ribo} n_{aa}\lambda^*}. \tag{33}$$

These relationships will be useful to solve for the some elongation factor optimal abundances below.

Optimal EF-Ts abundance

Differentiating equation 28 with respect to ϕ_{Tu} and ϕ_{Ts}, we get at the optimum:

$$\frac{1}{\ell_{ribo}} + \frac{\lambda^*}{k_{on}^{Tu}\left(\phi_{Tu^{GTP}}^*\right)^2}\left(\frac{\partial \phi_{Tu^{GTP}}}{\partial \phi_{Tu}}\right)^* = \frac{1}{\ell_{Tu}}\left(\frac{\partial \phi_{TC}}{\partial \phi_{Tu}}\right)^*,$$

$$\frac{1}{\ell_{ribo}} + \frac{\lambda^*}{k_{on}^{Tu}\left(\phi_{Tu^{GTP}}^*\right)^2}\left(\frac{\partial \phi_{Tu^{GTP}}}{\partial \phi_{Ts}}\right)^* = \frac{1}{\ell_{Tu}}\left(\frac{\partial \phi_{TC}}{\partial \phi_{Ts}}\right)^*.$$

By equation 33, the above leads to the additional condition at the optimum:

$$\left(\frac{\partial \phi_{Tu^{GTP}}}{\partial \phi_{Tu}}\right)^* = \left(\frac{\partial \phi_{Tu^{GTP}}}{\partial \phi_{Ts}}\right)^*.$$

Directly differentiating equation 29, and using equation 33, leads to:

$$\left(\frac{\partial \phi_{Tu^{GTP}}}{\partial \phi_{Tu}}\right)^* = 1 - \frac{k_{on}^{TC}\left(\phi_{TC}^*\right)^2}{\ell_{ribo} n_{aa}\lambda^*} = \left(\frac{\partial \phi_{Tu^{GTP}}}{\partial \phi_{Ts}}\right)^* = \frac{\ell_{Tu}\lambda^*}{k_{on}^{Ts}\left(\phi_{Ts}^* - \frac{\ell_{Ts}\lambda}{k_{cat}^{Ts}}\right)^2} - \frac{k_{on}^{TC}\left(\phi_{TC}^*\right)^2}{\ell_{ribo} n_{aa}\lambda^*}.$$

Therefore, the optimal abundance for EF-Ts is:

$$\phi_{Ts}^* = \sqrt{\frac{\ell_{Tu}\lambda^*}{k_{on}^{Ts}}} + \frac{\ell_{Ts}\lambda^*}{k_{cat}^{Ts}}. \tag{34}$$

Optimal EF-G abundance

The optimality condition for EF-G is complicated by the fact that EF-G free concentration appears in the solution for the steady-state ternary complex through the tRNA conservation equation 28. Differentiating the conservation tRNA equation, and using the optimality condition 31 (replacing a number of terms with the elongation time τ_{aa}, equation 30):

$$0 = -\frac{2}{\ell_{ribo}} + \frac{\lambda^* n_{aa}}{k_{on}^{TC}\left(\phi_{Tu}^*\right)^2}\left(\frac{\partial \phi_{TC}}{\partial \phi_G}\right)^* + \frac{1}{\ell_{Tu}}\left(\frac{\partial \phi_{TC}}{\partial \phi_G}\right)^* - \frac{\lambda^*}{k_{on}^{Tu}\left(\phi_{Tu^{GTP}}^*\right)^2}\left(\frac{\partial \phi_{Tu^{GTP}}}{\partial \phi_G}\right)^*. \tag{35}$$

Above, the right-hand portion corresponds to the additional constraint coming from the implication of EF-G in the steady-state concentration of the ternary complex. From the equation for $\phi_{Tu^{\mathrm{GTP}}}$ (equation 29), we have directly:

$$\left(\frac{\partial \phi_{Tu^{\mathrm{GTP}}}}{\partial \phi_G}\right)^* = -\left(\frac{\partial \phi_{TC}}{\partial \phi_G}\right)^*.$$

Substituting this in equation 35:

$$\frac{2}{\ell_{ribo}} = \left[\frac{1}{\ell_{Tu}} + \frac{\lambda^*}{k_{on}^{Tu}\left(\phi^*_{Tu^{\mathrm{GTP}}}\right)^2} + \frac{\lambda^* n_{aa}}{k_{on}^{TC}\left(\phi^*_{TC}\right)^2}\right]\left(\frac{\partial \phi_{TC}}{\partial \phi_G}\right)^*. \tag{36}$$

The derivative of the ternary complex with respect to EF-G at the optimum can be obtained from the original optimality condition 31, by carrying through the differentiation:

$$\left(\frac{\partial \phi_{TC}}{\partial \phi_G}\right)^* = \frac{k_{on}^{TC}}{n_{aa}}\left(\phi^*_{TC}\right)^2\left[\frac{1}{\ell_{ribo}\lambda^*} - \frac{1}{k_{on}^{G}\left(\phi^*_G - \frac{\ell_G \lambda^*}{k_{cat}^G}\right)^2}\right].$$

Substituting in equation 36, we arrive at a final equation for EF-G in terms of the concentration of other elongation factor and the optimal growth rate:

$$\frac{2}{\ell_{ribo}} = \lambda^*\left[1 + \frac{k_{on}^{TC}\left(\phi^*_{TC}\right)^2}{n_{aa}\ell_{Tu}\lambda^*} + \frac{k_{on}^{TC}\left(\phi^*_{TC}\right)^2}{n_{aa}k_{on}^{Tu}\left(\phi^*_{Tu^{\mathrm{GTP}}}\right)^2}\right]\left(\frac{1}{\ell_{ribo}\lambda^*} - \frac{1}{k_{on}^{G}\left(\phi^*_G - \frac{\ell_G\lambda^*}{k_{cat}^G}\right)^2}\right).$$

The optimal solution for EF-G is thus:

$$\phi^*_G = \sqrt{\frac{\ell_{ribo}\lambda^*}{k_{on}^G}\left(\frac{\Delta+1}{\Delta-1}\right)} + \frac{\ell_G\lambda^*}{k_{cat}^G} \geq \sqrt{\frac{\ell_{ribo}\lambda^*}{k_{on}^G}} + \frac{\ell_G\lambda^*}{k_{cat}^G}, \tag{37}$$

$$\text{where: } \Delta := \frac{k_{on}^{TC}\left(\phi^*_{TC}\right)^2}{n_{aa}\ell_{Tu}\lambda^*} + \frac{k_{on}^{TC}\left(\phi^*_{TC}\right)^2}{n_{aa}k_{on}^{Tu}\left(\phi^*_{Tu^{\mathrm{GTP}}}\right)^2}.$$

Note that given that the term Δ involves ϕ^*_{TC} and $\phi^*_{Tu^{\mathrm{GTP}}}$, and so the solution above is not a priori complete. However, using the approximate ternary complex concentration at the optimum (equation 12, derived in details in section Optimal EF-Tu and aaRS abundances), we have:

$$\Delta > \frac{k_{on}^{TC}\left(\phi^*_{TC}\right)^2}{n_{aa}\ell_{Tu}\lambda^*} \approx \frac{\ell_{ribo}}{\ell_{Tu}} \approx 18.5 \gg 1$$

This means that the lower bound for ϕ^*_G above (equation 37) is a good approximation: in the physiological regime, we can approximately neglect the indirect dependence of the ternary complex concentration on EF-G via the tRNA conservation equation. Hence, the approximate solution for the EF-G optimal abundance is (same for had we initially assumed that ϕ_{TC} was independent of ϕ_G, in which case the solution for EF-G can be obtained identically as that of release factors):

$$\phi^*_G \approx \sqrt{\frac{\ell_{ribo}\lambda^*}{k_{on}^G}} + \frac{\ell_G\lambda^*}{k_{cat}^G}.$$

Optimal EF-Tu and aaRS abundances

While simplifying relations were possible with EF-Ts and EF-G, allowing their solution (approximately) independently from the rest of the cycle, EF-Tu and aaRS are intricately connected through the tRNA cycle. We thus return to the tRNA conservation equation, equation 28. For notational simplicity, we group the catalytic step of the TC, EF-G binding, and EF-G catalytic action (translocation) in parameter k_{el}^{max} (these do not depend on ϕ_{Tu} and ϕ_{aaRS}) which we take to the be experimentally determined value of 22 s^{-1} (***Dai et al., 2016***). Further dropping the EF-Ts related and catalytic terms (will be added back at the end, they only contribute a fixed term at the optimum) in the equation for the free EF-Tu, we get:

$$\frac{\text{tRNA}_{tot}}{P\lambda} = \frac{n_{aa}}{k_{on}^{TC}\phi_{TC}} + \frac{2}{k_{el}^{max}} + \ldots \\ \frac{n_{aa}}{k_{on}^{aaRS}\left(\phi_{aaRS} - \frac{\ell_{aaRS}\lambda}{k_{cat}^{aaRS}}\right)} + \frac{1}{k_{cat}^{aaRS}} + \frac{1}{k_{on}^{Tu}\phi_{Tu^{\text{GTP}}}} + \frac{\phi_{TC}}{\ell_{Tu}\lambda}, \tag{38}$$

where $\phi_{Tu^{\text{GTP}}} = \phi_{Tu} - \phi_{TC}$ is the free EF-Tu concentration.

