Bacterial resource allocation to metabolism and protein translation: optimality and relationship to genome organization

Inaugural dissertation

for the attainment of the title of doctor
in the Faculty of Mathematics and Natural Sciences
at the Heinrich Heine University Düsseldorf

presented by

Xiao-Pan Hu
from Henan, China

Düsseldorf, 12 2021

from the institute for Computer Science
at the Heinrich Heine University Düsseldorf

Published by permission of the
Faculty of Mathematics and Natural Sciences at
Heinrich Heine University Düsseldorf

Supervisor: Prof. Dr. Martin Lercher
Co-supervisor: Prof. Dr. Oliver Ebenhöh

Date of the oral examination: 29.04.2022

Declaration

I declare under oath that I have compiled this dissertation independently and without any undue assistance by third parties under consideration of the "Fundamental principles for safeguarding good scientific practice at Heinrich-Heine-Universität Düsseldorf". Furthermore, neither this dissertation, nor a similar work, has been submitted to another faculty. I have not made any successful or unsuccessful attempts to obtain a doctoral degree.

Düsseldorf, December 2021

Xiao-Pan Hu

Table of Contents

Summary

Phenomenologically, the macromolecular composition of exponentially growing microorganisms largely dependents on their exponential growth rate, not on environmental details. For example, the cellular RNA content increases almost linearly with growth rate in *Escherichia coli*. What are the mechanisms behind these growth rate dependencies? This cumulative thesis examines if these growth rate dependencies can be explained as consequences of the optimal allocation of cellular resources.

Manuscript 1 studies the theoretically optimal expression of components of the *E. coli* protein translation machinery from first principles, and compares the predictions to published experimental data. Translation is the most expensive cellular process at high growth rates in bacteria, both in terms of the proteome fraction of the translation machinery and in terms of ATP usage. It has been suggested that translation components are expressed at optimal efficiency. But what does optimal efficiency mean in an evolutionary context? The cytosol density is near constant across growth conditions *in E. coli*. Thus, if more cytosolic dry mass is allocated to one cellular process, less is available for other processes. We thus hypothesized that the translation machinery has been optimized by natural selection such that its components together amount to the smallest possible mass concentration of all its components at the given growth rate. To test this hypothesis, I built a detailed mechanistic translation model and fully parametrized the model with kinetics constants reported in the literature. The model is constrained only by the physicochemical properties of the molecules and has no adjustable parameters. The growth rate-dependent concentrations of all modeled translational components, including ribosome, tRNAs, mRNA, elongation factor Tu, and elongation factor Ts are accurately predicted by minimizing the combined cost of the whole translation machinery at the given protein synthesis rate. Further, the resulting optimal configurations explain experimental data for the RNA/protein ratio and ribosome activity in both normal growth and antibiotics stress conditions. Minimizing alternative cost measures, such as carbon content, energy cost, and biosynthesis cost, leads to similar results. Thus, the translation machinery works close to optimal efficiency in *E. coli*.

Manuscript 2 examines the growth rate-dependent proteome efficiency of metabolic pathways. In manuscript 1, we found that the protein translation machinery is expressed for maximal efficiency in *E. coli*. However, other recent studies indicate that the overall proteome is not allocated in a way that achieves maximal efficiency. Especially at low growth rates, a substantial fraction of the proteome is unneeded for balanced cell growth. More than half of the total proteome is allocated to metabolism in *E. coli* growing on minimal media. Prior to our work, it was unclear if proteome allocation to different metabolic pathways is similarly efficient, or if some pathways are systematically closer to maximal proteome efficiency than others. In manuscript 2, the minimal proteome demand of individual pathways was predicted by minimizing the proteome cost with a modified version of flux balance analysis with molecular crowding. By comparing the predicted optimal proteome demand of individual pathways with the experimental data, I found that proteome efficiency can qualitatively explain the growth rate-dependent expression of biosynthesis pathways, glycolysis, and the pentose phosphate pathway, but is unable to explain the expression of other pathways. Unexpectedly, by mass, more than half of the metabolic pathways show a growth rate dependence opposite of that expected from optimal demand. Overall, growth rate-dependent proteome efficiency increases along the carbon flow through the metabolic network. While this work provides a bird's-eye view of proteome efficiency at the

pathway level, future work will have to elucidate why proteome allocation evolved this way, and how it gives rise to the widely used bacterial growth laws when averaging over sets of pathways.

Manuscript 3 builds on manuscript 1 by exploring an RNA composition growth law and its partial implementation through the genes' genomic positions in fast-growing bacteria. In contrast to the proteome composition, RNA composition is usually assumed to be independent of the growth rate, despite experimental evidence to the contrary. By minimizing the combined costs of the ribosome and ternary complex, I analytically derived an RNA growth law. This law describes how the optimal tRNA/rRNA ratio decreases monotonically with growth rate, consistent with experimental data from *E. coli* and other fast-growing microbes. In most of these species, rRNA genes are located closer to origin of replication than tRNA genes. Accordingly, the number of rRNA gene copies increases faster than the number of tRNA gene copies with increasing growth rate, a consequence of replication-associated gene dosage effects. The tRNA/rRNA gene copy ratio thus decreases with increasing growth rate, consistent with the RNA growth law. I conclude that the RNA growth law is partially implemented through the relative positions of tRNA and rRNA genes, indicating that natural selection on growth rate-dependent resource allocation patterns can influence the genome organization of bacteria.

In sum, the three manuscripts of this thesis quantify the optimality of growth rate-dependent allocation of bacterial resources into macromolecules involved in different biochemical pathways, linking optimal resource allocation to genome organization.

1 Introduction

1.1 Overview of the growth rate-dependent macromolecular composition in bacteria

In 1958, a pioneering work by Ole Maaløe's lab found that the content of DNA, RNA, and protein in dry weight as well as the cell mass itself are growth rate-dependent in *Salmonella* Typhimurium; these growth rate dependencies were not influenced by the detailed composition of the growth medium or the temperature (Schaechter *et al*, 1958). From the 1960s to the 1980s, substantial efforts were put into the quantification of the growth rate-dependent macromolecular composition of microbes (Neidhardt & Magasanik, 1960; Ecker & Schaechter, 1963; Rosset *et al*, 1966; Dennis & Bremer, 1974; Ikemura, 1981; Bremer & Dennis, 1996) and the interpretation of these growth rate-dependencies in the context of ribosome efficiency (Maaløe & Kjeldgaard, 1966; Koch, 1971, 1988; Ehrenberg & Kurland, 1984).

This thesis focuses on the growth rate-dependent macromolecular composition in *Escherichia coli*. **Fig. 1.1a** shows how the *E. coli* dry weight fraction allocated to protein and RNA changes with growth rate (Bremer & Dennis, 1996). Remarkably, the RNA dry mass fraction is a near-linear function of the growth rate.

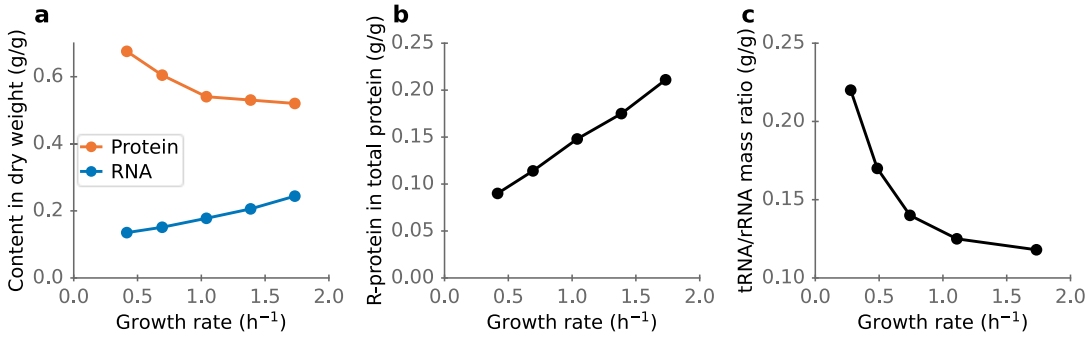


Fig. 1.1. Growth rate-dependent macromolecular composition of *E. coli*. (**a**) The dry mass fractions of protein and RNA. Data from (Bremer & Dennis, 1996). (**b**) Mass fraction of ribosomal protein (R-protein) in total protein. Data from (Bremer & Dennis, 1996). (**c**) Mass ratio between tRNA and rRNA. Data from (Dong *et al*, 1996).

As most of the RNA by mas is ribosomal (rRNA), the RNA and protein content in **Fig. 1.1a** suggests that the fraction of ribosomal protein in total protein changes with growth rate. Indeed, direct measurement of ribosomal protein synthesis rates found that the proteome fraction of ribosomal proteins is an almost linear function of growth rate (Bremer & Dennis, 1996) (**Fig. 1.1b**). A linear scaling of the ribosomal proteome fraction with growth rate was also found in many other species, e.g., *Salmonella* Typhimurium (Schaechter *et al*, 1958), *K. aerogenes* (Neidhardt & Magasanik, 1960), and *S. cerevisiae* (Metzl-Raz *et al*, 2017). Recently, this robust growth rate-dependence of the ribosome abundance has been framed as a bacterial "growth law" (Scott *et al*, 2010; Jun *et al*, 2018).

Similar to the proteome composition, the RNA composition in bacteria is also growth rate-dependent (Kjeldgaard & Kurland, 1963; Dong *et al*, 1996). As shown in **Fig. 1.1c**, the transfer RNA (tRNA) to rRNA ratio decreases with growth rate in *E. coli* (Dong *et al*, 1996).

What causes the growth rate-dependence of the proteome and RNA compositions in bacteria? The current thesis aims to answer this question from the view of optimal resource allocation at the molecular level, further venturing to find the possible imprint of optimal resource allocation on genome organization.

1.2 Proteome sectors

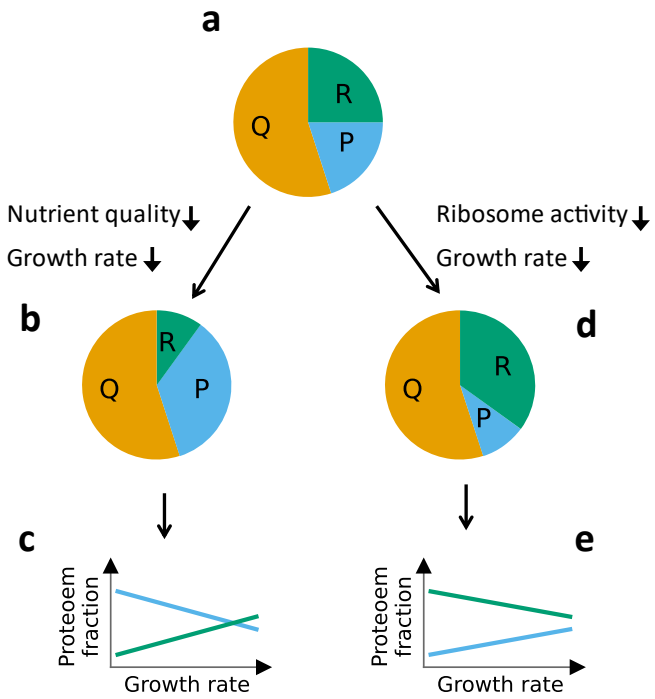


Fig. 1.2. A diagram of the proteome distributed across three sectors: the condition-independent Q-sector, the metabolic P-sector, and the translation R-sector. (**a**) Proteome sectors on rich nutrients without translation limitation. (**b**) and (**c**) Proteome sectors under nutrient limited conditions. (**d**) and (**e**) Proteome sectors under translation-limited conditions enacted through ribosome-targeting antibiotics.

Recently, the growth rate-dependent proteome composition has been characterized in *E. coli* by Terence Hwa's lab (Scott *et al*, 2010; You *et al*, 2013; Klumpp *et al*, 2013; Hui *et al*, 2015). In their theoretical description, the proteome is partitioned into several sectors. **Fig. 1.2** shows how a partitioning into 3 sectors, the simplest version of this theory. The three sectors are the fixed, non-growth-related Q-sector, the metabolic P-sector, and the translation R-sector. Briefly, proteins of the P-sector catalyze nutrients to produce precursors; proteins of the R sector synthesize protein from precursors produced by the P-sector; finally, the Q-sector is a fixed fraction of the proteome that is independent of growth conditions. During growth in a nutrient-rich medium, the partitioning of the proteome looks as shown in **Fig. 1.2a**. When the quality of the medium is reduced, the growth rate decreases and the cell increases its proteome allocation to the P-sector to compensate for the decreased nutrient quality (**Fig. 1.2b**). Because the P-sector increases and the Q-sector is constant, the R-sector must decrease under nutrient limitation. Varying nutrient quality results in a growth rate-dependent expression of the P-sector and the R-sector (**Fig. 1.2c**). Similar to nutrient-limited conditions, a translation limiting condition (a decrease in the effective ribosome activity, e.g., through ribosome-targeting antibiotics) also leads to growth rate decreases. To compensate for translation limitation, the cell increases the proteome fraction of

the R-sector (**Fig. 1.2d**). As the R-sector increases and Q-sector is constant, the P-sector will decrease. Varying the degree of translation inhibition results in a growth rate-dependent expression of the R-sector and the P-sector (**Fig. 1.2e**). At the mechanistic level, the scaling of the R-sector is regulated by the alarmone ppGpp (Scott *et al*, 2010, 2014), while the scaling of the P-sector is mainly regulated by cAMP-CRP and α-ketoglutarate (You *et al*, 2013). In models with more than three sectors, the scaling of the sectors follows a similar scheme to that shown in **Fig. 1.2** (Hui *et al*, 2015): Given varying degrees of a specific kind of growth limitation, the corresponding sector(s) will be up-regulated to compensate the limitation while other, uninvolved sectors will be down-regulated passively (Hui *et al*, 2015).

The empirical proteome sector theory is built on observed proteome allocation patterns and does not consider proteome efficiency (Hui *et al*, 2015). Please note that the scaling of R-sector under nutrient-limited conditions is an exception, which was found to be originated from maximal proteome efficiency (Scott *et al*, 2010; Klumpp *et al*, 2013). But why the cell needs these scalings? Are the scalings of other sectors also originated from maximal proteome efficiency?

1.3 Resource allocation on global and local optimum

What makes the macromolecular composition growth rate-dependent? It was suggested that these phenomena originate from global optimal resource allocation that maximizes growth rate (Bruggeman *et al*, 2020). In global optimal resource allocation, all cellular components are expressed at optimal levels, such that the cell can achieve its maximal growth rate on the given condition (Bruggeman *et al*, 2020). Nevertheless, several findings indicate that resource allocation is not globally optimized in different *E. coli* strains. First, a large fraction of the expressed proteome is unneeded for cell growth at low growth rates (O'Brien *et al*, 2016). Second, the growth rate of *E. coli* can increase by ~ 20-30% in a few hundred generations in adaptive laboratory evolution experiments on given minimal carbon media, even on glucose, the most frequently used carbon source for *E. coli* cultivation (Ibarra *et al*, 2002; Anand *et al*, 2019). Most of the observed changes were regulatory in nature, indicating that the changes in growth rate were mainly caused by resource re-allocation rather than by changes in protein functions (Anand *et al*, 2019). Thus, the growth rate is not maximized on batch growth conditions, even on glucose.

Though it is likely that resource allocation is not globally optimized, it is still possible that resource allocation into subsystems is optimized. For a subsystem, I define its optimal efficiency as the state in which the total cost of all its components is minimized at the given net flux through this subsystem. For example, the net flux through the translation machinery is equivalent to the protein synthesis rate, while the net flux through the metabolic network corresponds to the synthesis rate of precursors included in the biomass function. I describe resource allocation to subsystems that is optimal in this sense as "locally optimal" resource allocation. If all subsystems are expressed at their local optimum, resource allocation in the whole cell is globally optimized.

In this thesis, I test if different subsystems of *E. coli* are expressed at the local optimum, in other words, if the subsystem is expressed at the minimal level that can support its output on the given medium.

1.4 Resource allocation to translation

Translation is the most expensive cellular process at fast growth in bacteria. In *E. coli*, up to 50% of the dry weight (Bremer & Dennis, 2008) and 2/3 of ATP (Russell & Cook, 1995) are devoted to translation at its maximal growth rate.

1.4.1 Translation in three phases

Protein translation occurs in three phases: initiation, elongation, and termination (Rodnina, 2018).

In initiation, free 30S ribosomal subunit and 50S ribosomal subunit assemble on mRNA with the assistance of translation initiation factors and initiator tRNA. The initiation factors will dissociate from the ribosome-mRNA complex after the ribosome assembly is complete; and the initiator tRNA will be discharged after the ribosome moves to the next codon (Goyal *et al*, 2015). In brief, during initiation the ribosome is activated, and simultaneously the first amino acid of the nascent peptide is produced.

In elongation, a ternary complex (TC, the complex of elongation factor Tu (EF-Tu), charged tRNA, and GTP) carries an amino acid to the ribosome A-site (Rudorf *et al*, 2014; Rudorf & Lipowsky, 2015). Then, the ribosome catalyses the polymerization of the amino acid and releases the EF-Tu complex with GDP (EF-Tu·GDP). Next, the ribosome is translocated to the next codon with the help of elongation factor G (EF-G). Moreover, the GDP in EF-Tu·GDP will be replaced by GTP with the help of elongation factor Ts (EF-Ts), and the newly formed EF-Tu·GTP can bind with a new charged tRNA for the next round of elongation (Rodnina, 2018). Polymerization of one amino acid in elongation costs 4 ATPs: two ATPs for tRNA charging, one ATP for polymerization of the amino acid, and one ATP for ribosome translocation. At any given time during cellular growth, most of the ribosomes are in the elongating phase.

When the elongating ribosome reaches a stop codon, the ribosome releases the polypeptide and leaves the mRNA with the assistance of termination factors (Rodnina, 2018).

1.4.2 Growth rate-dependent expression of translation components

As shown in section 1.1, the expression of the ribosome increases almost linearly with the growth rate. Similar to the ribosome, the expression of other translation components – namely EF-Tu (Furano, 1975), EF-Ts (Gordon, 1970; Gordon & Weissbach, 1970), EF-G (Miyajima & Kaziro, 1978), and tRNA (Dong *et al*, 1996) – also increases with the growth rate. The protein synthesis rate increases roughly linearly with the growth rate, so more translation components are needed at higher growth rates. Puzzlingly, the growth dependence of the relative abundance of elongation factors compared to the ribosome differs between factors. The EF-Tu/ribosome ratio decreases with increasing growth rate, while the EF-Ts/ribosome ratio stays near constant in all growth conditions (Gordon, 1970; Gordon & Weissbach, 1970; Furano, 1975). These behaviors of translation factors were found decades ago, but we still do not have a quantitative model to explain these phenomena mechanistically.

1.4.3 What is the objective of translation optimization?

As translation is the most expensive process in fast growing *E. coli* cells, it is likely that natural selection acted to optimize the efficiency of translation. But what exactly is efficiency in the evolutionary context?

1.4.3.1 Constant ribosome activity theory

Since the ribosome is much larger than all other components in translation, it was suggested in the 1960s that ribosome activity is maximized in any given condition, ensuring that the cell can grow at its maximal growth rate (Maaløe & Kjeldgaard, 1966; Koch, 1971, 1988). Due to its simplicity and its ability to explain the expression of the ribosome from moderate to fast growth rates (Koch, 1988), this theory was very popular from the 1960s to 1980s (Maaløe & Kjeldgaard, 1966; Maaløe, 1969, 1979; Ingraham *et al*, 1983; Koch, 1988). However, experiments found that the ribosome activity increases with growth rate (Dennis & Bremer, 1974), as also evidenced by the growth rate-dependent expression of elongation factors (Furano, 1975).

1.4.3.2 Parsimonious usage of protein theory

Ehrenberg and Kurland proposed the parsimonious usage of translation-associated proteins, including ribosomal protein and EF-Tu, at the given protein synthesis rate (Ehrenberg & Kurland, 1984). In this case, the length of a protein was used as a proxy for its cost. This beautiful theory started to treat the translation machinery at the systems level. It considered not only the cost of the ribosome, but also the cost of EF-Tu. In this theory, the expression of both EF-Tu and ribosome increase with growth rate. Moreover, it also qualitatively predicts that the EF-Tu/ribosome ratio decreases with the growth rate. Recently, a theory of parsimonious protein usage together with the notion of proteome sectors was used by Klumpp *et al.* (Klumpp *et al*, 2013). With a diffusion limited rate of TC binding to the ribosome, both the abundance and the activity of ribosomes were predicted accurately. However, the abundance of EF-Tu predicted from this theory is much below the measured data. This suggests that protein cost alone is still not enough for fully understanding translation efficiency.

1.4.3.3 The hypothesis of parsimonious usage of the cytosol density

Besides the protein part, RNA is the other important part of translation components. About 2/3 of the ribosome and 1/3 of the TC by mass consist of RNA. We thus believe that the cost of RNA should be considered when discussing translation efficiency. Since protein and RNA have very different synthesis processes, how can we combine the costs of RNA and protein in a uniform framework?

Cellular dry mass per cell volume is approximately constant across environments and growth rates in *E. coli* (Nanninga & Woldringh, 1985), as is the total mass concentration in the cytosol (Kubitschek *et al*, 1984). If the cell allocates more of this limited mass concentration budget to one particular process, less is available for all other processes. The upper bound for the cytosolic mass concentration, beyond which diffusion becomes inefficient, is a fundamental constraint on cellular growth (Atkinson, 1969; Beg *et al*, 2007; Vazquez, 2010), and thus the molecular weight of a particular molecule can be an approximation to its cost. In this way, the cost of any given molecule can be determined easily.

We thus hypothesize that natural selection minimizes the total mass concentration of translation components utilized to achieve the required protein production rate in *E. coli*. A corresponding optimality principle has been used to understand the relationship between the concentrations of enzymes and their substrates (Dourado *et al*, 2017).

1.4.3.4 Alternative costs for translation components

Besides the protein cost and cytosol density cost, other alternative costs are also widely used in studying resource allocation: the carbon content, which is the number of carbon atoms of a given molecule (Beck *et al*, 2016); the energy cost or ATP cost, which is the total number of high-energy

phosphate bonds required for the synthesis of a given molecule (Akashi & Gojobori, 2002; Lynch & Marinov, 2015); and the synthesis cost, which is the total enzyme mass required for the synthesis of a given molecule (Noor *et al*, 2016).

The biological fitness of cells depends on many factors, and hence any simple assignment of fitness costs to molecules can only be approximate. Based on the assumption that natural selection has optimized a cost/benefit ratio for translation, with these alternative costs, we can test which cost approximates "fitness cost" better.

1.4.4 Aim of Manuscript 1

The aim of Manuscript 1 is to explore the organizing principle of the translation-related growth rate dependencies from first principles. To do this, I first built a mechanistic translation model at the molecular level. The model avoids any empirical growth rate-dependent parameters, such as a growth rate-dependent effective ribosome activity. All reactions in the model are explicitly and exclusively constrained by kinetic parameters retrieved from published papers. Then, I used the model to test if the total cost of the translation machinery is minimized at the given protein production rates in different nutritional environments. Specifically, (1) for a given growth condition, can the abundance of individual translation components be predicted by minimizing the total cost? (2) Across growth conditions, can the growth rate-dependence of the abundance of translation components and the ribosome activity be predicted by minimizing the total cost? (3) Which cost measure appears to be a better proxy for the likely evolutionary objective?

1.5 Resource allocation to metabolism

1.5.1 Open questions in metabolic proteome allocation

In Manuscript 1, we found that resource allocation to the translation machinery is at a local optimum, i.e., the concentrations of translational components are at the minimal required levels that can support the given protein synthesis rate. On the other hand, recent work indicates that the total proteome is not expressed for maximal efficiency in unevolved *E. coli* strains: (1) a large fraction of the proteome is unneeded at low growth rates in *E. coli* (O'Brien *et al*, 2016); and (2) growth rate can increase by ~ 20% in a few hundred generations in adaptive laboratory evolution experiments on minimal media (Ibarra *et al*, 2002), associated with reductions in the abundance of unused proteins (O'Brien *et al*, 2016). These findings indicate some heterogeneity across pathways in their proteome efficiency. In *E. coli*, more than half of the total proteome is allocated to metabolism on minimal media (Schmidt *et al*, 2016). Which pathways are expressed for maximal efficiency and which pathways are not? To answer these questions, we need both proteome data and a genome scale metabolic model for predicting the optimal efficiency of pathways.

1.5.2 Modeling of resource allocation in a genome scale metabolic model

1.5.2.1 Metabolic modeling with enzyme kinetics

Genome scale metabolic models usually contain thousands of reactions, metabolites, and gene-protein-reaction relationships (Fang *et al*, 2020). Ideally, the resource allocation to metabolism would be studied considering full reaction kinetics. However, a large fraction of kinetic parameters are still not available for *E. coli* (Bar-Even *et al*, 2011; Chang *et al*, 2021). Recently, MetabOlic Modeling with ENzyme kineTics (MOMENT), a modified version of flux balance analysis with metabolic crowding, was developed (Adadi *et al*, 2012). MOMENT does not model the full kinetics

of metabolism, instead, it models reaction rate as the product of its enzyme concentration and the enzyme turnover number (k_{cat}). Further, the sum of enzyme concentrations is constrained by the empirical total concentration of metabolic enzymes (the "C-budget"). With these constraints, the optimal concentrations of individual enzymes that maximize growth rate can be predicted in a given growth condition (Adadi *et al*, 2012).

Instead of maximizing the growth rate (as done by MOMENT), in this thesis, I predict the minimal proteome that satisfies the observed growth rate in the given condition. The predicted minimal proteome, hence, is the locally optimal proteome (see section 1.3). Due to the linear relationship between the C-budget and the growth rate in this modelling scheme (Beg *et al*, 2007; Adadi *et al*, 2012; Desouki, 2016), the optimal proteome in a given condition can be readily determined given two pairs of C-budgets and the corresponding growth rates.

1.5.2.2 Enzyme effective turnover number

The enzyme turnover numbers k_{cat} used in the original MOMENT are retrieved from BRENDA (Adadi *et al*, 2012; Chang *et al*, 2021). Because k_{cat}s were measured by different research groups with different methods, the k_{cat} values are commonly inconsistent (Bar-Even *et al*, 2011; Chang *et al*, 2021). Further, the k_{cat}s were measured *in vitro*, where the exact conditions might have been very different from the *in vivo* environment (Davidi *et al*, 2016). Recently, it was found that *in vivo* enzyme effective turnover numbers represent the cellular environment better than the *in vitro* k_{cat}s (Davidi *et al*, 2016; Heckmann *et al*, 2020). Thus, *in vivo* enzyme effective turnover numbers are used for simulations in this thesis whenever they are available.

1.5.2.3 Growth rate-dependent biomass composition

MOMENT maximizes the biomass objective function to simulate cell growth, just as FBA does. The biomass composition is constant in such models (Orth *et al*, 2010). However, it is known that the biomass composition changes with growth rate (see also section 1.1). On the one hand, cell size increases exponentially with growth rate under nutrient limitations (Si *et al*, 2017). As a consequence, the surface-to-volume ratio (S/V) of the cell can be expressed as a function of the growth rate, and thus the mass ratio of cell envelope components to cytosolic components is growth rate-dependent. On the other hand, in the cytosol, the RNA/protein ratio is growth rate-dependent, as more ribosomes are required at faster growth rates (Schaechter *et al*, 1958; Scott *et al*, 2010). With the growth rate-dependent S/V and RNA/protein ratios, and the original biomass composition of the corresponding subsets, the growth rate-dependent composition of biomass can be readily re-calculated.

In MOMENT or FBA, growth rate is the output of simulations. Introducing a growth rate-dependent biomass composition to MOMENT leads to a non-linear optimization problem. In contrast, in the framework of locally optimal resource allocation, both the growth rate and the growth rate-dependent biomass composition are given in advance (see section 1.5.2.1). The optimization problem can thus still be solved linearly.

Recently, Metabolism and Expression models (ME-models) (O'Brien *et al*, 2013) and Resource Balance Analysis (RBA) (Goelzer *et al*, 2015) were developed to study resource allocation into metabolism and gene expression machineries. In these models, the activity of gene expression machineries, e.g., RNA polymerase and ribosome, are functions of the growth rate (Goelzer *et al*, 2015). In the framework of locally optimal resource allocation, the growth rate is given in advance, and thus the activity of gene expression machineries is already determined before starting the simulation. The optimal abundance of gene expression machineries required for synthesizing the proteome allocated to metabolism can be readily calculated from the activities of gene expression

machineries and the predicted optimal metabolic enzyme abundances. Because we explicitly model the growth-rate dependence of the biomass composition, the optimal expression of gene expression machineries does not affect the optimal proteome for metabolism and these more complicated models are equals to the MOMENT in this context.

1.5.3 Aim of Manuscript 2

The aim of Manuscript 2 is to study the growth rate-dependent proteome efficiency of metabolic pathways. Specifically, this manuscript explores if the proteome allocation to different metabolic pathways can be explained by optimal proteome efficiency. Moreover, with the concept of locally proteome efficiency, it is possible to re-think the optimality of resource allocation and the limiting factors for growth rate.

1.6 RNA composition and gene positions

1.6.1 Growth rate-dependent RNA composition

In contrast to the proteome composition, the partitioning of bacterial RNA into messenger (mRNA), ribosomal (rRNA), and transfer (tRNA) RNA is often assumed to be growth rate-independent (Scott *et al*, 2010; Klumpp *et al*, 2013; O'Brien *et al*, 2013; Bosdriesz *et al*, 2015; Dai *et al*, 2016; Bremer & Dennis, 2008). However, it was found the tRNA/rRNA expression ratio decreases monotonically with growth rate in many microbes, including *E. coli* (Dong *et al*, 1996), *B. subtilis* (Doi & Igarashi, 1964), *Salmonella* Typhimurium (Rosset *et al*, 1966), *S. pyogenes* (Panos *et al*, 1965), *K. aerogenes* (Rosset *et al*, 1966), *N. crassa* (Alberghina *et al*, 1975), and *S. cerevisiae* (Waldron & Lacroute, 1975).

Can the growth rate-dependent tRNA/rRNA ratio be explained by optimal resource allocation? Indeed, in Manuscript 1, we found that locally optimal resource allocation to the translation machinery leads to a tRNA/ribosome ratio that is a decreasing function of growth rate. But the method used in Manuscript 1 cannot be extended to other species due to a lack of parameter values. However, most ribosomes are actively translating in the elongation state (Dai *et al*, 2016) and most tRNAs are binding with EF-Tu as the substrate for elongating ribosomes. Thus, the detailed model in Manuscript 1 can be simplified by considering only the elongation part, which can further be simplified into a single Michaelis-Menten type reaction (Klumpp *et al*, 2013). This simplified translation model can be readily solved analytically (Dourado *et al*, 2017) and can be applied across species.

1.6.2 Gene dosage depends on both growth rate and genomic position

In fast-growing prokaryotes, the doubling time can be even shorter than the genome replication time. For example, the minimal doubling time is about 20 minutes, whereas the DNA replication time is about 40 minutes in *E. coli* (Cooper & Helmstetter, 1968). To coordinate DNA replication and cell division, fast-growing prokaryotes must re-initiate chromosome replication before the previous replication round is complete. The higher the cellular growth rate is, the more replication rounds are needed in the cell. Thus, a gene's DNA copy number per cell (termed "gene dosage" hereafter) is growth rate-dependent. These phenomena are called replication-associated gene dosage effects (Couturier & Rocha, 2006).

Moreover, because DNA replication always starts at the origin and ends at the terminus of replication in prokaryotes, gene dosage also depends on the gene's relative position on the chromosome. When there are multiple replication rounds in the cell, genes near the origin of

replication (oriC) have more DNA copies than genes near the terminus of replication. Fast-growing *E. coli* cells can have up to three replication rounds simultaneously (Cooper & Helmstetter, 1968). In this case, during the D period of the cell cycle, the dosage of genes near the origin are eight times higher than those of genes near the terminus. Therefore, the dosage of a given gene is a function of both its relative position on the chromosome and the cell's growth rate.

With the DNA replication model developed by Cooper and Helmstetter (Cooper & Helmstetter, 1968) and later generalized by Bremer and Churchward (Bremer & Churchward, 1977), the average dosage of gene *i* ($\overline{X_i}$) in a population can be written as

$$\overline{X_i} = e^{\mu[C(1-position_i)+D]}. \quad (1.1)$$

Here, μ is the growth rate, *C* is the time needed for DNA replication, *D* is the time needed for DNA segregation after replication is complete, and $position_i$ is the relative position of gene *i* on the chromosome, calculated as the shortest distance between the gene and oriC on the circular chromosome divided by half of the length of the chromosome.

1.6.3 Relationship between gene position and gene expression

Since the DNA copy number of a gene directly influence its expression, gene position can impact gene expression through replication-associated gene dosage effect. This effect has been observed for protein coding genes, rRNA genes, and tRNA genes. First, chromosome rearrangements that shift highly expressed protein coding genes from the origin to the terminus of replication reduce the expression of the shifted genes (Campo *et al*, 2004; Louarn *et al*, 1985; Soler-Bistué *et al*, 2017). Second, rRNA operons near oriC have higher expression levels than operons near the terminus (Condon *et al*, 1992). Finally, tRNA dosage explains tRNA expression better than tRNA copy per chromosome (Ardell & Kirsebom, 2005).

1.6.4 Hypothesis of coordination between gene position and expression demand

It is widely recognized that highly expressed genes tend to be located near oriC. However, only highly expressed transcription- and translation-related genes are located near oriC, whereas other highly expressed genes are not (Couturier & Rocha, 2006). Why are other highly expressed genes not located near oriC?

Neither gene dosage nor gene expression are static, they both change with growth conditions. With increasing growth rate, the cell needs more transcription- and translation-related genes, as evidenced by the growth rate-dependent RNA polymerase and ribosomal proteins in *E. coli* (Bremer & Dennis, 2008). Since transcription- and translation-related genes are located near oriC, their dosage will increase faster than the average gene dosage with increasing growth rate. Thus, the dosage and expression are coordinated for transcription- and translation-related genes: dosage and expression simultaneously increase with growth rate; and, therefore, the genomic positions of these genes can help in regulating the growth rate-dependent expression of these genes.

Based on this speculation, I propose the hypothesis that natural selection for optimal proteome efficiency resulted in a gene's position being coordinated with its expression demand. **Fig. 1.3a** illustrates this hypothesis for two genes, gene A, which is located near oriC, and gene B, which is located near the terminus. With increasing growth rate, the relative dosage of gene A (dosage of gene A / average dosage of all genes in the genome) increases, whereas the relative dosage of gene B decreases (middle column). If the optimal demand of gene A increases with growth rate,

its genomic position near oriC can facilitate its optimal expression (right column). The instances for gene A are transcription- and translation-related genes, as these genes are expressed to increase robustly with growth rate. If the optimal demand of gene B decreases with growth rate, its genomic position near the terminus can facilitate its optimal expression (right column). Until now, there are no verified instances of such gene Bs. A preliminary analysis shows that the major outer membrane lipoprotein (lpp) is very likely to be an instance of gene B: lpp is one of the most highly expressed genes in *E. coli* (Wang *et al*, 2012), and it is located near the terminus. As a structural protein that is anchored to the inner layer of the outer membrane, its optimal demand decreases with growth rate (due to the robust relationship between the surface to volume (S/V) ratio and the growth rate, see section 1.5.2.3).

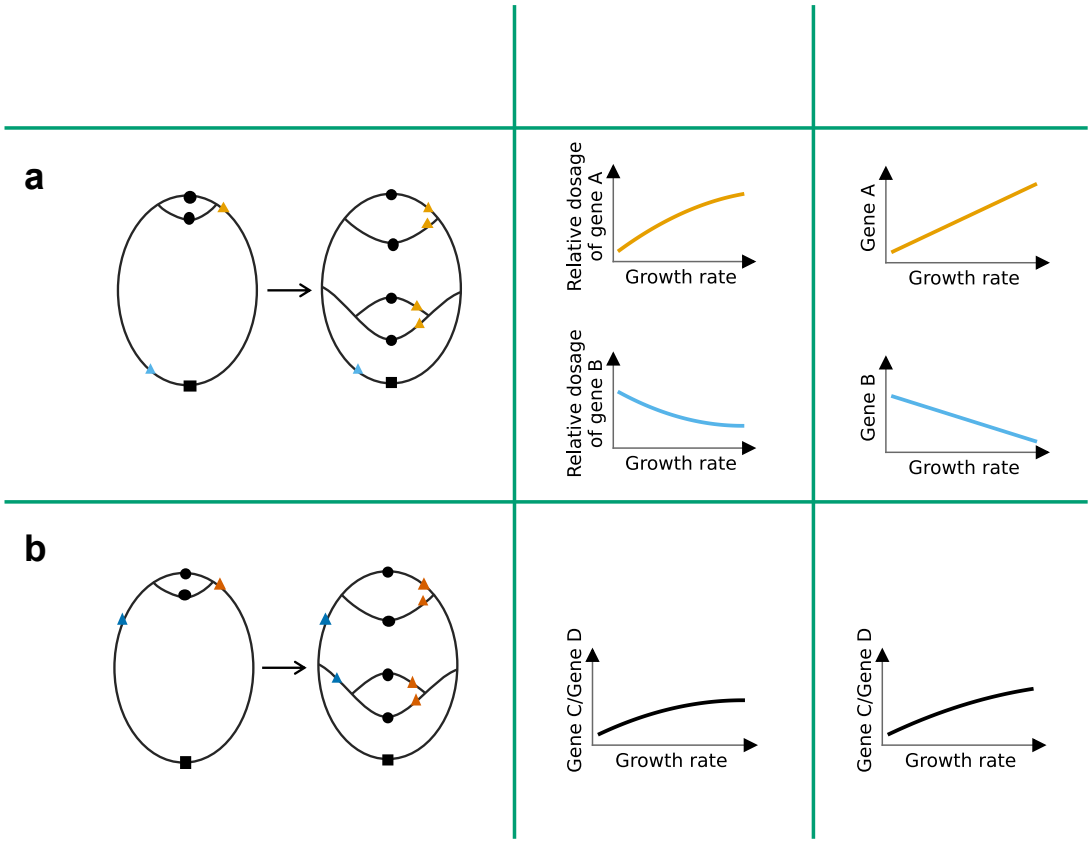


Fig. 1.3. Diagram of the coordination between gene position and expression demand. (**a**) Coordination between gene position and expression demand in individual genes. (**b**) Coordination between relative position and required expression ratio. Please note that the curves of gene dosage (middle column) and gene expression (right column) only indicate the growth rate-dependent trends and are not representing realistic data.

Importantly, gene dosage depends robustly on growth rate as it obeys equation (1.1); but the expression of genes may depend strongly on growth conditions, in particular on the composition of the medium. For example, amino acid synthesis enzymes are highly demanded on minimal carbon media, but on a medium supplied with amino acids, the expression of amino acid synthesis enzymes sharply decreases with increasing growth rate (Schmidt *et al*, 2016). Since bacteria always face changing environments, these condition-dependent metabolic genes are less likely to have a biased position on the chromosome. In summary, for a given gene, the more its expression depends on the growth rate rather than on other factors, the more likely it is to be located closer

to oriC or the terminus. This might explain why other other highly expressed genes are not located near oriC.

The hypothesis of coordination between gene position and expression demand can be applied to two genes, quantifying how their relative positions are coordinated with their expression demand ratio. With equation (1.1), the dosage ratio of two genes can be written as

$$\frac{\overline{X_i}}{\overline{X_j}} = e^{\mu C(position_j - position_i)}. \quad (1.2)$$

Here, $\overline{X_i}$ and $\overline{X_j}$ are the dosages of gene *i* and gene *j,* respectively; $position_i$ and $position_j$ are the relative positions of gene *i* and gene *j,* respectively. As illustrated in **Fig 1.3b**, if the expression ratio of gene C to gene D demanded by optimal cellular efficiency increases with increasing growth rate, a closer position of gene A than gene B to oriC can facilitate this growth rate-dependent expression ratio.

Both the species-specific maximal growth rate and the optimal, growth rate-dependent expression of a particular gene may influence its optimal position. The effect of gene position on the growth rate is less significant in slow-growing species than in fast-growing species, as replication-associated gene dosage effects highly depend on the maximal growth rate of a given species. As highly expressed genes are under stronger selection pressures than lowly expressed genes (Pál *et al*, 2001), it is also likely that highly expressed also have more biased genomic positions than lowly expressed genes. Thus, the coordination between gene position and expression demand may be less significant for lowly expressed genes or for genes in slow-growing species.

The hypothesis proposed in this section will be tested on tRNA and rRNA genes in this thesis. Future work will test this hypothesis on protein-coding genes.

1.6.5 Aim of Manuscript 3

The first aim of Manuscript 3 is to explore a universal growth law for RNA composition that results from the optimal translation efficiency hypothesis, which was proposed in Manuscript 1. The second aim is to test if the differential position of tRNA and rRNA genes on the chromosome is related to their growth rate-dependent expression ratio. Moreover, since replication-associated gene dosage effects highly depend on the maximal growth rate of a species, this work also tests if the relative positioning of tRNA and rRNA genes is only conserved in fast-growing species.

2 Manuscript 1. The protein translation machinery is expressed for maximal efficiency in *Escherichia coli*

This manuscript was adapted from the following publication:

Hu X-P, Dourado H, Schubert P, Lercher MJ. The protein translation machinery is expressed for maximal efficiency in *Escherichia coli*. *Nature Communications*. 2020;11: 5260. https://doi.org/10.1038/s41467-020-18948-x.

Contribution: I developed, implemented, and parameterized the model, performed the analyses, and drafted the manuscript.

Abstract

Protein synthesis is the most expensive process in fast-growing bacteria. Experimentally observed growth rate dependencies of the translation machinery form the basis of powerful phenomenological growth laws; however, a quantitative theory on the basis of biochemical and biophysical constraints is lacking. Here, we show that the growth rate-dependence of the concentrations of ribosomes, tRNAs, mRNA, and elongation factors observed in *Escherichia coli* can be predicted accurately from a minimization of cellular costs in a mechanistic model of protein translation. The model is constrained only by the physicochemical properties of the molecules and has no adjustable parameters. The costs of individual components (made of protein and RNA parts) can be approximated through molecular masses, which correlate strongly with alternative cost measures such as the molecules' carbon content or the requirement of energy or enzymes for their biosynthesis. Analogous cost minimization approaches may facilitate similar quantitative insights also for other cellular subsystems.

Introduction

Protein translation is central to the self-replication of biological cells. While the workings of its individual components are well understood, the translation apparatus is a complex machine with many degrees of freedom, where the same rate of protein production could be achieved with very different relative abundances of its components. Although a large body of quantitative experimental data on these abundances in *E. coli* across different growth conditions is available, it is still unclear according to which organizing principle(s) – if any – they are set by the cell. Given that translation is the energetically most expensive process in fast growing *E. coli* cells, accounting for up to 50% of the proteome (Bremer & Dennis, 1996) and 2/3 of cellular ATP consumption (Russell & Cook, 1995), it is likely that natural selection acted to optimize the efficiency of translation. But what exactly is "efficiency" in the evolutionary context?

In the late 1950s, it was hypothesized that ribosomes operate at a constant, maximal rate (Schaechter *et al*, 1958; Koch, 1988), consistent with the observed linear dependence of ribosome concentration on growth rate (Schaechter *et al*, 1958; Scott *et al*, 2010; Neidhardt & Magasanik, 1960; Maaløe, 1979). This hypothesis was later proven untenable, as the activity of ribosomes was observed to increase with growth rate (Forchhammer & Lindahl, 1971). Klumpp *et al.* (Klumpp *et al*, 2013) suggested that optimal translational efficiency corresponds to the parsimonious usage of translation-associated proteins, most notably ribosomal proteins, elongation factor Tu, and tRNA synthetases. While these authors were able to fit a coarse-grained phenomenological model to the data, their suggested evolutionary objective could also not explain the observed growth rate dependencies quantitatively (see **Text S2.1** for a discussion of Ref. (Klumpp *et al*, 2013) and of the phenomenological model of bacterial growth it is based on (Scott *et al*, 2010, 2014)). Thus, it is currently unclear to what extent translation has indeed been optimized by natural selection, and – if such optimization indeed occurred – whether its action can be expressed in terms of a simple objective function.

In principle, these questions would best be addressed in the context of a whole-cell model of balanced growth that combines mechanistic descriptions of metabolism and protein production. However, while such models have been described conceptually (Molenaar *et al*, 2009; Dourado & Lercher, 2020), kinetic parameterizations are unavailable for a majority of the relevant enzymatic reactions (Nilsson *et al*, 2017) , preventing a truly mechanistic description that combines metabolism and protein translation. Thus, we here focus on protein production alone, taking the experimentally observed output of translation (proteome production rate and composition in a given growth condition), the corresponding input (charged tRNAs), and the kinetics of individual translation reactions as given. We then use this mechanistic description of translation to find the combination of the concentrations of mRNA, ribosomes, elongation factors, and tRNA that results in minimal cellular costs in the given condition. Thus, our estimate of the optimal efficiency of the translation machinery is not based on the maximization of ribosome activity, but on the minimization of the combined cost of the complete translation machinery at an observed protein production output.

We base our cost definition on the experimental observation that cellular dry mass per cell volume is approximately constant across environments and growth rates in *E. coli* (Nanninga & Woldringh, 1985), as is the total mass concentration in the cytosol (Kubitschek *et al*, 1984). If the cell allocates more of this limited mass concentration budget to one particular process, less is available to other processes. The upper bound for the cytosolic mass concentration, beyond which diffusion becomes inefficient, is a fundamental constraint on cellular growth (Atkinson, 1969; Beg *et al*,

2007; Vazquez, 2010), and we thus use the cytosolic mass concentration of a particular molecule type as an approximation to its cost. Theoretical models of cellular growth that account for all major biochemical and biophysical constraints indicate that the limit on cellular dry mass indeed represents a dominant constraint on bacterial growth rates (Dourado & Lercher, 2020).

We hypothesize that to maximize the *E. coli* growth rate in a given environment, natural selection minimizes the total mass concentration of translation components utilized to achieve the required protein production rate. A corresponding optimality principle has been used to understand the relationship between the concentrations of enzymes and their substrates (Dourado *et al*, 2017).

We find that a theoretical minimization of the combined cellular costs of the translation machinery components indeed leads to accurate predictions for their abundances, the resulting elongation rate, and the RNA/protein ratio. In addition to molecular masses, we also examine four alternative cost measures for cellular components that have been explored in the literature: (i) their protein content (Ehrenberg & Kurland, 1984; Klumpp *et al*, 2013); (ii) their carbon content (Beck *et al*, 2016); and (iii) the energy (Mahmoudabadi *et al*, 2017; Weiße *et al*, 2015) or (iv) the amount of catalysts(Scott *et al*, 2010; Noor *et al*, 2016) required for their production. We find that these alternative cost measures are strongly correlated for the studied components of the translation machinery and lead to very similar predictions for their abundances; the only cost measure that leads to substantially different predictions is the protein content, which does not assign any cost to tRNA and mRNA molecules.

Results and Discussion

Cost minimization in a mechanistic model of translation

To test our hypothesis, we constructed a translation model consisting of 274 biochemical reactions, including 119 reactions with non-linear kinetics. **Fig. 2.1** shows the modeled reactions for a subset of the 61 codons and the 40 species of charged tRNAs; for details see Methods, Table S2.1, and Data S2.1. This mechanistic model accounts for the concentrations of mRNA, the ribosome, the different charged tRNAs, and the elongation factors Ts (EF-Ts) and Tu (EF-Tu). We fully parameterized the model with molecular masses and kinetic constants measured experimentally (Tadmor & Tlusty, 2008; Gromadski *et al*, 2002; Louie & Jurnak, 1985); the only exceptions are the translation initiation parameters, which were previously estimated from gene expression data (Tadmor & Tlusty, 2008), and the ribosomal Michaelis constant for the ternary complexes, which was previously estimated based on the diffusion limit (Klumpp *et al*, 2013). The model is based purely on biochemical and biophysical considerations; it contains no adjustable parameters, nor does it include any explicit growth rate dependencies.

For *E. coli* growing under different experimental conditions, we used measured growth rates and protein concentrations (Schmidt *et al*, 2016) to determine the required translation rate and the proportions of the different amino acids incorporated into the elongating proteins. At this required protein production rate, we minimized the combined cost of the translation machinery in our model, treating the concentrations of all components as free variables; the values of individual reaction fluxes result deterministically from these concentrations according to the respective rate laws (Methods). As the modeled kinetic rate laws are non-linear, all optimizations were performed numerically. In repeated optimization runs with two different solvers, we never found alternative optima, indicating that the optimization problem may be convex. The results shown in the main text and figures are based on the assumption that costs are proportional to molecular masses; results based on other cost functions are shown in **Figs. S2.5** to **S2.7**.

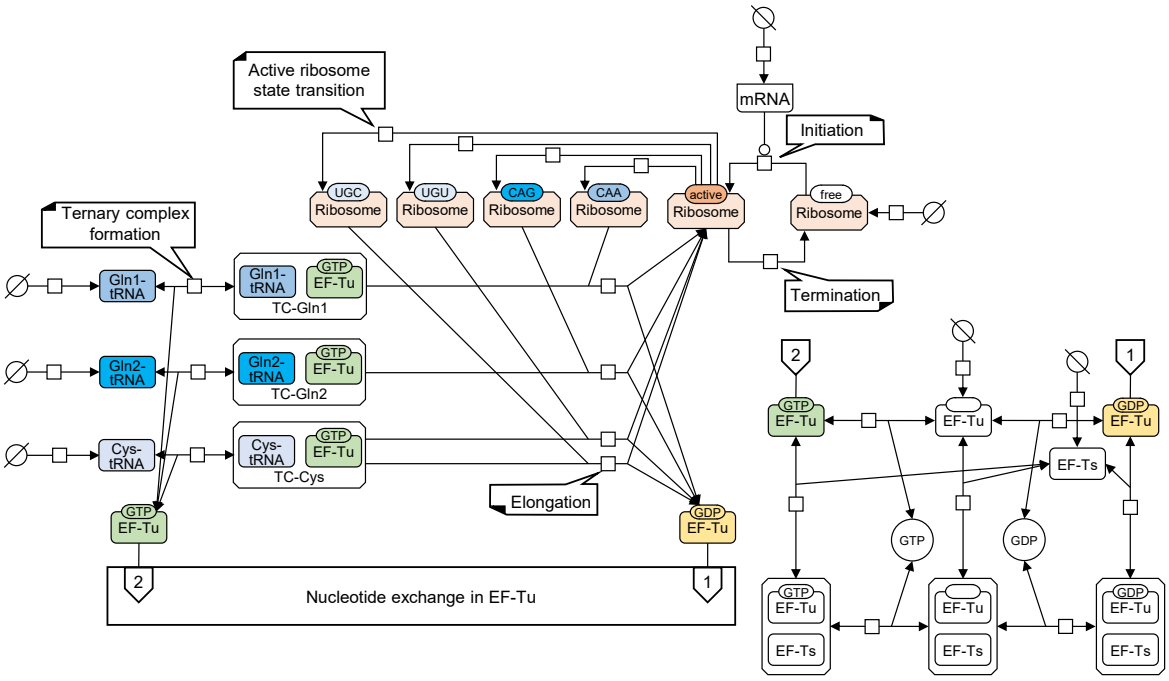


Fig. 2.1 Schematic overview of the translation model. A reduced pathway for elongation with amino acids cysteine (aminoacyl-tRNA Cys) and glutamine (aa-tRNA Gln1, Gln2) is represented in Systems Biology Graphical Notation. Initiation: free (unbound) ribosome gets converted to active ribosome, modulated by mRNA. Termination: active ribosome converts back to free ribosome at a rate fixed by the desired protein production rate. Active ribosome state transition: active ribosome instantaneously binds to codons (61 codons in full model, 4 here) at the fractions set by the specified proteome composition. Ternary complex formation: charged tRNAs (40 aa-tRNAs in full model, 3 here), replenished from a pool, combine with EF-Tu*GTP to form ternary complexes (40 TCs in the full model). Kinetic parameters of these reversible processes depend on the aa-tRNA. Elongation: labeled ribosome binds with the cognate TC to elongate the protein with the respective amino acid. The ribosome returns to its active state and EF-Tu*GDP is released. Other products of this reaction, such as deacylated tRNA, are not modelled. Nucleotide exchange (see right panel): EF-Tu*GDP is reactivated to EF-Tu*GTP in a sequence of steps modeled by reversible mass action kinetics. GTP and GDP pools are modeled with fixed concentrations. The nucleotide exchange is supported by EF-Ts, and the main flux is carried through the complexes formed by EF-Tu with EF-Ts.

Predicted concentrations agree with observations

We first compared our predictions to experimental data for exponentially growing *E. coli* in different conditions (Schmidt *et al*, 2016; Dong *et al*, 1996; Forchhammer & Lindahl, 1971; Valgepea *et al*, 2013; Skjold *et al*, 1973). **Fig. 2.2** shows the results for growth in a glucose-limited chemostat at growth rate μ = 0.35 h^{-1}; for other conditions, see **Fig. S2.1**. The mechanistic model accurately predicts the absolute concentrations of ribosomes, EF-Tu, EF-Ts, mRNA, and total tRNA in each condition. Predictions for individual tRNA concentrations are less accurate but are still mostly within a 2-fold error (**Fig. 2.2**, **Fig. S2.1**); the discrepancies may be due to the simplifying assumption of the same ribosomal Michaelis constant K_m for all tRNA species (Klumpp *et al*, 2013).

We next tested if this systems-level view on the total cost of translation explains the observed growth rate-dependencies of the expression of translation machinery components (Schmidt *et al*, 2016; Forchhammer & Lindahl, 1971; Skjold *et al*, 1973; Dong *et al*, 1996; Dai *et al*, 2016), of the elongation rate(Dai *et al*, 2016), and of the RNA/protein ratio (Dai *et al*, 2016; Scott *et al*, 2010), considering experimental data across 20 diverse conditions (14 minimal media, including 3 stress

conditions; 4 chemostats; and 2 rich media) (Schmidt *et al*, 2016). The predicted concentrations of ribosomes, EF-Tu, and EF-Ts increase with growth rate in line with experimental observations (**Fig. 2.3**).

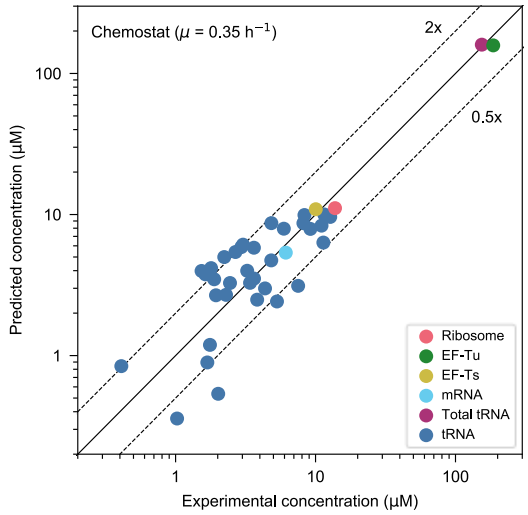


Fig. 2.2 Predicted vs. observed concentrations in a glucose-limited chemostat. Growth rate μ = 0.35 h^{-1} (for other conditions, see Fig. S2.1). The solid line shows the expected identity, whereas the upper and lower dashed lines show prediction errors of 2x and 0.5x, respectively. Predictions for ribosome, EF-Tu, EF-Ts, mRNA, and total tRNA are highly accurate, with Pearson's R^2 = 0.99 and geometric mean fold-error GMFE = 1.13, *i.e.*, predictions based purely on a physico-chemical model and the assumption of cost minimization are on average 13% off. Predictions for individual tRNA species are somewhat less accurate, GMFE = 1.64. Experimentally determined concentrations of the ribosome (averaged over all ribosomal proteins), EF-Tu, and EF-Ts are from Ref. (Schmidt *et al*, 2016). mRNA (Valgepea *et al*, 2013) and tRNA (Dong *et al*, 1996) concentrations are interpolated values based on growth rates.

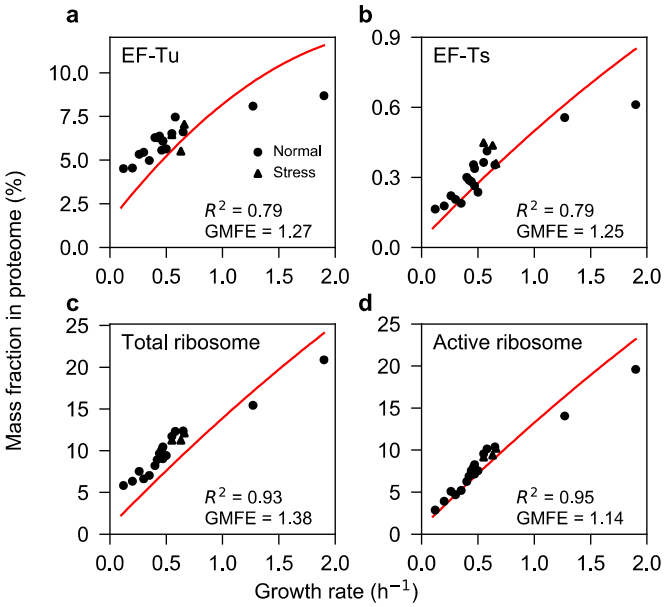


Fig. 2.3 Growth rate-dependence of predicted (red lines) and observed concentrations. (a) EF-Tu, R^2 = 0.79, GMFE = 1.27. **(b)** EF-Ts, R^2 = 0.79, GMFE = 1.25. **(c)** Total ribosome concentration (arithmetic means across ribosomal proteins). **(d)** Actively elongating ribosomes, estimated from data in panel (c) according to Ref. (Dai *et al*, 2016) (see Methods). Circles indicate normal conditions; triangles indicate stress conditions.

Predicted absolute abundances of EF-Tu (**Fig. 2.3a**), EF-Ts (**Fig. 2.3b**), and mRNA (**Fig. S2.2a**) account quantitatively for the experimental data (Schmidt *et al*, 2016; Dong *et al*, 1996; Forchhammer & Lindahl, 1971; Skjold *et al*, 1973; Valgepea *et al*, 2013), with average deviations (geometric mean fold-error) GMFE ≤ 27% for the elongation factors and GMFE = 6% for mRNA. At low growth rates, experimentally observed concentrations of EF-Tu (**Fig. 2.3a**) and tRNA (**Fig. S2.2b**) are higher than predicted. The model only includes charged (aminoacyl-) tRNA concentrations, and it is likely that the unknown fraction of uncharged tRNA explains at least part of this deviation. Overall, the largest deviations between observed concentrations and predictions are seen in the two non-minimal conditions, which also exhibit the fastest growth ($\mu > 1$ h^{-1}). A recent analytical study of balanced cellular growth indicates that these deviations may result from the influence of an increased growth-related dilution of cofactors and other intermediate metabolites, a phenomenon not included in our simulations (Dourado & Lercher, 2020).

Active and de-activated ribosome fractions

At low growth rates ($\mu < 0.3$ h^{-1}; **Fig. 2.3c**), observed ribosome concentrations exceed those predicted from cost minimization, a deviation consistent with a substantial reserve of deactivated ribosomes at low growth rates (Dai *et al*, 2016). Such deactivated ribosomes may provide fitness benefits in changing environments (Mori *et al*, 2017; Erickson *et al*, 2017), but cannot be maximally efficient in a constant environment and thus cannot be predicted by our optimization strategy. To allow a meaningful comparison between predictions and experiment, we thus estimated the experimental concentration of ribosomes actively involved in elongation (Methods). Cost minimization predicts these experimental estimates with high accuracy across the full range of assayed growth rates; observed values deviate from predictions on average by GMFE = 14% (**Fig. 2.3d**).

The remaining, non-active ribosome fraction comprises two parts: the deactivated ribosome reserve currently unavailable for translation (Dai *et al*, 2016), and free, potentially active ribosomes not currently bound to mRNA (see **Text S2.2** for the nomenclature on ribosome states). As our model quantifies the abundance of both active and free ribosomes, their subtraction from observed total ribosome concentrations provides an estimate of the deactivated ribosome reserve as a function of the growth rate (**Fig. 2.4**). While this reserve accounts for less than 20% of total ribosomes at fast to moderate growth, it reaches almost 50% at the lowest growth rate assayed in Ref. (Schmidt *et al*, 2016).

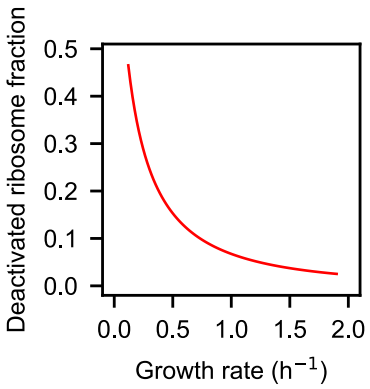


Fig. 2.4 Estimated fraction of deactivated ribosomes. The deactivated fraction reaches almost 50% for the lowest growth rate assayed in Ref. (Schmidt *et al*, 2016) and drops rapidly towards zero at higher growth rates.

RNA/protein ratio and elongation rate

A linear correlation between the RNA/protein ratio and growth rate was discovered in the 1950s (Schaechter *et al*, 1958; Neidhardt & Magasanik, 1960; Dennis & Bremer, 1974; Maaløe, 1979) and forms the basis of phenomenological bacterial growth laws (Klumpp *et al*, 2013; Dai *et al*, 2016; Scott *et al*, 2010). Relating the predicted total RNA (ribosomal RNA + tRNA + mRNA) with measured protein concentrations (Schmidt *et al*, 2016) indeed results in a near-linear relationship, accurately matching observed values at high to intermediate growth rates (μ > 0.3 h^{-1}; **Fig. 2.5a**). At lower growth rates, model predictions are slightly too low, likely because of the deactivated ribosome reserve (Dai *et al*, 2016) (**Fig. 2.4**). At low growth rates (μ = 0.12 h^{-1}), predictions of RNA and proteins allocated to an optimally efficient translation machinery (including deactivated ribosomes) account for 13% of total dry mass, rising almost linearly to 49% at high growth rates (μ = 1.9 h^{-1}; **Fig. S2.3**).

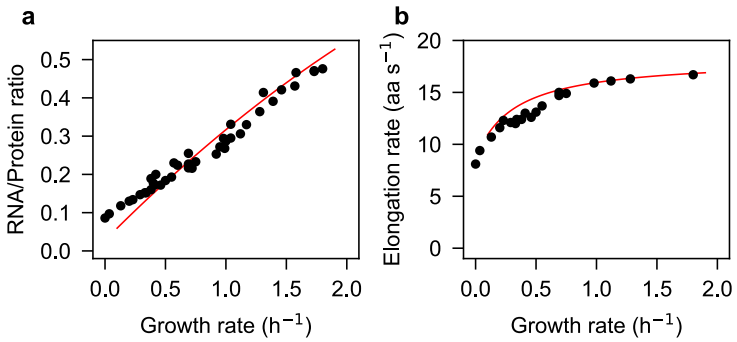


Fig. 2.5 Growth rate dependences of total RNA/protein ratio and ribosome activity. (a) Predicted total RNA concentration (mRNA + tRNA + rRNA) relative to observed total protein concentration at different cellular growth rates (red line) compared to experimental observations (Dai *et al*, 2016; Scott *et al*, 2010); R^2 = 0.97, GMFE = 1.12. **(b)** Predicted (red line) and experimentally determined (Dai *et al*, 2016) elongation rates of actively translating ribosomes (ribosome activities); R^2 = 0.93, GMFE = 1.06. At the lowest assayed growth rates, non-growth-related translation – which is not included in the model – may become comparable to growth-related translation; at these growth rates, the numerical optimization of our model did not converge (μ < 0.1 h^{-1}), and thus the red lines are not extended into this region.

The concentrations of the individual components of the translation machinery determine the average translation elongation rate (ribosomal activity), defined as the total cellular translation rate divided by the total active ribosome content (Erickson *et al*, 2017). The predicted elongation rates closely match the experimental data (Dai *et al*, 2016) over a broad range of growth rates (**Fig. 2.5b**).

Cost minimization predicts response to antibiotics

The expression of *E. coli*'s translation machinery reacts strongly to the exposure to antibiotics that inhibit the ribosome, such as chloramphenicol (Dai *et al*, 2016; Scott *et al*, 2010; Hui *et al*, 2015). The details of these changes can also be understood from our hypothesis of cost minimization. The concentrations of ribosomes and EF-Tu, the RNA/protein ratio, and the elongation rate of active ribosomes increase under chloramphenicol stress (**Fig. S2.4**); these changes partially compensate for the reduced fraction of active ribosomes. The concentration of EF-Ts instead decreases with increasing chloramphenicol concentration (**Fig. S2.4c**). EF-Ts contributes to translation by converting EF-Tu·GDP to EF-Tu·GTP, which then forms a ternary complex with

charged tRNA. Under chloramphenicol stress, fewer ternary complexes are turned over, and hence less EF-Ts is needed.

Alternative cost measures lead to similar results

The biological fitness of *E. coli* cells depends on many factors, and hence any simple assignment of fitness costs to molecules can only be approximate. The results presented so far are based on the assumption that costs are proportional to molecular masses. To test if alternative cost measures lead to consistent results, we repeated our calculations using four distinct costs that have been employed in the literature. In particular, it has been argued that the cost of protein expression lies in the process rather than the product (Stoebel *et al*, 2008), and we hence also explore synthesis costs of RNA and protein.

Across most conditions, we obtained very similar predictions for the concentrations of translation machinery components when our model assigned molecular costs based on the carbon content of the molecules (Beck *et al*, 2016), on the amount of energy (Mahmoudabadi *et al*, 2017; Weiße *et al*, 2015) spent on their production (ATP cost), or on the total investment into macromolecular catalysts (Noor *et al*, 2016) required for their production (synthesis cost) (**Figs. S2.5, S2.6**; we estimated ATP costs based on Refs. (Akashi & Gojobori, 2002; Lynch & Marinov, 2015), and calculated synthesis costs using flux balance analysis with molecular crowding (Adadi *et al*, 2012; Gelius-Dietrich *et al*, 2013; Desouki, 2016), see Methods). All components whose concentrations we predict consist of protein, RNA, or both, and all costs examined are approximately proportional to the lengths of RNA and protein molecules. Thus, the relative costs of all components are essentially a function of the RNA/protein cost ratio r, i.e., the cost of RNA per nucleotide divided by the cost of protein per amino acid. We assume that the cost of RNA per nucleotide is identical for tRNA and rRNA; the corresponding cost ratio is broadly similar between molecular masses (r = 3.0), carbon content (r = 2.0), ATP cost (r = 1.6), and synthesis costs (r = 1.7-2.1) across minimal growth conditions ($\mu < 1$ h^{-1}; **Fig. S2.7**). In contrast, assuming that costs are proportional to only the protein content of the molecular assemblies (Ehrenberg & Kurland, 1984; Klumpp *et al*, 2013) results in an RNA/protein cost ratio of zero. Predictions based on protein costs hence overestimate mRNA and tRNA concentrations (which cost nothing), resulting in corresponding underestimates of EF-Ts and especially EF-Tu concentrations (**Figs. S2.5, S2.6**).

We note that in rich medium (μ = 1.9 h^{-1}), the RNA/protein cost ratio for synthesis costs is much lower than across minimal media, falling to r = 0.21 for tRNA and rRNA (**Fig. S2.7**). This results in an overprediction of the observed tRNA concentration (Dong *et al*, 1996; Skjold *et al*, 1973; Forchhammer & Lindahl, 1971) by a factor of almost 2 (**Fig. S2.6**). Moreover, while the predicted tRNA concentration is also almost twice the predicted EF-Tu concentration, experimental estimates for tRNA and EF-Tu are very similar (Dong *et al*, 1996; Skjold *et al*, 1973; Forchhammer & Lindahl, 1971; Schmidt *et al*, 2016; Furano, 1975). We conclude that if the translation machinery has been optimized for efficiency at high growth rates by natural selection, the synthesis cost of its components is unlikely to have been central to this optimization.

Conclusions

In sum, cost minimization in a mechanistic bottom-up model of optimal translation efficiency, fully parameterized with known kinetic constants and molecular masses, accounts quantitatively for the concentrations of all molecule species involved without any adjustable parameters. The optimal concentrations of different components change differentially with growth rate, explaining the observed scaling of *E. coli*'s translation machinery composition, RNA composition, and

elongation rate. At least for the translation machinery part of the cellular economy, whose components consist largely of protein and RNA, approximate cost measures appear to be sufficient: several alternative cost measures provided predictions very similar to those based on molecular masses, emphasizing the generality of our findings.

We conclude that *E. coli*'s translation machinery works close to optimal cost efficiency. Accordingly, our findings are consistent with the hypothesis that natural selection has minimized a cost function similar to those examined here. Our results further support the idea that phenomenological growth laws of proteome composition (Scott *et al*, 2010; Klumpp *et al*, 2013; Hui *et al*, 2015; Dai *et al*, 2016) may have their root in the costs associated with the non-protein molecules involved in particular processes, and that their explicit inclusion in systems biology models of cellular growth (Klumpp *et al*, 2013; O'Brien *et al*, 2013; Goelzer *et al*, 2015; Tadmor & Tlusty, 2008) may eventually allow these models to abandon any reliance on phenomenological parameters.

Methods

Experimental concentrations of ribosomes, EF-Tu, and EF-Ts

We used molar concentrations (μM) in the model; thus, all experimental data were converted to molar concentrations. We first calculated the total protein density during exponential growth on a glucose minimal medium at growth rate μ = 0.58 h^{-1}. In this condition, the total protein mass per cell is 280 fg (Supplementary Note 3 in Ref. (Schmidt *et al*, 2016)), and cell volume is 1.90 fL (the cell volume 2.84 fL modified by a factor of 0.67 according to Supplementary Note 3 in Ref. (Schmidt *et al*, 2016)). Accordingly, the protein mass density on glucose is $\rho_{p,glc}$ = 147.15 $g \cdot L^{-1}$.

We then fitted a second-order polynomial function $\phi(\mu)$ to the fraction of total protein in dry mass provided in Ref. (Bremer & Dennis, 1996) across different growth rates μ. With $\phi(\mu)$, $\rho_{p,glc}$ = 147.15 $g \cdot L^{-1}$ at μ = 0.58 h^{-1}, and the observed constant dry mass density of *E. coli* across growth conditions(Nanninga & Woldringh, 1985; Kubitschek *et al*, 1984), we obtained the condition-specific total protein concentration, ρ_p, for all other growth conditions based on the respective observed growth rates.

For a given growth rate, μ', the total protein concentration $\rho_{p,\,\mu=\mu'}$ is given by

$$\rho_{\mathrm{p},\mu=\mu'} = \frac{\phi(\mu=\mu') \cdot \rho_{\mathrm{p,glc}}}{\phi(\mu=0.58)} \tag{2.1}$$

With the measured fraction of each protein ($f_{p,i}$) in the proteome (Schmidt *et al*, 2016), the molar concentration of each protein was then calculated as

$$c_i = \frac{\rho_{\mathrm{p}} \cdot f_{p,i}}{\mathrm{MW}_{p,i}} \quad , \tag{2.2}$$

where $\mathrm{MW}_{p,i}$ is the molecular weight of *protein* i in $g \cdot mole^{-1}$. The ribosome concentration was calculated as the arithmetic mean of the molar concentrations of all ribosomal proteins.

Experimental concentration of active ribosome

Active ribosomes are defined here as ribosomes engaged in peptide elongation. Dai *et al.* estimated the fraction of active ribosomes, f_{active}, in *E. coli* at different growth rates (Dai *et al*, 2016). We fitted a Michaelis-Menten type equation to their data, resulting in $f_{active} = \mu$ / (0.124 +

μ). For each total ribosome concentration $c_{ribosome}$ in Ref. (Schmidt *et al*, 2016), we then estimated the corresponding active ribosome concentration as $c_{active\text{-}ribosome} = f_{active} \cdot c_{ribosome}$.

Experimental concentrations of GTP and GDP

GTP and GDP concentrations are from Ref. (Bennett *et al*, 2009). We chose the data for growth on glucose for all simulations (See Text S2.3 for details).

Experimental concentration of mRNA

mRNA concentration was calculated from the data given in Ref. (Valgepea *et al*, 2013) as the ratio of mRNA copy number per cell and cell volume. To estimate the mRNA concentrations at the growth rates shown in Fig. 2.2 and Fig. S2.1, we fitted a second order polynomial to the mRNA concentration as a function of growth rate; we then read off the values at the required growth rates. mRNA concentrations were assayed only at growth rates between 0.11 h^{-1} and 0.49 h^{-1}, and we did not attempt to extrapolate values beyond this range.

Experimental concentration of tRNA

We collected three independent datasets of tRNA concentrations. Dataset1 (Dong *et al*, 1996) contains tRNA concentration for each individual tRNA, whereas both dataset2 (Forchhammer & Lindahl, 1971) and dataset3 (Skjold *et al*, 1973) contain only total tRNA concentrations. In each of these experiments, tRNA abundance was measured as the ratio of tRNA to ribosomal RNA (rRNA). We scaled these values to absolute tRNA concentrations assuming that the rRNA concentration corresponds to the ribosome concentration estimated from the proteomics data (see the subsection "Experimental concentrations of ribosomes, EF-Tu, and EF-Ts"). To estimate the tRNA concentrations at the growth rates shown in Fig. 2.2 and Fig. S2.1, we used the same fitting procedure as for mRNA.

Concentrations of individual tRNAs and relationship to model

Our model differentiates tRNAs by their anticodons (see below for details). Thus, 40 tRNAs were used to represent all elongator tRNAs. The tRNAs modeled in this work are listed in Table S2.1 together with their common names used in dataset1 (Dong *et al*, 1996) and their gene IDs.

In the experiments by Dong et al. (dataset1) (Dong *et al*, 1996), tRNAs were classified into 41 distinct sets based on two-dimensional polyacrylamide gel electrophoresis. We combined two tRNA sets corresponding to different tRNA weights if they have the same anticodon (*i.e.*, the pairs of Val2A + Val2B, Thr1 + Thr3, and Tyr1 + Tyr2). The experimenters could not distinguish between the tRNAs Gly1 and Gly2, as these have very similar molecular weights and isoelectric point; the same was true for Ile1 and Ile2. We estimated the individual concentrations of these four tRNAs based on the ratios 3:2 between Gly1 and Gly2 and 20:1 between Ile1 and Ile2 observed by Ikemura *et al.* (Ikemura, 1981).

To estimate the tRNA concentrations at the growth rates shown in Fig. 2.2 and Fig. S2.1, we fitted the concentration of each tRNA in dataset1 (measured for growth rates ranging from 0.28 h^{-1} to 1.73 h^{-1}) to a second order polynomial of growth rate and extended this function to the required range, 0.12 h^{-1} to 1.9 h^{-1}.

Combination of tRNAs that are predicted to be non-expressed

The predicted concentrations of 6 tRNAs are 0 μM; these are highlighted in red in Table S2.2. This result is a straightforward consequence of the model structure. The relationship between codons

and tRNAs is not one-to-one in *E. coli*. Consider a given codon (codon1) that has more than one cognate tRNA, say, tRNA1 and tRNA2. If tRNA2 is also the cognate tRNA of another codon (codon2), the predicted concentration of tRNA1 will be zero: for the same "price" (the same contribution to the limited total mass concentration), tRNA2 can service two codons, while tRNA1 can service only one. For example, codon GGG has two cognate tRNAs, gly1 and gly2; gly2 is also the cognate tRNA of codon GGA. Thus, both gly1 and gly2 can translate GGG, but gly2 can translate GGA, too, and is thus more valuable to the cell if we assume that both tRNAs are processed equally efficiently, as done in the model. Therefore, the predicted optimal concentration of gly1 will be zero (note that this might not occur in models that consider different ribosomal k_{cat} or K_m values for the two tRNAs). To compare our predictions to the experimental data (Dong *et al*, 1996), we combined tRNAs with predicted zero concentration with their co-functioning tRNAs in both the predictions and the experimental data. The resulting six combined tRNA pairs are: GLy1 + Gly2; Leu1 + Leu3; Leu4 + Leu5; Pro1 + Pro3; Ser2 + Ser1; Thr2 + Thr4. In all reported figures, the total number of tRNAs shown is thus 34.

Concentrations of amino acids and total protein

We first calculated the molecular weight of protein (MW_p) at the given condition. The molar concentration of protein, $c_{protein}$, is given by

$$c_{\mathrm{protein}} = \frac{\rho_{\mathrm{p}}}{\mathrm{MW_p}} \tag{2.3}$$

where ρ_p is the mass concentration of total protein at the given condition (equation (2.1)).

The concentration of amino acids encoded by *codon-i* is given by:

$$c_{codon-i} = f_{codon-i} \cdot L_{\mathrm{protein}} \cdot c_{\mathrm{protein}} \tag{2.4}$$

where $L_{protein}$ is the abundance-weighted average protein length at the given condition; $f_{codon\text{-}i}$ is the frequency of codon *i* in the genome, where each gene is weighted by its relative abundance in the proteome.

Note that for an amino acid AA_j encoded by multiple synonymous codons, $c_{codon-i}$ is not the total concentration of AA_j in cellular proteins, but only of the fraction encoded by $codon - i$; the total concentration of AA_j is obtained by summing the $c_{codon-i}$ values for all synonymous codons for AA_j.