This system is first solved numerically (Figure 3B). To close the equation in terms of uniquely ϕ_{TC}, we use our relationship for λ (equation 1), with:

$$\tau_{trl} = \langle \ell \rangle \left(\frac{n_{aa}}{k_{on}^{TC}\phi_{TC}} + \frac{1}{k_{el}^{max}} \right) + \tau_{ini} + \tau_{ter},$$

where as before k_{el}^{max} is the maximum rate of translation elongation (from reactions other than ternary complex diffusion) estimated from *in vivo* kinetic measurements (≈ 22 s^{-1} (***Dai et al., 2016***)), and $\tau_{ini} + \tau_{ter} \approx 0.5$ s the estimated time for the initiation and termination step ($\approx 5-10\%$ of the full translation cycle translation time), taken as fixed parameters here. Using this relationship for the translation time leads to the explicit relationship between growth and ternary complex concentration:

$$\lambda(\phi_{TC}) = \frac{\phi_{ribo}}{\ell_{ribo}} \left(\frac{k_{trl}\phi_{TC}}{\phi_{TC} + K_{TC}} \right), \text{ with } k_{trl} := \frac{\langle \ell \rangle k_{el}^{max}}{\langle \ell \rangle + k_{el}^{max}(\tau_{ini} + \tau_{ter})} \text{ and } K_{TC} := \frac{k_{trl} n_{aa}}{k_{on}^{TC}} \tag{39}$$

which is the same relationship as the one derived in (***Klumpp et al., 2013***), with the addition of the terms corresponding to the rest translation cycle. Substituting the explicit relationship between growth and ternary complex concentration above (equation 39) in the aaRS/EF-Tu tRNA cycle relationship (equation 38) closes the system for ϕ_{TC}. Numerical solution for this equation is presented in Figure 3B (see section Estimation of optimal abundances for other parameters).

The main conclusion from numerically solving the reduced system (equations 38 and 39) is that the EF-Tu/aaRS space is partitioned in two regimes, resulting from the separation of scale of reactions in the coarse-grained model. Specifically, $k_{on}^{Tu} \gg \frac{k_{on}^{TC}}{n_{aa}}$, so that any imbalance between the constituents of the ternary complex (charged tRNAs, free EF-Tu), results in stoichiometric unproductive excess of the component in surplus.

We can derive a relation for the "transition line" in the aaRS/EF-Tu space where both free charged tRNAs and free EF-Tu are at low concentrations. This corresponds to setting the (formally impossible) requirement $\phi_{Tu^{\text{GTP}}} \approx 0 \Rightarrow \phi_{TC} \approx \phi_{Tu}$ and [aatRNA] $\propto \frac{1}{k_{on}^{Tu}\phi_{Tu^{\text{GTP}}}} \approx 0$, i.e.,

$$\frac{\text{tRNA}_{tot}}{P\lambda\left(\bar{\phi}_{Tu}\right)} - \frac{n_{aa}}{k_{on}^{TC}\bar{\phi}_{Tu}} - \frac{2}{k_{el}^{max}} - \frac{\bar{\phi}_{Tu}}{\ell_{Tu}\lambda\left(\bar{\phi}_{Tu}\right)} = \frac{n_{aa}}{k_{on}^{aaRS}\left(\bar{\phi}_{aaRS} - \frac{\ell_{aaRS}\lambda(\bar{\phi}_{Tu})}{k_{cat}^{aaRS}}\right)} + \frac{1}{k_{cat}^{aaRS}}. \tag{40}$$

The $\bar{\cdot}$ signifies the transition line relationship between $\bar{\phi}_{Tu}$ and $\bar{\phi}_{aaRS}$, which is displayed in Figure 3B.

The heuristic to estimate the optimal EF-Tu concentration described in the main text can be extended to include the EF-Ts cycle. In particular, in the EF-Tu limited regime, with $\phi_{Tu^{GTP}} \approx 0$, we have (from equation 29):

$$\phi_{TC} \approx \phi_{Tu} - \frac{\ell_{Tu}\lambda}{k_{on}^{Ts}\left(\phi_{Ts} - \frac{\ell_{Ts}\lambda}{k_{cat}^{Ts}}\right)} - \frac{\ell_{Tu}\lambda}{k_{cat}^{Ts}} - \frac{\ell_{Tu}\lambda}{k_{cat}^{TC}}.$$

Substituting the above expression for ϕ_{TC} in the optimality condition (equation 32) for ϕ_{Tu}, we arrive at (using the optimal solution for EF-Ts, equation 34):

$$\phi_{Tu}^{*} \approx \sqrt{\frac{\ell_{ribo} n_{aa} \lambda^{*}}{k_{on}^{TC}}} + \sqrt{\frac{\ell_{Tu}\lambda^{*}}{k_{on}^{Ts}}} + \frac{\ell_{Tu}\lambda^{*}}{k_{cat}^{Ts}} + \frac{\ell_{Tu}\lambda^{*}}{k_{cat}^{TC}}.$$

Above, the last three terms (not appearing in equation 12) correspond to the additional diffusion of the EF-Ts cycle, and catalytic contributions.
Following the argument (see main text) that the optimal aaRS abundance should lie on the transition line (equation 40), we obtain:

$$\phi_{aaRS}^{*} \approx \frac{n_{aa}\lambda^{*}}{k_{on}^{aaRS}\Delta_{tRNA}^{*}} + \frac{\ell_{aaRS}\lambda^{*}}{k_{cat}^{aaRS}},$$

with Δ_t related to the excess tRNA (tRNAs remaining after subtracting tRNAs sequestered on the ribosome and TC from the total tRNA budget):

$$\Delta_{tRNA}^{*} := \frac{\text{tRNA}_{tot}}{P} - \frac{n_{aa}\lambda^{*}}{k_{on}^{TC}\phi_{TC}^{*}} - \frac{2\lambda^{*}}{k_{el}^{max}} - \frac{\phi_{TC}^{*}}{\ell_{Tu}} - \frac{\lambda^{*}}{k_{cat}^{aaRS}}, \text{ where } \phi_{TC}^{*} = \sqrt{\frac{n_{aa}\ell_{ribo}\lambda^{*}}{k_{on}^{TC}}}.$$

Interpretation of the sharp separation between aaRS and EF-Tu limited regimes

The sharp separation of the solution for ϕ_{TC} in two distinct regimes (EF-Tu limited, and aaRS limited, illustrated in Figure 3B), can be intuitively understood from a geometrical viewpoint. For the simplicity of the argument (not strictly necessary), neglecting the short initiation and termination times in equation 39, and using $\text{tRNA}_{tot} = \frac{t\phi_{ribo}P}{\ell_{ribo}}$ (with t the tRNA to ribosome molar ratio). The tRNA conservation condition, equation 38, can then be rewritten as (binding-limited regime):

$$\underbrace{(t-1)\frac{\phi_{ribo}}{\ell_{ribo}}}_{\text{tRNA budget}} - \underbrace{\frac{\phi_{TC}}{\ell_{Tu}}}_{\text{ternary complex}} - \underbrace{\frac{\lambda(\phi_{TC})}{k_{el}^{max}}}_{\text{A-site tRNA}} = \lambda(\phi_{TC})\left[\underbrace{\frac{n_{aa}}{k_{on}^{aaRS}\phi_{aaRS}}}_{\text{uncharged tRNA}} + \underbrace{\frac{1}{k_{on}^{Tu}\left(\phi_{Tu} - \phi_{TC}\right)}}_{\text{free charged tRNA}}\right]$$

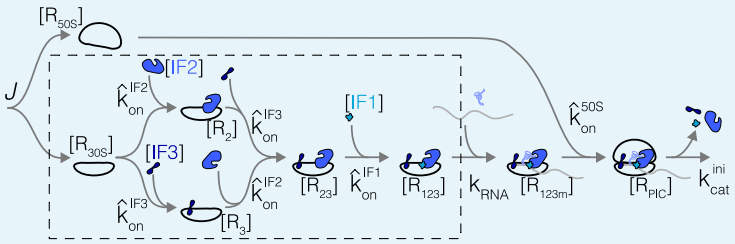
At given abundance of EF-Tu (ϕ_{Tu}) and aaRS (ϕ_{aaRS}), the solution for ϕ_{TC} is obtained when equality in the above equation is reached. The behavior of the various terms with ϕ_{TC} is illustrated for different values of ϕ_{aaRS} and ϕ_{Tu} in Figure 3-Figure supplement 1: the number of uncharged tRNAs (pink line in Figure 3-Figure supplement 1) is a decreasing function of aaRS, and free charged tRNA (red line in Figure 3-Figure supplement 1) are dependent on ϕ_{Tu}. Specifically, the free charged tRNA contribution, due to the rapid association rate k_{on}^{Tu} (codon agnostic) between charged tRNAs and EF-Tu (red line), is negligible except for a very narrow range where $\phi_{TC} \approx \phi_{Tu}$, at which point a sharp divergence occurs. This rapid divergence bounds the solution for ϕ_{TC} at the total EF-Tu concentration.
The aaRS limited regime corresponds to conditions in which the uncharged tRNA contribution (pink line) intersects the available tRNA budget (full black line), lower left in Figure 3-Figure supplement 1. In contrast, the EF-Tu limited regime corresponds to conditions in which

the free charged tRNA (red line) intersects the tRNA budget, upper right in Figure 3-Figure supplement 1. The sharpness of the transition between the two regime arises from the near vertical divergence of the free charged tRNA contribution.

Appendix 4

Translation initiation

Translation initiation is also relatively complex compared to translation termination. In contrast with other steps of the translation cycle, binding of factors necessary for the process (IF1, IF2, IF3, initiator tRNA) do not occur in a strict sequential order, leading to a "heterogeneous assembly landscape" (***Gualerzi and Pon, 2015***; ***Chen et al., 2016***) more complex to model. However, one assembly pathway is kinetically favored (***Milón et al., 2012***). We take this favored assembly pathway as our kinetic scheme (Appendix 4 Figure 1, note that binding of tRNA/mRNA are coarse-grained to a single even without loss of generality). We provide some evidence below that taking a more complex assembly pathway would minimally affect the predicted optimal initiation factor abundances.



Appendix 4 Figure 1. Simplified kinetic scheme for translation initiation. Reactions in dashed box correspond to sub-system solved in detail first (section Sub-pathway without subunits joining). Variables are labeled on the scheme.

The reactions in our simplified schemes are:

$$\xrightarrow{J} R_{30S} + R_{50S},$$

$$R_{30S} + IF3 \xrightarrow{\hat{k}_{on}^{IF3}} R_3,$$

$$R_{30S} + IF2 \xrightarrow{\hat{k}_{on}^{IF2}} R_2,$$

$$R_3 + IF2 \xrightarrow{k_{on}^{IF2}} R_{23},$$

$$R_2 + IF3 \xrightarrow{\hat{k}_{on}^{IF3}} R_{23},$$

$$R_{23} + IF1 \xrightarrow{\hat{k}_{on}^{IF1}} R_{123},$$

$$R_{123} \xrightarrow{k_{RNA}} R_{123m},$$

$$R_{123m} + R_{50S} \xrightarrow{k_{on}^{50S}} R_{PIC},$$

$$R_{PIC} \xrightarrow{k_{cat}^{ini}} IF1 + IF2 + IF3,$$

with corresponding mass action equations:

$$\begin{aligned}
\frac{d[R_{30S}]}{dt} &= J - \hat{k}_{on}^{IF2}[R_{30S}][IF2] - \hat{k}_{on}^{IF3}[R_{30S}][IF3],\\
\frac{d[R_2]}{dt} &= \hat{k}_{on}^{IF2}[R_{30S}][IF2] - \hat{k}_{on}^{IF3}[R_2][IF3],\\
\frac{d[R_3]}{dt} &= \hat{k}_{on}^{IF3}[R_{30S}][IF3] - \hat{k}_{on}^{IF2}[R_3][IF2],\\
\frac{d[R_{23}]}{dt} &= \hat{k}_{on}^{IF2}[R_3][IF2] + \hat{k}_{on}^{IF3}[R_2][IF3] - \hat{k}_{on}^{IF1}[R_{23}][IF1],\\
\frac{d[R_{123}]}{dt} &= \hat{k}_{on}^{IF1}[R_{23}][IF1] - k_{RNA}[R_{123}],\\
\frac{d[R_{123m}]}{dt} &= k_{RNA}[R_{123}] - \hat{k}_{on}^{50S}[R_{123m}][R_{50S}],\\
\frac{d[R_{PIC}]}{dt} &= \hat{k}_{on}^{50S}[R_{123m}][R_{50S}] - k_{cat}^{ini}[R_{PIC}],\\
\frac{d[R_{50S}]}{dt} &= J - \hat{k}_{on}^{50S}[R_{123m}][R_{50S}],\\
\frac{d[IF1]}{dt} &= -\hat{k}_{on}^{IF1}[R_{23}][IF1] + k_{cat}^{ini}[PIC],\\
\frac{d[IF2]}{dt} &= -\hat{k}_{on}^{IF2}\left([R_{30S}] + [R_3]\right)[IF2] + k_{cat}^{ini}[PIC],\\
\frac{d[IF3]}{dt} &= -\hat{k}_{on}^{IF3}\left([R_{30S}] + [R_2]\right)[IF3] + k_{cat}^{ini}[PIC],
\end{aligned}$$

and conservation equations:

$$\begin{aligned}
IF1_{tot} &= [IF1] + [R_{123}] + [R_{123m}] + [R_{PIC}],\\
IF2_{tot} &= [IF2] + [R_2] + [R_{23}] + [R_{123}] + [R_{123m}] + [R_{PIC}],\\
IF3_{tot} &= [IF3] + [R_3] + [R_{23}] + [R_{123}] + [R_{123m}] + [R_{PIC}],\\
[R_{50S}] &= [R_{30S}] + [R_2] + [R_3] + [R_{23}] + [R_{123}] + [R_{123m}].
\end{aligned}$$

We assume the steady-state concentrations of small and large ribosomal subunits to be equal.

Sub-pathway without subunits joining

The system of equation is complicated by the second branch of the pathway corresponding to 50S subunit binding. However, in the regime $\sqrt{\frac{\ell_{IF}}{\ell_{ribo}}\frac{\hat{k}_{on}^{50S}}{\hat{k}_{on}^{IF}}} \ll 1$ (which is realized because of the large size of the ribosome and slower association rate constant for the large subunit compared to the initiation factors again due to size), the effect of this branch is to add a term to the optimal abundance equal to the concentration of species R_{123m} (see derivation in section Pathway including subunits joining). We focus here on the solution of the part of the reaction scheme boxed in Appendix 4 Figure 1. This sub-scheme corresponds to:

$$\begin{aligned}
&\xrightarrow{J} R_{30S},\\
R_{30S} + IF3 &\xrightarrow{\hat{k}_{on}^{IF3}} R_3,\\
R_{30S} + IF2 &\xrightarrow{\hat{k}_{on}^{IF2}} R_2,\\
R_3 + IF2 &\xrightarrow{\hat{k}_{on}^{IF2}} R_{23},\\
R_2 + IF3 &\xrightarrow{\hat{k}_{on}^{IF3}} R_{23},\\
R_{23} + IF1 &\xrightarrow{\hat{k}_{on}^{IF1}} R_{123},\\
R_{123} &\xrightarrow{k_{RNA}} R_{123m}.
\end{aligned}$$

$$
\begin{aligned}
\frac{\mathrm{d}[R_{30S}]}{\mathrm{dt}} &= J - \hat{k}_{on}^{IF2}[R_{30S}][IF2] - \hat{k}_{on}^{IF3}[R_{30S}][IF3],\\
\frac{\mathrm{d}[R_2]}{\mathrm{dt}} &= \hat{k}_{on}^{IF2}[R_{30S}][IF2] - \hat{k}_{on}^{IF3}[R_2][IF3],\\
\frac{\mathrm{d}[R_3]}{\mathrm{dt}} &= \hat{k}_{on}^{IF3}[R_{30S}][IF3] - \hat{k}_{on}^{IF2}[R_3][IF2],\\
\frac{\mathrm{d}[R_{23}]}{\mathrm{dt}} &= \hat{k}_{on}^{IF2}[R_3][IF2] + \hat{k}_{on}^{IF3}[R_2][IF3] - \hat{k}_{on}^{IF1}[R_{23}][IF1],\\
\frac{\mathrm{d}[R_{123}]}{\mathrm{dt}} &= \hat{k}_{on}^{IF1}[R_{23}][IF1] - k_{RNA}[R_{123}],\\
\frac{\mathrm{d}[IF1]}{\mathrm{dt}} &= -\hat{k}_{on}^{1}[R_{23}][IF1] + k_{RNA}[R_{123}],\\
\frac{\mathrm{d}[IF2]}{\mathrm{dt}} &= -\hat{k}_{on}^{IF2}([R_{30S}] + [R_3])[IF2] + k_{RNA}[R_{123}],\\
\frac{\mathrm{d}[IF3]}{\mathrm{dt}} &= -\hat{k}_{on}^{IF3}([R_{30S}] + [R_2])[IF3] + k_{RNA}[R_{123}],
\end{aligned}
$$

with conservation equations:

$$
\begin{aligned}
IF1_{tot} &= [IF1] + [R_{123}],\\
IF2_{tot} &= [IF2] + [R_2] + [R_{23}] + [R_{123}],\\
IF3_{tot} &= [IF3] + [R_3] + [R_{23}] + [R_{123}],
\end{aligned}
$$