For simulations under chloramphenicol stress, $c_{protein}$ and $c_{codon\text{-}i}$ are not available. We approximated their values by the corresponding concentrations for growth on glucose in the absence of the antibiotic.

Mass fraction of translation machinery in total dry weight

The dry mass fraction of the translation machinery shown in Fig. S2.3 includes ribosome, mRNA, charged tRNAs, EF-Tu, and EF-Ts; it does not include GDP, GTP, free tRNA, tRNA-synthetases, and elongation factor G (FusA). We converted from protein fractions to mass fractions of total dry weight using the relationship between total protein mass and dry mass discussed in the subsection "Experimental concentrations of ribosomes, EF-Tu, and EF-Ts".

Experimental estimate. The mass fraction of the translation machinery in total dry weight is the sum of two parts: (1) protein and (2) RNA.

(1) We calculated the mass fraction of translational proteins (including ribosomal protein, EF-Tu, and EF-Ts) in dry weight from the proteomics data in Ref. (Schmidt *et al*, 2016).

(2) We fitted the reported total RNA/protein ratio in Refs. (Scott *et al*, 2010; Dai *et al*, 2016) to a second order polynomial of growth rate. We then used this fitted function to calculate the RNA/protein ratio at the growth rates assayed by Schmidt *et al.* (Schmidt *et al*, 2016), and multiplied this ratio with the dry mass fraction of protein.

Theoretical prediction. The predicted dry mass fraction of the translation machinery was the ratio of the total mass concentration of the translation machinery (including free ribosome, active ribosome, EF-Tu, EF-Ts, charged tRNA, and mRNA) to the total dry mass density. The total dry mass density was estimated as $\rho_{p,glc}$ / ϕ(0.58 h^{-1}) = 147.15 g·L^{-1} / 0.631 = 233.30 g·L^{-1}. For the prediction including de-activated ribosome concentrations (dashed line in Fig. S2.3), we added estimates of de-activated ribosome concentrations according to Fig. 2.4 of the main text.

Molecular weights

Molecular weights of ribosome, charged tRNAs (aa-tRNAs), EF-Tu, EF-Ts, and the ternary complexes (TC, EF-Tu·GTP·aa-tRNA) were calculated from their sequences. The stoichiometry of ribosomal proteins and RNAs in the ribosome was obtained from the EcoCyc database (Keseler *et al*, 2017); the stoichiometry of all components is 1 except for RplL, for which it is 4.

We used an average mRNA to represent the total mRNA. The molecular weight of an average mRNA ($\mathrm{MW}_{\mathrm{mRNA}}$) is the sum of two parts: (1) the molecular weight of the coding sequence (CDS) of mRNA ($\mathrm{MW}_{\mathrm{mRNA\text{-}CDS}}$), which was calculated from protein-expression-weighted mRNA length and nucleotide composition of *E. coli* protein-coding sequences in each growth condition (Schmidt *et al*, 2016); (2) the weight of the untranslated region (UTR) of mRNA ($\mathrm{MW}_{\mathrm{mRNA\text{-}UTR}}$), which was calculated from the average nucleotide composition of the genome and the typical length of UTR. The length of UTR was assumed to be 85 nt, a typical length of the untranslated region in *E. coli* (Kim *et al*, 2012). Thus,

$$\mathrm{MW}_{\mathrm{mRNA}} = \mathrm{MW}_{\mathrm{mRNA-CDS}} + \mathrm{MW}_{\mathrm{mRNA-UTR}}. \quad (2.5)$$

In the simulations of translation under antibiotic stress, the molecular weight of chloramphenicol was set to 0.

Alternative costs of translation machinery components

The hypothesis underlying our analysis is that the components of the translation machinery are expressed to minimize the total cost of translation at a given protein production rate. For the calculations underlying the figures of the main text, we assumed that molecular costs are proportional to molecular mass concentrations. To test alternative cost measures proposed in the literature, we estimated the costs of each translation machinery component in terms of (1) its carbon content (Beck *et al*, 2016) (**carbon cost**); (2) the total number of high-energy phosphate bonds required for its production (**ATP cost**); (3) the total enzyme mass required for its production (**synthesis cost**) (Noor *et al*, 2016); and (4) its protein content (Ehrenberg & Kurland, 1984; Klumpp *et al*, 2013) (**protein content**). Some of these cost measures are condition-dependent. The estimated costs are provided in Data S2.2; the RNA/protein cost ratios are compared in Fig. S2.7 and listed in Table S2.3. All costs were estimated per component (i.e., per molecule or per macromolecular complex). Note that all components whose costs are considered in the model consist of protein, RNA, or both.

Carbon cost as an alternative cost measure

The carbon cost of a component is its total number of carbon atoms.

ATP cost as an alternative cost measure

The ATP cost of a component is the number of high-energy phosphate bonds (denoted ~P) that were invested into its production. The ATP costs includes (1) the ATP invested into the synthesis of the precursors (nucleoside triphosphates or amino acids) and (2) the ATP cost of polymerization during RNA transcription or protein translation.

The ATP cost of amino acid synthesis. For cells growing on minimal carbon media, the ATP cost of amino acid production was obtained from Ref. (Akashi & Gojobori, 2002). Since the ATP costs of a given amino acid are very similar across minimal media with different carbon sources (Akashi & Gojobori, 2002), we used the ATP costs for amino acid synthesis on glucose for all minimal media considered. For *E. coli* growing on glycerol + amino acids (Schmidt *et al*, 2016), the ATP production cost for amino acids was assumed to be zero.

The polymerization cost of protein. The polymerization cost of protein is 4 ATP per amino acid: two ATP for tRNA charging, 1 ATP for EF-Tu in elongation, and 1 ATP for elongation factor G (EF-G) in elongation.

The ATP cost of NTP synthesis. For cells growing on minimal carbon media, the *de novo* synthesis cost of NTP was obtained from Ref. (Lynch & Marinov, 2015). The glycerol + amino acids medium used in the proteomics study (Schmidt *et al*, 2016) also contains adenine and uracil; here, we assumed that the synthesis of ATP and GTP starts from adenine and that the synthesis of UTP and CTP starts from uracil. PRPP (5-phospho-α-D-ribose 1-diphosphate), whose production consumes 29 ATP (Akashi & Gojobori, 2002), was considered to be the donor of ribose to the synthesis of NTPs. Finally, we estimated the total energy (~P) costs in the glycerol + amino acids medium for ATP, UTP, CTP, and GTP as 31, 31, 29, and 32, respectively. We did not attempt to estimate the ATP cost of NTP synthesis in the LB condition, as it is not clear to what extent NTPs are taken up from the medium.

The polymerization cost of RNA. The polymerization cost of RNA is 0 ATP per base, as no high-energy phosphate bonds beyond those of the polymerized NTPs are required.

The degradation cost of mRNA. The degradation rates of tRNA and rRNA are much lower than their production rates, and hence we did not account for their degradation. In contrast, mRNA is degraded much more quickly than tRNA and rRNA, and we thus considered the influence of degradation on the mRNA polymerization cost. At steady state, all degraded mRNA (in the form of nucleoside mono-phosphates, NMPs) is assumed to be recycled to re-transcribe mRNA. These recycled NMPs require two ~P to form NTPs, so the mRNA recycling cost ($\text{cost}_{\text{NTP-deg}}$) is 2 ~P per NTP. At a given mRNA concentration $c_{\text{mRNA-NTP}}$ (in units of NTPs built into mRNA), the production of mRNA must offset the combination of mRNA degradation and mRNA dilution by cellular growth at rate μ. The rate of ATP consumption for this production is thus given by

$$v_{\text{cost-ATP-mRNA}} = \mu \cdot \text{cost}_{\text{NTP}} \cdot c_{\text{mRNA-NTP}} + k_{\text{deg}} \cdot \text{cost}_{\text{NTP-deg}} \cdot c_{\text{mRNA-NTP}} \qquad (2.6)$$

where cost_{NTP} is the synthesis cost of NTPs (estimated above, "*The ATP cost of NTP synthesis"*) and k_{deg} is the mRNA degradation rate constant. k_{deg} *is* calculated from mRNA half-life (t_{half}), $k_{\text{deg}} = \ln(2) / t_{\text{half}}$, with t_{half} = 5 min for all growth conditions (Bernstein *et al*, 2002). To obtain the ATP cost per NTP in mRNA, we must divide this rate by $\mu \cdot c_{\text{mRNA-NTP}}$:

$$\mathrm{cost}_{\mathrm{NTP-mRNA}} = \frac{v_{\mathrm{cost-ATP-mRNA}}}{\mu \cdot c_{\mathrm{mRNA-NTP}}} = \mathrm{cost}_{\mathrm{NTP}} + \frac{k_{\mathrm{deg}}}{\mu} \cdot \mathrm{cost}_{\mathrm{NTP-deg}} = \mathrm{cost}_{\mathrm{NTP}} + \frac{2\,k_{\mathrm{deg}}}{\mu} \quad (2.7)$$

Thus, mRNA degradation dominates the ATP cost of mRNA at very low growth rates, but becomes insignificant at growth rates higher than the mRNA degradation rate.

Synthesis cost as an alternative cost measure

The synthesis cost is the total macromolecular dry mass, which includes transporters, enzymes, RNA polymerase, and ribosome, that is needed to synthesize each component of the translation machinery. To estimate the synthesis cost of each component, we first estimated the macromolecular dry mass that is needed to synthesize one millimole of amino acid (cost_{AA}) and one millimole of nucleotide (cost_{nucl}). Based on these estimates, the synthesis cost of a protein is $\mathrm{cost}_{protein} = L_{protein} \cdot \mathrm{cost}_{AA}$, where $L_{protein}$ is the protein length in amino acids, and the synthesis cost of an RNA molecule is $\mathrm{cost}_{RNA} = L_{RNA} \cdot \mathrm{cost}_{nucl}$, where L_{RNA} is the length of the RNA molecule in nucleotides.

We calculated cost_{AA} and cost_{nucl} using ccFBA (Gelius-Dietrich *et al*, 2013; Desouki, 2016), which is an implementation of the MOMENT (Adadi *et al*, 2012) algorithm for flux balance analysis with molecular crowding, featuring an improved treatment of co-functional enzymes. Briefly, ccFBA assigns each enzyme a constant catalytic rate (k_{cat}) and molecular weight, and then finds the flux distribution that maximizes biomass production while not exceeding a threshold on the total enzyme mass.

Synthesis cost of protein. We first added a protein synthesis reaction to the iML1515 model implemented in ccFBA. The stoichiometric coefficient of each amino acid consumed in this reaction was set to its proportion in the biomass reaction of the iML1515 model. In *E. coli*, there are approx. 9 TCs per ribosome at high growth rates (Dong *et al*, 1996; Furano, 1975); we thus designated the ribosome plus 9 ternary complexes (TCs) as the "enzyme" of the protein synthesis reaction. We parameterized this "enzyme" with k_{cat} = 22 s^{-1} and molecular weight = 2933.241 kD (Bremer & Dennis, 1996). We then set the objective function to the rate of protein production, v_{AA}, instead of the biomass production rate v_{bio}. We simulated different growth conditions by only allowing the model to import nutrients available in the respective medium, setting the lower bound of the corresponding exchange reactions to -1000 ($\mathrm{mmol \cdot g_{DW}^{-1} \cdot h^{-1}}$). For cells growing on LB medium, the lower bound of all exchange reactions was set to -1000, i.e., all metabolites for which there is a transporter in the model can be taken up. We maximized the protein synthesis rate v_{AA}, given a limit on the dry mass fraction of macromolecules involved in metabolism (a "budget") of C=0.27. As the lower bound chosen for the exchange rates was very high, the optimizations were constrained by C, i.e., the optimal v_{AA} value is the maximal rate of protein production with this macromolecular budget. Accordingly, the synthesis cost of amino acids is cost_{AA} = 0.27 / v_{AA}, expressed as the dry mass fraction required to produce 1 $\mathrm{mmol \cdot g_{DW}^{-1} \cdot h^{-1}}$ of amino acids. The computed protein synthesis costs for all 20 conditions considered are shown in Table S2.4.

Synthesis cost of stable RNA. To estimate the cost per nucleotide of stable RNA production, cost_{nucl}, we implemented an analogous algorithm. We first added an RNA synthesis reaction to the iML1515 model (Monk *et al*, 2017) and set this reaction as the objective function. The stoichiometry of NTPs consumed in this reaction was set to the corresponding fractions of NTPs in the biomass reaction of the iML1515 model. We designated the RNA polymerase (molecular weight = 389.11 kD) as the enzyme catalyzing this reaction. For stable RNA (tRNA and rRNA) synthesis, the turnover rate of the RNA polymerase (RNA-P) (Bremer & Dennis, 1996) is $k_{RNAP\text{-}sRNA}$

= 85 s^{-1} . We maximized the flux of the RNA synthesis reaction, v_{sRNA}, constrained by the macromolecular budget C = 0.27. The synthesis cost of stable RNA per nucleotide was then calculated as $cost_{nucl\text{-}sRNA} = 0.27 / v_{sRNA}$, expressed as the dry mass fraction required to produce 1 $mmol \cdot g_{DW}^{-1} \cdot h^{-1}$ of NTP in stable RNA. The computed stable RNA synthesis costs for all 20 conditions considered are listed in Table S2.4.

Synthesis cost of mRNA. In *E. coli*, mRNA transcription is slower than stable RNA transcription. The turnover rate of RNA-P for mRNA is $k_{RNAP\text{-}mRNA} = 66\ s^{-1}$ (i.e., the fastest rate of mRNA transcription has been observed to be about three times the maximal translation rate) (Proshkin *et al*, 2010). As for the calculation of the ATP cost of mRNA, we need to account for mRNA degradation at rate k_{deg} (which is much faster than the degradation of stable RNA). We again assumed that nucleotides from degraded mRNA are re-used by RNA-P to synthesize mRNA. As before, we note that at a given mRNA concentration c_{mRNA}, the production of mRNA must offset the combination of mRNA degradation and mRNA dilution by cellular growth at rate μ. The rate of mRNA production by the RNA-P is thus

$$v_{mRNA} = \mu \cdot c_{mRNA} + k_{deg} \cdot c_{mRNA} \tag{2.8}$$

To obtain the concentration of RNA-P necessary to catalyze this rate, we need to divide this expression by $k_{RNAP\text{-}mRNA}$:

$$c_{\mathrm{RNAP-mRNA}} = \frac{v_{\mathrm{mRNA}}}{\mu} = \mathrm{c_{mRNA}}\left(1 + \frac{k_{\mathrm{deg}}}{\mu}\right) \tag{2.9}$$

Thus, the concentration of RNA-P required for mRNA production is larger by a factor $\left(1 + k_{\mathrm{deg}}/\mu\right)$ when accounting for mRNA degradation than it would be otherwise. To account for mRNA degradation, we thus set the effective turnover number of RNA-P to $k_{\mathrm{eff-mRNA}} = k_{\mathrm{RNAP-mRNA}}/\left(1 + k_{\mathrm{deg}}/\mu\right)$. We maximized the flux of the mRNA synthesis reaction, v_{mRNA}, constrained by the macromolecular budget C = 0.27. The synthesis cost of mRNA per nucleotide was then calculated as $cost_{nucl\text{-}mRNA} = 0.27 / v_{mRNA}$, expressed as the dry mass fraction required to produce 1 $mmol \cdot g_{DW}^{-1} \cdot h^{-1}$ of NTP in mRNA. The computed costs for all 20 conditions considered are provided in Table S2.4.

Protein content as an alternative cost measure

The protein content was considered as the only relevant cost of the translation machinery in two previous models (Ehrenberg & Kurland, 1984; Klumpp *et al*, 2013). These models accurately predicted the ribosomal protein fraction in total protein, whereas they substantially underestimated the EF-Tu proteome fraction compared to measurements data (Ehrenberg & Kurland, 1984; Klumpp *et al*, 2013).

Following these earlier works, we calculated the protein cost of a component as the number of amino acid residues it incorporates. Since mRNA and tRNA have zero protein content, this definition assigns no costs to their expression, potentially leading to a prediction of infinite concentrations. To avoid such pathological predictions, we set the protein content costs of mRNA and charged tRNAs to small, arbitrary values in the model (mRNA: 10 amino acid residues; tRNA: 5 amino acid residues).

Model overview

The mechanistic translation model encompasses the processes of translation initiation, elongation, termination, nucleotide exchange in EF-Tu, and ternary complex (TC) formation. Fig. 2.1 of the

main text illustrates the modeled reaction; for better readability, the figure shows only a subset of codon / charged tRNA combinations. In total, the model includes 274 reactions.

Translation initiation

During initiation, mRNA converts free ribosomes to active ribosomes:

$$\mathrm{Ribosome}_{\mathrm{free}} \rightarrow \mathrm{Ribosome}_{\mathrm{active}}. \qquad (\mathrm{r1})$$

Translation initiation consists of multiple elementary reactions (Milón & Rodnina, 2012). However, a recent study found that at steady state, the kinetics of initiation effectively follow Michaelis–Menten kinetics, with mRNA in the enzyme position (with concentration c_{mRNA}) and free (unbound) ribosomes in the substrate position (with concentration $c_{\mathrm{ribo-free}}$) (Borkowski *et al*, 2016),

$$v_{\mathrm{tl-init}} = k_{\mathrm{cat-mRNA}} \cdot c_{\mathrm{mRNA}} \cdot \frac{c_{\mathrm{ribo-free}}}{K_{\mathrm{M-ribo}} + c_{\mathrm{ribo-free}}}, \qquad (2.10)$$

with $k_{\mathrm{cat-mRNA}}$ = 1.33 s^{-1} and $K_{\mathrm{M\text{-}ribo}}$ = 8.5 μM from Ref. (Tadmor & Tlusty, 2008).

We assume that the turnover number of mRNA for ribosome binding ($k_{\mathrm{cat-mRNA}}$) is growth rate-independent, and hence that the observed growth rate-dependent activity is due to changes in the concentration of free ribosomes available for initiation (Espah Borujeni *et al*, 2014). Thus, we used the maximal reported mRNA activity as an estimate of $k_{\mathrm{cat-mRNA}}$.

Ternary complex formation and EF-Tu nucleotide exchange

Ternary complex formation and nucleotide exchange in EF-Tu are the processes by which the translation machinery recycles its substrates, the ternary complexes, for elongation.

TC formation. The binding of EF-Tu·GTP to charged-tRNA (aa-tRNA) forms the ternary complex (*ternary complex formation* in Fig. 2.1), which is the substrate of translation elongation.

Nucleotide exchange in EF-Tu. EF-Tu·GDP is released after the formation of a new peptide bond. Elongation factor Ts (EF-Ts) binds to EF-Tu·GDP and induces the exchange of GDP for GTP (right panel in Fig. 2.1).

The individual steps of these two processes are modelled with mass action kinetics (Gromadski *et al*, 2002; Louie & Jurnak, 1985).

In the implementation of the model, we divide each reversible reaction into two irreversible reactions. The TC formation reaction is a set of reactions that include the binding of 40 aa-tRNAs to Tu·GTP, and its rate constants depend on which amino acid is involved (Louie & Jurnak, 1985) (for the parameter values see Data S2.1.)

Elongation

Elongation is a very complex process (Rudorf *et al*, 2014; Vieira *et al*, 2016). For simplicity, we model elongation as a single reaction with an active ribosome as the enzyme and TC as the substrate, which was proposed by Klumpp *et al.* (Klumpp *et al*, 2013). In this single reaction model, TC is discriminated by the anticodon and all TCs are treated with the same activity. Michaelis–Menten kinetics are used to describe the reaction rate (Klumpp *et al*, 2013). There are 40 anticodons in total for all elongator tRNAs in *E. coli*; accordingly, our model uses 40 tRNAs to represent all tRNAs.

At steady state, the total translation rate (per cytosolic volume) of each codon remains constant. We do not model the translation of a whole protein. Instead, we decompose protein synthesis into the translation of 61 codons (see also "Modeling" below).

Active ribosome state. We distinguish active ribosomes according to their binding codons (codon in ribosome A site). Thus, there are 61 types of active ribosome in the model, distinguished in the model by subscripts indicating the codon currently presented by the ribosome:

$$\mathrm{Ribosome}_{\mathrm{active}} \rightarrow \mathrm{Ribosome}_{codon\text{-}i} \quad ; \tag{r2}$$

here, *codon-i* is one of the 61 codons and $\mathrm{Ribosome}_{codon\text{-}i}$ is the active ribosome that binds *codon-i*. This reaction is constrained by mass balance, but is considered to be instantaneous. We assume that when an active ribosome binds with a specific codon, it only translates the codon's cognate tRNA.

In the model, there are 61 codons and 40 tRNAs, and the relation between tRNA and codon is not one-to-one. Based on the number of cognate tRNAs, we partition the 61 codons into 2 classes: class 1 codons have one cognate tRNA, whereas class 2 codons have two cognate tRNAs. The lists of class 1 and class 2 codons are provided in Table S2.5.

For class 1 codons (n = 51), the elongation reaction is:

$$\mathrm{TC}_{codon\text{-}i} \xrightarrow{\mathrm{Ribosome}_{codon\text{-}i}} \mathrm{EF\text{-}Tu{\cdot}GDP} + \mathrm{tRNA}_{codon\text{-}i} + \mathrm{aa}_{codon\text{-}i}, \tag{r3}$$

where $TC_{codon\text{-}i}$ is the cognate TC of *codon-i*, $\mathrm{tRNA}_{codon\text{-}i}$ is the released free tRNA, and $\mathrm{aa}_{codon\text{-}i}$ symbolizes the amino acid that was just appended to the growing peptide. The $\mathrm{tRNA}_{codon\text{-}i}$ and $\mathrm{aa}_{codon\text{-}i}$ are included here for completeness, but are not included explicitly in the optimized model, as they do not influence the results once appropriate exchange reactions have been added. Simultaneously, the $\mathrm{Ribosome}_{codon\text{-}i}$ is converted to $\mathrm{Ribosome}_{\mathrm{active}}$, which is ready to participate in the next round of elongation,

$$\mathrm{Ribosome}_{codon\text{-}i} \rightarrow \mathrm{Ribosome}_{\mathrm{active}}. \tag{r4}$$

For simplicity, we combine r3 and r4 into the new reaction r5, with $\mathrm{Ribosome}_{codon\text{-}i}$ as a substrate and $\mathrm{Ribosome}_{\mathrm{active}}$ as a product (the same for r6 and r7),

$$\mathrm{TC}_{codon\text{-}i} + \mathrm{Ribosome}_{codon\text{-}i} \xrightarrow{\mathrm{Ribosome}_{codon\text{-}i}} \mathrm{Ribosome}_{\mathrm{active}} + \mathrm{EF\text{-}Tu{\cdot}GDP} + \mathrm{tRNA}_{codon\text{-}i} + \mathrm{aa}_{codon\text{-}i}. \tag{r5}$$

The translation rate of *codon-i* is described by Michaelis-Menten kinetics,

$$v_{tl-codon-i} = c_{ribo-codon-i} \cdot k_{\mathrm{cat-ribo}} \cdot \frac{c_{TC-codon-i}}{c_{TC-codon-i} + K_{\mathrm{M-TC}}}, \tag{2.11}$$

where $c_{ribo\text{-}codon\text{-}i}$ is the concentration of ribosomes that present *codon-i* ($\mathrm{Ribosome}_{codon\text{-}i}$), $c_{TC\text{-}codon\text{-}i}$ is the concentration of cognate TC of *codon-i* ($\mathrm{TC}_{codon\text{-}i}$), $k_{\mathrm{cat\text{-}ribo}}$ = 22 s^{-1} , and $K_{\mathrm{M\text{-}TC}}$ = 3 μM (parameters from Ref. (Klumpp *et al*, 2013)).

For class 2 codons (n = 10), the active ribosome can translate two TCs and thus there are two reactions:

$$\mathrm{TC}_{codon\text{-}i\text{-}1} + \mathrm{Ribosome}_{codon\text{-}i} \xrightarrow{\mathrm{Ribosome}_{codon\text{-}i}} \mathrm{Ribosome}_{\mathrm{active}} + \mathrm{EF\text{-}Tu{\cdot}GDP} + \mathrm{tRNA}_{codon\text{-}i\text{-}1} + \mathrm{aa}_{codon\text{-}i} \tag{r6}$$

and

$$\mathrm{TC}_{codon\text{-}i\text{-}2} + \mathrm{Ribosome}_{codon\text{-}i} \xrightarrow{\mathrm{Ribosome}_{codon\text{-}i}} \mathrm{Ribosome}_{\mathrm{active}} + \mathrm{EF\text{-}Tu{\cdot}GDP} + \mathrm{tRNA}_{codon\text{-}i\text{-}2} + \mathrm{aa}_{codon\text{-}i}. \quad \text{(r7)}$$

The translation rate of *codon-i* is the sum of these two reactions:

$$v_{tl-codon-i} = v_1 + v_2, \quad (2.12)$$

with

$$v_1 = c_{ribo-codon-i} \cdot k_{\mathrm{cat-ribo}} \cdot \frac{c_{TC-codon-i-1}}{(c_{TC-codon-i-1} + c_{TC-codon-i-2}) + K_{\mathrm{M-TC}}} \quad (2.13)$$

and

$$v_2 = c_{ribo-codon-i} \cdot k_{\mathrm{cat-ribo}} \cdot \frac{c_{TC-codon-i-2}}{(c_{TC-codon-i-1} + c_{TC-codon-i-2}) + K_{\mathrm{M-TC}}}. \quad (2.14)$$

Termination

In termination, an active ribosome is converted to a free ribosome:

$$\mathrm{Ribosome}_{\mathrm{active}} \rightarrow \mathrm{Ribosome}_{\mathrm{free}}. \quad \text{(r8)}$$

The termination rate is equal to the protein synthesis rate at steady state:

$$v_{\mathrm{term}} = v_{\mathrm{protein-syn}} = \mu \cdot c_{\mathrm{protein}}, \quad (2.15)$$

where c_{protein} is the absolute concentration of protein measured experimentally.

Modeling exchange reactions

Besides the reactions mentioned above, we also add exchange reactions that allow the influx of free ribosome, charged tRNA, mRNA, EF-Tu, EF-Ts, and GTP into the system. We also add exchange reactions that allow efflux of GDP out of the system.

Modeling antibiotic stress

To model chloramphenicol (cm) stress, we add the exchange reaction for chloramphenicol. All forms of ribosome can be inhibited by chloramphenicol:

$$\mathrm{Ribosome} + \mathrm{cm} \leftrightarrow \mathrm{Ribosome \cdot cm}, \quad \text{(r9)}$$

where Ribosome includes both free and active ribosomes. The reaction rate is described by mass action kinetics with k_{on} = 0.00057 μM·s^{-1} (0.034 μM·min^{-1}) and k_{off} = 0.0014 s^{-1} (0.084 min^{-1}) (Harvey & Koch, 1980).

For simplicity, we make the following assumptions:

(1) chloramphenicol diffuses freely across cell membranes, such that the intracellular concentration of free chloramphenicol is the same as that in the medium;

(2) chloramphenicol bound to an active ribosome (ribosome·cm) causes the ribosome·cm complex to dissociate quickly from the mRNA, thus not affecting further translation of the mRNA.

We assume that the reaction of ribosome and chloramphenicol binding is at steady state (dynamic equilibrium), and thus the active ribosome concentration will be constant. However, not all active ribosomes will be able to successfully finish translation, as the active ribosome can be inhibited by chloramphenicol during translation. Thus, to estimate the production rate of functional proteins, we need to estimate the probability that the active ribosome can finish translation without chloramphenicol inhibition.

To calculate this probability, we use the method proposed in Ref. (Dai *et al*, 2016). The probability of chloramphenicol binding to a ribosome in a given time unit is:

$$k_{\mathrm{hit}} = k_{\mathrm{on}} c_{\mathrm{cm}}, \tag{2.16}$$

where k_{on} is the binding constant. The probability that the ribosome is bound n times in the time interval t follows the Poisson distribution

$$P(n) = e^{-k_{\mathrm{hit}} t_{\mathrm{tl}}} \frac{k_{\mathrm{hit}} t_{\mathrm{tl}}}{n!}, \tag{2.17}$$

where t_{tl} is the experimental measured translation time of the translated gene. The probability that the ribosome can finish translation without inhibition by chloramphenicol is $P(0)$:

$$P_{\mathrm{tl}} = e^{-k_{\mathrm{hit}} t_{\mathrm{tl}}}. \tag{2.18}$$

Here, t_{tl} is the time for *LacZ* and t_{tl} = 72 s (from Ref. (Dai *et al*, 2016)). For simplicity, we assume that all codon-presenting active ribosomes (61 forms of active ribosome) have the same P_{tl}, and thus the effective concentration of the ribosome presenting *codon-i* is

$$c_{ribo-eff-i} = P_{\mathrm{tl}} c_{ribo-i}. \tag{2.19}$$

Under inhibition by chloramphenicol, this effective ribosome concentration of *codon-i* ($c_{ribo\text{-}eff\text{-}i}$) replaces the ribosome concentration of ($c_{ribo\text{-}i}$) in the model.

Model optimization

We assume translation at steady state and use a constraint-based optimization model. The constraints are given by the above equations and by the requirement of a given total cellular rate of protein synthesis (estimated as the product of growth rate and experimental proteome composition at this growth rate). For simplicity, we decomposed protein synthesis into the translation of 61 codons, and so the constraint on protein synthesis rate is implemented as 61 individual equations, each representing the translation rate of one codon. At steady state, the codon translation rate equals the dilution rate of amino acids incorporated at protein positions encoded by *codon-i*, i.e., the growth rate μ multiplied by the concentration of amino acids coded by *codon-i* in proteome data:

$$v_{tl-codon-i} = \mu \cdot c_{codon-i} \ ; \tag{2.20}$$

note that if multiple codons encode the same amino acids, then the total concentration of that amino acid in cellular proteins is the sum over the $c_{codon-i}$ values for the individual codons.

Given these constraints, we minimize the total mass concentration of the modeled translation apparatus, consisting of ribosome, EF-Tu, EF-Ts, mRNA, GTP, GDP, and charged tRNAs (aa·tRNA),

$$\sum_{m \in C} \mathrm{MW}_m \cdot c_m \quad , \tag{2.21}$$

where the molecule types together form the set *C*, MW_m is the molecular weight, and c_m is the concentration of molecule type *m*. We do not minimize the contributions of GTP and GDP, who participate in multiple other cellular processes (Feist *et al*, 2007) and whose concentrations are thus unlikely to be dominated by translation; their concentrations are consequently fixed to experimentally observed values in our simulations (see Text S2.3).

Let *R* be the set of reactions that together comprise translation. We also consider the dilution of the molecules involved in these reactions due to cellular volume growth at rate μ:

$$\mathbf{S} \cdot \mathbf{v}(\mathbf{c}) - \mu \cdot \mathbf{c} = 0, \tag{2.22}$$

where **S** is the stoichiometric matrix for the reactions in *R*, **c** is a vector of the concentrations c_m, and $\mathbf{v}(\mathbf{c})$ is the corresponding vector of reaction rates v_i, with the concentration-dependent kinetics described above.

Thus, we solve the non-linear constrained optimization problem:

$$\min_{\mathbf{c}} \quad \sum_{m \in C} \mathrm{MW}_m \cdot c_m \,. \tag{2.23}$$

subject to:

$$\mathbf{S} \cdot \mathbf{v}(\mathbf{c}) - \mu \cdot \mathbf{c} = 0,$$

$$v_{tl-codon-i} = \mu \cdot c_{codon-i} \quad \text{for } i\text{=1, …, 61}$$

$$v_{\mathrm{term}} = \mu \cdot c_{\mathrm{protein}}$$

We formulated the optimization problem in GAMS and used the BARON global solver (Tawarmalani & Sahinidis, 2005) on NEOS Server (Czyzyk *et al*, 1998) with 8 hours as the time limit to solve this problem. Because the problem is non-linear, it is not clear *a priori* if it is convex, in which case only a single optimum would exist. The problem is conceptionally similar to the one studied by Noor *et al.* (Noor *et al*, 2016) , and thus convexity is conceivable. A first optimum was returned by the solver within at most a few minutes for all optimizations performed. However, to guard against the existence of overlooked alternative optima, we allowed the search to continue for a total of 8 hours in each case. In addition, we repeated all simulation with the global non-linear solver in Lingo 13 (LINDO Systems, Inc., https://www.lindo.com/index.php/products/lingo-and-optimization-modeling). No alternative optima were ever found; we thus have no evidence of non-convexity.

To assess the alternative costs, the molecular weight of each component was replaced by the corresponding cost measure (see the section "*Alternative costs of translation machinery components*"). The costs for each component are shown in Data S2.2, and the corresponding predictions are shown in Data S2.3.

Code availability

The optimization problem, including the model and its parameterization, is provided as an SBML file (Data S2.4) and as a GAMS input file (Data S2.5, with protein production requirements set to those for growth on minimal glucose medium). In addition, the model has been submitted to Biomodels (MODEL2006210001; https://www.ebi.ac.uk/biomodels/MODEL2006210001).

Supplementary information

Supplementary Data and **Tables** are available online at https://doi.org/10.1038/s41467-020-18948-x.

Supplementary Texts

Text S2.1. The coarse-grained optimization models by Scott et al. and Klumpp et al.

In Ref. (Scott *et al*, 2014), Scott *et al.* use a phenomenological model based on the „bacterial growth laws" most prominently described in Ref. (Scott *et al*, 2010). This model considers two proteome sectors, a ribosomal sector with proteome fraction Φ_R and a metabolic sector with proteome fraction Φ_P, responsible for the production of the amino acids consumed by the ribosomal sector. A constraint relates these two proteome sectors to the maximally available proteome fraction for protein synthesis, Φ_R^{max}, which is assumed to be constant: $\Phi_R + \Phi_P = \Phi_R^{max}$. Thus, the model of Scott *et al.* has only one free parameter, the proteome fraction allocated to the ribosomal sector, Φ_R. For a given Φ_R, the growth rate (which is defined by the rate of protein production) is set by three phenomenological parameters: the "translational efficiency" γ (translation rate per ribosomal proteome fraction), assumed to be constant; the "nutritional efficiency" ν (amino acid production rate per metabolic proteome fraction), assumed to be condition-dependent; and a constant proteome fraction of inactive ribosomes, Φ_R^{min}. Maximizing the growth rate under these constraints results in an optimal ribosomal proteome fraction.

The approach by Scott *et al.* aims to answer broadly the same question as explored in the present work: given that cellular resources are limited, what is the optimal way to allocate them in order to allow fast growth? Scott *et al.* approach this question by maximizing the growth rate while assuming constant translational efficiency γ and fraction of inactive ribosomes Φ_R^{min}. The parameters are derived from fits to coarse-grained experimental data. With this approach, Scott *et al.* show that under relatively simple assumptions, an optimal allocation of proteome mass to translation and metabolism exists, and the relationship between the ribosomal proteome fraction and the growth rate is qualitatively similar to that observed experimentally, i.e., is linear. While in subsequent publications of the same group, the model of Scott *et al.* has been shown to be very powerful at explaining growth-related phenomena, it requires parameters fitted to experimental data, and the mechanistic basis of its components are unclear. In particular, there is no clear explanation for the existence and size of the "offset" of the ribosomal proteome fraction at zero growth rate in this model, Φ_R^{min}.

While our model minimizes the cost of translation rather than maximizing growth rate, our approach is mathematically equivalent to a maximization of growth rate under a constraint on the total cost and under certain additional assumptions (such as a constant amino acid composition of the proteome across growth rates). Both our approach and that of Scott *et al.* vary some condition-dependent parameters (the nutritional efficiency for Scott *et al.*, the proteome mass

and composition in our manuscript) and then optimize an aspect of cellular resource allocation. However, in contrast to Scott *et al.*, we are not interested in the relative global resource allocation between translation and biosynthesis based on a schematic, coarse-grained model, but in a mechanistic explanation of the quantitative pattern of resource allocation across different components of the translation machinery.

Building on the same phenomenological bacterial growth laws (Scott *et al*, 2010) as Scott *et al.* (Scott *et al*, 2014), Klumpp *et al.* (Klumpp *et al*, 2013) also analysed the composition of the translation machinery. Noting that this machinery includes not only ribosomes, but also other highly expressed proteins – most notably elongation factors (Schmidt *et al*, 2016) and tRNA synthetases – Klumpp *et al.* argued that a full appreciation of the efficiency of protein synthesis requires the inclusion of the cost of these translation components. They extended the phenomenological, coarse-grained model of Ref. (Scott *et al*, 2010) into four proteome sectors, including a ribosomal (Rb) and a translation-associated (T) sector. Assuming co-regulation of the Rb and T sectors and fitting three phenomenological constants to the data, they were able to approximate the growth rate dependence of ribosome concentration and elongation speed in *E. coli* (Klumpp *et al*, 2013). However, the experimentally observed ratio between the protein concentrations in the T- and Rb-sectors, Φ_T/Φ_{Rb} , deviates from the postulated constant ratio (see Fig. 2.3D in Ref. (Klumpp *et al*, 2013)), indicating shortcomings of this phenomenological theory.

Klumpp *et al.* also attempted to determine an optimal growth rate dependence of the ratio between T- and Rb-sectors, by treating both proteome fractions as independent parameters when numerically optimizing the growth rate of their coarse-grained model cell. However, the results predicted a ratio Φ_T/Φ_{Rb} that was substantially smaller than that observed (see Fig. 2.4C in Ref. (Klumpp *et al*, 2013)), indicating that translation in *E. coli* is either not organized optimally, or that the objective optimized by natural selection differs from the proteome allocation examined by Klumpp *et al.*. Comparing the objective functions used by Klumpp *et al.* and in the present work, we note that proteins make up 1/3 of the ribosome, but 2/3 of the ternary complex (by mass). Thus, the ternary complex appears much more expensive to the cell when considering protein mass than when considering total mass, explaining why optimization of protein allocation results in smaller predictions of the Φ_T/Φ_{Rb} ratio (Klumpp *et al*, 2013; Ehrenberg & Kurland, 1984).

Text S2.2. Ribosome states

The ribosome is the most central component of translation, and the ribosome states in our model are slightly different from those used in proteome partitioning models (Scott *et al*, 2010; Klumpp *et al*, 2013; Dai *et al*, 2016). In this section, we will discuss the difference and the rationality of ribosome states in our model.

Briefly, our model contains active ribosomes and free ribosomes. **Active ribosomes** are bound to mRNA and actively involved in elongating peptide chains. **Free ribosomes** are responsible for translation initiation; they are available for binding to mRNA and comprise a subset of the inactive ribosomes in proteome partitioning models.