This system can be solved as with the previous schemes. In steady-state, we find for concentrations in terms of the free concentrations $[IF2]$ and $[IF3]$:

$$
[R_{123}] = \frac{J}{k_{RNA}},\ [IF1] = IF1_{tot} - \frac{J}{k_{RNA}},\ [R_{23}] = \frac{J}{\hat{k}_{on}^{IF1}[IF1]},\ [R_{30S}] = \frac{J}{\hat{k}_{on}^{IF2}[IF2] + \hat{k}_{on}^{IF3}[IF3]},
$$

$$
[R_2] = \frac{\hat{k}_{on}^{IF2}[IF2]}{\hat{k}_{on}^{IF3}[IF3]}\left(\frac{J}{\hat{k}_{on}^{IF2}[IF2] + \hat{k}_{on}^{IF3}[IF3]}\right),\ [R_3] = \frac{\hat{k}_{on}^{IF3}[IF3]}{\hat{k}_{on}^{IF2}[IF2]}\left(\frac{J}{\hat{k}_{on}^{IF2}[IF2] + \hat{k}_{on}^{IF3}[IF3]}\right),
$$

and the coupled equations for $[IF2]$ and $[IF3]$ that need to be solved:

$$
\begin{aligned}
IF2_{tot} &= [IF2] + \frac{\hat{k}_{on}^{IF2}[IF2]}{\hat{k}_{on}^{IF3}[IF3]}\left(\frac{J}{\hat{k}_{on}^{IF2}[IF2] + \hat{k}_{on}^{IF3}[IF3]}\right) + \frac{J}{\hat{k}_{on}^{IF1}[IF1]} + \frac{J}{k_{RNA}},\\
IF3_{tot} &= [IF3] + \frac{\hat{k}_{on}^{IF3}[IF3]}{\hat{k}_{on}^{IF2}[IF2]}\left(\frac{J}{\hat{k}_{on}^{IF2}[IF2] + \hat{k}_{on}^{IF3}[IF3]}\right) + \frac{J}{\hat{k}_{on}^{IF1}[IF1]} + \frac{J}{k_{RNA}}.
\end{aligned}
\tag{41}
$$

As for translation termination (section Coarse-grained translation termination time) and elongation (section Coarse-grained translation elongation time), summing the ribosome containing species:

$$
\begin{aligned}
[R_{ini}] &= [R_{30S}] + [R_2] + [R_3] + [R_{23}] + [R_{123}],\\
&= J\left(\frac{1}{\hat{k}_{on}^{IF2}[IF2]} + \frac{1}{\hat{k}_{on}^{IF3}[IF3]} - \frac{1}{\hat{k}_{on}^{IF2}[IF2] + \hat{k}_{on}^{IF3}[IF3]} + \frac{1}{\hat{k}_{on}^{IF1}[IF1]} + \frac{1}{k_{RNA}}\right),
\end{aligned}
$$

allows us to read the initiation time directly (recast in proteome fraction units):

$$
\tau_{ini} = \frac{1}{k_{on}^{IF2}\phi_{IF2}^{free}} + \frac{1}{k_{on}^{IF3}\phi_{IF3}^{free}} - \frac{1}{k_{on}^{IF2}\phi_{IF2}^{free} + k_{on}^{IF3}\phi_{IF3}^{free}} + \frac{1}{k_{on}^{IF1}\phi_{IF1}^{free}} + \frac{1}{k_{RNA}}.
\tag{42}
$$

The above is the time can be used in the optimality condition (equation 5). Note that the parallel nature of the reactions with IF2 and IF3 leads to a reduction compared to a purely sequential pathway (negative term above decreasing the total initiation time, as expected if multiple reactions can occur in parallel).

Given that binding of IF1 occurs last in this scheme, its free concentration takes a simple form ($\phi_{IF1}^{free} = \phi_{IF1} - \frac{\ell_{IF1}\lambda}{\langle\ell\rangle k_{RNA}}$). In contrast, computing the free IF2 and IF3 concentrations

requires solving the non-linear coupled system, equations 41. Recasting these in units of proteome fraction:

$$\tilde{\phi}_{IF2} = \phi_{IF2}^{free} + \frac{\lambda \ell_{IF2}}{\langle \ell \rangle k_{on}^{IF3} \phi_{IF3}^{free}} \left(\frac{k_{on}^{IF2} \phi_{IF2}^{free}}{k_{on}^{IF2} \phi_{IF2}^{free} + k_{on}^{IF3} \phi_{IF3}^{free}} \right),$$
$$\tilde{\phi}_{IF3} = \phi_{IF3}^{free} + \frac{\lambda \ell_{IF3}}{\langle \ell \rangle k_{on}^{IF2} \phi_{IF2}^{free}} \left(\frac{k_{on}^{IF3} \phi_{IF3}^{free}}{k_{on}^{IF2} \phi_{IF2}^{free} + k_{on}^{IF3} \phi_{IF3}^{free}} \right),$$

with $\tilde{\phi}_{IF2} := \phi_{IF2} - \frac{\ell_{IF2}\lambda}{\langle \ell \rangle k_{RNA}} - \frac{\ell_{IF2}\lambda}{\langle \ell \rangle k_{on}^{IF1} \phi_{IF1}^{free}}$, and similarly for $\tilde{\phi}_{IF3}$. We show now that the terms coupling the two equations for ϕ_{IF2}^{free} and ϕ_{IF2}^{free} (bracketed above) are small at the optimum. Indeed, based on results in simpler schemes (self-consistency confirmed below), we expect at the optimum:

$$\phi_{IF2}^{free,*} \sim \sqrt{\frac{\ell_{ribo}\lambda^*}{\langle \ell \rangle k_{on}^{IF2}}} \text{ and } \phi_{IF3}^{free,*} \sim \sqrt{\frac{\ell_{ribo}\lambda^*}{\langle \ell \rangle k_{on}^{IF3}}}.$$

Hence, we expect the two terms at the optimum in the coupled equations above to compare as (e.g., in the free IF2 equation):

$$\frac{\phi_{IF2}^{free,*}}{\left(\frac{\lambda^* \ell_{IF2}}{\langle \ell \rangle k_{on}^{IF3} \phi_{IF3}^{free,*}} \right)} \sim \frac{\ell_{ribo}}{\ell_{IF2}} \sqrt{\frac{k_{on}^{IF3}}{k_{on}^{IF2}}} \gg 1,$$

coming from the large size of the ribosome compared to the initiation factors. In addition, the derivative of the coupling terms, which appear in the optimality condition and therefore in identifying the optimal abundances, are all of the form $\frac{\lambda^* \ell_{IF}}{\langle \ell \rangle k_{on}^{IF} (\phi_{IF}^{free})^2}$ compared to the main term. This scales scales as $\ell_{IF} \ell_{ribo}^{-1} \ll 1$ at the self-consistent solution. Hence, neglecting the coupling is justified as an approximate solutions near the optimum, and we obtain for the free concentrations of IFs:

$$\phi_{IF1}^{free} = \phi_{IF1} - \frac{\ell_{IF1}\lambda}{\langle \ell \rangle k_{RNA}},$$
$$\phi_{IF2}^{free} \approx \phi_{IF2} - \frac{\ell_{IF2}\lambda}{\langle \ell \rangle k_{RNA}} - \frac{\ell_{IF2}\lambda}{\langle \ell \rangle k_{on}^{IF1} \phi_{IF1}^{free}},$$
$$\phi_{IF3}^{free} \approx \phi_{IF3} - \frac{\ell_{IF3}\lambda}{\langle \ell \rangle k_{RNA}} - \frac{\ell_{IF3}\lambda}{\langle \ell \rangle k_{on}^{IF1} \phi_{IF1}^{free}}.$$