Proteome partitioning models distinguish between active ribosomes and inactive ribosomes. **Active ribosomes** have exactly the same meaning as in our model: they are engaged in elongation. At steady-state growth, the protein synthesis rate can be written as: $v_{protein\_syn} = \mu P = f_{active} \cdot k_{eff} \cdot R$, where μ is the growth rate, P is the total protein concentration (measured in amino acids per volume), f_{active} is the fraction of active ribosomes among total ribosomes, k_{eff} is the turnover number of ribosomes during elongation, and R is the concentration of ribosomes. By measuring μ,

k_{eff}, and the ratio between R and P (estimated through the RNA/Protein ratio and the fraction of rRNA in total RNA), Dai *et al.* estimated the fraction of active ribosomes as a function of growth rate (Dai *et al*, 2016). In the view of protein partitioning models, the **inactive ribosomes** comprise all ribosomes not actively engaged in elongation. Inactive ribosomes include not only ribosomes available for initiation (**free ribosomes**), but also ribosomes that are unavailable for initiation (**unused, or deactivated, ribosomes**) (Scott *et al*, 2010; Klumpp *et al*, 2013; Dai *et al*, 2016). In this work, we modeled both initiation and elongation, and thus both free and active ribosomes (but not deactivated ribosomes) are included.

Our model is carefully built on first principles. All reactions are explicitly and exclusively constrained by reaction parameters and steady state growth; we avoid any empirical growth rate-dependent parameters, such as a growth rate-dependent fraction of active ribosomes or effective ribosome activity. In other words, our model contains only reactions for which we know why and how they occur. The mechanism leading to a fraction of deactivated ribosomes is not clear. Deactivated ribosomes facilitate faster transitions between growth environments that support different growth rates (Mori *et al*, 2017), a phenomenon that cannot be predicted with steady-state models such as ours. Moreover, the true fraction of deactivated ribosomes has not been measured experimentally. Thus, we did not attempt to predict the total concentration of ribosomes (including deactivated ribosomes), and only compared our predictions for active ribosome concentrations to experimental estimates.

Text S2.3. Impact of GTP and GDP concentrations on the predictions

GTP and GDP are involved in many intracellular reactions (Feist *et al*, 2007), and we thus do not expect to predict their concentrations in this translation model. In our model, GTP and GDP are involved in nucleotide exchange by elongation factor Tu (Gromadski *et al*, 2002) (see Methods). The concentrations of GTP and GDP may influence the rates of some reactions directly. In this section, we assess the impact of the assumed GTP and GDP concentrations on the predictions, examining three pairs of concentrations (Bennett *et al*, 2009) resulting from growth of *E. coli* K-12 on different media. Note that GTP and GDP concentrations (Bennett *et al*, 2009) and proteome data were collected for different strains of *E. coli* K-12 (NCM3722 and BW25113, respectively).

The GTP/GDP measurements were done for growth on acetate (c_{GTP} = 1250 μM; c_{GDP} = 18 μM), glycerol (c_{GTP} = 2690 μM; c_{GDP} = 23 μM), and glucose (c_{GTP} = 4900 μM; c_{GDP} = 680 μM); all three conditions also appear in our simulations. We first simulated growth on acetate and on glycerol with GTP and GDP concentrations measured for *E. coli* cells growing on the same media. Next, we replaced the GTP and GDP concentrations with the data for glucose and repeated the simulations (Fig. S2.8). Despite the large differences in GTP and GDP concentrations, the results obtained are very similar. For both acetate and glycerol growth, geometric mean fold-errors (GMFE) are below 1.03 (Fig. S2.8), *i.e.*, the predicted concentrations of the individual components of the translation machinery are on average less than 3% higher or lower in the two sets of predictions. Thus, GTP and GDP concentrations appear to have only a minor influence on the predictions. Because glucose is the reference condition for the protein expression data (Schmidt *et al*, 2016), we used the concentration of GTP and GDP for growth on glucose for all predictions in this study.

Supplementary Figures

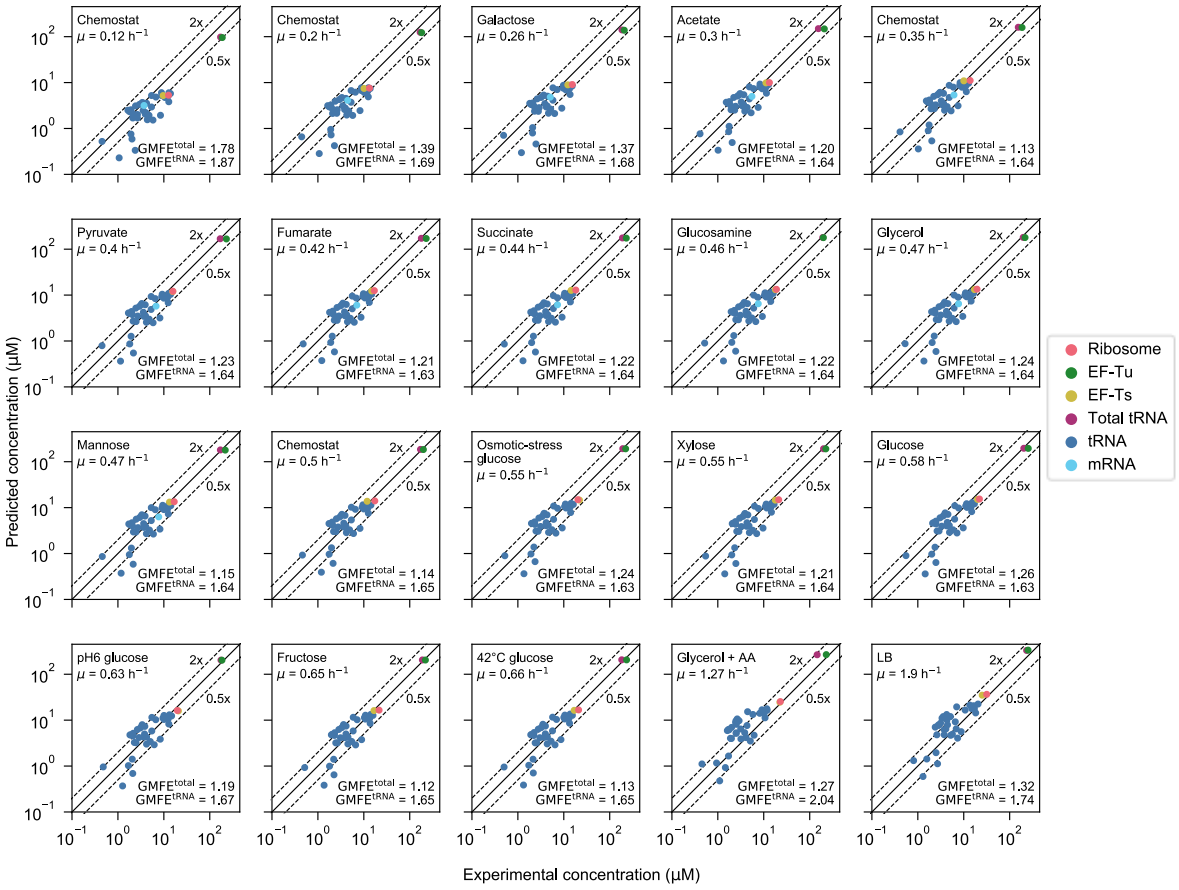


Fig. S2.1. The predicted optimal concentrations of the components of the translation machinery agree with experimental estimates across 20 growth conditions on different media and in chemostats with a minimal glucose medium (sorted by ascending growth rate). The conditions are those under which protein concentrations were measured in Ref. (Schmidt *et al*, 2016). mRNA (Valgepea *et al*, 2013) and tRNA (Dong *et al*, 1996) were assayed in conditions with growth rates that differ from those of the proteomics experiment. To plot mRNA and tRNA data in the same panels, we fitted second order polynomial regression models to the available data for mRNA and tRNA concentrations, respectively, and then used the regressions to estimate the concentrations at the growth rates shown in the panels. Absolute mRNA concentration (Valgepea *et al*, 2013) was only assayed for growth rates between 0.11 h^{-1} and 0.49 h^{-1}, and we did not attempt to extrapolate mRNA concentrations beyond this range.

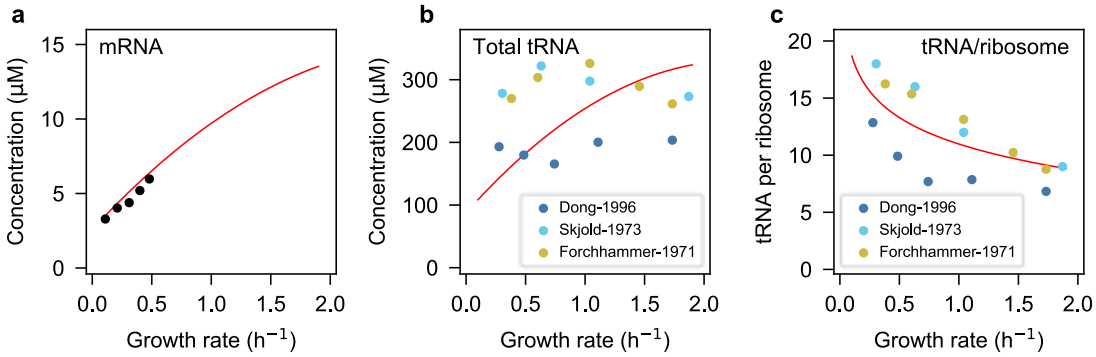


Fig. S2.2. The concentrations of the major non-ribosomal RNA pools predicted from cost minimization are consistent with experimental observations. (**a**) mRNA (Valgepea *et al*, 2013), R^2 = 0.97, GMFE = 1.06. (**b**) Total tRNA data from Dong *et al.* (Dong *et al*, 1996) (summed over individual tRNAs), Forchhammer *et al.* (Forchhammer & Lindahl, 1971), and Skjold *et al.* (Skjold *et al*, 1973); combined R^2 = 0.27, GMFE = 1.30. (**c**) number of tRNAs per ribosome from the same datasets as in (b).

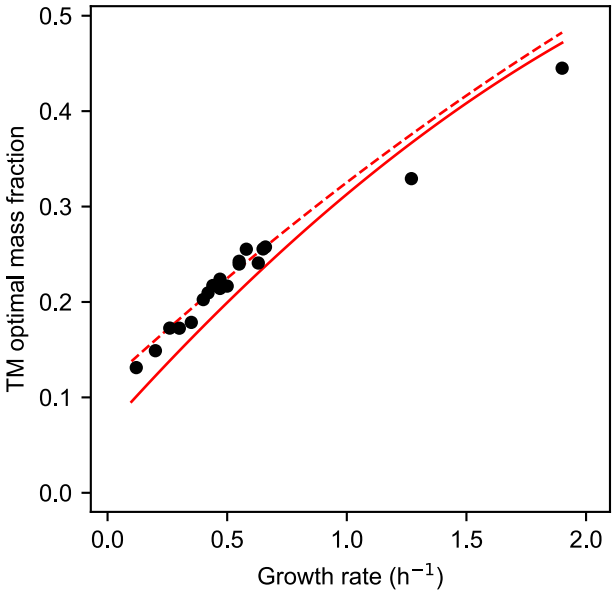


Fig. S2.3. Theoretically optimal resource allocation to the translation machinery as a fraction of total dry mass increases almost linearly with growth rate. The solid red line indicates the model predictions, without accounting for deactivated ribosomes. The dashed line indicates the predicted optimal mass fraction when we additionally include the fraction of deactivated ribosomes, which cannot be predicted by a steady-state model but which we estimated from experimental observations (Fig. 2.4 of the main text; see Methods for details). Experimental data (points) is the sum over the observed concentrations of translation associated proteins (Schmidt *et al*, 2016) (ribosomal proteins, EF-Tu, EF-Ts) and RNA (Scott *et al*, 2010; Dai *et al*, 2016) (ribosomal RNA, tRNA, mRNA; interpolated to the same growth rates as in the protein measurements, see Methods). Note that the mass fraction of the translation machinery does not include GDP, GTP, free tRNA, tRNA-synthetases, and elongation factor G (FusA).

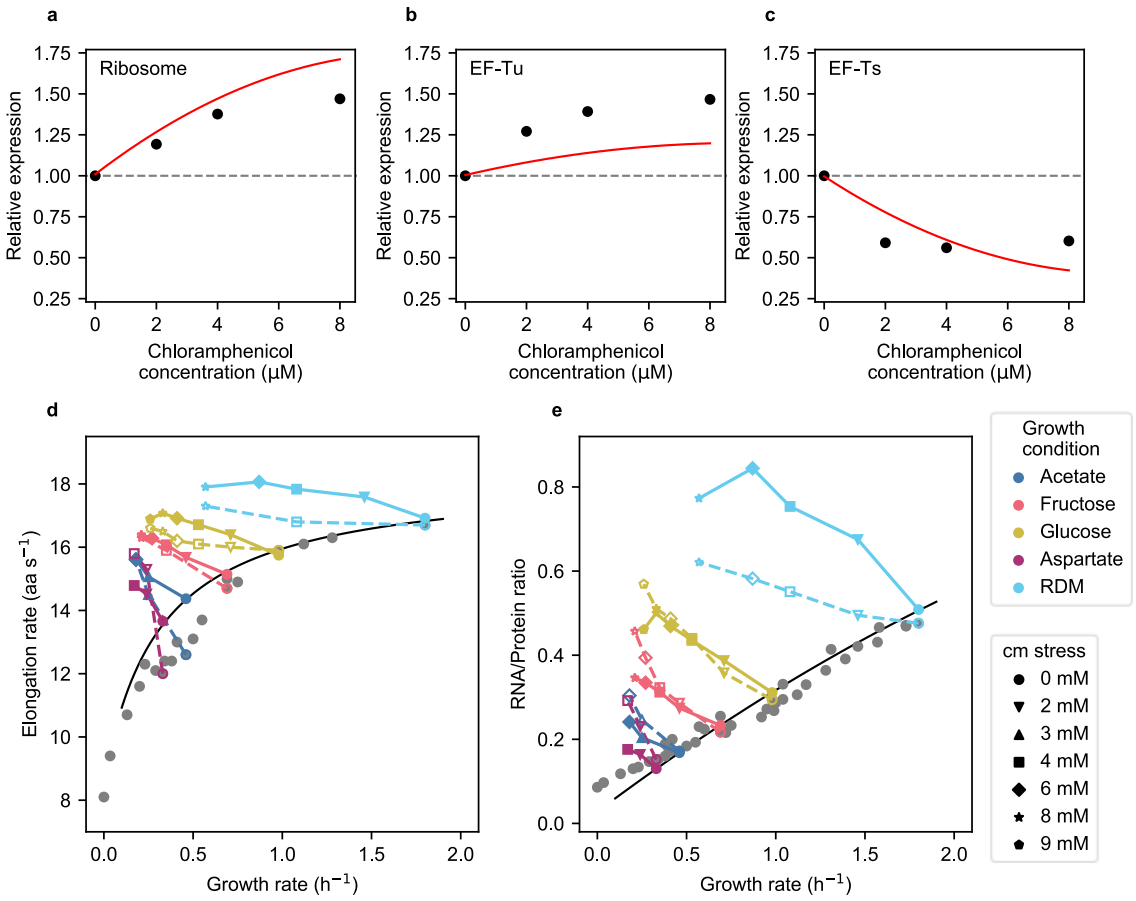


Fig. S2.4. Optimality of the translation machinery under chloramphenicol stress. Model predictions (red lines) of relative changes in the concentrations of **(a)** ribosome, **(b)** EF-Tu, and **(c)** EF-Ts under increasing chloramphenicol stress are qualitatively consistent with experimental data (Hui *et al*, 2015) (**a**, **b**, **c** show the results for growth on glucose). Predicted **(d)** elongation rates and **(e)** RNA/protein ratios under chloramphenicol stress are also qualitatively consistent with experimental data (Dai *et al*, 2016). Grey dots indicate experimental elongation rates without chloramphenicol stress; the black line marks the corresponding (non-stressed) predictions. Different symbols indicate varying chloramphenicol concentrations, while colours indicate growth conditions (different nutrients). Dashed lines connect experimental elongation rates (open symbols) under chloramphenicol stress on the same nutrient; solid lines connect the corresponding elongation rate predictions (filled symbols). Chloramphenicol concentrations were varied from 0 mM to 9 mM. In both predictions and experiment, elongation rates increase with growing chloramphenicol stress, with faster increases under progressively poorer nutrient conditions. The overestimated RNA/protein ratio on rich defined medium (RDM) likely reflects the fact that ribosome is inhibited less by chloramphenicol *in vivo* than theoretical calculations predict (see Fig. N1 in Ref. (Dai *et al*, 2016)). The predictions are functions of the growth rate and of chloramphenicol concentration; the non-smoothness of the prediction lines likely arise from experimental uncertainties in the corresponding values.

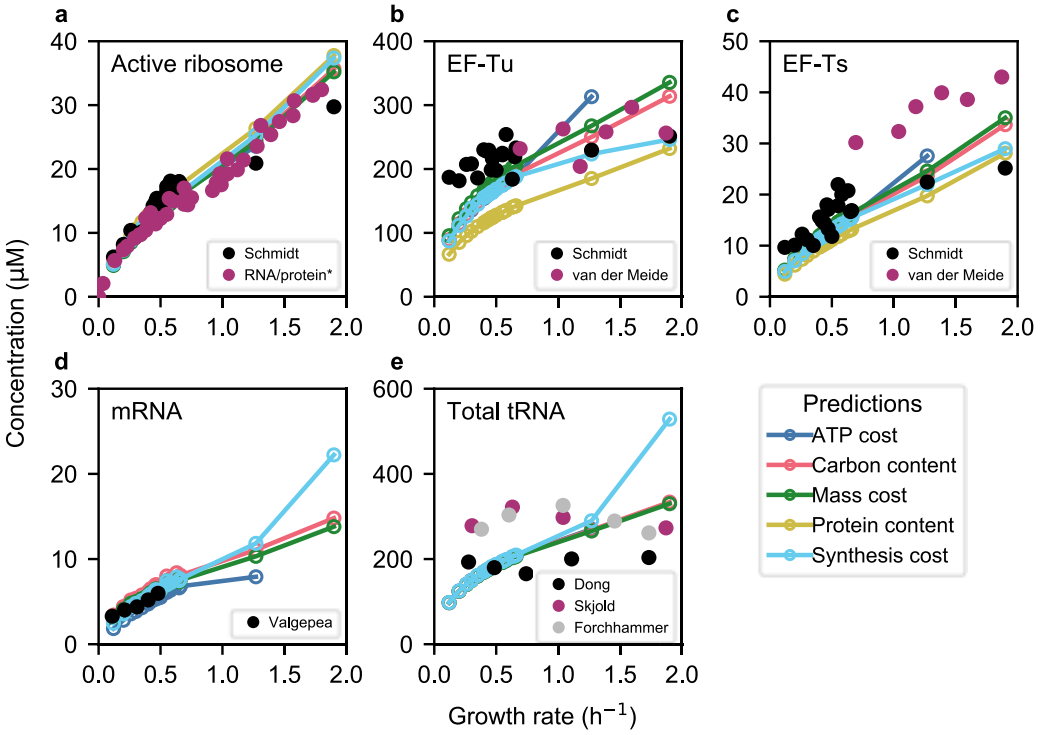
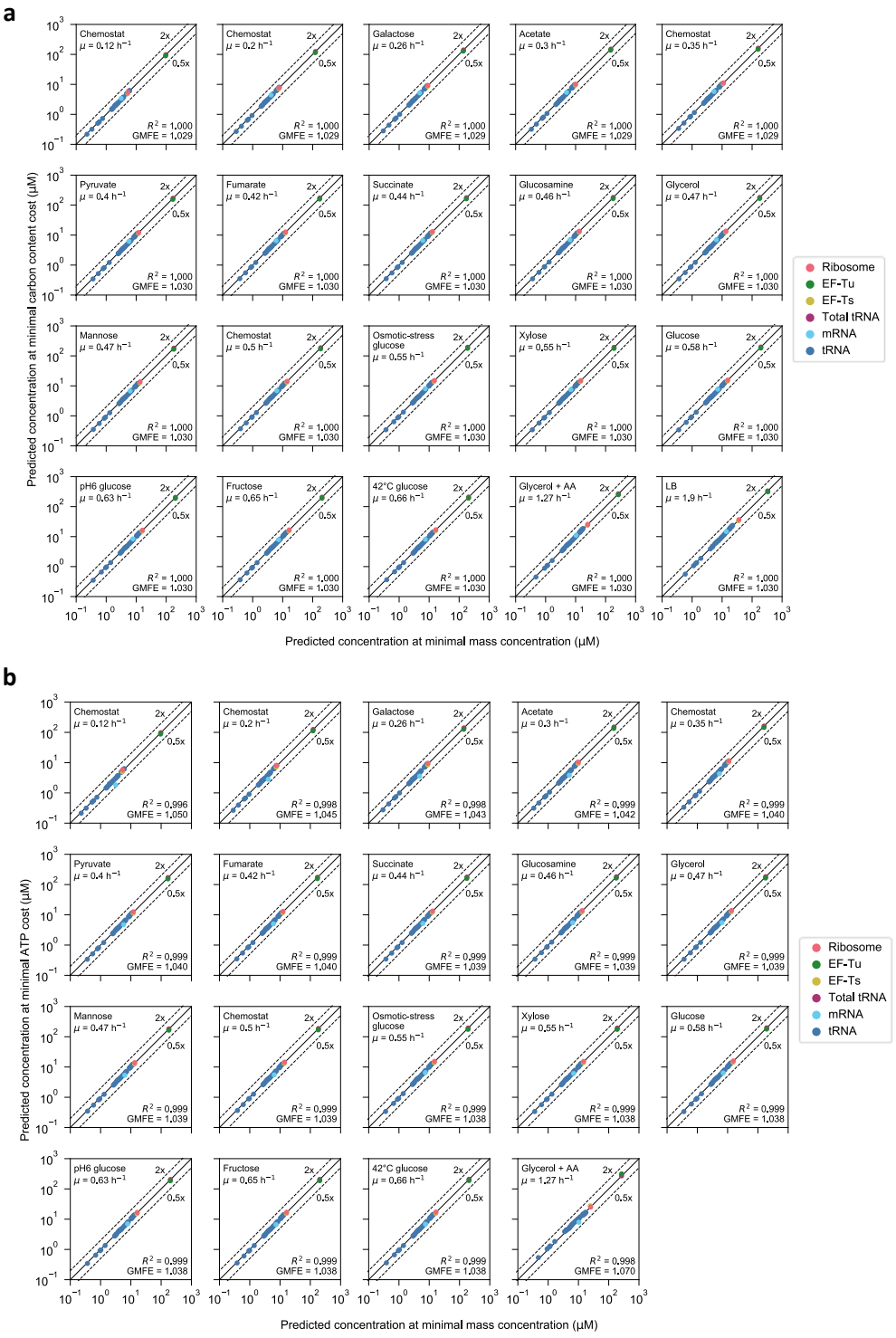


Fig. S2.5. Different cost definition lead to broadly similar growth rate dependencies. The coloured lines show predictions based on minimizing the total mass density (as in the main text), carbon content, protein mass, and synthesis cost of the components of the translation machinery, respectively. The panels compare the predictions to experimental estimates for (**a**) active ribosomes (based on proteomics (Schmidt *et al*, 2016), black dots, and RNA/protein ratios, red dots); (**b**) EF-Tu; (**c**) EF-Ts; (**d**) mRNA; and (**e**) total tRNA. As it is unclear how to calculate ATP costs in the LB medium (μ = 1.9 h^{-1}), no results for ATP costs are shown for this condition.

To derive the molar concentration of active ribosomes from reported RNA/protein mass ratios for panel (**a**), we used the tRNA/ribosome ratios reported in the experimental papers on tRNA concentrations (Dong *et al*, 1996; Skjold *et al*, 1973; Forchhammer & Lindahl, 1971) to calculate the rRNA in total RNA (f_{rRNA}) (as mRNA is a very small fraction of total RNA by mass (~ 2-5%), we ignored its contribution in this calculation). With f_{rRNA}, the RNA/Protein mass ratio (Scott *et al*, 2010; Dai *et al*, 2016), and the mass fraction of protein in the ribosome, we calculated the fraction of ribosomal protein in total protein. Then, with equation (2.1) and the equation for the active ribosome fraction ($f_{active} = \mu / (0.124 + \mu)$) we calculated the active ribosome concentration.


a
Chemostat μ = 0.12 h⁻¹
Chemostat μ = 0.2 h⁻¹
Galactose μ = 0.26 h⁻¹
Acetate μ = 0.3 h⁻¹
Chemostat μ = 0.35 h⁻¹
Pyruvate μ = 0.4 h⁻¹
Fumarate μ = 0.42 h⁻¹
Succinate μ = 0.44 h⁻¹
Glucosamine μ = 0.46 h⁻¹
Glycerol μ = 0.47 h⁻¹
Mannose μ = 0.47 h⁻¹
Chemostat μ = 0.5 h⁻¹
Osmotic-stress glucose μ = 0.55 h⁻¹
Xylose μ = 0.55 h⁻¹
Glucose μ = 0.58 h⁻¹
pH6 glucose μ = 0.63 h⁻¹
Fructose μ = 0.65 h⁻¹
42°C glucose μ = 0.66 h⁻¹
Glycerol + AA μ = 1.27 h⁻¹
LB μ = 1.9 h⁻¹
2x
0.5x
R² = 1.000
GMFE = 1.029
GMFE = 1.030
Predicted concentration at minimal carbon content cost (μM)
Predicted concentration at minimal mass concentration (μM)
Ribosome
EF-Tu
EF-Ts
Total tRNA
mRNA
tRNA
b
Chemostat μ = 0.12 h⁻¹
Chemostat μ = 0.2 h⁻¹
Galactose μ = 0.26 h⁻¹
Acetate μ = 0.3 h⁻¹
Chemostat μ = 0.35 h⁻¹
Pyruvate μ = 0.4 h⁻¹
Fumarate μ = 0.42 h⁻¹
Succinate μ = 0.44 h⁻¹
Glucosamine μ = 0.46 h⁻¹
Glycerol μ = 0.47 h⁻¹
Mannose μ = 0.47 h⁻¹
Chemostat μ = 0.5 h⁻¹
Osmotic-stress glucose μ = 0.55 h⁻¹
Xylose μ = 0.55 h⁻¹
Glucose μ = 0.58 h⁻¹
pH6 glucose μ = 0.63 h⁻¹
Fructose μ = 0.65 h⁻¹
42°C glucose μ = 0.66 h⁻¹
Glycerol + AA μ = 1.27 h⁻¹
2x
0.5x
R² = 0.996
GMFE = 1.050
R² = 0.998
GMFE = 1.045
GMFE = 1.043
R² = 0.999
GMFE = 1.042
GMFE = 1.040
GMFE = 1.039
GMFE = 1.038
R² = 0.998
GMFE = 1.070
Predicted concentration at minimal ATP cost (μM)
Predicted concentration at minimal mass concentration (μM)
Ribosome
EF-Tu
EF-Ts
Total tRNA
mRNA
tRNA

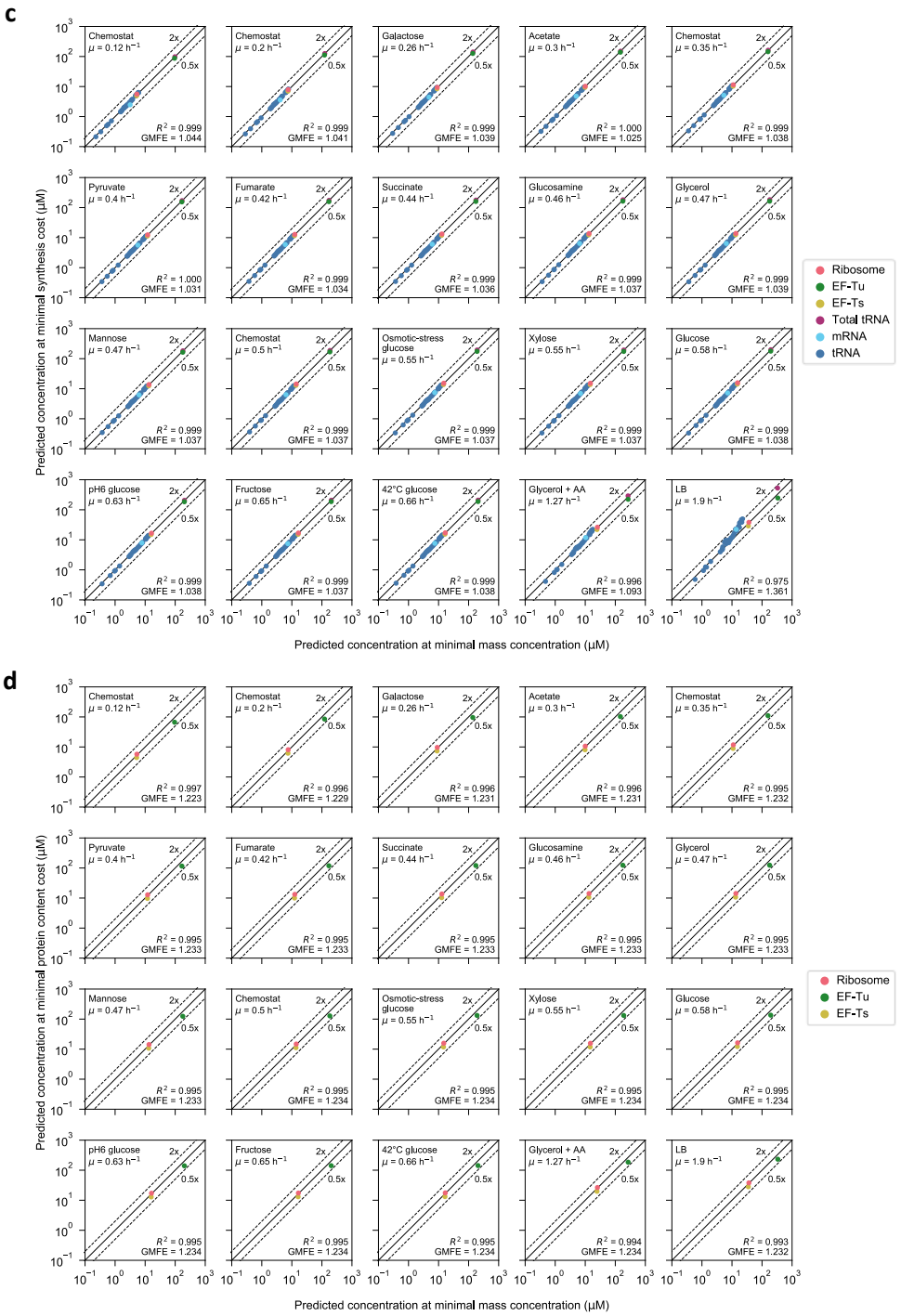


Fig. S2.6. Comparison of predicted concentrations at minimal mass concentration with alternative cost measures. Alternative cost measures based on (**a**) carbon content, (**b**) ATP cost of synthesis, and (**c**) the macromolecular investment into the synthesis lead to very similar predictions of the concentrations of the translation machinery components as mass concentration costs. A cost measure based on the protein content (**d**) does not assign costs to mRNA and tRNA and can hence not predict their concentrations. Each sub-panel corresponds to one growth condition assayed in Ref. (Schmidt *et al*, 2016). A data point shows the predicted concentration for one component based on an alternative cost measure vs. the predicted concentration based on the mass concentration cost employed for Figs. 2.2-2.5 in the main text. In the bottom right corner of each sub-panel, we provide the square of Pearson's correlation coefficient on log-scale, R^2, and the geometric mean fold error, GMFE. As it is not clear how to estimate ATP costs in the rich medium (LB), we made no predictions for this condition in (b).

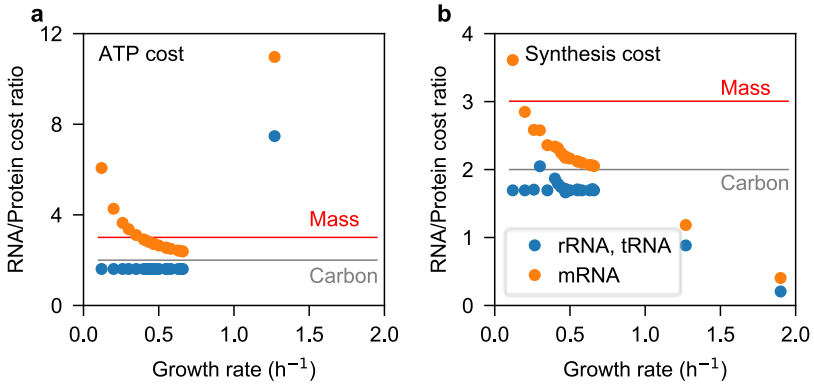


Fig. S2.7. The cost of RNA per nucleotide, divided by the cost for the synthesis of protein per amino acid, plotted against the growth rate μ. (**a**) ATP cost of synthesis; (**b**) total required catalyst mass of synthesis (synthesis cost). The horizontal red line shows the RNA/protein cost ratio for the mass concentration cost, the horizontal grey line the cost ratio based on the carbon content. For ATP and synthesis costs, the RNA/protein cost ratios are different between stable RNA (tRNA, rRNA) and mRNA, as for mRNA we additionally consider degradation. The row of identical rRNA, tRNA cost ratios at low growth rates in (b) is for chemostat conditions with a minimal glucose medium.

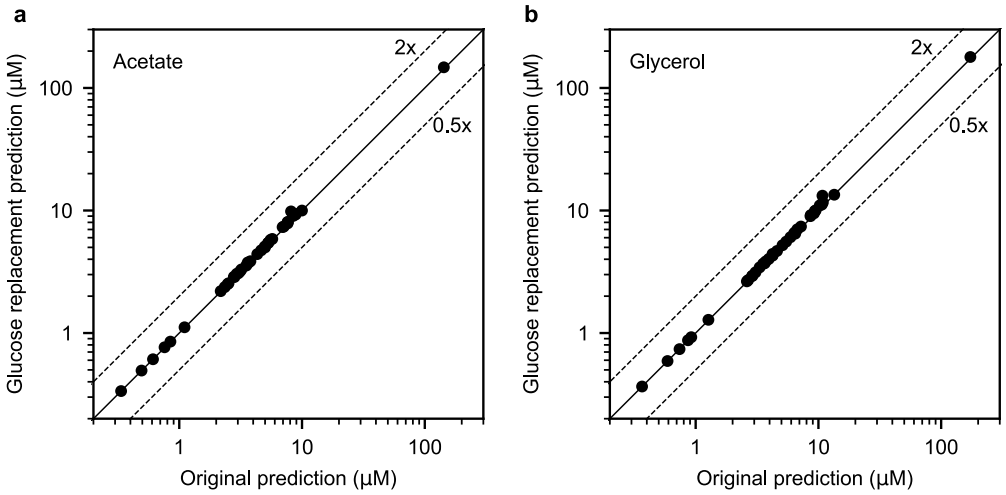


Fig. S2.8. Impact of GTP and GDP concentrations on model predictions. (a) Growth on acetate (geometric mean fold-error GMFE = 1.028). **(b)** Growth on glycerol (*GMFE* = 1.030). Each datapoint represents the concentration of one model component (ribosome, EF-Tu, EF-TS, aa-tRNA). x-axes show predictions using the GTP and GDP concentrations measured for the corresponding medium; y-axes show predictions when instead assuming the GTP and GDP concentrations measured for growth on glucose. GMFE measures the mean deviation from the identity line on the log-log plot; GMFE = 1 indicates perfect identity. The very low GMFE values indicate that *in vivo* GTP and GDP concentration has a very small effect on our model.

3 Manuscript 2. Proteome efficiency of metabolic pathways in *Escherichia coli* increases along the carbon flow

Xiao-Pan Hu[1], Martin J. Lercher[1]*

[1] Institute for Computer Science and Department of Biology, Heinrich Heine University, D-40225 Düsseldorf, Germany

* Corresponding author

E-mail: martin.lercher@hhu.de (MJL)

Contribution: I designed the study, performed the analyses, and drafted the manuscript.

Abstract

Understanding the allocation of the cellular proteome to different cellular processes is central to unraveling the organizing principles of bacterial physiology. In previous work, we found that proteome allocation to protein translation is maximally efficient, i.e., it represents the minimal allocation of dry mass able to sustain the observed protein production rate. However, recent studies on bacteria have demonstrated that overall, the proteome is not allocated for maximal efficiency, i.e., resource allocation to many proteins exceeds the minimal level required to support the observed growth rate. While these findings indicate some heterogeneity across pathways in their proteome efficiency, systematic studies at the pathway level are lacking. Here, we systematically analyze the proteome efficiency of metabolic pathways, which together account for more than half of the *E. coli* proteome during exponential growth. Comparing the predicted and observed proteome allocation to different metabolic pathways across growth conditions, we find that the most costly biosynthesis pathways – those for amino acid biosynthesis and cofactor biosynthesis – are expressed for near optimal efficiency. Overall, proteome efficiency increases along the carbon flow through the metabolic network: proteins involved in pathways of carbon uptake and central metabolism tend to be highly over-abundant, while proteins involved in anabolic pathways and in protein translation are much closer to the expected minimal abundance across conditions. Further, it appears that proteome efficiency alone is not enough to explain the utilization of alternative metabolic pathways, such as the switch from respiration to aerobic fermentation or from the PEP-glyoxylate cycle to the TCA cycle. Our work thus provides a bird's-eye view of metabolic pathway efficiency, demonstrating systematic deviations from optimal cellular efficiency at the network level.

Introduction

Proteins account for more than half of the cell dry mass in *E. coli* (Bremer & Dennis, 2008) and drive most biological processes. How and why proteome is allocated to different cellular processes and pathways is a vital question for understanding the principles behind bacterial physiology (Basan, 2018). Proteome allocation into different groups of genes is growth rate-dependent (Peebo *et al*, 2015). When partitioning the proteome into specific, coarse-grained "sectors", the corresponding proteome fractions follow simple, empirical growth laws, increasing or decreasing linearly with the growth rate μ (Scott *et al*, 2010; You *et al*, 2013; Klumpp *et al*, 2013; Hui *et al*, 2015). For example, the proteome fraction allocated to the ribosome and ribosome-affiliated proteins (the R-sector (Klumpp *et al*, 2013)) scales as a linear function of growth rate under nutrient limiting conditions (Scott *et al*, 2010).