Substituting these in the expression for the initiation time, equation 42, and using the optimality condition (equation 5, we find that no simple solution exist for the non symmetric case of $k_{on}^{IF2} \neq k_{on}^{IF3}$. Since the on-rates should be similar for IF2 and IF3 (difference in size should only lead to modest difference in on-rates coefficient, by roughly $(\ell_{IF2}/\ell_{IF3})^{1/3} \approx 1.7$ assuming Stokes scaling), the symmetric case is approximately correct. We report the symmetric solution for simplicity. The final optimal solutions for the three factors for the sub-scheme solved here is:

$$\phi_{IF1}^* \approx \sqrt{\frac{\ell_{ribo}\lambda^*}{\langle \ell \rangle k_{on}^{IF1}} \left[1 + \frac{\ell_{IF2} + \ell_{IF3}}{\ell_{ribo}} \right]} + \frac{\ell_{IF1}\lambda^*}{\langle \ell \rangle k^{ini}}, \tag{43}$$
$$\phi_{IF2}^* \approx \sqrt{\frac{3}{4}} \sqrt{\frac{\ell_{ribo}\lambda^*}{\langle \ell \rangle k_{on}^{IF2}}} + \frac{\ell_{IF2}}{\langle \ell \rangle} \sqrt{\frac{\ell_{ribo}\lambda^*}{\langle \ell \rangle k_{on}^{IF1}}} + \frac{\ell_{IF2}\lambda^*}{\langle \ell \rangle k^{ini}},$$
$$\phi_{IF3}^* \approx \sqrt{\frac{3}{4}} \sqrt{\frac{\ell_{ribo}\lambda^*}{\langle \ell \rangle k_{on}^{IF3}}} + \frac{\ell_{IF3}}{\langle \ell \rangle} \sqrt{\frac{\ell_{ribo}\lambda^*}{\langle \ell \rangle k_{on}^{IF1}}} + \frac{\ell_{IF3}\lambda^*}{\langle \ell \rangle k^{ini}}.$$

The form of the solution is again similar to that derived for the simpler translation termination scheme (c.f., equation 20), with three differences, each of which has an intuitive interpretation. First, the factor $\left[1+\frac{\ell_{IF2}+\ell_{IF3}}{\ell_{ribo}}\right]$ in the IF1 solution arises as a result of IF1 binding being last in our initiation pathway. Indeed, IF1 concentration also influences free IF2 and IF3 concentration, leading to additional selective pressure to increase its abundance. In effect, the molecular species waiting for IF1 to diffuse to its target is not only the ribosome, but the ribosome with IF2 and IF3 bound, and a total amino acid weight $\ell_{ribo} \rightarrow \ell_{ribo}+\ell_{IF2}+\ell_{IF3}$. Second, the factor of $\sqrt{3/4} \approx 0.87 < 1$ for IF2 and IF3 (corresponding to the symmetric case), arising from the parallel pathway for IF2 and IF3 rendering the process more efficient. We therefore see that the correction from having multiple reactions in parallel is modest (0.87 vs. 1). The third difference to the simpler case of translation termination are the second terms for IF2 and IF3, corresponding to the additional delay incurred by binding of IF1. These come from the assumed sequential nature of our initiation scheme (Appendix 4 Figure 1). In such cases, factors binding earlier have to be present at higher abundances to account for their wait times for later binding events. The exact form of this correction term would be different for more complex assembly pathways (but would be captured by average delays from other factor binding).

Pathway including subunits joining

The solutions above (equations 43) are for the reduced scheme (boxed in Appendix 4 Figure 1). The full solutions includes the delay arising from 50S subunit binding. Including subunit joining requires the solution of an additional equation for the steady-state concentration of species with all three initiation factors, mRNA and initiator tRNA waiting for subunit joining (species R_{123m} in Appendix 4 Figure 1, denoted ϕ_{123m} in units of proteome fraction). The equation to solve for ϕ_{123m} can be obtained from the 50S ribosome subunit conservation equation:

$$\frac{\lambda}{k_{on}^{50S}\phi_{123m}} = \frac{\lambda}{k_{on}^{IF2}\phi_{IF2}^{free}} + \frac{\lambda}{k_{on}^{IF3}\phi_{IF3}^{free}} - \frac{\lambda}{k_{on}^{IF2}\phi_{IF2}^{free}+k_{on}^{IF3}\phi_{IF3}^{free}} + \frac{\lambda}{k_{on}^{IF1}\phi_{IF1}^{free}} + \frac{\lambda}{k_{RNA}} + \frac{\langle\ell\rangle\phi_{123m}}{\ell_{30S}}.$$

ϕ_{123m} appears in the equations for the free concentration of the initiation factors (from the conservation equations), and also leads to the appearance of a new term in the expression for the initiation time τ_{ini} (equation 42) corresponding to this step: $\frac{\langle\ell\rangle\phi_{123m}}{\ell_{30S}\lambda}$.
These two additions, resulting from the parallel branch of 50S joining, can be simplified due to a separation of scales between the various terms. For large initiation factor concentrations, the corresponding mass action terms in the equation for ϕ_{123m} negligibly contribute to the solution. In this regime, the new term involving ϕ_{123m} in the initiation time τ_{ini} does not alter the form the optimal abundances of IF1, IF2, and IF3 beyond adding a constant term. Hence, in the regime of high free IF concentration, the optimality condition has the same form as derived in the previous section.We can therefore obtain ϕ_{123m} assuming large IF concentration, denoted ϕ_{123m}^{∞}:

$$\phi_{123m}^{\infty} = \frac{\ell_{30S}}{\langle\ell\rangle}\left(-\frac{\lambda}{2k_{RNA}} + \sqrt{\frac{1}{4}\left(\frac{\lambda}{k_{RNA}}\right)^2 + \frac{\langle\ell\rangle\lambda}{\ell_{30S}k_{on}^{50S}}}\right)$$

This solution will be self-consistent provided (for all initiation factors):

$$\frac{\lambda^*}{k_{on}^{IF}\phi_{IF}^{free,*}} \ll \frac{\lambda^*}{k_{RNA}} + \frac{\langle\ell\rangle\phi_{123m}^{\infty}}{\ell_{30S}} = \frac{\lambda^*}{2k_{RNA}} + \sqrt{\frac{1}{4}\left(\frac{\lambda^*}{k_{RNA}}\right)^2 + \frac{\langle\ell\rangle\lambda^*}{\ell_{30S}k_{on}^{50S}}},$$

It therefore suffices to show:

$$\frac{\lambda^*}{k_{on}^{IF}\phi_{IF}^{free,*}} \ll \sqrt{\frac{\langle \ell \rangle \lambda^*}{\ell_{30S} k_{on}^{50S}}}.$$

Using our optimality condition on $\phi_{IF}^{free,*}$ (equation 43) assuming no contribution from ϕ_{123m} (self-consistency), and converting association rates in units $\mu M^{-1}s^{-1}$, the above condition reduces to:

$$\sqrt{\frac{\ell_{IF}}{\ell_{ribo}} \frac{\hat{k}_{on}^{50S}}{\hat{k}_{on}^{IF}}} \ll 1.$$

The self-consistency condition is met both because initiation factors are smaller than ribosomes $\ell_{IF} \ll \ell_{ribo}$, and because the on-rate for subunit joining is lower than initiation factor binding ($\hat{k}_{on}^{50S} \ll \hat{k}_{on}^{IF}$), given again the size differences. The solution, including the contribution from ribosome subunits joining is then:

$$\phi_{IF1}^* \approx \sqrt{\frac{\ell_{ribo}\lambda^*}{\langle \ell \rangle k_{on}^{IF1}} \left[1 + \frac{\ell_{IF2} + \ell_{IF3}}{\ell_{ribo}}\right]} + \frac{\ell_{IF1}}{\ell_{30S}}\phi_{123m}^{\infty} + \frac{\ell_{IF1}\lambda^*}{\langle \ell \rangle}\left(\frac{1}{k_{RNA}} + \frac{1}{k_{cat}^{ini}}\right),$$

$$\phi_{IF2}^* \approx \sqrt{\frac{3}{4}}\sqrt{\frac{\ell_{ribo}\lambda^*}{\langle \ell \rangle k_{on}^{IF2}}} + \frac{\ell_{IF2}}{\langle \ell \rangle}\sqrt{\frac{\ell_{ribo}\lambda^*}{\langle \ell \rangle k_{on}^{IF1}}} + \frac{\ell_{IF2}}{\ell_{30S}}\phi_{123m}^{\infty} + \frac{\ell_{IF2}\lambda^*}{\langle \ell \rangle}\left(\frac{1}{k_{RNA}} + \frac{1}{k_{cat}^{ini}}\right),$$

$$\phi_{IF3}^* \approx \sqrt{\frac{3}{4}}\sqrt{\frac{\ell_{ribo}\lambda^*}{\langle \ell \rangle k_{on}^{IF3}}} + \frac{\ell_{IF3}}{\langle \ell \rangle}\sqrt{\frac{\ell_{ribo}\lambda^*}{\langle \ell \rangle k_{on}^{IF1}}} + \frac{\ell_{IF3}}{\ell_{30S}}\phi_{123m}^{\infty} + \frac{\ell_{IF3}\lambda^*}{\langle \ell \rangle}\left(\frac{1}{k_{RNA}} + \frac{1}{k_{cat}^{ini}}\right),$$

where for k_{RNA} much faster than the association between the subunits, $\phi_{123m}^{\infty} \approx \sqrt{\frac{\ell_{30S}\lambda^*}{\langle \ell \rangle k_{on}^{50S}}}$.

Appendix 5

Estimation of optimal abundances

To compare prediction from our parsimonious framework (Table 2) requires specific values of kinetic parameters. We use empirical measurements together with scaling relations to estimate these kinetic parameters.

Catalytic rates for many enzymes have been measured *in vitro*, but the obtained values can be sharply incompatible with kinetic parameters that have been measured in the cell. An example is the class tRNA synthetases. Tallying the measured k_{cat} for all wild-type *E. coli* aaRSs (***Jeske et al., 2019***), we find a median value of $k_{cat}^{aaRS} \approx 3\ s^{-1}$, and 80% of reported value below 6 s^{-1}. The total molar concentration of aaRSs in the cell is comparable to the total number of ribosomes, and the per-step elongation speed of ribosome is above 15 s^{-1} (***Dai et al., 2016***; ***Johnson et al., 2020***). Hence, the absolute minimum catalytic rate to sustain the translation elongation flux needs to obey $k_{cat}^{aaRS} > 15\ s^{-1}$, which is much higher than most *in vitro* measured values. To avoid the difficulties in estimating catalytic parameters, and to derive a lower bound on factor abundance from our model, we focus on the binding component (related to the associate rate) of our predictions, assuming large catalytic rates ($k_{cat} \to \infty$).

To estimate association rates $\hat{k}_{on}$, we scaled the measured *in vivo* association rate for the ternary complex, $\hat{k}_{on}^{TC} = 6.4\ \mu M^{-1}s^{-1}$ (***Dai et al., 2016***) by diffusion of the respective components, i.e., $\hat{k}_{on}^{AB}/\hat{k}_{on}^{TC} = (D_A + D_B)/(D_{TC} + D_{ribo})$, where D_i is the diffusion coefficients for the molecular species i. While the *in vivo* diffusion coefficient for a number of component of the translation apparatus exist (***Bakshi et al., 2012***; ***Sanamrad et al., 2014***; ***Volkov et al., 2018***; ***Plochowietz et al., 2017***), several factors do not have measured diffusion coefficients. For these, we used the cubic root scaling from the Stokes-Einstein relation (***Nenninger et al., 2010***), see Appendix 5 Table 1.

To motivate this approach, we can compare for the ternary complex the measured $\hat{k}_{on}^{TC}$ to the simple Smoluchowski relation corresponding to the diffusion-limited association rate constant for perfectly absorbing spheres: $\hat{k}_{on}^{diff} = 4\pi DR$, where D is the relative diffusion coefficients of the two reactants and R the capture radius. Diffusion coefficients of various proteins of different sizes have been measured in the cell (***Elowitz et al., 1999***; ***Kumar et al., 2010***; ***Nenninger et al., 2010***), including for components of the translation machinery (***Bakshi et al., 2012***; ***Sanamrad et al., 2014***; ***Plochowietz et al., 2017***; ***Volkov et al., 2018***). In particular, Volkov et al report an EF-Tu diffusion coefficient (and a similar measurement for a major diffusive state of tRNAs) of $\approx 3\ \mu m^2 s^{-1}$. Since ribosomes are nearly immobile, this can be used in an estimate of k_{on}^{diff}. Taking $R \approx 2$ nm (***Klumpp et al., 2013***) as the rough size of the capture radius for the ternary complex, we get $\hat{k}_{on}^{TC,diff} \approx 45\ \mu M^{-1}s^{-1}$. *In vivo* estimates based on kinetic measurements of elongation (***Dai et al., 2016***) find $\hat{k}_{on}^{TC} = 6.4\ \mu M^{-1}s^{-1}$. Multiple features likely explain this difference, notably as orientational constraints (***Schlosshauer and Baker, 2004***), finite off-rate possibly requiring multiple binding events before productive encounter, or possibly because the ternary complex needs to sample multiple non-cognate sites to find a cognate target thereby slowing its diffusive search. This comparison emphasizes that the idealized diffusion-limited Smoluchowski regime is not physiologically applicable, and motivates our scaling approach. The later comes at the price of assuming similar molecular properties leading to decrease of the association rates for the other tIFs. These could be further refined via e.g., structural modeling (***Schlosshauer and Baker, 2004***), or upon new in vivo rate constant measurements.

Additional measured quantities required to compute our estimates are: the measured growth rate $\lambda^* = 5.5 \times 10^{-4}\ s^{-1}$ (21 min doubling time, average of fast growth conditions), the

tRNA concentration (estimated from the tRNA to ribosome ratio of 6.5 (***Dong et al., 1996***) using: $\text{tRNA}_{tot} = (\text{tRNA/ribo})\phi_{ribo}P/\ell_{ribo}$), the maximum per-codon elongation rate, excluding ternary complex diffusion, $k_{el}^{max} = 22\ \text{s}^{-1}$ (***Dai et al., 2016***) (used to estimate the number of tRNAs sequestered on ribosomes and therefore the excess tRNA number in the optimum for aaRS, see equations 18 and 38), the in-protein amino acid concentration $P = 2.6$ M (***Klumpp et al., 2013***; ***Bremer and Dennis, 2008***).

For the fast growth average, results displayed in Figure 4 listed in Supplementary File 2. Additional predictions in individual conditions are shown in Figure 4-Figure supplement 1, with numerical values for measured and predicted values listed in Supplementary Files 1 to 4. For predictions in different growth conditions/species, we used used the measured growth rates in the corresponding conditions (values listed in Supplementary Files 1 and 3), and association rate constants estimated based on *E. coli* data (Appendix 5 Tables 1 to 3), and the tRNA abundance (only needed for the prediction of aaRS) at the corresponding growth rate in *E. coli* from (***Dong et al., 1996***). As a result of the lack of quantitation of tRNA abundance in other species, these values were used for *B. subtilis, V. natriegens* and *C. crescentus*, and should be interpreted with caution given possible difference in cellular physiology for these species.

Factor	Number of codon per protein	Diffusion coefficient ($\mu m^2\ s^{-1}$)
Ribosome	$\ell_{ribo} = 7336$	$D_{ribo} = 0.05 \pm 0.01$
30S subunit	$\ell_{30S} = 3108$	$D_{subunits} = 0.2 \pm 0.1$
TC	$\ell_{TC} = 630^{\#}$	$D_{TC} = 3 \pm 0.5$
tRNA	N/A	$D_{tRNA} = 8 \pm 1$
IF1	$\ell_{IF1} = 72$	$D_{IF1} = D_{TC}\sqrt[3]{\frac{\ell_{TC}}{\ell_{IF1}}}$
IF2	$\ell_{IF2} = 890$	$D_{IF2} = D_{TC}\sqrt[3]{\frac{\ell_{TC}}{\ell_{IF2}}}$
IF3	$\ell_{IF3} = 180$	$D_{IF3} = D_{TC}\sqrt[3]{\frac{\ell_{TC}}{\ell_{IF3}}}$
EF-G	$\ell_{G} = 704$	$D_{G} = D_{TC}\sqrt[3]{\frac{\ell_{TC}}{\ell_{G}}}$
EF-Ts	$\ell_{Ts} = 283$	$D_{Ts} = D_{TC}\sqrt[3]{\frac{\ell_{TC}}{\ell_{Ts}}}$
EF-Tu	$\ell_{Tu} = 394$	$D_{Tu} = D_{TC}\sqrt[3]{\frac{\ell_{TC}}{\ell_{Tu}}}$
aaRS	$\ell_{aaRS} = 987^{\dagger}$	$D_{aaRS} = D_{TC}\sqrt[3]{\frac{\ell_{TC}}{\ell_{aaRS}}}$
RF1/RF2	$\ell_{RFI} = 362$	$D_{RFI} = D_{TC}\sqrt[3]{\frac{\ell_{TC}}{\ell_{RFI}}}$
RF4	$\ell_{RF4} = 185$	$D_{RF4} = D_{TC}\sqrt[3]{\frac{\ell_{TC}}{\ell_{RF4}}}$

Appendix 5 Table 1. Protein sizes (number of codons) and diffusion coefficients. Unless otherwise noted, number of codons per protein are taken for *E. coli* (***Keseler et al., 2016***) (ribosome size taken from (***Wittmann, 1982***)). $^{\#}$For the ternary complex, the total mass of tRNA+EF-Tu was converted to an equivalent amino acid length for the diffusion constant scaling estimate. †For aaRS, the size for the summed aaRSs is, from the coarse graining, $\ell_{aaRS} = \sum_i \phi_{aaRS,i} / \sum_i (\phi_{aaRS,i}/\ell_{aaRS,i})$, here with proteome fractions estimated from ribosome profiling (***Li et al., 2014***) in *E. coli* and sizes accounting for varying complex stoichiometries. Measured diffusion coefficients are taken from: (***Bakshi et al., 2012***; ***Sanamrad et al., 2014***) for the ribosome, from (***Plochowietz et al., 2017***; ***Volkov et al., 2018***) for tRNAs, and from (***Volkov et al., 2018***) for the TC.