Why does the proteome composition scale with the growth rate? Protein is the most abundant and costly macromolecule in bacterial cells. It has thus been speculated that the proteome composition is adjusted to the specific growth condition to maximize the growth rate (Bruggeman *et al*, 2020). If this were true, all protein concentrations would be at the minimal level required to sustain the observed cellular growth rate. This simple assumption has been widely used in computational models of cellular growth (Adadi *et al*, 2012; Dourado & Lercher, 2020; Goelzer *et al*, 2015, 2011; O'Brien *et al*, 2013; Beg *et al*, 2007; Molenaar *et al*, 2009). However, even if proteome allocation had evolved to be maximally efficient, it is not obvious that this efficiency would simply maximize the instantaneous growth rate. Instead, it appears likely that proteome allocation has evolved to maximize cellular fitness in unpredictable, dynamic environments with varying nutrients and involving periods of famine and stresses (Bruggeman *et al*, 2020). Indeed, recent experimental work indicates that the proteome is not expressed for maximal efficiency in unevolved *E. coli* strains, at least not in the naïve sense of maximizing the instantaneous growth rate. First, a large fraction of the expressed proteome is unneeded for the current environment, especially at low growth rates (O'Brien *et al*, 2016). Second, the growth rate can increase by approx. 20% in a few hundred generations in adaptive laboratory evolution experiments on minimal media (Ibarra *et al*, 2002), a process associated with reductions in the abundance of unused proteins (O'Brien *et al*, 2016). Finally, the fluxes through some cellular processes, e.g., nutrient transport and energy production, are not limited by specific proteins in these pathways at low growth rates (Belliveau *et al*, 2021). Thus, *E. coli* proteome allocation seems not to be globally optimized for maximizing the instantaneous growth rate.

On the other hand, proteome allocation to at least one cellular pathway – protein translation – is optimized for maximal efficiency at the given protein synthesis rate (Hu *et al*, 2020; Hu & Lercher, 2021; Lalanne & Li, 2021; Belliveau *et al*, 2021). This indicates that while the global allocation of proteins is not always optimized for maximal growth rate, the proteome allocation to some cellular pathways is at a local optimum – i.e., the individual pathway utilizes the minimal protein mass required to support the observed pathway output. In *E. coli* growing on minimal media, more than half of the proteome by mass is metabolic enzymes (Schmidt *et al*, 2016). Computational models can predict the optimally efficient proteome allocation to each metabolic pathway (Adadi *et al*, 2012; Goelzer *et al*, 2015; O'Brien *et al*, 2013, 2016), and quantitative proteomics data is available for *E. coli* growing on a wide range of minimal media with different carbon sources (Schmidt *et al*, 2016).

Here, we exploit these resources to compare experimental data in diverse minimal media conditions (Schmidt *et al*, 2016) to the predicted optimal pathway expression at the observed

growth rate. We find that pathways differ systematically in how much excess protein mass is allocated to them compared to the local optimum, with decreasing excesses over optimal allocation along the carbon flow from nutrient import to protein production.

Results and Discussion

Modeling proteome allocation with linear enzyme kinetics and growth rate-dependent biomass composition

To analyze local pathway efficiency, we first predict the local optima of all metabolic enzymes with an improved version of FBA with molecular crowding (Beg *et al*, 2007; Adadi *et al*, 2012). We modelled *E. coli* metabolism with the constraint-based *i*ML1515 model (Monk *et al*, 2017). The standard model assumes a constant composition of biomass across conditions. As the RNA/protein mass ratio (Scott *et al*, 2010) and the cell surface/volume ratio (Si *et al*, 2017) can be expressed as functions of growth rate under the investigated conditions, we re-formulated the biomass function of *i*ML1515 with growth rate-dependent contents of RNA, protein, and cell envelope components (murein, lipopolysaccharides, and lipid) (See Methods, **Fig. S3.1**).

We performed calculations using MOMENT (MetabOlic Modeling with ENzyme kineTics) (Adadi *et al*, 2012; Desouki, 2016; Heckmann *et al*, 2018), a version of flux balance analysis (FBA) with molecular crowding (Beg *et al*, 2007). Similar to other constraint-based approaches (O'Brien *et al*, 2013; Goelzer *et al*, 2015), MOMENT estimates the enzyme concentration required to support a given flux v_i as $[E_i]= v_i / k_i$, where k_i is the effective turnover number of the enzyme. This effective turnover number is assumed to be constant across conditions, a zero-order approximation to the true growth rate-dependence (Dourado *et al*, 2021). Maximal *in vivo* effective enzyme turnover number ($k_{app,max}$) represent turnover in the cellular environment better than *in vitro* estimates of enzyme turnover numbers (k_{cat}) (Davidi *et al*, 2016; Heckmann *et al*, 2020). We thus parameterized the reactions of the *i*ML1515 model with the $k_{app,max}$ from Ref. (Heckmann *et al*, 2020) by replacing the original k_{cat} (Desouki, 2016) when $k_{app,max}$ was available. For reactions with neither $k_{app,max}$ nor k_{cat} , the enzyme turnover number predicted by machine learning from (Heckmann *et al*, 2020) is used in the simulation. Most enzymes in the metabolic model have measured parameters ($k_{app,max}$ or k_{cat}): enzymes with $k_{app,max}$ or k_{cat} account for ~70% of total enzyme by mass in the whole metabolic network, and account for ~80% when excluding transport reactions (**Fig. S3.2**).

With the growth rate-dependent biomass function and updated enzyme turnover numbers, we identified the minimal total mass concentration of enzymes and transporters (in units of g/g_{DW}) that can support the observed growth rate on the given carbon source (see Methods). Thus, our predictions do not reflect globally optimal resource allocation, but quantify the minimal proteome allocation into pathways required to sustain the observed growth rate (local optimality). Note that the calculation of required concentrations assumes that all enzymes are fully saturated with their products; this means that our estimates provide a lower bound of proteome allocation into pathways, which is expected to deviate increasingly from the actual demand at lower growth rates (Dourado *et al*, 2021).

Metabolic pathways differ systematically in their proteome efficiency

Following earlier work (O'Brien *et al*, 2016), we first compared the predicted minimal required proteome with experimental data across the whole metabolic network on minimal media with different sources assayed in Ref. (Schmidt *et al*, 2016). As *E. coli* uses different central metabolic

reactions on glycolytic and gluconeogenic carbon sources and most of the proteome data in Ref. (Schmidt *et al*, 2016) were measured on glycolytic carbon sources, we focused on the proteome efficiency of metabolic pathways on glycolytic carbon sources here (results for gluconeogenic carbon sources are shown in **Table S3.1**). We classify proteins according to their experimental and predicted expression. An individual protein is labeled as:

- "*shared*" if its presence is predicted under local optimality and is confirmed in the experiment (these proteins were labeled "utilized" in Ref. (O'Brien *et al*, 2016));
- "*measured-only*" if it is found in the experiment but predicted to be absent (these proteins were labeled "un-utilized" in Ref. (O'Brien *et al*, 2016));
- "*predicted-only*" if its presence is predicted but not confirmed in the experiment.

The predicted abundances of proteins not found experimentally (*predicted-only* proteins) account for only a very small fraction of the total predicted proteome (<1%) in all studied pathways except for nutrient transport and proteins without specified pathways in this study ("others") (**Fig. S3.3**). We thus do not show the *predicted-only* proteins in the following figures.

Metabolic enzymes account for a decreasing fraction of the proteome with growth rate, with observed proteome fractions ranging from 67% to 53% (**Fig. S3.4**). In agreement with earlier work (O'Brien *et al*, 2016), we found that the total abundance of *shared* proteins – those required for maximally efficient growth – increases with growth rate, but far exceeds the predicted globally optimal abundance especially at lower growth rates (**Fig. S3.4**).

To explore the local proteome efficiency of pathways, we assigned the metabolic proteins of the *i*ML1515 *E. coli* model to individual pathways according to their functions, further arranged into four coarse-grained sets:

(1) transporters, which shuttle metabolites across the outer or inner membrane (based on the *i*ML1515 model annotation).

(2) Precursors and energy generation pathways (based on the EcoCyc database pathway ontology term "generation of precursor metabolites and energy" (Keseler *et al*, 2017)). This category is further divided into six "pathways": glycolysis, pentose phosphate pathway, TCA cycle, glyoxylate bypass, energy production (comprising electron transfer chains and ATP biosynthesis), and the remainder of central metabolism (comprising all other genes involved in "generation of precursor metabolites and energy" that are not in the previously listed pathways).

(3) Biosynthesis pathways (based on the EcoCyc database pathway ontology term "biosynthesis" (Keseler *et al*, 2017)). This category is further divided into 5 sets of pathways: amino acid biosynthesis, nucleotide biosynthesis, cofactors biosynthesis, cell wall component biosynthesis (comprising lipid, peptidoglycan, and lipopolysaccharide synthesis pathways), and all other biosynthesis enzymes.

(4) Others, all enzymes in the *i*ML1515 model not included in (1)-(3).

To assess the proteome efficiency of pathways, we consider four aspects, described below. **Table 3.1** shows the pathway proteome efficiency measures on glycolytic carbon sources, which are discussed in the following subsections.

(a) Considering all *shared* pathway proteins – those that are predicted to be active and found experimentally – we calculated the Pearson correlation coefficient *r* between their combined mass concentrations in the locally optimal prediction and in the experimental measurements. For

locally optimal proteome allocation and under the assumption of constant enzyme saturation, this correlation should approach r=1, independent of enzyme kinetic parameter values.

(b) The geometric mean fold-error (GMFE) of predicted vs. observed protein concentrations of the pathway's *shared* proteins. The GMFE shows by which factor the observed concentrations deviate from predicted values on average.

(c) The experimentally observed mass fraction of *measured-only* proteins of the pathway. This is the proteome fraction that makes no contribution to growth according to our predictions.

(d) The squared Pearson's correlation coefficient between predicted and measured abundances across individual proteins in a given growth condition. While measures (a)- (c) assess optimality at the pathway level, this last measure quantifies the relationships between proteins within the pathway: a correlation coefficient close to 1 indicates that all proteins are equally close to – or equally distant from – the optimal prediction. Note that in contrast to measure (a), the comparison across individual proteins relies strongly on the accuracy of the individual turnover numbers. As the latter are only known approximately, we expect these estimates to be noisy.

Table 3.1. Proteome efficiency of pathways.

Pathway	Pathway expression (for *shared* proteins) (n = 14)[α]			*measured-only* fraction[β] (median across 14 conditions)	individual *shared* proteins; median across 14 conditions[ν]		
	Pearson's *r*	*p*	GMFE		r^2	*p*	N^{δ}
Measures (a)-(d)	(a)	(a)	(b)	(c)	(d)	(d)	(d)
Biosynthesis	**0.91**	**4.8×10^{-06}**	**1.70**	**0.26**	**0.45**	**4.2×10^{-31}**	**226**
Amino acid	0.88	3.7×10^{-05}	1.40	0.30	0.45	1.1×10^{-10}	72
Nucleotide	0.82	3.7×10^{-04}	3.32	0.23	0.15	0.05	28
Envelope	0.66	0.01	1.88	0.14	0.38	2.3×10^{-05}	40
Cofactor	0.91	4.9×10^{-06}	1.24	0.11	0.59	4.1×10^{-15}	72
Biosynthesis others	0.78	1.1×10^{-03}	2.91	0.25	0.46	5.4×10^{-05}	29
Central metabolism	**0.16**	**0.60**	**2.32**	**0.31**	**0.15**	**3.3×10^{-03}**	**56**
Glycolysis	0.79	6.9×10^{-04}	2.21	0.08	0.35	0.05	11
Pentose phosphate pathway	0.85	1.3×10^{-04}	1.30	0.39	0.32	0.24	6
TCA cycle	-0.65	0.01	6.40	0.10	0.38	0.03	12
Glyoxylate shunt	-	-	-	1	-	-	0
Energy generation	-0.15	0.61	1.63	0.06	0.11	0.08	28
Central metabolism others	0.67	9.4×10^{-03}	1.56	0.55	0.98	0.10	3
Transporters	**-0.75**	**1.9×10^{-03}**	**3.39**	**0.997**	**0.13**	**0.64**	**4**
Others	**0.06**	**0.84**	**1.79**	**0.91**	**0.16**	**0.03**	**30**
Total	**0.85**	**1.4×10^{-04}**	**1.79**	**0.52**	**0.35**	**1.7×10^{-30}**	**309**

[α] Values reflect the local optimality of complete pathways across conditions. n = 14 indicates the number of glycolytic carbon sources analyzed.

[β] Mass fraction of *measured-only* (un-predicted but observed) proteins relative to all proteins in the pathway

[ν] These columns reflect the local optimality compared across individual proteins within each pathway at a given growth condition; values are medians across the n = 14 glycolytic growth conditions.

[δ] Number of proteins in each pathway or pathway set.

The most expensive biosynthesis pathways are consistent with optimality

The biosynthesis pathways utilize precursors and energy generated by central metabolism to produce building blocks of macromolecules. The predicted proteome fractions of these pathways are almost linear functions of the growth rate (**Fig. 3.1**), as mostly the same reactions are used for biosynthesis across the studied minimal conditions. Overall, we find a strong correlation and good quantitative agreement between predicted and observed abundances of *shared* enzymes (r^2=0.83, p=4.8x10^{-6}; GMFE=1.70; **Table 3.1**).

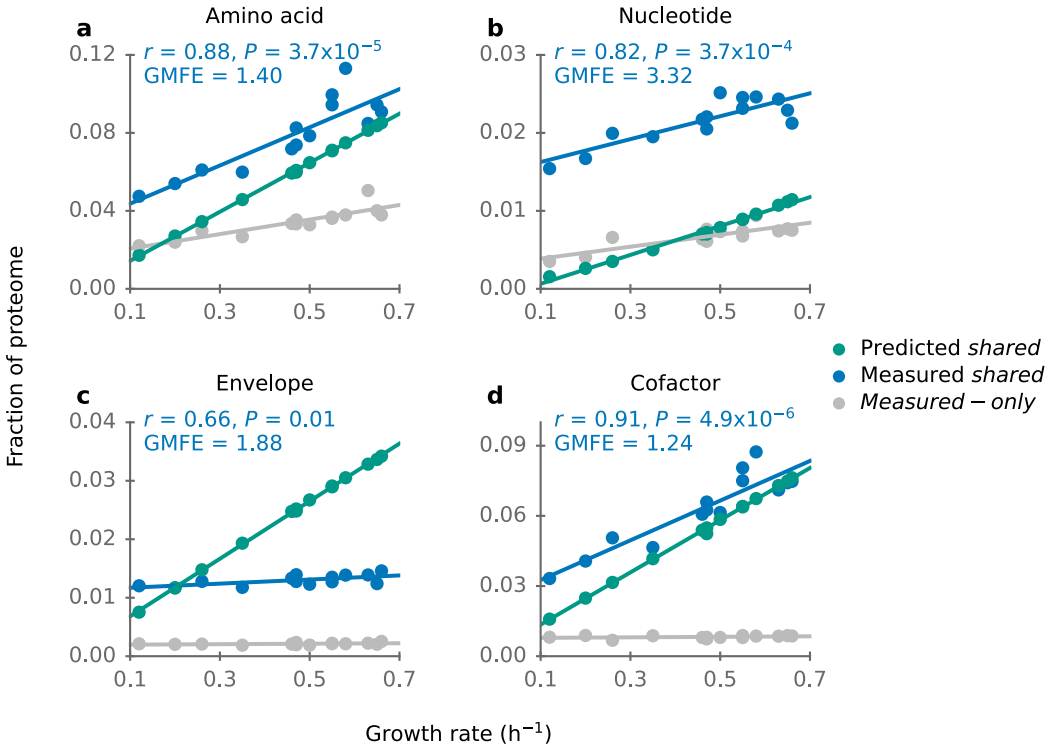


Fig. 3.1. Experimentally observed and predicted proteome fractions of biosynthesis pathways across glycolytic carbon sources. See Fig. S3.5a for biosynthetic proteins not covered here.

The amino acid biosynthesis pathways account for ~15% of the proteome at high growth rates. Predicted and observed abundances of *shared* proteins are strongly correlated (**Fig. 3.1a** and **Table 3.1**; Pearson's r^2 = 0.77, p = 3.7x10^{-6}; GMFE = 1.40). At lower growth rates, we expect decreasing enzyme saturation (Dourado *et al*, 2021) and thus a progressively stronger underestimation of the required proteome by the model; accordingly, **Fig. 3.1a** appears to be consistent with optimal expression of the *shared* proteins of amino acid biosynthesis pathways. Moreover, the prediction of individual *shared* proteins is also significantly correlated with measured data (r^2 = 0.45, p = 1.1x10^{-10}). While the cellular investment into amino acid biosynthesis pathways is thus consistent with optimal resource allocation when considering only proteins predicted to be active, the cell also makes a sizeable investment into *measured-only* proteins of these pathways; *measured-only* proteins account for 30% of the total proteome allocated to these pathways.

The proteome fraction allocated to nucleotide biosynthesis pathways is less than half of that dedicated to amino acid synthesis (**Fig. 3.1b**). While the predicted and observed abundances of *shared* proteins are strongly correlated (r^2 = 0.67, p = 3.7x10^{-4}), their magnitudes differ by more

than 3-fold (GMFE = 3.32; **Fig. 3.1b** and **Table 3.1**). Moreover, the expression of individual enzymes in this pathway cannot be well explained by the predictions (r^2 = 0.15, p = 0.05; **Table 3.1**).

Cell envelope biosynthesis pathways encompass lipid, peptidoglycan, and lipopolysaccharide (LPS) biosynthesis. While predicted and observed expression of *shared* enzymes in these pathways show a statistically significant correlation (r^2=0.44, p=0.01; **Table 3.1**), their growth rate dependence differs markedly. The observed proteome allocation is almost constant across growth conditions; in contrast, the predicted proteome allocation increases linearly with growth rate (**Fig. 3.1c**). It is noteworthy that this disagreement does not stem from an incorrect assumption of constant biomass composition across conditions: our model explicitly accounts for the changing biomass fractions of cell envelope components (Methods). Proteome allocation into *measured-only* proteins of these pathways (14%) is negligible. Theoretically, the predicted optimal proteome allocation should provide a lower limit on the required proteome investment; that predictions exceeds observed proteome allocation for cell envelope biosynthesis at faster growth indicates that one or more enzymes were assigned turnover numbers that are much lower than the true values.

The proteome allocation to cofactor biosynthesis pathways (~10% of the proteome at fast growth) is similarly high as that to amino acid biosynthesis. As for the amino acid biosynthesis pathways, **Fig. 3.1d** indicates that the observed proteome fraction of *shared* proteins may be consistent with optimal proteome allocation to cofactor biosynthesis, once we take the decreasing enzyme saturation at lower growth rates into account (Dourado *et al*, 2021) (r^2 = 0.83, p = 4.9×10^{-6}; GMFE = 1.24; r^2 = 0.59, p = 4.1×10^{-15} for the prediction of individual enzymes; **Table 3.1**). In contrast to the amino acid biosynthesis pathways, proteome investment into *measured-only* proteins is low, accounting for only 11% of the pathway proteome.

In sum, proteome efficiency varies substantially across biosynthesis pathways. On one hand, proteome investment into cofactor and amino acid biosynthesis pathways appears to be consistent with optimal efficiency. On the other hand, proteome allocation to nucleotide, envelope, and other biosynthesis pathways appears to be sub-optimal. While observed proteome investment only increases by roughly two-fold for amino acid, nucleotide, and cofactor biosynthesis and shows almost no increase in envelope and other biosynthesis pathways, predicted investment increases by almost a factor of 5.5 (which is the fold-change of growth rate across the examined conditions). As suggested above, this discrepancy may in part be explained by expected changes in enzyme saturation across growth rates (Dourado *et al*, 2021); however, it may also indicate substantial deviations from optimality also for the most expensive biosynthesis pathways.

Central metabolism: precursor metabolite and energy generation pathways appear not to be regulated for optimality

The pathways grouped under the term “central metabolism” produce precursor metabolites and energy for all other cellular processes. In sum, the enzymes of central metabolism show little systematic variation with growth rate, and their abundance is at most weakly correlated with the predicted concentrations (r^2=0.03, p=0.60; GMFE=2.32).

Glycolysis is among the central pathways whose enzyme abundances increase markedly with growth rate and are strongly correlated with predicted values (**Fig. 3.2a**; r^2 = 0.62, p = $6.9\text{x}10^{-4}$); *measured-only* proteins account for only a very small fraction of the pathway (8% on average). However, protein levels are substantially higher than predicted (GMFE = 2.21). Moreover, the

predicted expression of individual proteins is only weakly correlated with measured protein expression in glycolysis (r^2 = 0.35, p = 0.05). A potential reason for these discrepancies between observations and predictions is that most of the reactions in glycolysis are reversible, while the simple approximation for enzyme activity used here (k_{cat}) cannot capture the demand of enzymes close to thermodynamic equilibrium (Bar-Even *et al*, 2012). Moreover, many of the enzymes in glycolysis are regulated allosterically (Diether *et al*, 2019), and may hence act at lower activities than expected without regulation.

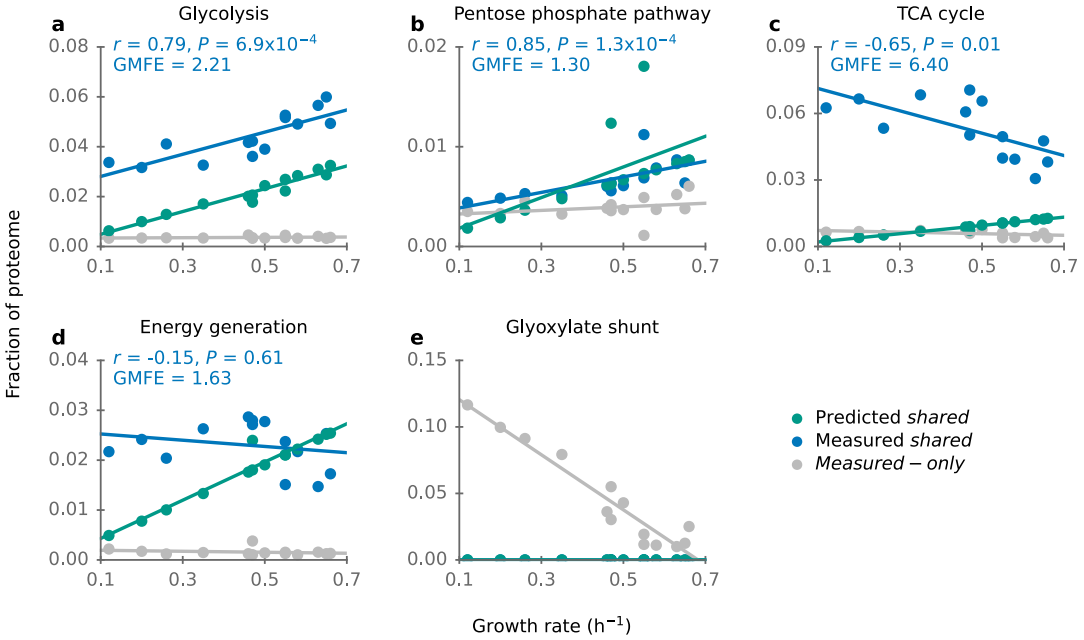


Fig. 3.2. Experimentally observed and predicted proteome fractions of central metabolic pathways. See Fig. S3.5b for central metabolic proteins not covered here.

The pentose phosphate pathway accounts for only ~1% of the proteome. While the measured abundance of its *shared* proteins is close to and strongly correlated with the predictions, (**Fig. 3.2b**; r^2 = 0.72, p = 1.3x10^{-4}; GMFE = 1.3), *measured-only* proteins account for 39% of the pathway proteome. Moreover, the correlation of predicted abundances of individual proteins with measured protein abundances is not statistically significant (r^2 = 0.32, p = 0.24).

Enzyme expression in the TCA cycle is decidedly non-optimal. The expression of *shared* enzymes decreases with growth rate, while predictions indicate it should increase (**Fig. 3.2c**; r = -0.65, r^2 = 0.42, p = 0.011). In addition, enzyme abundance is massively higher than predicted across all growth rates (GMFE = 6.4). At the same time, *measured-only* proteins account for only a very small fraction of the pathway (10%), and the abundances of individual proteins are also correlated with measured data (r^2 = 0.38, p = 0.03).

The proteome fraction allocated to energy generation pathways (electron transport chain and ATP synthase) is almost growth rate-independent, while predictions increase with growth rate (**Fig. 3.2d**). Similar to the TCA cycle, *measured-only* proteins make up only a small fraction of the pathway (6%). Whereas *E. coli* fully oxidizes carbon sources to CO_2 at low growth rates under aerobic conditions (aerobic respiration), at high growth rates it only partially oxidizes some carbon sources (in particular glucose and fructose), leading to products such as lactate or acetate (aerobic fermentation – overflow metabolism). Along with the metabolic switch from aerobic respiration

to aerobic fermentation, the TCA cycle is gradually switched off (Basan *et al*, 2015). In our predictions, aerobic fermentation is more efficient than aerobic respiration for all conditions.

We were surprised to find that the proteins of the glyoxylate shunt (aceA, aceB, and glcB) are highly abundant at low growth rates (~12% of the proteome at μ = 0.12 h^{-1}; **Fig. 3.2e**). This high abundance at low growth rates is not specific to the BW25113 strain – it is mirrored in the MG1655 strain (**Fig. S3.6a**) (Valgepea *et al*, 2013; Peebo *et al*, 2015). Fluxomics data shows that across many conditions with low growth rates, flux into the glyoxylate shunt is roughly equal to the flux into the TCA cycle (Fischer & Sauer, 2003; Nanchen *et al*, 2006; Gerosa *et al*, 2015; Haverkorn van Rijsewijk *et al*, 2011; Ishii *et al*, 2007; Rui *et al*, 2010) (**Fig. S3.6b**). In contrast, the model predicts the glyoxylate shunt to be inactive except on acetate.

In sum, proteome allocation to the pathways of central metabolism is not well explained by optimal proteome efficiency alone, at least not as far as can be discerned with the type of model employed here. This is particularly true for the metabolic switches from aerobic respiration to aerobic fermentation and from the glyoxylate shunt to the TCA cycle.

Transporters and other pathways

The carbon source is the only nutrient that differs between the minimal media used in the proteomics experiments. To make the transporters comparable across conditions, we thus excluded inner and outer membrane transporters for all carbon sources used in the studied conditions and analyzed only the transporters for other metabolites (Methods). In stark contrast to all other pathways, the vast majority of transporters – more than 90% – are *measured-only*, i.e., the experimentally observed proteins are not part of the predicted optimal proteome (**Fig. 3.3**). Thus, transporter expression cannot be predicted from minimal proteome demand. *E. coli* does not possess external nutrient sensors, and requires the uptake of small amounts of chemicals to detect their presence. It is hence likely that many *measured-only* transporters are expressed for sensory functions to detect environmental changes, rather than to contribute to optimal growth in steady state. Moreover, *E. coli* possesses alternative transporters with different kinetics for some nutrients (for example, *E. coli* has 7 inner membrane glucose transporters (Fuentes *et al*, 2013; Monk *et al*, 2017)); prediction errors due to inaccurate enzyme turnover numbers are likely responsible for at least some fraction of the *measured-only* proteins, just as they are the likely cause of the substantial proteome fraction of *predicted-only* proteins.

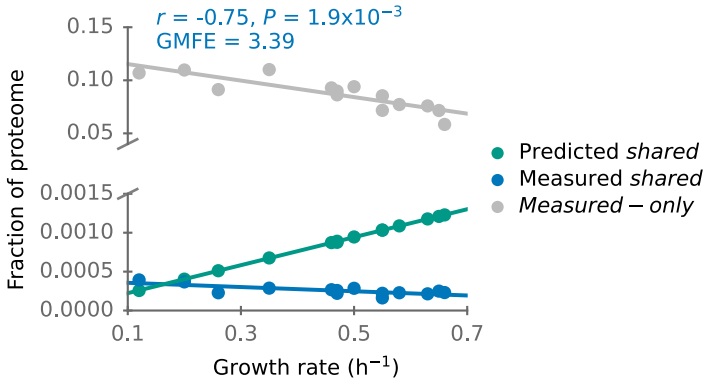


Fig. 3.3. Experimentally observed and predicted proteome fractions of transporters. Note the change in scale at the y-axis break.

The situation is very similar for those proteins that cannot be assigned to one of the pathways described above (other proteins). While the abundance of those proteins predicted to be active under optimality is very similar to the expected abundance, a much higher proteome fraction is allocated to *measured-only* proteins (**Fig. S3.5c**).

Proteome efficiency increases towards the core of biosynthesis

Combining the above results for sets of pathways, we find that proteome efficiency gradually increases along the flow of nutrients and carbon (**Fig. 3.4**). It is lowest at the interface to the environment: the vast majority of proteome allocation to transport is for proteins not expected to contribute to growth (*measured-only*) (**Fig. 3.3**). Proteome efficiency in central metabolism varies across pathways (**Fig. 3.2**): proteome allocation to glycolysis and the pentose phosphate pathway scales qualitatively as predicted from optimal demand; proteome allocation to energy production is independent of optimal demand; and the growth rate-dependent proteome allocation to the TCA cycle and glyoxylate shunt are opposite to optimal expectations. Of all metabolic sectors, the biosynthesis pathways show the best agreement between experiments and optimal predictions across growth rates (**Fig. 3.1**). Still, the predicted proteome allocation to biosynthesis pathways increases by 5.5-fold, identical to the fold change of growth rates (from 0.12 $h^{-1}$to 0.66 h^{-1}), while the observed proteome allocation changes by at most ~2 fold (for amino acid biosynthesis and cofactor biosynthesis). Finally, most of the proteome not included in one of the above metabolic categories are translation-related proteins, which have previously been shown to be expressed at maximal efficiency (Hu *et al*, 2020). Unexpectedly, by mass, about half of the metabolic pathways (including TCA cycle, glyoxylate shunt, transporters, and “the others”) show a growth rate dependence opposite of that expected from optimal demand.

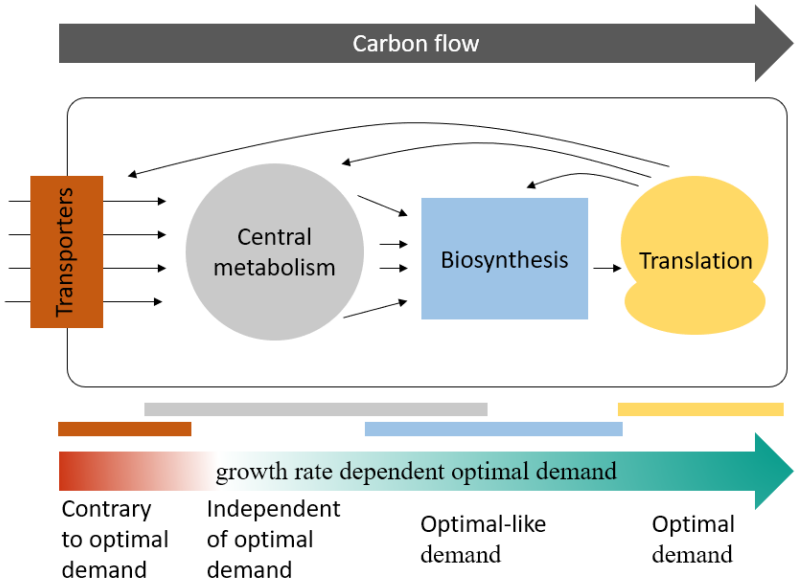


Fig. 3.4. Growth rate-dependent proteome efficiency is lowest at the interface to the environment and increases along the carbon flow.

The relationship between nutrient/carbon flow and growth rate-dependent proteome efficiency can be understood from the topology of the metabolic network. Transporters represent the metabolic interface of the cell to the environment. In the absence of external sensors, the expression of a transporter for a potential nutrient is a necessary condition for its detection by the cell; thus, non-optimal transporters serve an important cellular function unrelated to steady-state

growth. Central metabolism acts as a hub that connects all other pathways. When carbon/nutrients are transported to the cell, they are either themselves metabolites of central metabolism, or they need to be degraded (by catabolism) to such metabolites. For this reason, optimal proteome allocation to central metabolism is strongly environment-dependent. For example, when the main carbon source changes from pyruvate to glucose, *E. coli* metabolism immediately reverses its flux pattern from gluconeogenesis to glycolysis (Link *et al*, 2013). Keeping a certain fraction of the enzymes of central metabolism and of transporters in "standby" for environmental changes will thus be beneficial in transitions between cellular physiological states. Moreover, their optimal expression would require detailed, environment-dependent regulation, which may be difficult to achieve without substantial cellular investment into sensing and regulation.

In contrast, optimal resource allocation into translation and the biosynthesis (anabolic) pathways, which synthesize building blocks for the cell, is largely independent of nutrients across minimal environments, and depends almost exclusively on the growth rate. Their optimal regulation thus requires only a sensor for growth rate itself, and can be implemented relatively easily. This fact likely explains the observation of bacterial growth laws (Scott *et al*, 2010), which appear to be related to the expression of the regulatory molecule ppGpp (Potrykus *et al*, 2011). At the same time, our observations are consistent with a reserve of unused biosynthesis enzymes at low growth rates (**Fig. 3.1**), which can benefit the cell in fluctuating conditions (Mori *et al*, 2017; Korem Kohanim *et al*, 2018).

Utilization of alternative pathways cannot be explained by optimal proteome efficiency

With increasing growth rate, metabolic fluxes may shift between alternative pathways. For example, energy production from glucose switches from aerobic respiration to aerobic fermentation (overflow metabolism) (Basan *et al*, 2015). Consistent with previous studies, we found that the PEP-glyoxylate cycle (Fischer & Sauer, 2003; Gerosa *et al*, 2015) gradually switched to the TCA cycle with increasing growth rate (**Fig. S3.6**).

Neither aerobic respiration nor the glyoxylate shunt are used in the predicted flux distributions. In constraint-based models, overflow metabolism emerges when an additional growth-limiting constraint becomes active (de Groot *et al*, 2020). While it is likely that overflow metabolism is rooted in a limit on proteome investment into catabolic enzymes (Schuster *et al*, 2015; Basan *et al*, 2015), this effect cannot be reproduced without empirical adjustments to the model. One way of enforcing aerobic fermentation is to impose a decrease in proteome usage and an increase in energy production with increasing growth rate (Basan *et al*, 2015; Zeng & Yang, 2019); an alternative is to allocate a constant empirical mass of proteins to energy production (Chen & Nielsen, 2019).

The PEP-glyoxylate cycle, which contains the glyoxylate shunt, represents an alternative route to the TCA cycle (Fischer & Sauer, 2003). Compared to the TCA cycle, the PEP-glyoxylate cycle produces an additional NADH instead of one NADPH (Fischer & Sauer, 2003). Since NADPH is a common cofactor in anabolic pathways in *E. coli*, it was suggested that the cell should choose the pathway which can produce more NADPH (the TCA cycle) at high growth rates (Fischer & Sauer, 2003). But the interconversion between NADPH and NADH is a very common process in *E. coli* (Sauer *et al*, 2004), and it is not clear how the small difference in pathway output (1 NADPH vs. 1 NADH) could explain the massive resource allocation (~ 12% of the proteome) into the glyoxylate shunt at low growth rates. A recent study showed that overexpression of glyoxylate shunt

enzymes can reduce the lag time when *E. coli* experiences a transition from a glycolytic carbon source to a gluconeogenic carbon source (Basan *et al*, 2020). However, it is still challenging to develop mechanistic models that explain the growth rate-dependent expression of alternative pathways and lag times from first principles.

Resource allocation in *E. coli* is not optimized for instantaneous growth

An unsolved mystery is why *E. coli* grows at radically different rates on carbon sources with very similar pathway usages. For example, the observed growth rate on galactose is much lower than that on glucose (0.26 h^{-1} vs. 0.58 h^{-1}) (Schmidt *et al*, 2016), although the only proteins expected to be differentially utilized are transporters and the galactose degradation pathway; all other active metabolic pathways are the same. Accordingly, proteome efficiency alone is insufficient to explain the growth rate difference: assuming a total protein concentration of 0.2 g/g_{DW}, the predicted optimal growth rate is 0.80 h^{-1} on galactose and 0.87 h^{-1} on glucose. This observation indicates that *E. coli*'s pathway usage is not regulated for maximal growth on galactose. A potential reason may be that *E. coli* might rarely encounter galactose over prolonged time periods in natural environments, so that it is more beneficial for the cell to invest already in proteins that may accelerate growth in the next environment (Basan *et al*, 2020).

What actually limits the total proteome that can be allocated to different cellular processes? Historically, protein translation has been assumed to be the rate-limiting factor for bacterial growth, as the output of translation matches ribosome abundance very well across growth rates (Koch, 1988; Maaløe & Kjeldgaard, 1966). However, as shown above, the near-optimal expression of the ribosome and other biosynthesis pathways only means that the pathway is expressed close to the locally optimal level at the given pathway output; for example, ribosomal proteins are optimally abundant to achieve the observed protein translation rates, but global shifts in resource allocation – including higher investment into ribosomes – would facilitate faster growth. Moreover, many pathways – in particular transporters and central metabolism – are expressed independent of or even contrary to optimal demand. Thus, the tight relationship between ribosome content and growth rate (Koch, 1988) only means that the translation machinery is optimized for the observed growth rate, but does not mean that the growth rate is fundamentally limited by the cell's ability to make ribosomes. Instead, *E. coli* appears to limit its growth rate actively on poor carbon sources by allocating resources to pathways that do not contribute to maximal instantaneous growth.

Methods

Growth rate-dependent biomass composition

The original biomass composition in the *i*ML1515 model is very similar to that of the iAF1260 model, formulated for a doubling time of 40 min or μ = 1.04 h^{-1} (Feist *et al*, 2007). However, biomass composition varies across growth rates. The two most significant changes are those of the RNA/protein mass ratio and the cell volume, which determines the surface/volume ratio (S/V). Both ratios can be expressed as functions of the growth rate; accordingly, we estimated the growth rate-dependent biomass fraction of RNA, protein, and cell envelope components (including murein, lipopolysaccharides, and lipid) as functions of the growth rate, as described below.

We first fitted experimental data for the RNA/protein mass ratio ($\frac{\mathrm{m_{RNA}}}{\mathrm{m_{protein}}}$) (Scott *et al*, 2010; Dai *et al*, 2016) and the surface/volume ratio (S/V) (Si *et al*, 2017) to linear functions of the growth rate, resulting in the relationships

$$\frac{m_{RNA}}{m_{protein}}(\mu) = 0.223\mu + 0.08 \quad , \tag{3.1}$$

$$\frac{S}{V}(\mu) = -0.1895\mu + 7.952 \quad . \tag{3.2}$$

Assuming that the biomass contribution of cell envelope components ($m_{envelope}$) is proportional to the surface/volume ratio gives

$$\frac{m_{envelope}(\mu = \mu_1)}{m_{envelope}(\mu = \mu_2)} = \frac{\frac{S}{V}(\mu = \mu_1)}{\frac{S}{V}(\mu = \mu_2)} \quad . \tag{3.3}$$

The growth rate-dependent biomass fraction of cell envelope components ($m_{envelope}$) can then be estimated by equation (3.3) given equation (3.2) and $m_{envelope}$ at μ = 1.04 h^{-1}. The relative composition of murein, lipopolysaccharides, and lipid was assumed to be constant.