Factors involved in reaction	Variable	Used expression for association rate constant
Ternary complex and ribosome	$\hat{k}_{on}^{TC}$	6.4 ± 0.6 μM^{-1}s^{-1} (***Dai et al., 2016***)
EF-G and ribosome	$\hat{k}_{on}^{G}$	$\hat{k}_{on}^{TC}(D_G + D_{ribo})/(D_{TC} + D_{ribo})$
aaRS And tRNAs	$\hat{k}_{on}^{aaRS}$	$\hat{k}_{on}^{TC}(D_{tRNA} + D_{aaRS})/(D_{TC} + D_{ribo})$
EF-Ts and ribosome	$\hat{k}_{on}^{Ts}$	$\hat{k}_{on}^{TC}(D_{Ts} + D_{ribo})/(D_{TC} + D_{ribo})$
EF-Tu and tRNAs	$\hat{k}_{on}^{Tu}$	$\hat{k}_{on}^{TC}(D_{tRNA} + D_{Tu})/(D_{TC} + D_{ribo})$
IF1 and 30S subunit	$\hat{k}_{on}^{IF1}$	$\hat{k}_{on}^{TC}(D_{IF1} + D_{subunit})/(D_{TC} + D_{ribo})$
IF2 and 30S subunit	$\hat{k}_{on}^{IF2}$	$\hat{k}_{on}^{TC}(D_{IF2} + D_{subunit})/(D_{TC} + D_{ribo})$
IF3 and 30S subunit	$\hat{k}_{on}^{IF3}$	$\hat{k}_{on}^{TC}(D_{IF3} + D_{subunit})/(D_{TC} + D_{ribo})$
50S and 30S subunits	$\hat{k}_{on}^{50S}$	$\hat{k}_{on}^{TC}(D_{subunit} + D_{subunit})/(D_{TC} + D_{ribo})$
RF1/RF2 and ribosome	$\hat{k}_{on}^{RFI}$	$\hat{k}_{on}^{TC}(D_{RFI} + D_{ribo})/(D_{TC} + D_{ribo})$
RF4 and ribosome	$\hat{k}_{on}^{RF4}$	$\hat{k}_{on}^{TC}(D_{RF4} + D_{ribo})/(D_{TC} + D_{ribo})$

Appendix 5 Table 2. Expression used to estimate the association rate constants for our predictions (Table 2). Diffusion coefficients are listed in Appendix 5 Table 1.

Parameter	Value	Description
P	2.6 ± 0.5 M	In-protein amino acid concentration in the cell.
λ	$(5.5 \pm 0.6) \times 10^{-4}$ s^{-1}	Average fast growth, see Supplementary File 1.
$\langle \ell \rangle$	200 ± 10	Average number of codons per protein (equation 16).
n_{aa}	20 ± 2	Rescaling factor in elongation model (see equation 26).
k_{el}^{max}	22 ± 2 s^{-1}	Maximal translation elongation rate.
$\sqrt{1+\delta}$	1.05 ± 0.01	Factor in three stop codon model (see equation 23)
t :=tRNA/ribosome	6.5 to 11	Values taken listed in Supplementary Files 2 and 4.
tRNA$_{tot}$	$t\phi_{ribo}P/\ell_{ribo}$	Total tRNA abundance, estimated from tRNA/ribosome.

Appendix 5 Table 3. Additional parameters used to obtain numerical values for predictions. For the doubling times (growth rates) and tRNA to ribosome ratios used for in individual growth conditions considered, see Supplementary Files 2 and 4. P is taken from (***Klumpp et al., 2013***), k_{el}^{max} from (***Dai et al., 2016***), and the tRNA/ribosome ratios from (***Dong et al., 1996***).

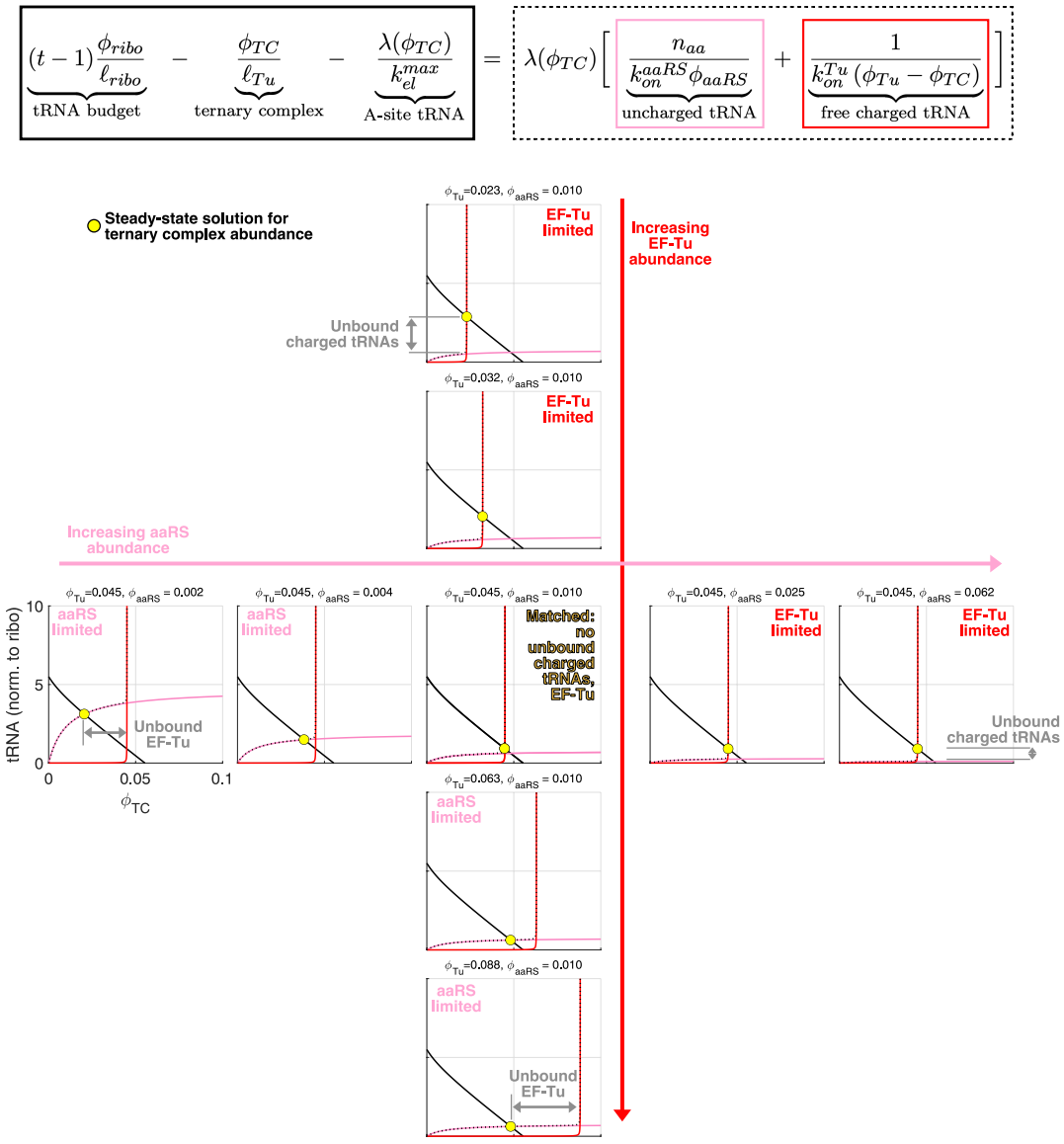


Figure 3–Figure supplement 1. Geometrical interpretation of the sharpness of the separation of the aaRS limited and EF-Tu limited regimes. Each graph corresponds to a different combination of aaRS and EF-Tu abundance. The solution for ϕ_{TC} (yellow circle) corresponds to the intersection of the full (tRNA budget minus TC concentration and ribosome bound tRNAs) and dashed (all remaining tRNA contributions) black lines. Red and pink lines correspond to the free uncharged and charged tRNAs respectively. Because of the rapid divergence of the free charged tRNA term (red) at $\phi_{TC} = \phi_{Tu}$, the system shifts from being limited by aaRS-limited (pink line intersecting full black line) to being EF-Tu limited (red line intersect full black line) over a very narrow range in aaRS or EF-Tu expression change. The central graph corresponds to the abundance of EF-Tu and aaRS matched (no unbound charged tRNAs or EF-Tu), and falls on the transition line of Figure 3