The biomass fractions of cellular components other than RNA, protein, and cell envelope components (m_{others}) were assumed to be independent of the growth rate. The sum of RNA and protein is given by:

$$m_{RNA} + m_{protein} = 1 - m_{others} - m_{envelope} \tag{3.4}$$

Combining equation (3.1) and (3.4), the content of RNA and protein can be calculated for all conditions (Fig. S3.1). The relative contributions of individual nucleotides to total RNA and of individual amino acids to total protein were assumed to be growth rate-independent.

Implementation of MOMENT

To perform flux balance analysis with molecular crowding, we used ccFBA (Desouki, 2016), which implements the MOMENT algorithm (Adadi *et al*, 2012) with an improved treatment of co-functional enzymes (Heckmann *et al*, 2018). For enzymes for which maximal *in vivo* effective enzyme turnover numbers ($k_{app,max}$) were available from (Heckmann *et al*, 2020), we used these to replace the original *in vitro* k_{cat} values.

Instead of maximizing the growth rate at a given nutrient condition, we solved the complementary optimization problem that estimates the minimal required proteome (C) able to support the observed growth rate on the given carbon source. In the constraint-based type of model employed here, there is a linear relationship between proteome investment and predicted growth rate, $C = a\mu + b$ for two constants *a,b*. Note that due to a non-zero non-growth-related maintenance energy term included in the model, *b*>0. The constants *a* and *b* can be determined by any two pairs of proteome budget and growth rate. Here, we set *C*=0.1 g/g_{DW} ($C_{0.1}$) and *C*=0.2 g/g_{DW} ($C_{0.2}$) and calculated the predicted growth rates (denoted as $\mu_{0.1}$ and $\mu_{0.2}$, respectively).

Given the observed growth rate (μ'), the minimal required proteome at μ' (C') can be written as

$$C' = a\mu' + b. \tag{3.5}$$

For a given protein *i*, its minimal demand at the observed growth rate μ' ($p_{i,\mu'}$) in units of g/g_{DW} can be expressed as

$$p_{i,\mu'} = \frac{C'}{C_{0.1}} p_{i,\mu_{0.1}} \tag{3.6}$$

with $p_{i,\mu_{0.1}}$ the minimal demand for protein *i* at $C_{0.1}$.

With the protein content in dry mass at μ' ($m_{protein,\mu'}$) estimated in equation (3.4), the proteome fraction of protein *i* at μ' ($m_{i,\mu'}$) can be written as

$$m_{i,\mu'} = \frac{p_{i,\mu'}}{m_{protein,\mu'}}. \tag{3.7}$$

Pathway membership

Proteins were characterized as transporters if the corresponding genes are assigned to transport reactions according to the *i*ML1515 annotation (Monk *et al*, 2017).

Proteins are labeled as biosynthetic enzymes based on the EcoCyc pathway ontology annotation "biosynthesis" (Keseler *et al*, 2017). Pathways in this category are: (1) Amino acid biosynthesis ("Amino Acid Biosynthesis" in EcoCyc), (2) nucleotide biosynthesis ("Nucleoside and Nucleotide Biosynthesis"), (3) cofactors ("Cofactor, Carrier, and Vitamin Biosynthesis"), and (4) cell wall components ("Cell Structure Biosynthesis and Fatty Acid and Lipid Biosynthesis"), including lipid, peptidoglycan, and LPS. All other biosynthetic enzymes are merged into (5) other biosynthetic pathways.

Treatment of enzymes involved in the nucleotide salvage pathway: In the range of studied growth rates, the transcription of mRNA accounts for more than half of the total RNA transcription (Bremer & Dennis, 2008). The half-life of mRNA is very short (~ 5.5 min) (Bernstein *et al*, 2002) compared to the doubling time, and degraded mRNA will be reused through the nucleotide salvage pathway. However, our model only predicts the expression of *de novo* biosynthesis pathways. To make the prediction comparable with the observed data, the nucleotide salvage pathway was thus excluded from "nucleotide biosynthesis pathway".

Enzymes are designated as being involved in precursors and energy generation according to the EcoCyc pathway ontology annotation "Generation of Precursor Metabolites and Energy". Pathways in this category are: (1) glycolysis, (2) Pentose Phosphate Pathways, (3) TCA cycle, (4) glyoxylate bypass (EcoCyc does not list a pathway for the glyoxylate shunt; the three genes classified as glyoxylate shunt are aceA, aceB, and glcB), (5) energy production ("Electron Transfer Chains and ATP biosynthesis"), and (6) other enzymes.

The pathway ontology used for these classifications was downloaded from EcoCyc (Keseler *et al*, 2017) on 13. January 2021.

Supplementary information

Supplementary Figures

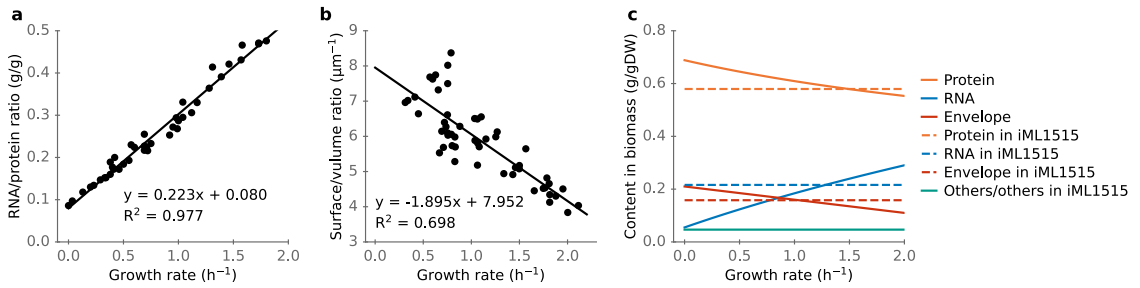


Fig. S3.1. Growth rate-dependent biomass composition. (**a**) Growth rate-dependent RNA/protein mass ratio; data from (Scott *et al*, 2010; Dai *et al*, 2016). (**b**) Growth rate-dependent surface/volume ratio (wildtype strain in non-stressed growth conditions); data from (Si *et al*, 2017). **(c)** Growth rate-dependent biomass composition, re-formulated considering the growth rate-dependent RNA/protein ratio shown in panel a and the surface/volume ratio shown in panel b (see Methods). The category "Envelope" includes murein, lipopolysaccharides, and lipids. The content of protein and cell envelope decrease with increasing growth rate, while the RNA content increases sharply with growth rate due to the increased abundance of ribosomal RNA and tRNA.

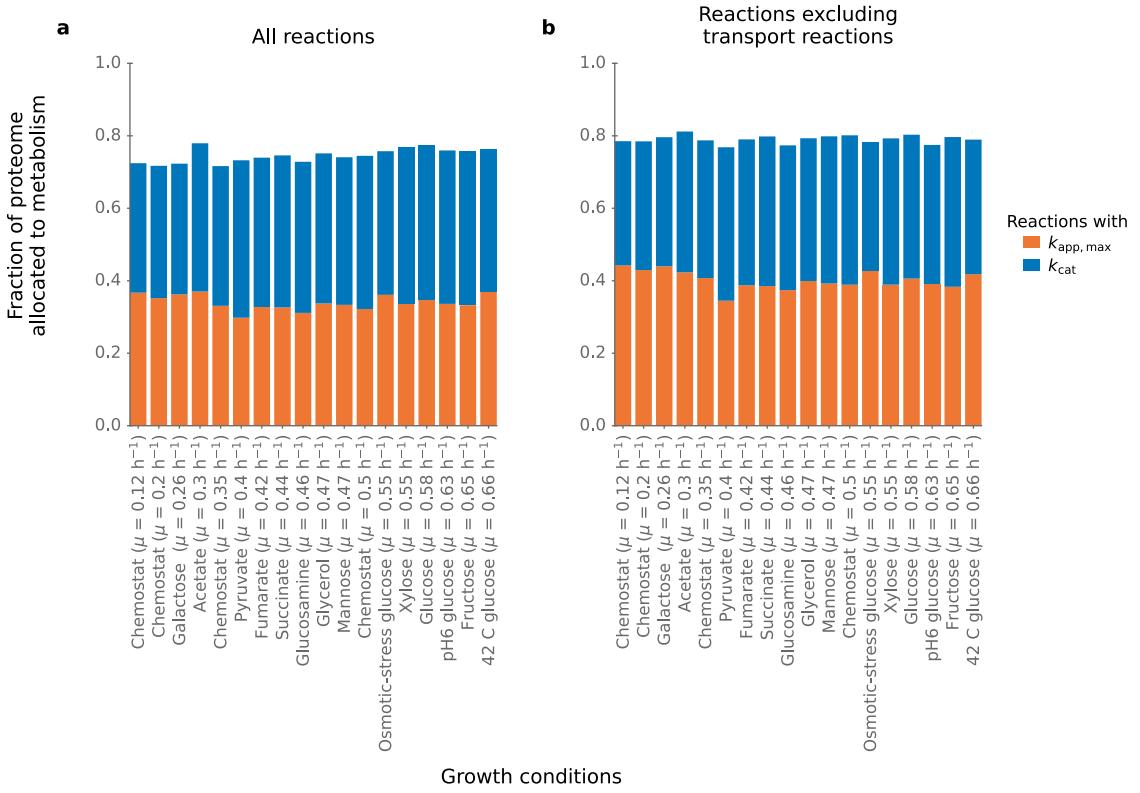


Fig. S3.2. Reactions with measured enzyme turnover numbers ($k_{app,max}$ and k_{cat}) cover more than 70% of proteome allocated to metabolism. Different colors show the proteome fractions of enzymes with *in vivo* enzyme turnover number ($k_{app,max}$) and in vitro turnover number (k_{kcat}), respectively.

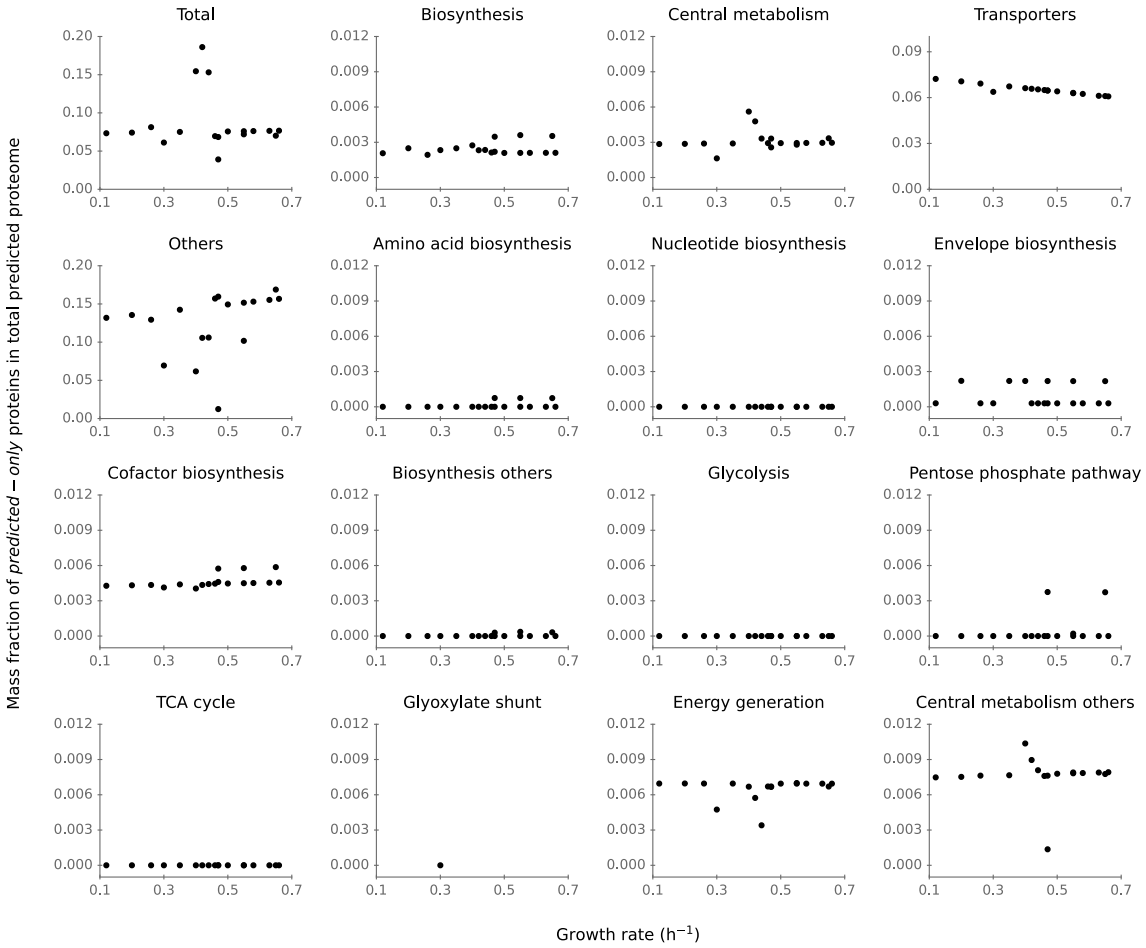


Fig. S3.3. The mass fraction of predicted-only proteins in total predicted proteome for the assessed pathways. Except transporters and proteins cannot be assigned to particular pathways (Others), the predicted-only proteins accounts less than 1% of the total predicted proteome in the given pathway.

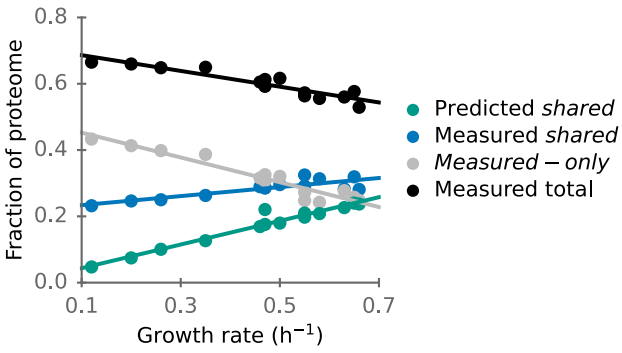


Fig. S3.4. Proteome efficiency of the whole metabolic network. With increasing growth rate, predicted and experimentally observed *shared* metabolic proteome fractions increase, while the *measured-only* metabolic proteome fraction decreases.

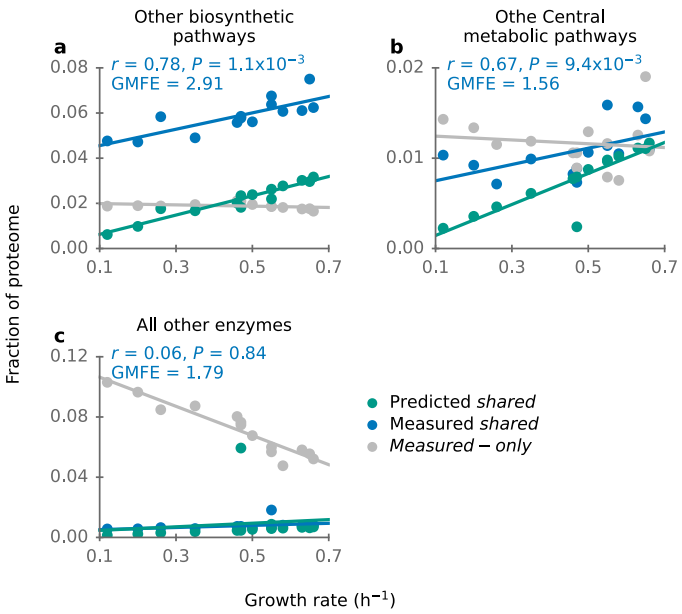


Fig. S3.5. The proteome efficiency of rest proteins. (**a**) Proteome investment into biosynthetic proteins that are not part of the four pathway sets in Fig. 3.1. This set is by definition heterogeneous. While its predicted and observed expression of *shared* enzymes are strongly correlated (r^2 = 0.61, p = 1.1x10^{-3}), the corresponding values differ on average by almost 3-fold (GMFE = 2.91). However, the expression of individual enzymes can be well explained by prediction (r^2 = 0.46, p = 5.4×10^{-5}). (**b**) Proteome investment into central metabolism that are not part of the five pathway sets in Fig. 3.2. The prediction increases with growth rate whereas the measured data is independent of growth rate. (**c**) All the other enzymes in *i*ML1515 model. While the abundance of those proteins predicted to be active under optimality is very similar to the expected abundance, a much higher proteome fraction is allocated to *measured-only* proteins.

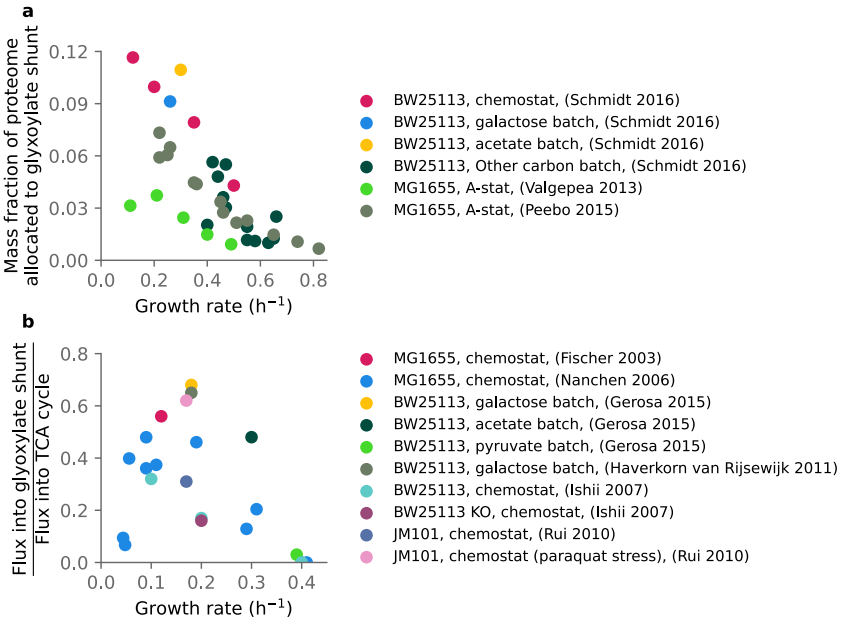


Fig. S3.6. Growth rate-dependent abundance and activity of glyoxylate shunt. (**a**) Proteome allocated to glyoxylate shunt decreases with increasing growth rate; data from (Schmidt *et al*, 2016; Valgepea *et al*, 2013; Peebo *et al*, 2015). (**b**) Flux diverted to glyoxylate shunt from TCA cycle decreases with increasing growth rate; data from (Fischer & Sauer, 2003; Nanchen *et al*, 2006; Gerosa *et al*, 2015; Haverkorn van Rijsewijk *et al*, 2011; Ishii *et al*, 2007; Rui *et al*, 2010).

Table S3.1. Proteome efficiency of pathways on both glycolytic and gluconeogenic carbon sources.

Pathway	Pathway expression (for *shared* proteins) (n = 18)[α]			*measured-only* fraction[β] (median across 18 conditions)	individual *shared* proteins; median across 18 conditions[γ]		
	Pearson's *r*	*p*	GMFE		r^2	*p*	n^{δ}
Measures (a)-(d)	(a)	(a)	(b)	(c)	(d)	(d)	(d)
Biosynthesis	**0.88**	**1.2×10^{-06}**	**1.77**	**0.27**	**0.44**	**1.7×10^{-30}**	**229**
Amino acid	0.87	2.9×10^{-06}	1.39	0.31	0.45	8.6×10^{-11}	72
Nucleotide	0.76	2.6×10^{-04}	3.42	0.23	0.15	0.05	28
Envelope	0.62	6.3×10^{-03}	1.81	0.14	0.38	2.3×10^{-05}	40
Cofactor	0.91	1.2×10^{-07}	1.23	0.11	0.62	5.5×10^{-16}	70
Biosynthesis others	0.35	0.16	3.88	0.26	0.44	8.5×10^{-05}	29
Central metabolism	**0.07[ϵ]**	**0.79**	**2.39**	**0.31**	**0.15**	**3.2×10^{-03}**	**56**
Glycolysis	0.86	4.0×10^{-06}	3.44	0.08	0.30	0.08	11
Pentose phosphate pathway	0.43	0.07	1.92	0.38	0.30	0.34	5
TCA cycle	-0.21[ϵ]	0.42	5.80	0.10	0.38	0.03	12
Glyoxylate shunt	-	-	-	1	-	-	0
Energy generation	-0.06	0.80	1.57	0.05	0.13	0.06	28
Central metabolism others	0.83	2.5×10^{-05}	1.50	0.55	0.99	0.06	3
Transporters	**-0.70**	**1.2×10^{-03}**	**3.30**	**0.997**	**0.10**	**0.69**	**4**
Others	**0.07**	**0.78**	**1.62**	**0.92**	**0.14**	**0.05**	**30**
Total	**0.71**	**9.6×10^{-04}**	**1.81**	**0.52**	**0.35**	**3.4×10^{-30}**	**308**

[α] Values reflect the local optimality of complete pathways across conditions. *n* = 18 indicates the number carbon (both glycolytic and gluconeogenic) sources analyzed.

[β] Mass fraction of *measured-only* (un-predicted but observed) proteins relative to all proteins in the pathway.

[γ] These columns reflect the local optimality compared across individual proteins within each pathway at a given growth condition; values are medians across the *n* = 18 all growth conditions.

[δ] Number of proteins in each pathway or pathway set.

[ϵ] Data on acetate was excluded for calculating Pearson's *r* because it is an outlier.

4 Manuscript 3. An optimal growth law for RNA composition and its partial implementation through ribosomal and tRNA gene locations in bacterial genomes

This manuscript was adapted from the following publication:

Hu X-P, Lercher MJ. An optimal growth law for RNA composition and its partial implementation through ribosomal and tRNA gene locations in bacterial genomes. *PLOS Genetics*. 2021;17: e1009939. https://doi.org/10.1371/journal.pgen.1009939.

Contribution: I designed the study, performed the analyses, and drafted the manuscript.

Abstract

The distribution of cellular resources across bacterial proteins has been quantified through phenomenological growth laws. Here, we describe a complementary bacterial growth law for RNA composition, emerging from optimal cellular resource allocation into ribosomes and ternary complexes. The predicted decline of the tRNA/rRNA ratio with growth rate agrees quantitatively with experimental data. Its regulation appears to be implemented in part through chromosomal localization, as rRNA genes are typically closer to the origin of replication than tRNA genes and thus have increasingly higher gene dosage at faster growth. At the highest growth rates in *E. coli*, the tRNA/rRNA gene dosage ratio based on chromosomal positions is almost identical to the observed and theoretically optimal tRNA/rRNA expression ratio, indicating that the chromosomal arrangement has evolved to favor maximal transcription of both types of genes at this condition.

Author summary

Unlike the proteome composition, RNA composition is often assumed to be independent of growth rate in bacteria, despite experimental evidence for a growth rate dependence in many microbes. In this work, we derived a growth rate-dependent optimal tRNA/rRNA concentration ratio by minimizing the combined costs of ribosome and ternary complex at the required protein production rate. The predicted optimal tRNA/rRNA expression ratio, which is a monotonically decreasing function of growth rate, agrees with experimental data for *E. coli* and other fast-growing microbes. This indicates the existing of an RNA composition growth law. Due to the presence of partially replicated chromosomes, gene dosage is higher for those genes whose DNA is replicated earlier, an effect that becomes stronger at higher growth rates. Because rRNA genes are located closer to origin of replication than tRNA genes in fast-growing species, the tRNA/rRNA gene dosage ratio scales with growth rate in the same direction as the optimal tRNA/rRNA expression ratio. Thus, it appears that the RNA growth law is – at least in part – implemented simply through the genomic positions of tRNA and rRNA genes. This finding indicates that growth rate-dependent optimal resource allocation can influence the genomic organization in bacteria.

Introduction

The systematic change of the coarse-grained composition of bacterial proteomes with growth rate (Schaechter *et al*, 1958; Bremer & Dennis, 2008) can be quantified through phenomenological growth laws (Scott *et al*, 2010; You *et al*, 2013). The most prominent growth law describes an apparently linear increase of the ribosomal protein fraction with growth rate (Schaechter *et al*, 1958; Scott *et al*, 2010). These laws have been successfully applied to the prediction of a range of phenotypic observations (Scott *et al*, 2010; Klumpp *et al*, 2013; Dai *et al*, 2016; Erickson *et al*, 2017; Mori *et al*, 2017). Recently, it has been argued that they arise from an optimal balance between the cellular investment into catalytic proteins and their substrates (Dourado *et al*, 2021).

In contrast to the proteome composition, the partitioning of bacterial RNA into messenger (mRNA), ribosomal (rRNA), and transfer (tRNA) RNA is often assumed to be growth rate-independent (Scott *et al*, 2010; Klumpp *et al*, 2013; O'Brien *et al*, 2013; Bosdriesz *et al*, 2015; Dai *et al*, 2016; Bremer & Dennis, 2008). For example, the assumption of a constant RNA composition has been used to estimate an empirical relationship for the macromolecular cellular composition across bacterial species (Kempes *et al*, 2012, 2016). However, experimental evidence from multiple species suggests that the tRNA/rRNA expression ratio decreases monotonically with growth rate (Kjeldgaard & Kurland, 1963; Doi & Igarashi, 1964; Panos *et al*, 1965; Rosset *et al*, 1966; Sykes & Young, 1968; Alberghina *et al*, 1975; Waldron & Lacroute, 1975; Furano, 1975; Dong *et al*, 1996), suggesting the existence of a bacterial growth law for RNA composition.

The regulatory implementation of bacterial growth laws is generally assumed to arise from a small number of major transcriptional regulators such as ppGpp (Potrykus *et al*, 2011; Zhu & Dai, 2019) and cAMP (You *et al*, 2013; Towbin *et al*, 2017). However, growth rate-dependent transcriptional regulation could also be implemented through chromosomal gene positioning. In many prokaryotes, the cellular doubling time can be even shorter than the time required for genome replication. To coordinate DNA replication and cell division, fast-growing prokaryotes re-initiate DNA replication before the previous round of replication is complete. In this case, genes closer to oriC have more DNA copies than genes further away in the genome, a phenomenon described as replication-associated gene dosage effects (below, we use "gene dosage" to refer to the growth rate-dependent average DNA copy number per cell of a given gene). Prokaryotic genes are non-

randomly located on multiple levels (Rocha, 2008; Pang & Lercher, 2017; Gao *et al*, 2017), with highly expressed genes biased towards the origin of replication (oriC) (Couturier & Rocha, 2006). The latter observation is thought to facilitate high expression levels at fast growth due to replication-associated gene dosage effects (Yoshikawa *et al*, 1964; Cooper & Helmstetter, 1968; Bremer & Churchward, 1977). Indeed, chromosome rearrangements that shift highly expressed genes from the origin to the terminus of replication reduce fitness (Kothapalli *et al*, 2005; Campo *et al*, 2004; Hill & Gray, 1988; Louarn *et al*, 1985; Soler-Bistué *et al*, 2017).

rRNA forms the central part of the catalyst of peptide elongation, while tRNA forms the core of the substrate; together, they account for the bulk of cellular RNA (Bremer & Dennis, 2008). Their cytosolic concentrations at different growth rates in *E. coli* are well described by an optimality assumption (Dourado *et al*, 2021; Dourado & Lercher, 2020; Hu *et al*, 2020). Moreover, chromosomal gene positions in *E. coli* are known to affect the expression of both tRNA and rRNA genes (Condon *et al*, 1992; Ardell & Kirsebom, 2005); both types of genes are located closer to oriC in fast- compared to slow-growing bacteria, with rRNA genes positioned closer to oriC than tRNA genes in most examined fast-growing bacteria (Couturier & Rocha, 2006).

Based on these previous observations, we hypothesize (i) that the relative expression of tRNA and rRNA can be described by a bacterial growth law that arises from optimal resource allocation and (ii) that this growth law is at least partially implemented through the relative chromosomal positioning of tRNA and rRNA genes.

Results and discussion

An RNA growth law resulting from maximal efficiency of translation

Cellular dry mass density appears to be approximately constant across conditions (Kubitschek *et al*, 1984; Oldewurtel *et al*, 2021). Dry mass may thus be considered a limiting resource (Dourado *et al*, 2021; Hu *et al*, 2020) if the dry mass density is occupied by one particular molecule, less will be available for all other molecules. In terms of dry mass allocation, translation is the most expensive process in fast-growing bacteria (Bremer & Dennis, 2008; Russell & Cook, 1995). Thus, at a given protein synthesis rate, it is likely that natural selection will act to minimize the summed dry mass density of all translational components. As evidenced by comparison of diverse data to a detailed biochemical model of translation, the allocation of cellular resources across components of the *E. coli* translation system minimizes their total dry mass concentration at a given protein production rate (Hu *et al*, 2020). This result indicates that natural selection indeed favored the parsimonious allocation of cellular resources to the translation machinery in *E. coli*.

To generalize this optimization hypothesis to other species, we here analyze a coarse-grained translation model that only considers peptide elongation, where the active ribosome acts as an enzyme that converts ternary complexes (TC), consisting of elongation factor Tu (EF-Tu), GTP, and charged tRNA, into an elongating peptide chain following Michaelis-Menten kinetics (**Fig. 4.1A**) (Klumpp *et al*, 2013; Wong *et al*, 2018). In exponential, balanced growth at rate μ with cellular protein concentration $[P]$, the total rate of protein production is $v = \mu \cdot [P]$. We derived the optimal concentration ratio between TC (with molecular mass m_{TC}) and ribosome (*R*, with molecular mass m_{R}) at this production rate by minimizing their combined mass concentration, $M_{\mathrm{total}} = m_{\mathrm{TC}}[TC] + m_{\mathrm{R}}[R]$ (Methods):

$$\frac{[TC]}{[R]} = \frac{a \cdot k_{\mathrm{cat}}}{\sqrt{a \cdot \mu \cdot [P] \cdot k_{\mathrm{on}}^{\mathrm{diff}} + k_{\mathrm{cat}}}} \quad ; \tag{4.1}$$

here, $a = \frac{m_{\mathrm{R}}}{m_{\mathrm{TC}}} = 33.1$ is the ratio of molecular weights of ribosome and TC; k_{cat} is the turnover number of the ribosome; and $k_{\mathrm{on}}^{\mathrm{diff}}$ is the diffusion-limited binding constant of TC to ribosome (Klumpp *et al*, 2013), which can be treated as a constant if the cell density is approximately constant across species.

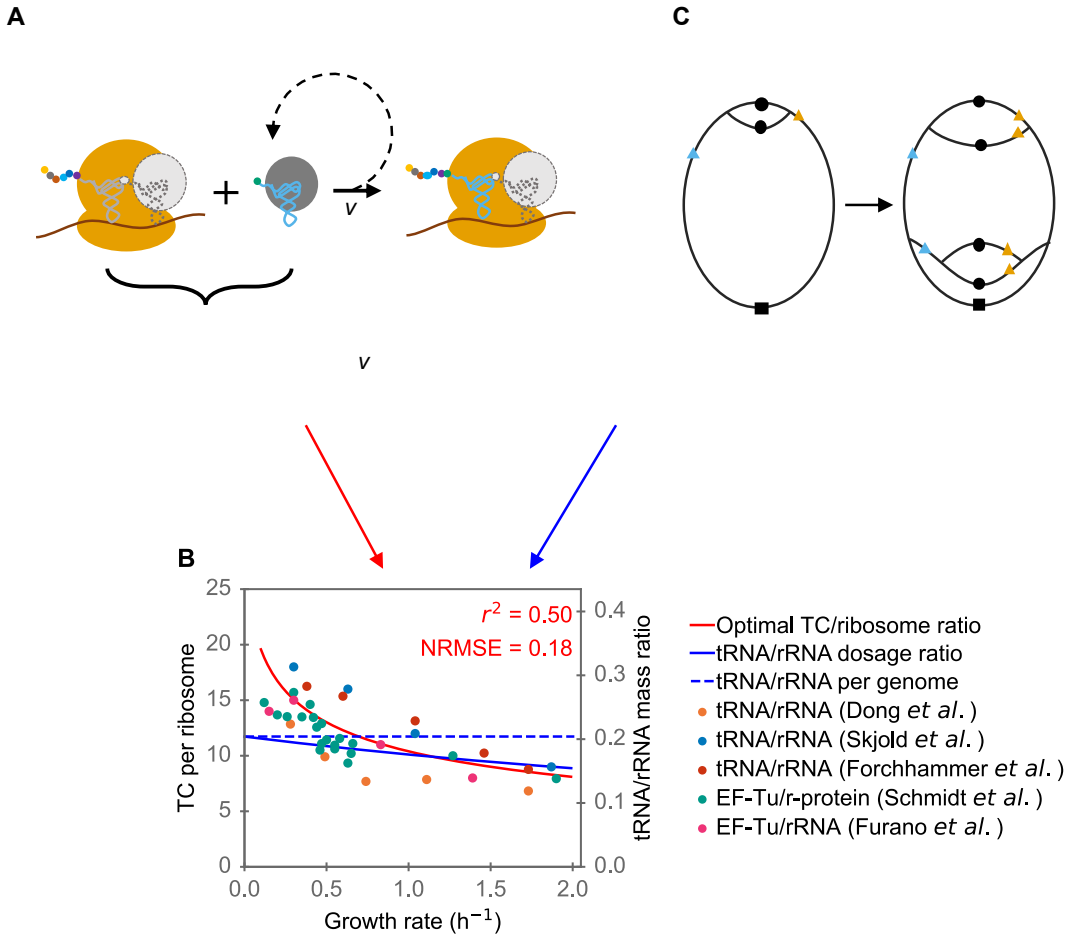


Fig. 4.1. The RNA growth law and its implementation through gene positions. (**A**) Coarse-grained protein translation model, following Michaelis-Menten kinetics with the active ribosome as catalyst and TC as substrate. The optimal TC/ribosome expression ratio is derived by minimizing the combined mass concentration of ribosome and TC at the given protein synthesis rate *v*. (**B**) Different experimental estimates of TC/ribosome expression ratios in *E. coli* (points, colors indicate the data source) are consistent with the optimal ratio according to equation (4.1) (red line) (Pearson's r^2 = 0.50; NRMSE = 0.18). The dashed blue line indicates the genomic tRNA/rRNA ratio, the solid blue line indicates the tRNA/rRNA gene dosage ratio estimated from equation (4.21). (**C**) A schematic diagram showing the dosage ratio of two genes as a function of growth rate. If rRNA genes are located on average closer to oriC than tRNA genes – which is the case in *E. coli* – then the dosage of rRNA genes will increase faster with increasing growth rate than that of the tRNA genes; consequently, the tRNA/rRNA gene dosage ratio becomes a decreasing function of growth rate (solid blue curve in panel B).

For a given genome, *a* and k_{cat} are constants (Klumpp *et al*, 2013; Dai *et al*, 2016). Moreover, the cellular protein concentration [*P*] (in terms of amino acid residues) appears to be similar across

most species (Milo, 2013) and shows only minor variations across growth rates in those bacteria where it has been tested (Erickson *et al*, 2017; Dauner & Sauer, 2001; Hanegraaf & Muller, 2001). Thus, equation (4.1) predicts that in any given species, the TC/ribosome expression ratio is a monotonically decreasing function of the growth rate μ. Since most cellular EF-Tu and tRNA are present in the form of TCs (Klumpp *et al*, 2013), hereafter, the TC concentration is assumed to be approximately equal to the concentrations of EF-Tu and tRNA.

To calculate the optimal TC/ribosome expression ratio in *E. coli*, we use the measured protein concentration [*P*] (Schmidt *et al*, 2016), set the turnover number k_{cat} to the maximal observed translation rate (Bremer & Dennis, 2008), and set $k_{\text{on}}^{\text{diff}}$ to the diffusion limit of the TC (Klumpp *et al*, 2013) (Methods; see also Ref. (Hu *et al*, 2020)). **Fig. 4.1B** compares the optimal predictions (red line) to experimental datasets for *E. coli* that estimated the TC/ribosome expression ratio based on ratios of tRNA/rRNA (Dong *et al*, 1996; Skjold *et al*, 1973; Forchhammer & Lindahl, 1971), EF-Tu/rRNA (Furano, 1975), and EF-Tu/ribosomal proteins (Schmidt *et al*, 2016) (Table S4.1). The Pearson correlation between observed and fitted data is $r^2 = 0.50$, $P = 5.9\times10^{-7}$ (root-mean-square error normalized by observed mean, NRMSE = 0.18); these measures have to be interpreted against the variability between the diverse datasets. Consistent with the predictions, all experimental estimates of the TC/ribosome expression ratio are approximately two-fold higher at low compared to high growth rates. As the TC and ribosome constitute the two major components of cellular RNA (Bremer & Dennis, 2008), we conclude that the optimal TC/ribosome expression ratio according to equation (4.1) represents a bacterial growth law for RNA composition:

$$\frac{M_{\text{tRNA}}}{M_{\text{rRNA}}} = r \cdot \frac{k_{\text{cat}}}{\sqrt{a \cdot \mu \cdot [P] \cdot k_{\text{on}}^{\text{diff}}} + k_{\text{cat}}} \quad , \tag{4.2}$$

where M_{tRNA} and M_{rRNA} are the cellular mass of tRNA and rRNA, respectively, and r=0.58 is the ratio of the tRNA mass fraction of a TC and the rRNA mass fraction of the bacterial ribosome (Methods).

The proteome degradation rate in *E. coli* is typically 0.02-0.04 h^{-1} (Pine, 1970, 1973; Calabrese *et al*, 2021), which is much smaller than the maximal growth rate. Accordingly, including protein degradation into the model only affects the predictions at very low growth rates in *E. coli* (**Fig. S4.1**). In contrast, protein degradation may have a large impact on the RNA growth law for species with degradation rates comparable to their maximal growth rates. Further, while our model assumes that all tRNA and ribosome are active, there is evidence for a substantial fraction of de-activated ribosomes and TCs at low growth rates in *E. coli* (Hu *et al*, 2020). This approximation may contribute to the discrepancy between our predictions and data at low growth rates.

In previous work by Klumpp *et. al*., the optimal TC/ribosome expression ratio was predicted by considering protein mass instead of dry mass as the limiting resource (Klumpp *et al*, 2013); these authors identified the proteome fractions allocated to ribosomes and TCs that maximize growth rate in a very similar model of protein translation to that used here. This optimal proteome allocation results in a substantial lower predicted TC/ribosome expression ratio compared to the experimentally observed data (**Fig. S4.2**). Our hypothesis of parsimonious dry mass allocation, which considers RNA and protein masses equally, explains the observed TC/ribosome expression ratio much better than optimal proteome allocation alone.

The RNA growth law is partially implemented through genomic positions in *E. coli*

Above, we have shown the existence of an RNA growth law in *E. coli*, reflecting a decrease of the optimal tRNA/ribosome expression ratio with increasing growth rate. Given that the genomic position of rRNA genes is typically closer to oriC than that of tRNA genes in bacteria (Couturier & Rocha, 2006), we hypothesize that this growth rate-dependence may – at least in part – be implemented through replication-associated gene dosage effects.