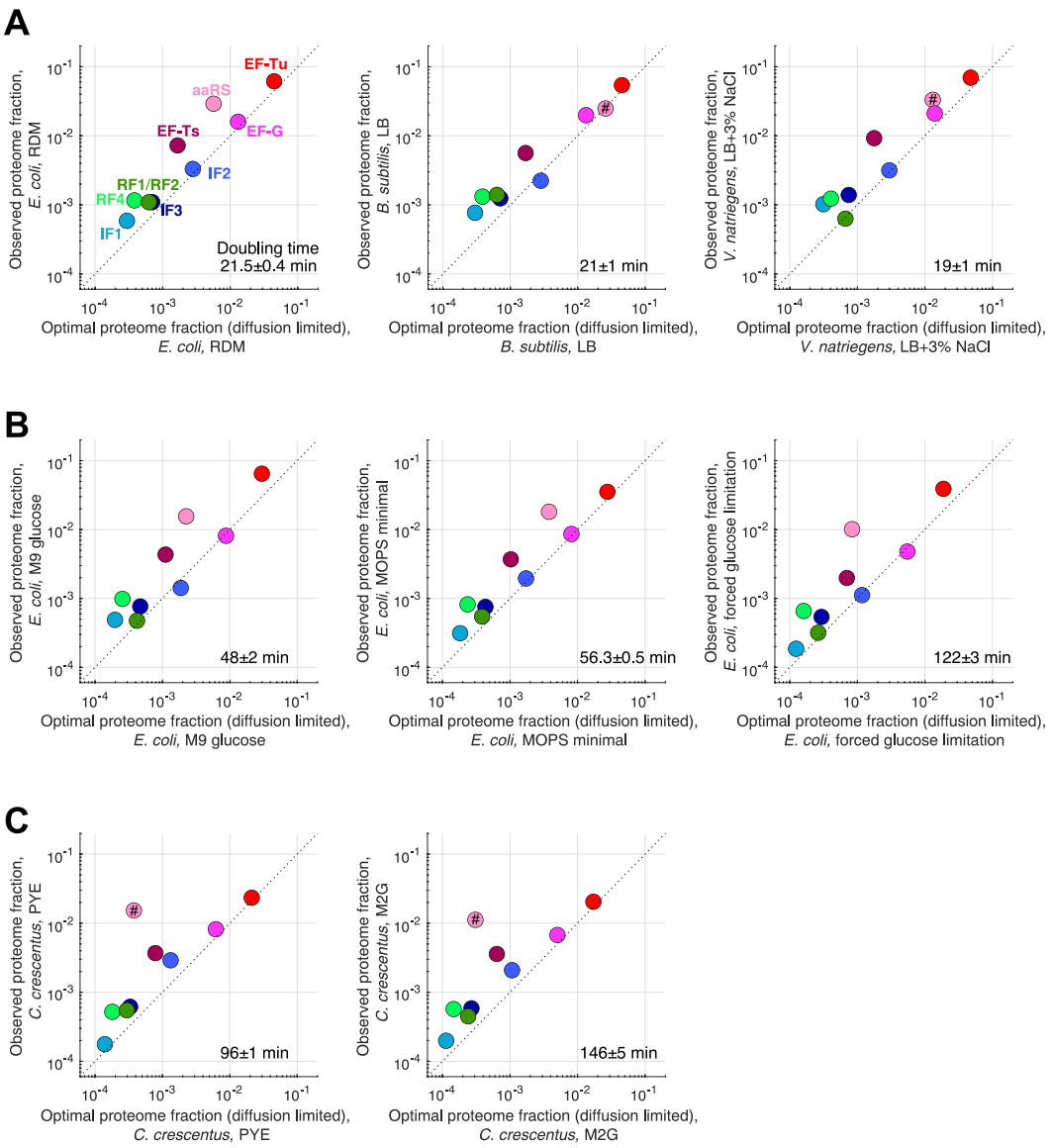


Figure 4–Figure supplement 1. Measured (ribosome profiling) and predicted (diffusion-limited estimates) proteome fraction for core translation factors in individual conditions corresponding to different ribosome profiling datasets included in our analysis (see Supplementary Files 1 to 4). Doubling time for each condition is indicated. (A) individual fast growing species (see Figure 4 for the average). (B) Slower growth conditions in *E. coli*. (C) *C. crescentus* datasets. Predictions of aaRS in species other than *E. coli* are marked by # to indicate that we used *E. coli* tRNA abundance measurements from (***Dong et al., 1996***) to make prediction for this tIF these other species.

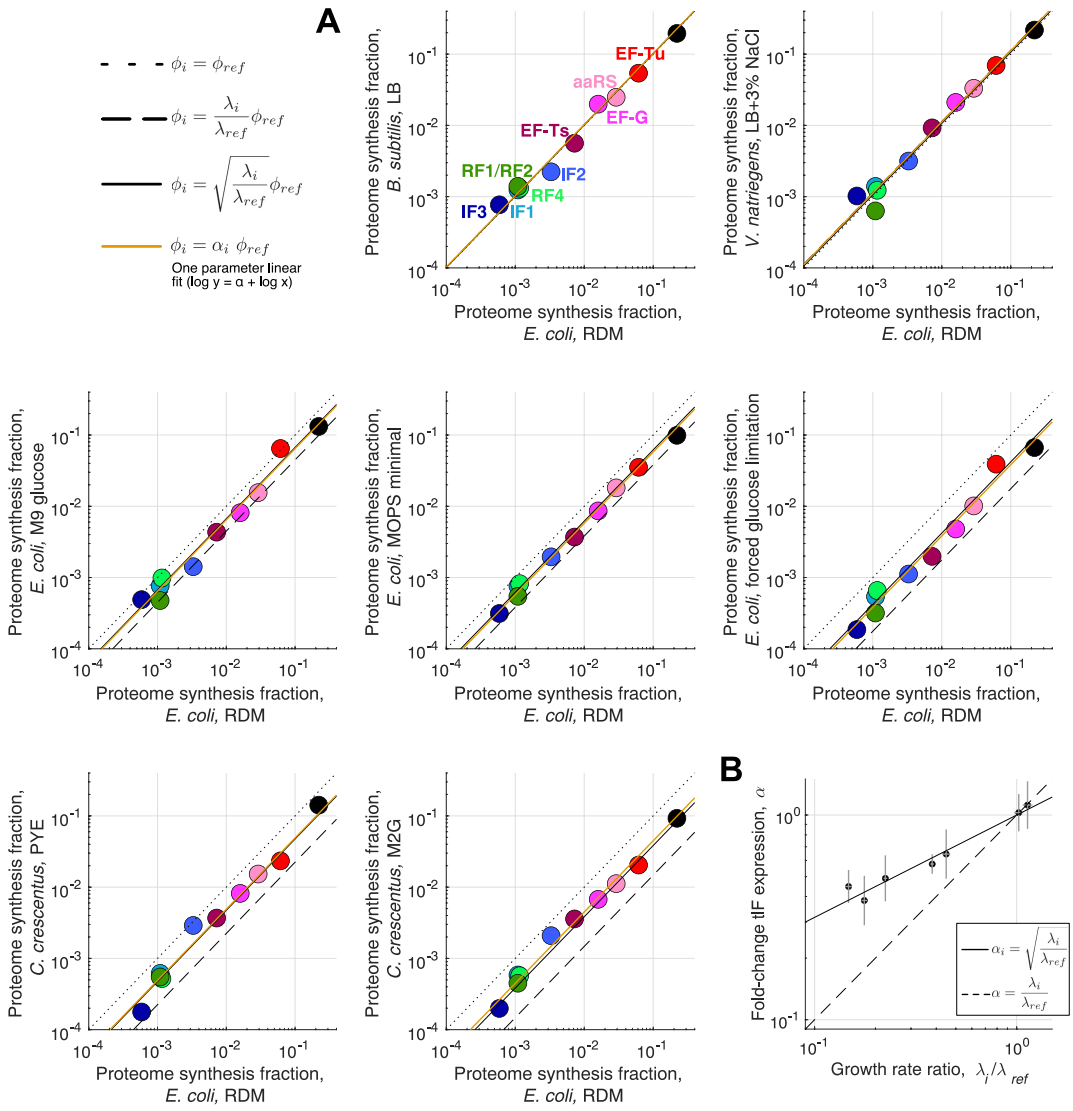


Figure 4–Figure supplement 2. Expression stoichiometry of core translation factors in different species and at different growth rates. (A) Comparison of measured (ribosome profiling) proteome fraction for core translation factors across different species and growth conditions (same conditions as Figure 4-Figure supplement 1). All conditions are compared to the *E. coli* RDM dataset (reference: ref, condition of interest: i). Dotted line correspond to $\phi_i = \phi_{ref}$, dashed line to $\phi_i = (\lambda_i/\lambda_{ref})\phi_{ref}$ and full black line to $\phi_i = \sqrt{\lambda_i/\lambda_{ref}}\phi_{ref}$ (the parameter free prediction from the binding-limited regime of the model, optimal abundance $\propto \sqrt{\lambda}$). Orange line corresponds to the one parameter fit $\log \phi_i = \alpha_i + \log \phi_{ref}$ (excluding aaRS, not expected to follow the square root scaling, and ribosomes), corresponding to the scaling of all factor's abundance. (B) Best one-parameter fit α_i (scale factor) from (A) as a function of the growth rate ratio λ_i/λ_{ref}. Square root scaling: full line. Linear scaling: dashed line. Uncertainties on the growth ratio are propagated from uncertainties of the respective growth rates. Uncertainties in α_i are 95% confidence interval from the linear fits in (A).