To test our hypothesis, we used the model developed by Bremer and Churchward (Bremer & Churchward, 1977) to quantify the dosage ratio of two genes at growth rate μ,

$$\frac{\overline{X_i}}{\overline{X_j}} = e^{C \cdot \mu \cdot (position_j - position_i)} \quad ; \tag{4.3}$$

here, for gene *i*, $\overline{X_i}$ is the dosage and $position_i$ is the position; *C* is the time required to complete one round of chromosome replication (see Methods for details, and see **Text S4.1** for the effect of a growth rate-dependent C period on the dosage ratio for tRNA and rRNA genes). Clearly, the dosage ratio of two genes with different chromosomal positions is a monotonous function of μ. As shown schematically in **Fig. 4.1C**, if a rRNA gene is located closer to oriC than a tRNA gene, the tRNA/rRNA gene dosage ratio (reflecting chromosomal copy numbers) will be a decreasing function of growth rate, just as the optimal tRNA/rRNA expression ratio (reflecting RNA production; **Fig. 4.1B**).

Consistent with a (partial) implementation of the RNA growth law through genomic positioning, the rRNA genes are, on average, located closer to oriC than tRNA genes in *E. coli*, with genomic position 0.20 ± 0.17 (mean ± standard deviation) for rRNA genes and 0.45 ± 0.27 for tRNA genes (see **Fig. S4.3** for the distributions). The difference in genomic positions between tRNA and rRNA genes results in a growth rate-dependent tRNA/rRNA gene dosage ratio (solid blue curve in **Fig. 4.1B**) that agrees qualitatively with the optimality predictions from equation (4.2) (to calculate the dosage ratio across multiple genes, we used equation (4.21), a generalized version of equation (4.3), see Methods). For comparison, **Fig. 4.1B** also shows the constant tRNA/rRNA genomic ratio, i.e., the ratio of gene copy numbers per complete chromosome (dashed blue line).

As all necessary parameters are available for *E. coli*, we can make quantitative predictions for the tRNA/rRNA expression ratio without adjustable parameters. It is notable that according to **Fig. 4.1B**, the tRNA/rRNA gene dosage ratio at high growth rates ($1\ \mathrm{h}^{-1} \leq \mu \leq 2\ \mathrm{h}^{-1}$) is very close to the optimal tRNA/rRNA expression ratio, which corresponds to about 9 tRNAs per ribosome (**Fig. 4.1B**). This result is consistent with the notion that at the highest growth rates, both tRNA and rRNA genes are transcribed at the maximal possible rate, such that their relative expression is dominated by gene dosage effects in these conditions. The expression of both tRNA and rRNA operons is regulated by the P1 promoter, which is repressed by ppGpp; at near-maximal growth rates, ppGpp concentrations are low, and the P1 promoter works near its maximal capacity (Gourse *et al*, 1996). In contrast, at low growth rates, P1 is repressed by ppGpp, and thus gene dosage can only partially explain the tRNA/rRNA expression ratio in these conditions.

The RNA growth law in fast-growing microbes beyond *E. coli*

The approximate Michaelis-Menten form of the rate law for peptide elongation, on which the RNA composition growth law is based, arises from the structure of the detailed elongation process (Wong *et al*, 2018). As this process is shared by all living cells (Wong *et al*, 2018), we expect that the RNA composition growth law, equation (4.2), also holds for other fast-growing microbes (with *a*=40.3 and *r*=0.59 in eukaryotes, Methods). To test this hypothesis, we collected all available

tRNA/rRNA expression ratios in microbes (**Fig. 4.2A** and Table S4.2). Note that if protein concentration [*P*] and the cellular dry mass density are indeed approximately constant across species (Milo, 2013), then equations (4.1) and (4.2) contain a single species-specific parameter, k_{cat}.

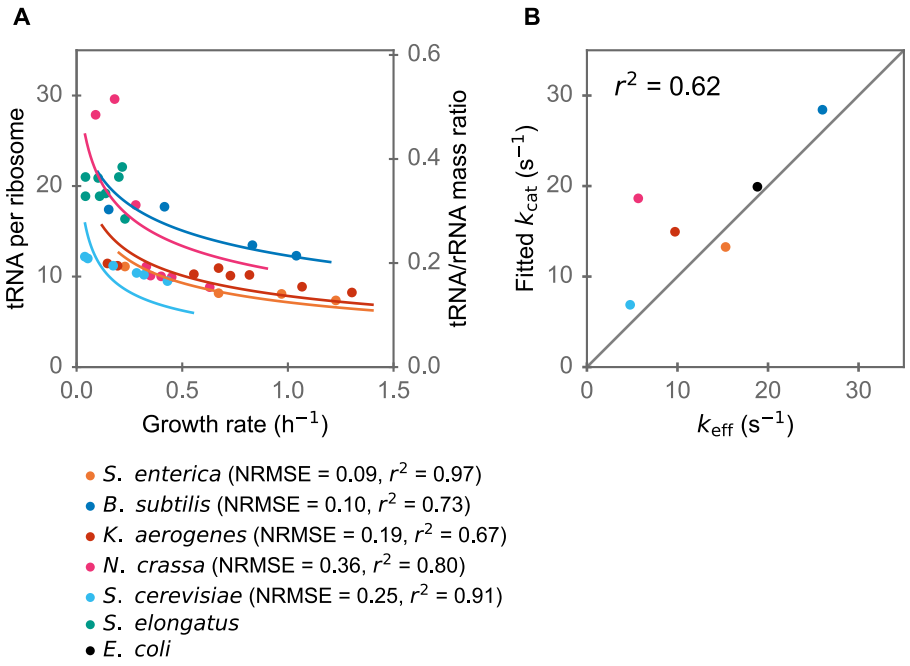


Fig. 4.2. The RNA growth law across species. (**A**) Experimentally observed tRNA/ribosome expression ratios in different microbes decrease with growth rate, consistent with the predicted optimal tRNA/ribosome expression ratio. For each species except *S. elongatus*, which is a slow-growing species and shows no systematic growth rate dependence, we fitted equation (4.1) to the data by varying the single adjustable parameter k_{cat} (solid lines; the numbers in parentheses after the species names quantify the agreement between the fitted lines and the data). Note that the y-axis on the right-hand side is based on the tRNA/rRNA mass ratio for bacteria. For eukaryotic microbes, the tRNA/rRNA mass ratio should be scaled by a factor of 0.84 according to equation (4.15). (**B**) Comparison of fitted k_{cat} and effective ribosome turnover number k_{eff}.

For six out of the seven datasets in **Figs. 4.1B** and **4.2A**, the tRNA/ribosome expression ratio decreases with increasing growth rate. The only exception, the cyanobacterium *Synechococcus elongatus*, has a much smaller maximal growth rate (μ_{max} = 0.23 h^{-1}) than the other species, and its tRNA/rRNA expression ratio does not show a clear growth rate-dependence (**Fig. 4.2A**, Spearman's ρ = -0.01, P = 0.98) (Mann & Carr, 1973). It is conceivable that slow-growing species do not fully optimize their translation machinery composition, as a near-optimal constant TC/ribosome expression ratio may incur a lower fitness cost than the expression of a regulatory system for growth rate-dependent optimal expression.

To verify the implementation of the RNA growth law in the remaining, fast-growing species, we used our model to estimate k_{cat} by fitting the measured tRNA/rRNA expression ratio to equation (4.1) (solid lines in **Fig. 4.2A**). Independently, we also estimated the effective ribosome turnover number (k_{eff}) through the relationship $\mu \cdot [P] = k_{eff} \cdot [R]$, using measured values for μ, $[P]$, and $[R]$ (Table S4.3; fitting was performed for all species excluding *S. elongatus*, in which the tRNA/rRNA expression ratio is independent of the growth rate and thus a fitting procedure would be meaningless). **Fig. 4.2B** shows a close correspondence between the k_{cat} values estimated via equation (4.1) and the effective turnover numbers (Pearson's r^2 = 0.62, P = 0.063). Given that the

tRNA/rRNA expression ratios used for fitting equation (4.1) were measured with different experimental methodologies by different groups, we do not expect a perfect correlation; that our model still explains 62% of the variation appears to strongly support our analyses. We thus conclude that equation (4.2) describes a universal RNA growth law for fast-growing bacterial species.

Implementation of the RNA growth law through tRNA and rRNA genomic positions across bacteria

Next, we asked if other bacteria also show a differential distribution of tRNA and rRNA genes along the chromosome that is consistent with a partial implementation of the RNA growth law through replication-associated gene dosage effects. As a strong selection pressure toward optimal tRNA/ribosome expression ratios is expected mainly in fast-growing species (**Fig. 4.2A**), we surveyed gene positions in bacteria for which maximal growth rates are available (Vieira-Silva & Rocha, 2010). In *E. coli*, the summed time of DNA replication (C period, ~ 40 min) and cell division (D period, ~20 min) (Cooper & Helmstetter, 1968) is approximately 1 h. Given that these times will be roughly similar in many other species, we assume that species with substantially larger doubling times are unlikely to perform multiple simultaneous rounds of replication, while cells with shorter doubling times will likely perform multiple replication rounds simultaneously and hence experience stronger replication-associated gene dosage effects. Accordingly, we classified bacteria with doubling times $\leq 1\ \mathrm{h}$ (i.e., $\mu_{\max} \geq 0.69\ \mathrm{h}^{-1}$) as fast-growing species, and bacteria with doubling times > 1 h as slow-growing species.

As shown in **Fig. 4.3A** and **4.3B** (orange points), we found that in fast-growing species, rRNA and tRNA genes are generally located in the vicinity of oriC, at relative positions < 0.5 (0.5 is located 0.25 genome lengths to either side of oriC, halfway between oriC and the terminus of replication; for each genome represented in **Fig. 4.3**, the positions are the arithmetic means across the corresponding genes). This observation is consistent with previous analyses (Couturier & Rocha, 2006; Vieira-Silva & Rocha, 2010). Moreover, we found that both rRNA and tRNA genes tend to be located ever closer to oriC with increasing $\mu_{\max}$ (correlation with $\mu_{\max}$ for $position_{\mathrm{rRNA}}$: Spearman's $\rho = -0.59$, $P = 9.2 \times 10^{-6}$, *P*-value calculated based on phylogenetically independent contrasts (Felsenstein, 1985) to control for phylogenetic non-independence between datapoints: $P_{\mathrm{ic}} = 0.04$; for $position_{\mathrm{tRNA}}$: $\rho = -0.40$, $P = 4.6 \times 10^{-3}$, $P_{\mathrm{ic}} = 2.1 \times 10^{-4}$). In slow-growing species, rRNA genes still tend to be close to oriC (**Fig. 4.3A**, blue points; one sample Wilcoxon signed rank test, $P = 2.8 \times 10^{-10}$), while tRNA genes are distributed around the midpoint between oriC and the terminus (**Fig. 4.3B**, blue points; one sample Wilcoxon signed rank test, *P* = 0.11).

As expected from our hypothesis of a partial implementation of the RNA growth law through replication-associated gene dosage effects, we found that rRNA genes are closer to oriC than tRNA genes in most slow-growing and in all but one fast-growing bacteria (**Fig. 4.3C**; note that the one exception has a small genome of only 1.8 Mb). Accordingly, the tRNA/rRNA expression ratio that would be obtained if regulation was exclusively through gene dosage would be a decreasing function of growth rate, in qualitative agreement with the optimality predictions from equation (4.2). This finding, together with our detailed analysis of individual species (**Figs. 4.1B** and **4.2**), supports our hypothesis that natural selection has fine-tuned the positions of tRNA and rRNA genes to match the RNA growth law for optimally efficient translation in fast-growing species.

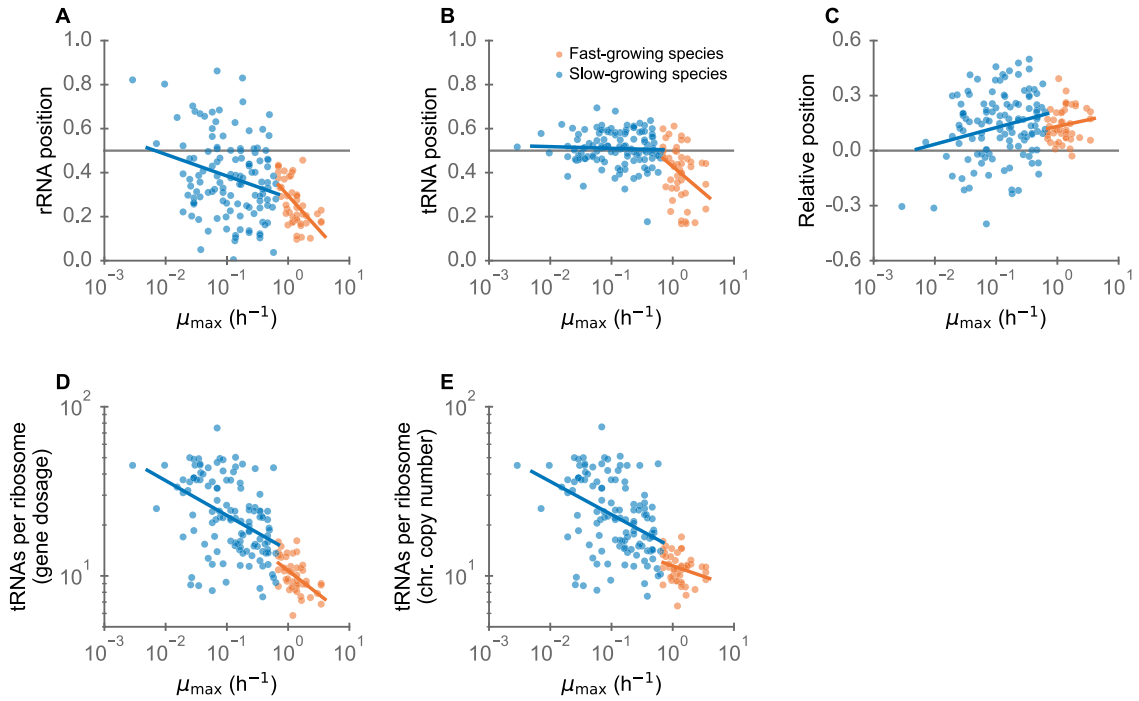


Fig. 4.3. The genomic positions of rRNA and tRNA genes implement the RNA growth law in fast-growing species. (**A**) Arithmetic means of the rRNA positions for individual genomes as a function of μ_{max}. The horizontal grey line (position 0.5) marks the midpoint between origin and terminus of replication. (**B**) Same for tRNA. (**C**) Relative positions between tRNA and rRNA genes ($position_{tRNA}$ - $position_{rRNA}$). (**D**) tRNA/rRNA gene dosage ratios. (**E**) Genomic tRNA/rRNA ratios (per chromosome). Blue points indicate slow growing species (with blue linear regression line), orange points indicate fast-growing species (with orange linear regression line).

The maximal growth rate μ_{max} is not the only factor that affects the strength of replication-associated gene dosage effects. At the same DNA replication rate, smaller genomes need less time to replicate than larger genomes. Thus, at the same growth rate, bacteria with smaller genomes are expected to have fewer replication forks in the cell, and hence experience weaker gene dosage effects. **Text S4.2** explores the influence of genome size on the positioning of tRNA and rRNA genes; here, we only provide a brief summary. Consistent with the above notions, in fast-growing species, we found that the position of rRNA genes is negatively correlated with genome size, i.e., there appears to be less selection pressure toward positioning rRNA genes close to oriC in smaller genomes. At the same time, the relative genomic position of tRNA and rRNA genes is positively correlated with genome size in fast-growing species, again indicating lower selection pressures toward specific genomic positions is smaller genomes. However, in a combined statistical model, μ_{max} remains the main predictor of tRNA and of rRNA positions in fast-growing species, with only marginal contributions from genome size. It is conceivable that the effective population size – which influences the efficiency of natural selection – also influences the genomic positions of tRNA and rRNA genes. However, we found no evidence for such an influence (**Fig. S4.4**).

For the multi-species dataset, we have so far only considered the genomic positions. We now turn our attention to the resulting tRNA/rRNA gene dosage ratio at the reported maximal growth rate. According to equation (4.1), faster growing species need a lower TC/ribosome expression ratio at maximal growth. We indeed find statistically highly significant negative correlations between the predicted tRNA/ribosome gene dosage ratio (equation (4.21)) and μ_{max} (**Fig. 4.3D**; slowly growing species: $\rho = -0.44,\ P = 2.8 \times 10^{-7}, P_{\mathrm{ic}} = 6.0 \times 10^{-4}$; fast-growing species: $\rho = -0.49,\ P =$

$4.3 \times 10^{-4}, P_{\mathrm{ic}} = 0.037$) (see **Text S4.3** for the treatment of tRNA genes; these calculations assume a constant DNA replication rate k_{rep} = 1000 s^{-1} across species, see **Text S4.1** for species-specific replication rate k_{rep}).

While slowly growing species show a wide range of tRNA/ribosome gene dosage ratios, the ratio in fast-growing species shows a much tighter distribution (*F*-test for equality of variances: $P < 10^{-15}$). In slow-growing species, the effects of replication-associated gene dosage effects are weak: the tRNA/ribosome gene dosage ratios are almost identical to the corresponding chromosomal copy number ratios (**Fig. 4.3E**). In fast-growing species, the chromosomal tRNA/rRNA gene copy number ratios show a distribution that is similarly tight as that for the corresponding gene dosage ratios (*F*-test for equality of variances: $P < 10^{-15}$). As expected, species harbor increasingly more tRNAs and ribosomal genes with increasing μ_{max}; consistent with the RNA growth law, this effect also leads to a negative correlation between the number of tRNA genes and the tRNA/ribosome (gene dosage and genomic) ratios (**Fig. S4.5**): at higher maximal growth rates, bacteria have more tRNA genes, but the number of ribosomal genes increases even faster. In contrast to the rRNA and tRNA gene positions (**Fig. 4.3A** and **4.3B**) and the gene dosage ratios (**Fig. 4.3D**), the tRNA/rRNA chromosomal copy number ratios show no strong systematic dependence on μ_{max} in fast-growing species (**Fig. 4.3E**, $\rho = -0.24,\ P = 0.10, P_{\mathrm{ic}} = 0.36$). Interestingly, we also find no statistically significant dependence of the relative position (*position*$_{\mathrm{tRNA}}$ - *position*$_{\mathrm{rRNA}}$) on μ_{max} in fast-growing species (**Fig. 4.3C**, $\rho = 0.15,\ P = 0.31, P_{\mathrm{ic}} = 0.15$).

All these findings indicate that in fast-growing species, not only the absolute numbers of rRNA and tRNA genes, but also the relative numbers of tRNA and rRNA genes (tRNA/rRNA gene dosage ratio and tRNA/rRNA genomic ratio) are tightly constrained, consistent with the optimization of the translation machinery composition according to the RNA growth law and its implementation through replication-associated gene dosage effects.

Impact of the RNA growth law on cell growth and genome organization

Above, we describe and explain a systematic dependence of RNA composition on growth rate in fast-growing bacteria. Why then does the assumption of a growth rate-independent RNA composition work well in theoretical models for the growth of *E. coli* under various perturbations (Scott *et al*, 2010; O'Brien *et al*, 2013; Bosdriesz *et al*, 2015; Dai *et al*, 2016)? We derived the RNA growth law from an assumption of parsimonious dry mass utilization by the protein translation machinery, in our simple model represented by TCs and ribosomes. As detailed in **Text S4.4**, we find that at intermediate to high growth rates in *E. coli*, the optimal combined mass concentration of ribosomes and TCs is very similar to the combined mass concentration under the assumption of a constant tRNA/rRNA expression ratio, with a 4.4% difference at μ = 0.2 h^{-1} and much smaller differences at higher growth rates (**Fig. S4.11**). Thus, except at the lowest growth rates, the optimal RNA composition will only have a small impact on predictions of cellular growth rates. However, even growth rate differences on the order of 1% or less are highly relevant in evolutionary terms for natural bacterial populations, explaining why we find systematic evidence for the optimal expression of ribosomes and TCs (**Figs. 4.1B** and **4.2A**) and the differential genomic positions of rRNA and tRNA genes (**Fig. 4.3A-4.3C**) across bacterial species.

Model limitations

The derivation of the RNA growth law, equation (4.2), is based on a coarse-grained protein translation model, where the ribosome acts as a catalyst that consumes TCs according to

irreversible Michaelis-Menten kinetics. This coarse-grained model ignores many details of the molecular processes contributing to protein translation, such as the rate parameters for individual sub-processes (Milón & Rodnina, 2012; Rudorf *et al*, 2014) and the occurrence of traffic jams of ribosomes co-translating the same mRNA (Tuller *et al*, 2010). Following earlier work (Klumpp *et al*, 2013), we absorb the effects of these detailed processes on the translation rate into the effective ribosomal turnover number, k_{cat}, which we treat as a species-specific constant. The agreement between the predictions derived from the coarse-grained model and experimental data (in particular **Figs. 4.1B** and **4.2B**) indicate that these simplifications represent an appropriate approximation.

One important parameter not explicitly considered here is temperature. At cold stress, the DNA replication rate becomes much slower in *E. coli* (Atlung & Hansen, 1999). Experimental data shows that at low temperatures, the gene dosage ratio is almost constant across growth rates in *E. coli* (**Fig. S4.6**). In our analyses, we only considered species-specific optimal growth temperatures, appropriate for the experimental data underlying **Figs. 4.1** and **4.2**, and for the maximal growth rates considered in **Fig. 4.3**. It appears not unlikely that the fine-tuned coordination between tRNA and ribosome expression breaks down at temperatures far away from optimal growth conditions.

Moreover, we here consider only the average genomic positions of tRNA and rRNA genes. While the optimal scaling of the tRNA/rRNA expression ratio (equation (4.2)) with growth rate is independent of codon frequencies, it is still conceivable that selection pressure toward specific genomic positions is stronger for tRNA genes whose products decode more abundant codons. However, we found no such systematic dependence across genomes (**Text S4.3**).

Conclusions

In sum, the tRNA/ribosome expression ratio appears to be tightly constrained across fast-growing bacteria. At fast growth, its regulation is likely dominated by replication-associated gene dosage effects, implemented through the relative chromosomal positioning of tRNA and ribosomal RNA genes. The objective of this regulation is to not only increase the expression of TCs and ribosomes with growth rate, but to also adjust their relative concentrations according to the RNA composition growth law quantified by equations (4.1) and (4.2).

Methods

Derivation of the optimal TC/ribosome expression ratio

In recent work, we have shown that the growth rate-dependent composition of the translation machinery in *E. coli* is accurately described by predictions based on detailed reaction kinetics and the numerical minimization of the total mass of all participating molecules (Hu *et al*, 2020). This minimization was motivated by the observation that the cellular dry mass density is approximately constant across growth conditions (Kubitschek *et al*, 1984). Accordingly, if part of the dry mass density is occupied by one particular molecule type, less will be available for all other molecule types. This reasoning assumes that cellular dry mass is a growth-limiting resource; considering other growth-limiting resources, such as the minimization of the energy consumed or the enzyme mass required for the production of the different molecules led to almost identical results (Hu *et al*, 2020).

Here, we consider a much simpler representation of the elongation step of protein translation, which can be modeled as an enzymatic reaction following Michaelis-Menten kinetics (Klumpp *et*

al, 2013). In this case, the minimization of the combined mass concentration of ribosome and TC can be performed analytically, as demonstrated by Dourado *et al.* (Dourado *et al*, 2021); following this work, we here briefly summarized the derivation of the optimal TC/ribosome expression ratio.

In the coarse-grained protein translation model (Klumpp *et al*, 2013), the protein synthesis rate v can be expressed as

$$v = k_{\mathrm{cat}}[R]\frac{[TC]}{K_m + [TC]}. \tag{4.4}$$

Here, k_{cat} is the effective turnover number of the ribosome, and K_{m} is the ribosome's Michaelis constant for TC. The combined cytosol mass density of ribosome and TC is given by

$$c = [R]\cdot m_{\mathrm{R}} + [TC]\cdot m_{\mathrm{TC}}\,, \tag{4.5}$$

where m_{R} is the molecular weight of the ribosome, and m_{TC} is the molecular weight of the TC. We can express the ribosome concentration $[R]$ as a function of v by rearranging equation (4.4),

$$[R] = \frac{v}{k_{\mathrm{cat}}}\left(\frac{K_{\mathrm{m}}}{[TC]} + 1\right). \tag{4.6}$$

Substituting equation (4.6) into equation (4.5), we have

$$c = \frac{v}{k_{\mathrm{cat}}}\left(\frac{K_{\mathrm{m}}}{[TC]} + 1\right)\cdot \mathrm{m}_{\mathrm{R}} + [TC]\cdot m_{TC}\,. \tag{4.7}$$

At a given protein production rate v, c is now only a function of the TC concentration. The minimal c can then be obtained by setting the derivative of equation (4.7) with respect to [*TC*] to zero,

$$\frac{dc}{d[TC]} = m_{TC} - \frac{m_R K_{\mathrm{m}} v}{k_{\mathrm{cat}}}\frac{1}{[TC]^2} = 0\,. \tag{4.8}$$

With the ribosome/TC mass ratio $a = m_R/m_{TC}$, the optimal $[TC]$ can be expressed as

$$[TC] = \sqrt{\frac{m_R K_{\mathrm{m}} v}{m_{TC} k_{\mathrm{cat}}}} = \sqrt{\frac{a K_{\mathrm{m}} v}{k_{\mathrm{cat}}}}. \tag{4.9}$$

Substituting equation (4.9) into equation (4.6), the optimal ribosome concentration $[R]$ can be expressed as

$$[R] = \frac{v}{k_{cat}} + \sqrt{\frac{K_{\mathrm{m}} v}{a k_{\mathrm{cat}}}}. \tag{4.10}$$

Thus, the TC/ribosome concentration ratio can be written as

$$\frac{[TC]}{[R]} = \frac{a\cdot\sqrt{k_{\mathrm{cat}}\cdot K_{\mathrm{m}}}}{\sqrt{a\cdot v} + \sqrt{k_{\mathrm{cat}}\cdot K_{\mathrm{m}}}}\,. \tag{4.11}$$

At steady state, the protein production rate *v* is equal to rate of protein dilution by volume growth,

$$v = \mu \cdot [P], \tag{4.12}$$

with growth rate μ and total cellular protein concentration [*P*] (in units of amino acids per volume).

As the binding between the ribosome and the TC is limited by the diffusion of the TC, K_{m} can be approximated through $K_{\mathrm{m}} \approx k_{\mathrm{cat}}/k_{\mathrm{on}}^{\mathrm{diff}}$, with $k_{\mathrm{on}}^{\mathrm{diff}}$ the diffusion-limited binding constant of the TC to the ribosome (Klumpp *et al*, 2013). Thus, equation (4.11) can be rewritten as (equation (4.1) of the main text)

$$\frac{[TC]}{[R]} = \frac{a \cdot k_{\mathrm{cat}}}{\sqrt{a \cdot \mu \cdot [P] \cdot k_{\mathrm{on}}^{\mathrm{diff}}} + k_{\mathrm{cat}}} \quad . \tag{4.13}$$

In *E. coli,* the molecular weight of the ribosome is 2307.0 kDa and the molecular weight of a TC is 69.6 kDa (Hu *et al*, 2020), thus $a = 33.1$. For a single TC, $K_{\text{m-singleTC}}$ = 3 μM (Klumpp *et al*, 2013); the effective number of TC (Klumpp *et al*, 2013) is 34 (the predicted expressed tRNA in Ref. (Hu *et al*, 2020)), and thus $K_{\mathrm{m}} = 34 \cdot K_{\text{m-singleTC}}$ = 102 μM. k_{cat} = 22 s^{-1} is the observed maximal translation rate of a ribosome (Klumpp *et al*, 2013), and $k_{\mathrm{on}}^{\mathrm{diff}} \approx k_{\mathrm{cat}}/K_{\mathrm{m}}$ =0.216 $\mu M^{-1}s^{-1}$.

The protein concentration [*P*] is calculated from *E. coli* proteome expression data (Schmidt *et al*, 2016) and cell volume (Volkmer & Heinemann, 2011) for growth on glucose,

$$[P] = \frac{\sum_i N_i L_i}{V_{\mathrm{cell}} N_{\mathrm{A}}} \quad , \tag{4.14}$$

where N_i is the copy number per cell and L_i the length of protein *i* (Schmidt *et al*, 2016), V_{cell} is the cell volume (Volkmer & Heinemann, 2011), and N_{A} is the Avogadro constant. In a more recent publication (Radzikowski *et al*, 2016), the authors of Ref. (Volkmer & Heinemann, 2011) re-measured the volume of cells by super-resolution microscopy and found that cell volume was overestimated in Ref. (Volkmer & Heinemann, 2011) by a factor of 0.67^{-1} for growth on glucose. We thus modified cell volume by a factor of 0.67 relative to the values in Ref. (Volkmer & Heinemann, 2011), resulting in $[P] = 1.16 \times 10^6$ μM.

By multiplying the left-hand side of equation (4.13) with the molecular weight ratio of tRNA to rRNA, we obtain the tRNA and rRNA mass ratio (equation (4.2) of the main text),

$$\frac{M_{\mathrm{tRNA}}}{M_{\mathrm{rRNA}}} = \frac{[TC] \cdot m_{\mathrm{tRNA}}}{[R] \cdot m_{rRNA}} = r \cdot \frac{k_{\mathrm{cat}}}{\sqrt{a \cdot \mu \cdot [P] \cdot k_{\mathrm{on}}^{\mathrm{diff}}} + k_{\mathrm{cat}}} \quad , \tag{4.15}$$

with

$$r = a \cdot \frac{m_{\mathrm{tRNA}}}{m_{rRNA}} \quad . \tag{4.16}$$

Here, m_{tRNA} is the molecular mass of tRNA, m_{rRNA} is the total mass of RNA in one ribosome, and *r* is the ratio of the tRNA mass fraction of a TC and the rRNA mass fraction of the ribosome. For bacteria, we use data from *E. coli* ($m_{\mathrm{tRNA}} = 25.8\ kDa$, $m_{\mathrm{rRNA}} = 1480\ kDa$), resulting in $a = 33.1$ and *r* = 0.58. For eukaryotes, we use data from *S. cerevisiae*, resulting in *a* = 40.3 and *r* = 0.59; the molecular weights of the ribosome (3044.4 kDa), rRNA (1750 kDa), TC (75.6 kDa), and tRNA

(25.6 kDa) were calculated from the respective sequences according to the *Saccharomyces* Genome Database (Cherry *et al*, 2012).

Gene positions

The chromosomal position of the center of the origin of replication (oriC) for different genomes was obtained from the DoriC database (version 10.0) (Luo & Gao, 2019). The start and end positions of rRNA and tRNA genes were downloaded from the RefSeq database (Release 93, downloaded on April 09, 2019); gene locations were defined as the midpoint between gene start and end. We defined gene position as the relative distance of a gene to oriC, calculated as the shortest distance between the gene and oriC on the circular chromosome, divided by half the length of the chromosome. Gene position ranges from 0 to 1.

Maximal growth rate dataset

Minimal doubling times $\tau_{\min}$ (in hours) were obtained from Ref. (Vieira-Silva & Rocha, 2010) and were converted to maximal growth rates as $\mu_{\max} = \frac{\ln(2)}{\tau_{\min}}$. For the analyses, we only used species for which we additionally had genome annotation and oriC location, and which had only one chromosome. The final trimmed dataset contains 170 species (Table S4.4).

For 35 out of the 170 species, more than one oriC has been annotated (Luo & Gao, 2019). However, we found that all oriCs are very close on the chromosome in these species: the maximal distance between two oriCs is much less than 1% of the chromosome length (the maximal distance between two oriCs is 0.0035, equal to 0.18% of the chromosome length). Thus, different oriCs are expected to have a negligible effect on gene position and we randomly selected one of the oriCs to calculate gene position.

Phylogenetically independent contrasts

16S rDNA sequences was aligned with MUSCLE (Edgar, 2004) embedded in MEGA X (Kumar *et al*, 2018). A phylogenetic tree was built using maximum likelihood methods with MEGA X with default parameters (Kumar *et al*, 2018). The phylogenetic tree was rooted by the minimal ancestor deviation method (Tria *et al*, 2017). We calculated phylogenetically independent contrasts (Felsenstein, 1985) with the pic function in ape package (Paradis & Schliep, 2019) in R (R Core Team, 2020). To control for phylogenetic non-independence between data points for different species, we then performed statistical tests on these contrasts (P_{ic} values).

Gene dosage

We used the Cooper-Helmstetter model (Cooper & Helmstetter, 1968; Bremer & Churchward, 1977) to calculate gene dosage. The model is briefly summarized below. Let *C* be the time required to replicate the chromosome; let *D* be the time between the termination of a round of replication and the next cell division; let τ be the doubling time. The average dosage of gene *i* ($\overline{X_i}$) per cell is then given by:

$$\overline{X_i} = 2^{\frac{C(1-position_i)+D}{\tau}} \quad , \tag{4.17}$$

where $position_i$ is the genomic position of gene *i*. With

$$\tau = \frac{\ln(2)}{\mu} \quad , \tag{4.18}$$

$$\overline{X_i} = e^{\mu[C(1-position_i)+D]} \tag{4.19}$$

The gene dosage ratio of two genes ($\overline{X_i}/\overline{X_j}$) is then (equation (4.3) of the main text)

$$\frac{\overline{X_i}}{\overline{X_j}} = e^{\mu C(position_j - position_i)}. \tag{4.20}$$

Each genome contains multiple tRNA and rRNA genes. In this case, we use the ratio of the total gene dosages,

$$\frac{\sum \overline{X}_{\mathrm{tRNA}}}{\sum \overline{X}_{\mathrm{ribosome}}} = \frac{\sum \overline{X}_{\mathrm{tRNA}}}{\frac{1}{n}\sum \overline{X}_{\mathrm{rRNA}}} = \frac{\sum e^{\mu[C(1-position_{tRNA})+D]}}{\frac{1}{n}\sum e^{\mu[C(1-position_{rRNA})+D]}} = \frac{\sum e^{\mu C(1-position_{\mathrm{tRNA}})}}{\frac{1}{n}\sum e^{\mu C(1-position_{\mathrm{rRNA}})}} \quad , \tag{4.21}$$

where *n* is the number of rRNA genes per ribosome. Since one ribosome contains three rRNA genes (5S, 16S, and 23S rRNA), n = 3.

We assumed a constant DNA replication rate of $k_{\mathrm{rep}} = 1000\ \mathrm{bp\ s^{-1}}$ (Couturier & Rocha, 2006) to calculate the C-period as

$$C = \frac{L_{\mathrm{genome}}}{2k_{\mathrm{rep}}} \quad , \tag{4.22}$$

with L_{genome} the length of the given genome.

Supplementary information

Supplementary Tables are available online at https://doi.org/10.1371/journal.pgen.1009939.

Supplementary Figures

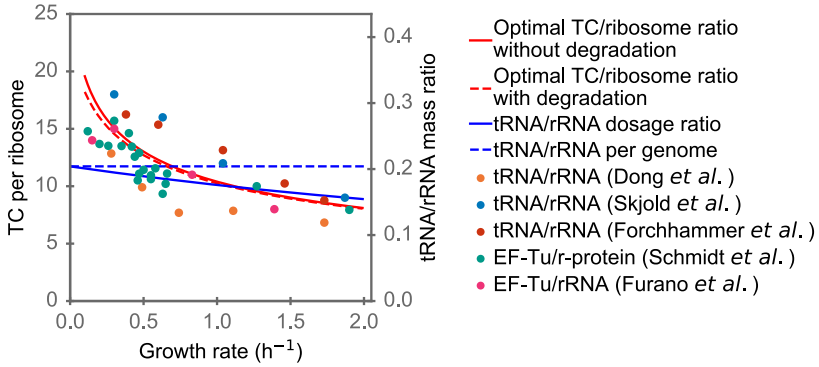


Fig. S4.1. The effect of protein degradation on the RNA growth law in *E. coli*. The solid red curve shows the RNA composition growth law as given by equation (4.1), which ignores protein degradation, whereas the dashed red curve shows the RNA composition growth law with protein degradation included. Protein degradation rate typically ranges from 0.02 h^{-1} to 0.04 h^{-1} (Pine, 1970, 1973) (see estimation below) and is higher at low growth rates than at high growth rates (Pine, 1970, 1973; Calabrese *et al*, 2021). We here assumed a constant protein degradation rate (k_{deg} = 0.04 h^{-1}); its inclusion is equivalent to moving the optimal TC/ribosome ratio 0.04 h^{-1} to the left. As the degradation rate is relatively small compared to the maximal growth rate of *E. coli* (Pine, 1970, 1973), protein degradation affects the optimal TC/ribosome ratio only at very low growth rates, where the degradation rate becomes comparable to the growth rate.

Estimation of the protein degradation rate: Publications (Pine, 1970, 1973) report the fraction of degraded protein (f_{deg}) after time t. The protein degradation function can be written as $1 - f_{\mathrm{deg}} = e^{-k_{\mathrm{deg}} \cdot t}$. With this equation, k_{deg} was estimated to range from 0.005 h^{-1} to 0.04 h^{-1} according to the data reported in (Pine, 1973) and to range from 0.025 h^{-1} to 0.03 h^{-1} according to the data reported in (Pine, 1970). To be conservative, we here chose k_{deg} = 0.04 h^{-1}, the largest reported degradation rate.

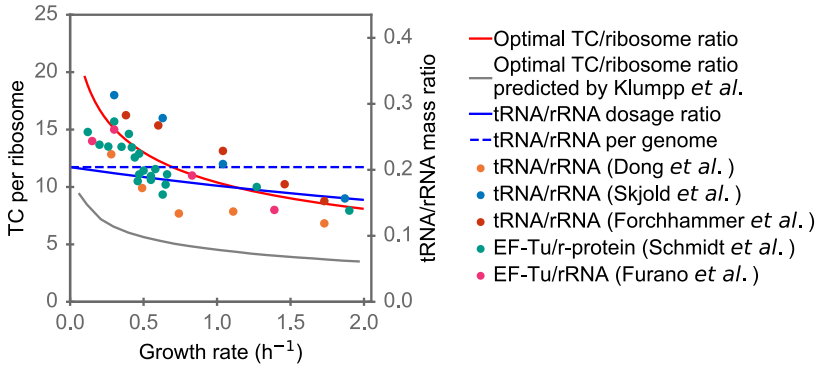


Fig. S4.2. Comparison of the optimal TC/ribosome expression ratio with a previously reported optimal TC/ribosome ratio for *E. coli* (Klumpp *et al*, 2013). Klumpp *et. al* predicted the optimal TC/ribosome expression ratio by identifying the proteome fractions of ribosome and TC that maximize growth rate in a coarse-grained, phenomenological model of cellular growth (Klumpp *et al*, 2013). This optimal proteome allocation pattern results in a predicted TC/ribosome ratio (gray line, extracted from Fig. 4C in (Klumpp *et al*, 2013)) that is substantially lower than the experimentally observed data. Here, we considered the optimal resource allocation into cellular dry mass (RNA and protein combined; red curve), motivated by the near-constant cellular dry mass density across growth conditions. The optimal dry mass allocation explains the experimentally observed TC/ribosome expression ratios much better than the optimal proteome allocation considered in Ref. (Klumpp *et al*, 2013). Protein accounts for roughly 1/3 of the ribosome mass, whereas it accounts for roughly 2/3 of the ternary complex mass. If only the protein cost is considered, the ternary complex appears much more expensive to the cell, resulting in a lower predicted ternary-complex/ribosome ratio.

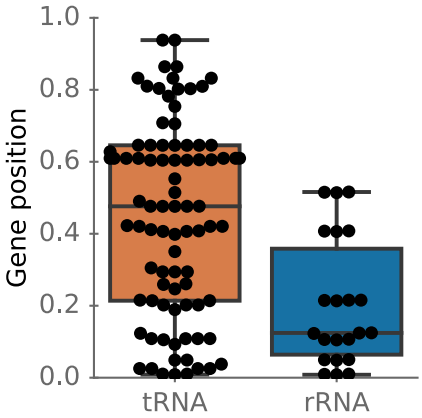


Fig. S4.3. Gene positions of rRNA and tRNA genes in *E. coli*. rRNA genes are located closer to oriC than tRNA genes in *E. coli*, with genomic position 0.20 ± 0.17 (mean ± standard deviation) for rRNA genes and 0.45 ± 0.27 for tRNA genes.

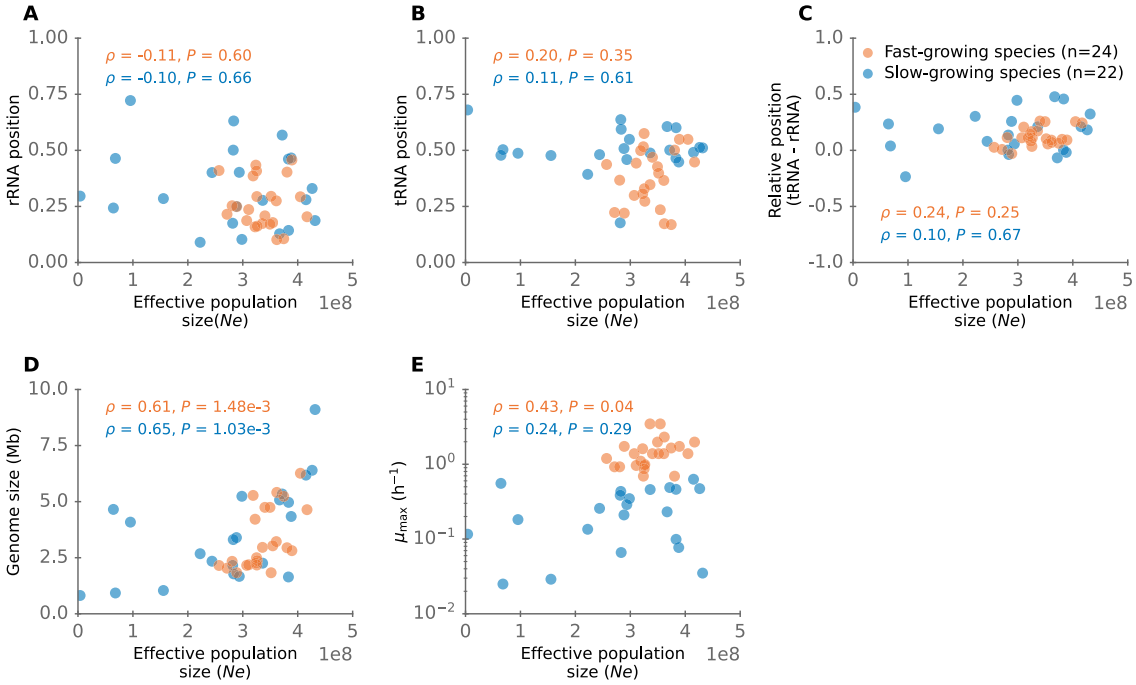


Fig. S4.4. Effective population size (*Ne*) is not correlated with the positions of tRNA and rRNA genes. The effective population sizes for 46 out of the 170 species were obtained from Ref. (Bobay & Ochman, 2018). As shown in (**A-C**), we found no statistically significant Spearman rank correlations ($P > 0.1$) between N_e and the different positions considered in our study (position of rRNA genes, tRNA genes, and the relative position, $position_{tRNA} - position_{rRNA}$) when analyzing fast- and slow-growing species ($P > 0.1$, see colored text in A-C). In contrast, we found statistically significant correlations between N_e and genome size (**D**) and, for fast-growing species, between N_e and maximal growth rate μ_{max} (**E**).

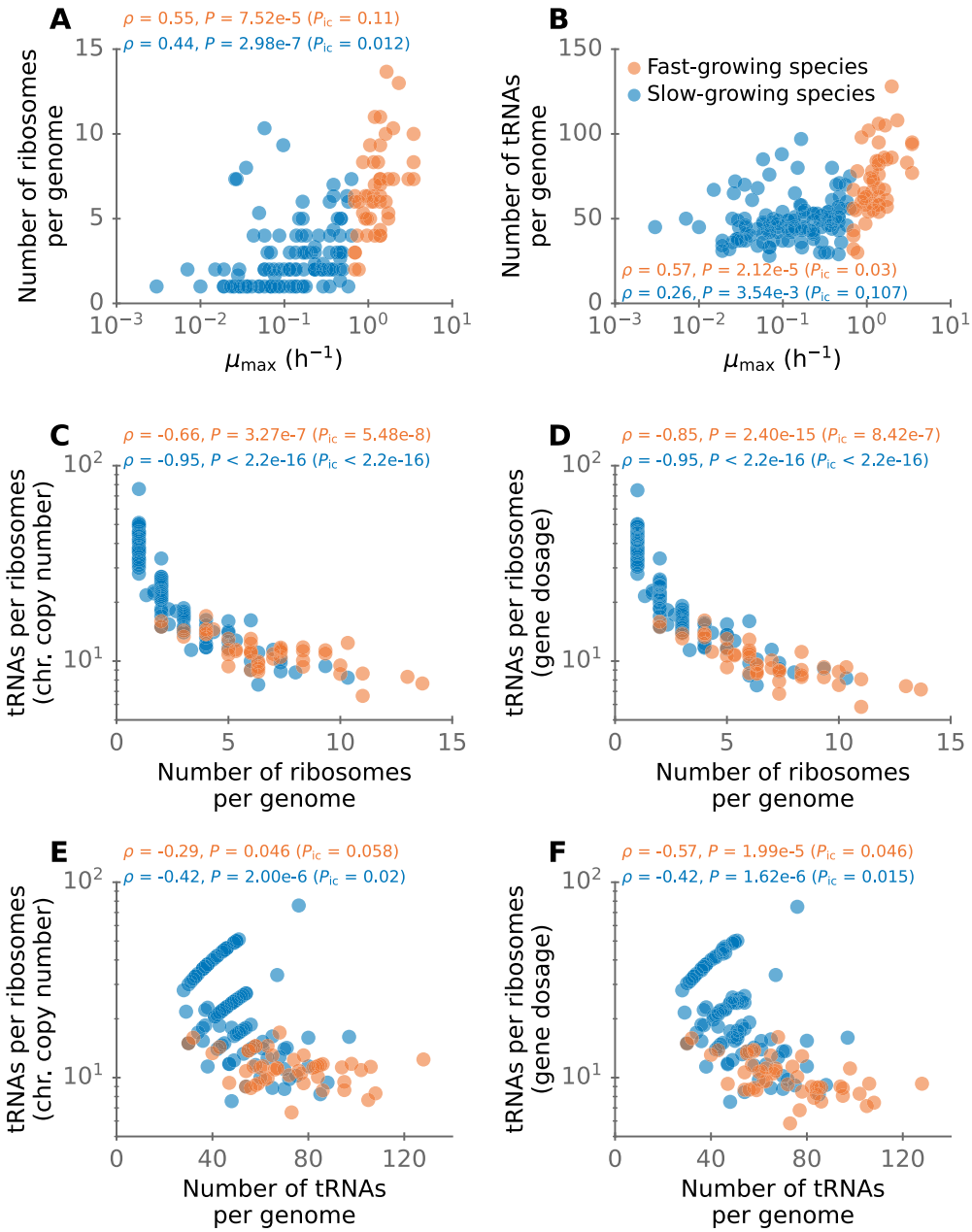


Fig. S4.5. The numbers of genomically encoded rRNA genes and tRNA genes are positively correlated with μ_{max} and negatively correlated with tRNA/ribosome ratios. With increasing maximal growth rate μ_{max}, bacterial genomes harbor more rRNA genes (**A**) and tRNA genes (**B**). (**C**-**F**) show the relationship between tRNA/ribosome dosage and genomic ratios and the number of ribosome and tRNA genes in the genome. While a negative correlation between the tRNA/ribosome ratios and the number of rRNA genes would trivially occur also if the numbers of tRNA and rRNA genes were independent (C and D), a negative correlation of the tRNA/ribosome ratios with the number of tRNA genes would not be expected in this case (E and F); the latter observation thus provides strong support for our hypothesis that the relative number of tRNA and rRNA genes are constrained to optimize translation efficiency.

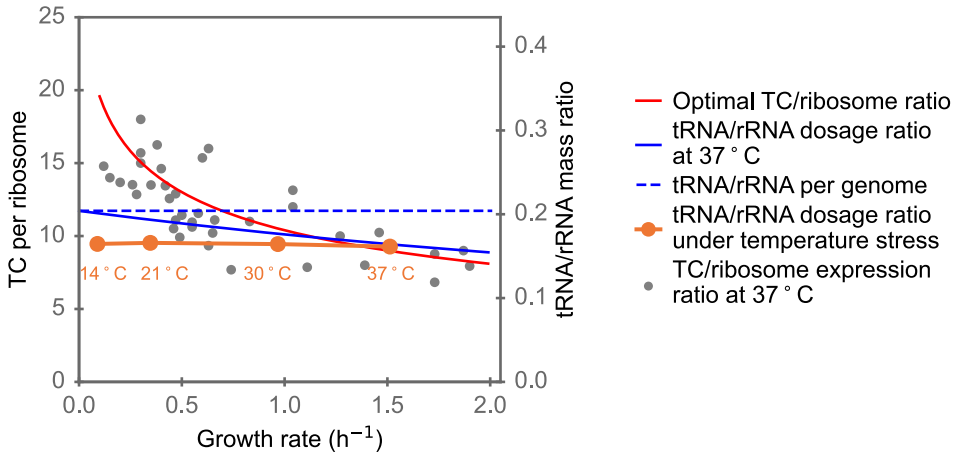


Fig. S4.6. The tRNA/rRNA gene dosage ratio is independent of growth rate under temperature stress conditions. The orange dots show experimental measurements at 14 °C, 21 °C, 30 °C, and 37 °C in the same medium (Atlung & Hansen, 1999). The DNA replication rate became slow at low temperatures, and multiple replication rounds were observed even at low growth rates. This effect made the rRNA/tRNA dosage ratio almost independent of growth rate. As we could not find experimental data for the tRNA/ribosome expression ratio under temperature stress conditions, it is unclear if the tRNA/ribosome expression ratio is still optimal under temperature stress conditions.

Supplementary Texts

Text S4.1. Assessment of tRNA/rRNA gene dosage ratio calculated by a constant DNA replication rate (k_{rep} = 1000 s^{-1})

In the main text, we used a constant DNA replication rate (k_{rep} = 1000 s^{-1}) for all species to calculate the C period and the tRNA/rRNA gene dosage ratio (Fig. 4.1B and Fig. 4.3D). This is an approximation, as the DNA replication rate (1) is growth rate-dependent in a given species (Michelsen *et al*, 2003; Zheng *et al*, 2020) and (2) is species-specific and depends on the maximal growth rate (μ_{max}) across species. Here, we assess how the assumption of a constant k_{rep} = 1000 s^{-1} affects the tRNA/rRNA dosage ratio in the results shown in Fig. 4.1B and Fig. 4.3D.

The effect of a growth rate-dependent C period on the tRNA/rRNA dosage ratio in E. coli

In *E. coli*, the C period is almost constant at growth rates above 0.7 h^{-1} but increases with decreasing growth rate below 0.7 h^{-1} (Fig. S4.7A; data from Refs. (Michelsen *et al*, 2003; Zheng *et al*, 2020)). However, as shown in Fig. S4.7B, when calculating the tRNA/rRNA dosage ratio using the experimentally observed rate-dependent C period, the results are very similar to those calculated under the assumption of a constant C period and DNA replication rate (k_{rep} = 1000 s^{-1}) (calculated by equation (4.21)). Thus, setting k_{rep} = 1000 bp s^{-1} appears to be an acceptable approximation for the tRNA/rRNA dosage ratio calculation.

The effect of a species-specific replication rate on the tRNA/rRNA gene dosage ratio across species

In a literature search, we found estimates of the replication rate or C period for 5 species in our dataset, including one slow-growing species and four fast-growing species (Fig. S4.8A and Table S4.5). We fitted a linear model for the dependence of the replication rate on μ_{max} (both on log scale; Fig. S4.8A) and used this linear model to estimate the μ_{max}-dependent DNA replication rate for all species in our dataset. The tRNA/ribosome gene dosage ratio estimated using a μ_{max}-

dependent replication rate (Table S4.4) is very similar to the dosage estimated using a constant replication rate k_{rep} = 1000 s^{-1} (Fig. S4.8B; R^2 = 0.973). Further, the dependence of the tRNA/ribosome gene dosage ratio on the maximal growth rate μ_{max} when considering a growth rate-dependent replication rate (Fig. S4.8C, Spearman's rank correlation $\rho = -0.47$, $P = 7.0 \times 10^{-4}$ for fast-growing species; $\rho = -0.47$, $P = 5.5 \times 10^{-8}$ for slow-growing species) is very similar to that shown in Fig. 4.3D . Thus, the dependence of the DNA replication rate on μ_{max} does not appear to affect our conclusions.

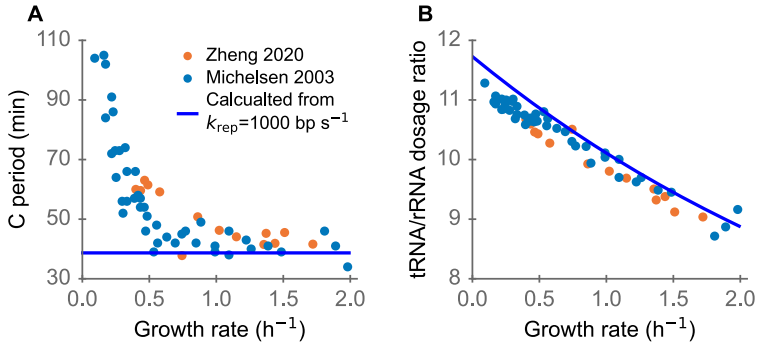


Fig. S4.7. The effect of a growth rate-dependent C period on the tRNA/rRNA dosage ratio in *E. coli*. (**A**) Comparison of the C period calculated from constant k_{rep} = 1000 bp s^{-1} (solid blue line) and experimental estimates of the C period (orange and blue dots (Michelsen *et al*, 2003; Zheng *et al*, 2020)). (**B**) The growth rate-dependent C period has only a minor effect on the predicted tRNA/rRNA dosage ratio. The blue line shows the dosage ratios calculated under the assumption of a constant C period. The orange and blue dots show dosage ratios calculated using the respective experimental C period values shown in (A).

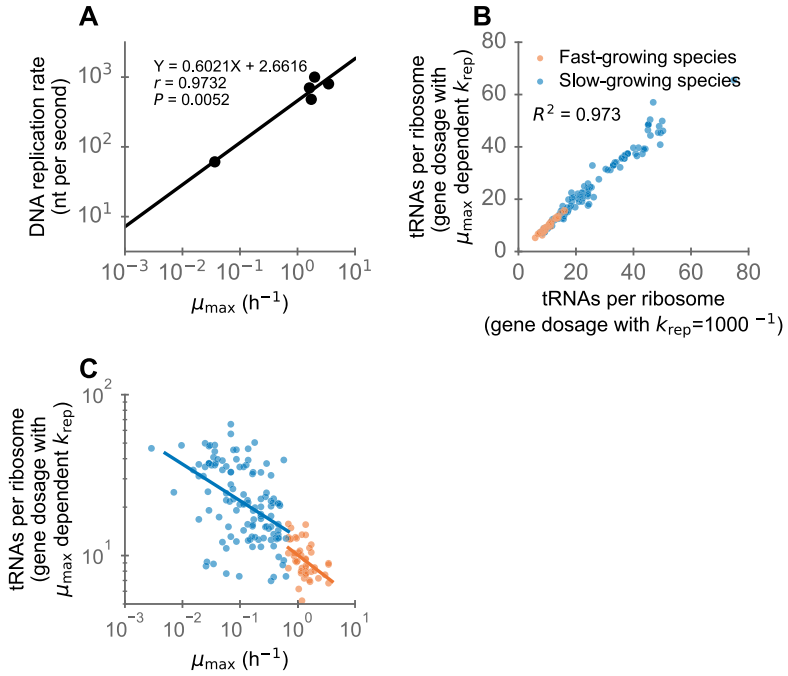


Fig. S4.8. The effect of a μ_{max}-dependent DNA replication rate on the tRNA/ribosome gene dosage ratio. (**A**) linear regression on log-log scale to predict the DNA replication rate from μ_{max}. (**B**) Comparison of tRNA/rRNA gene dosage ratios calculated using a constant DNA replication rate (x-axis) and using a μ_{max}-dependent replication rate. (**C**) Relationship between the tRNA/rRNA gene dosage ratio calculated using a μ_{max}-dependent replication rate and μ_{max}.

Text S4.2. Genome size affects gene position

At the same DNA replication rate, bacteria with smaller genomes need less time to replicate their DNA than bacteria with larger genomes; at the same growth rate, they hence have fewer simultaneous replication rounds in the cell. For example, *Streptococcus pneumoniae* (with a 2.1 Mb genome) does not need multiple replication rounds at its fast growth rate (Slager *et al*, 2014). Thus, replication-associated gene dosage effects will be less important in bacteria with small genomes, and the positions of tRNA and rRNA genes might play less important roles in these species.

In fast-growing species, we indeed found a statistically significant negative correlation between genome size and rRNA gene positions (Fig. S4.9A; Spearman's ρ = -0.40, P = 0.005). In contrast, we did not find a significant correlation between genome size and tRNA gene positions (Fig. S4.9B; Spearman ρ = 0.04, P = 0.80). As expected, a positive correlation was found between genome size and the relative position ($position_{tRNA} - position_{rRNA}$) in fast-growing species (Fig. S4.9C; Spearman's ρ = 0.48, $P = 5.1\times10^{-4}$). We found no statistically significant correlations in slow-growing species ($P > 0.1$ for Fig. S4.9A-C).

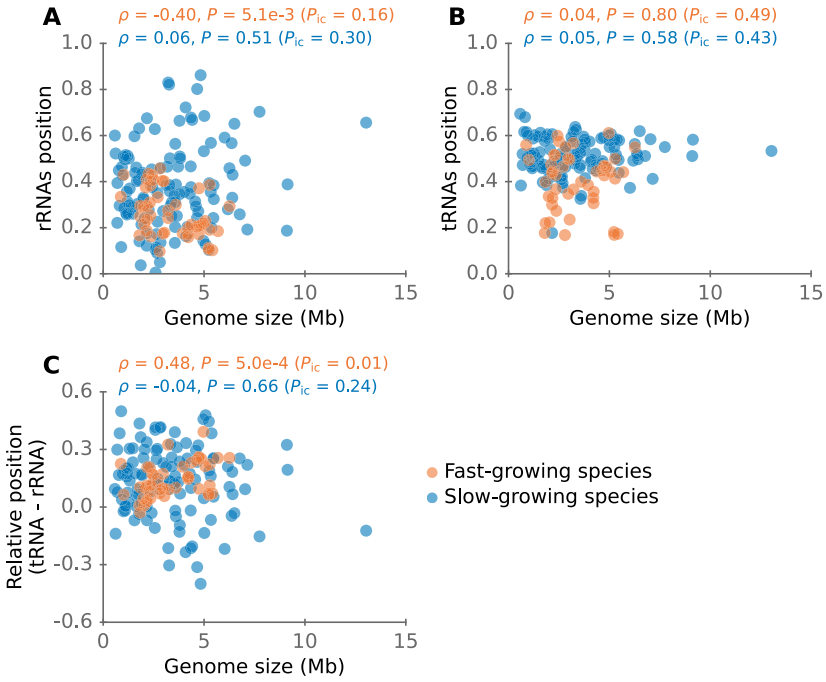


Fig. S4.9. The relationship between genome size and the positions of tRNA and rRNA genes. (**A**) In fast-growing species, rRNA genes in larger genome tend to be closer to oriC. (**B**) tRNA position is not correlated with genome size in both slow- and fast-growing species. (**C**) In fast-growing species, the relative position ($position_{tRNA} - position_{rRNA}$) is positively correlated with genome size, indicating that gene dosage plays a more important role in establishing the optimal tRNA/ribosome expression ratio in these species.

To control for the effects of genome size in our analysis of the dependence of genomic positions on μ_{max}, we analyzed linear regression models of the form

$$position \sim \log(\mu_{max}) + genome\_size \ ,$$

where position is either $position_{tRNA}$, $position_{rRNA}$, or the relative position $position_{tRNA} - position_{rRNA}$. For fast-growing species, the inclusion of genome size as an additional regressor leads to small increases of R^2, from 0.32 to 0.36 (P = 0.07 for genome size) for rRNA position, and from 0.13 to

0.15 (P = 0.24 for genome size) for tRNA position. Thus, the maximal growth rate remains the main predictor of the two positions, with only minor contributions by genome size.

Text S4.3. The positions of tRNA genes are not correlated with the corresponding codon frequencies.

In the main text, we showed that the genomic averages of tRNA gene positions are affected by the maximal growth rate of a species (Fig. 4.3B) and constrained by the optimal RNA growth law, which posits that tRNA genes are on average farther from oriC than rRNA genes (equation (4.1) and Fig. 4.3C). When calculating the tRNA/rRNA gene dosage ratio (Fig. 4.3D), we treated all tRNAs equally. However, tRNAs decode codons with different frequencies. While the optimal scaling of the tRNA/rRNA expression ratio (equation (4.2)) with growth rate is independent of codon frequencies, it is still conceivable that selection pressure toward specific genomic positions is stronger for tRNA genes whose products decode more abundant codons. We thus asked if there is a systematic dependence of genomic positions on the frequencies of the cognate codons.

As the anticodons of tRNAs are not annotated in RefSeq for some species (Pruitt *et al*, 2012), we identified the anticodon using tRNAscan-SE 2.0 (Chan *et al*, 2021) and used the wobble paring rule to find the cognate codon(s) of a given tRNA. Then, a tRNA's cognate codon frequency was calculated as the summed frequencies of all its cognate codon(s).

We tested if there is a correlation between a tRNA's position and its cognate codon frequency in a given species. Please note that the analysis here is different from Fig. 4.3B in the main text. In the main text, we tested if the average position of tRNA genes tends to be close to oriC in fast-growing species. Here, we test if individual tRNAs with higher codon frequencies tend to be located closer to oriC in a given genome.

Statistically significant correlations (P < 0.05) between tRNA genomic position and the cognate codon frequency were found in 21 out of 170 species (Fig. S4.10A). 6 out of the 21 species show positive correlations, while the remaining 15 species show negative correlations. This means that in only 15 species those tRNAs that can decode more codons (by codon frequency) tend to be located closer to the origin of replication.

We also tested for a correlation between tRNA copy number (the number of tRNA genes with the same anticodon) and cognate codon frequency. We found that tRNAs with higher copy numbers tend to have higher cognate codon frequencies in most fast-growing species (Fig. S4.10B; P < 0.05 in 45 out of 48 species).

We thus conclude that in fast-growing species, the copy number of a tRNA gene but not its genomic position is strongly affected by its cognate codon frequency; this finding supports our equal treatment of tRNAs when calculating tRNA gene dosage.

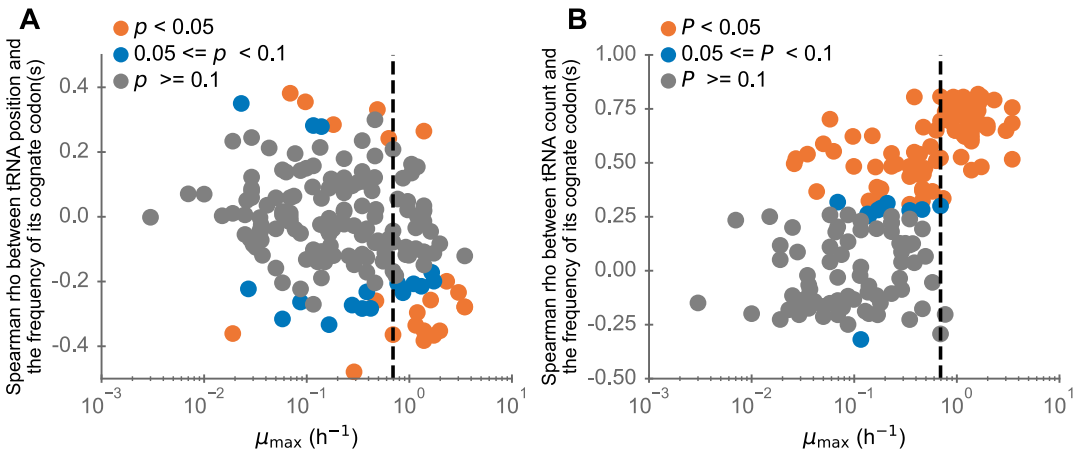


Fig. S4.10. Copy numbers of tRNA genes but not their positions are correlated with codon frequency. Each datapoint represents Spearman's rank correlation coefficient ρ for a single genome. Point color indicates statistical significance. (**A**) Correlation between tRNA position and cognate codon frequency. (**B**) Correlation between tRNA genomic copy number and cognate codon frequency.

Text S4.4. The effect of assuming a constant TC/ribosome expression ratio instead of an optimal ratio

Several recent modeling studies have assumed that the TC/ribosome expression ratio is constant (Klumpp *et al*, 2013; Bremer & Dennis, 2008; Scott *et al*, 2010; O'Brien *et al*, 2013; Bosdriesz *et al*, 2015; Dai *et al*, 2016). In contrast, we found that the optimal TC/ribosome expression ratio is growth rate-dependent (equation (4.1)), a relationship consistent with experimental data across species (Fig. 4.1B and Fig. 4.2A). In this section, we estimate by how much growth rate predictions are expected to change when accounting for this growth rate dependence. To be independent of any specific model, we estimated the associated change in the cellular cost for translation, where – consistent with the assumptions of our optimality estimate, equation (4.1) – we used the cytosol density as a proxy for cost; detailed predictions derived from this cost measure have been found to be in good agreement with experimental data for the *E. coli* translation machinery (Hu *et al*, 2020).

We first calculated the optimal dry mass per volume occupied by ribosomes and TCs as a function of growth rate according to our model. Assuming a constant protein concentration across conditions in *E. coli*, we first calculated the molar concentrations of ribosome and TC based on equations (4.9) and (4.10), respectively.

For comparison, we then calculated the concentrations of ribosome and TC under the constant assumption with $\mu \cdot [P] = k_{eff} \cdot [R]$, where μ is the growth rate, $\cdot[P]$ is the protein concentration, $[R]$ is the ribosome concentration, and k_{eff} is the effective turnover rate of the ribosome. The constant assumption assumes that the rRNA:tRNA mass ratio is 86:14 (Bremer & Dennis, 2008), resulting in a constant TC:ribosome molar ratio of 9.6:1. Further, $k_{eff} = k_{cat} \cdot K_M / ([TC] + K_M)$, with k_{cat} (22 s^{-1}) the ribosome turnover rate, $[TC] = 9.6\ [R]$ the concentration of TC, and K_M = 122 µM the Michaelis constant (see Methods). The ribosome concentration $[R]$ can then be obtained by solving $\mu \cdot [P] = k_{eff} \cdot [R]$.

Using the respective molecular weights, we converted both estimates for the ribosome and TC concentrations into a combined cytosolic mass density of the core translation components. Comparing these combined mass densities, we found that the total dry mass allocated to

translation is very similar between the two calculations at high growth rates; substantial differences are only found at low growth rates (Fig. S4.11).

Thus, the assumption of a constant TC/ribosome expression ratio will lead to very similar growth rate predictions at moderate to fast growth rates. However, translation is a very expensive process for fast-growing cells, and even these small differences will have a substantial effect on evolution in natural populations. The efficiency of natural selection in large bacterial populations likely explains why the optimal expression of the ribosome and TC (Fig. 4.1B and Fig. 4.2) and the differential genomic position of rRNA and tRNA genes (Fig. 4.3) are consistent with experimental data across species.

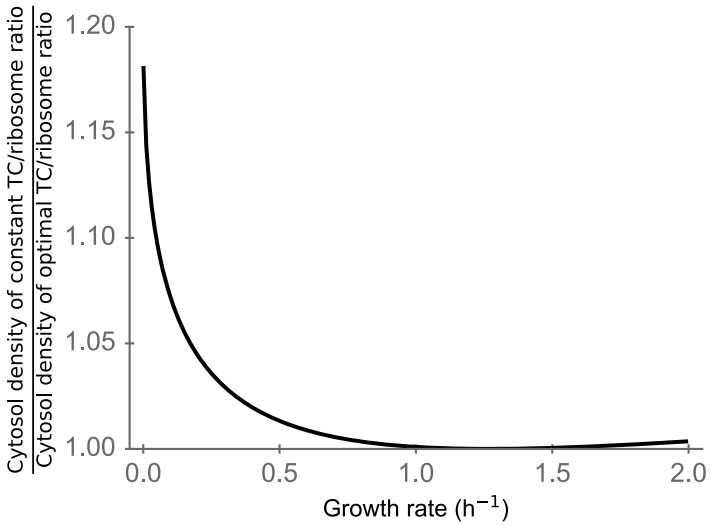


Fig. S4.11. Cost difference (in terms of dry mass density) between the optimal TC/ribosome expression ratio and a constant TC/ribosome expression ratio in *E. coli* as a function of growth rate.

5 Outlook

5.1 Resource allocation out of steady state growth

In this thesis, the optimal expression of a pathway (or a protein) is approximated by its minimal required level at the given growth rate. Manuscript 2 reveals the extent to which metabolic pathways differ in their proteome efficiency. Specifically, the expression of transporters, TCA cycle, and glyoxylate shunt, and "the other" proteins scales contrary to optimal demand.

It should be noted that all the proteome data used in this thesis are from unevolved *E. coli* strains. This means these cells are still adapted to fluctuating growth conditions. As parts of the unneeded proteome can facilitate fast adaptation to new environments (Mori *et al*, 2017; Korem Kohanim *et al*, 2018), the "unneeded" (overabundant) proteome at balanced growth conditions may in fact be "needed" under fluctuating growth conditions. Many quantitative questions are still open with respect to the overabundant proteome. Why does the cell express the specific amount of overabundant proteome in any given growth condition? Is there a relationship between the amount of overabundant proteome and the frequency or duration with which the cell faced a given growth condition in its evolution? Using the theoretically maximal growth rate as a proxy for nutrient quality, how much of the overabundant proteome can be explained by nutrient quality? Answering these questions can help in understanding the principles of resource allocation out of steady state.

5.2 Towards a comprehensive understanding of the importance of gene position

In section 1.6, I proposed the hypothesis that the positions of genes are coordinated with expression demand. In Manuscript 3, this hypothesis was validated on the optimal tRNA/rRNA ratio and the relative positions of tRNA and rRNA genes. The hypothesis in section 1.6 only relies on the "expression level" demand. Recent studies found that the positions of genes are also coordinated with the cell cycle. For example, the locations of two important sporulation network genes (*spo0F* and *kinA*) on opposite sides of the *B. subtilis* chromosome results in a transient imbalance of their gene dosage during replication, which further allows cells to decide between sporulation and vegetative growth in each cell cycle (Narula *et al*, 2015). Future work may consider the cell cycle-dependent demand of genes as a factor for shaping chromosomal gene positions.

The importance of gene positions is still not well appreciated. At present, the physiological consequences of gene positions have only been tested for very few genes (Slager & Veening, 2016). For example, rearrangement of ribosomal genes to the terminus of replication reduces growth rate (Soler-Bistué *et al*, 2015, 2017). To have a better understanding of the importance of gene positions on cellular processes, further work may study gene positions across species to find how many genes have biased positions. Further, the hypothesis proposed in this thesis may help in understanding why the genomic positions of genes are conserved.

List of symbols, abbreviations, and acronyms

μ	Exponential growth rate
μ_{max}	Maximal exponential growth rate
τ_{min}	Minimal doubling time
aa-tRNA	Aminoacyl-tRNA
C	Time of C period; the time required for chromosome replication
D	Time of D period; the time between termination of replication and cell division
EF-Ts	Translation elongation factor Ts
EF-Tu	Translation elongation factor Tu
FBA	Flux balance analysis
GMFE	Geometric mean fold error
k_{cat}	Enzyme turnover number
K_M	Michaelis constant
MOMENT	Metabolic Modeling with enzyme kinetics
oriC	Origin of replication
TC	Ternary complex of EF-Tu, GTP, and aa-tRNA

Acknowledgements

I would like to thank my supervisor, Professor Martin Lercher, for providing me the opportunity to study in the Computational Cell Biology group and for guiding me to an interesting research field. I would also like to thank Martin for his constant support. Whenever I find an interesting topic or have an idea, he always encourages me to go further and always provides me with great guidance. With his support, I have had a happy PhD time in Düsseldorf.

I would like to thank my colleagues, Hugo Dourado, Deniz Sezer, Peter Schubert, and Mayo Röttger for helpful discussions and insightful comments on my work.

Thanks to all members of the group for having great times together.

Thanks to Han Dang-Klein (from JUNO of HHU) for helping me in my first days in Germany.

感谢我的妻子刘丹和我的女儿蹦蹦，你们的支持和陪伴使我倍感幸福。

References

Adadi R, Volkmer B, Milo R, Heinemann M & Shlomi T (2012) Prediction of microbial growth rate versus biomass yield by a metabolic network with kinetic parameters. *PLoS Comput Biol* 8: e1002575

Akashi H & Gojobori T (2002) Metabolic efficiency and amino acid composition in the proteomes of *Escherichia coli* and *Bacillus subtilis*. *Proc Natl Acad Sci U S A* 99: 3695–700

Alberghina FAM, Sturani E & Gohlke JR (1975) Levels and rates of synthesis of ribosomal ribonucleic acid, transfer ribonucleic acid, and protein in *Neurospora crassa* in different steady states of growth. *J Biol Chem* 250: 4381–4388

Anand A, Chen K, Yang L, Sastry AV, Olson CA, Poudel S, Seif Y, Hefner Y, Phaneuf PV, Xu S, *et al* (2019) Adaptive evolution reveals a tradeoff between growth rate and oxidative stress during naphthoquinone-based aerobic respiration. *Proc Natl Acad Sci* 116: 25287–25292

Ardell DH & Kirsebom LA (2005) The genomic pattern of tDNA operon expression in *E. coli*. *PLoS Comput Biol* 1: e12

Atkinson DE (1969) Limitation of metabolite concentrations and the conservation of solvent capacity in the living cell. *Curr Top Cell Regul* 1: 29–43

Atlung T & Hansen FG (1999) Low-temperature-induced DnaA protein synthesis does not change initiation mass in *Escherichia coli* K-12. *J Bacteriol* 181: 5557–5562

Bar-Even A, Flamholz A, Noor E & Milo R (2012) Rethinking glycolysis: On the biochemical logic of metabolic pathways. *Nat Chem Biol* 8: 509–517

Bar-Even A, Noor E, Savir Y, Liebermeister W, Davidi D, Tawfik DS & Milo R (2011) The moderately efficient enzyme: Evolutionary and physicochemical trends shaping enzyme parameters. *Biochemistry* 50: 4402–4410

Basan M (2018) Resource allocation and metabolism: the search for governing principles. *Curr Opin Microbiol* 45: 77–83

Basan M, Honda T, Christodoulou D, Hörl M, Chang Y-F, Leoncini E, Mukherjee A, Okano H, Taylor BR, Silverman JM, *et al* (2020) A universal trade-off between growth and lag in fluctuating environments. *Nature* 584: 470–474

Basan M, Hui S, Okano H, Zhang Z, Shen Y, Williamson JR & Hwa T (2015) Overflow metabolism in *Escherichia coli* results from efficient proteome allocation. *Nature* 528: 99–104

Beck AE, Hunt KA, Bernstein HC & Carlson RP (2016) Interpreting and designing microbial communities for bioprocess applications, from components to interactions to emergent properties. In *Biotechnology for Biofuel Production and Optimization* pp 407–432. Amsterdam: Elsevier B.V., Amsterdam

Beg QK, Vazquez A, Ernst J, de Menezes MA, Bar-Joseph Z, Barabási A-LA-L & Oltvai ZN (2007) Intracellular crowding defines the mode and sequence of substrate uptake by *Escherichia coli* and constrains its metabolic activity. *Proc Natl Acad Sci U S A* 104: 12663–12668

Belliveau NM, Chure G, Hueschen CL, Garcia HG, Kondev J, Fisher DS, Theriot JA & Phillips R (2021) Fundamental limits on the rate of bacterial growth and their influence on proteomic composition. *Cell Syst* 12: 924-944.e2

Bennett BD, Kimball EH, Gao M, Osterhout R, Van Dien SJ & Rabinowitz JD (2009) Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*. *Nat Chem Biol* 5: 593–9

Bernstein J a, Khodursky AB, Lin P-H, Lin-Chao S & Cohen SN (2002) Global analysis of mRNA decay and abundance in *Escherichia coli* at single-gene resolution using two-color fluorescent DNA microarrays. *Proc Natl Acad Sci U S A* 99: 9697–702

Bobay L-M & Ochman H (2018) Factors driving effective population size and pan-genome evolution in bacteria. *BMC Evol Biol* 18: 153

Borkowski O, Goelzer A, Schaffer M, Calabre M, Ma der U, Aymerich S, Jules M & Fromion V (2016) Translation elicits a growth rate-dependent, genome-wide, differential protein production in *Bacillus subtilis*. *Mol Syst Biol* 12: 870–870

Bosdriesz E, Molenaar D, Teusink B & Bruggeman FJ (2015) How fast-growing bacteria robustly tune their ribosome concentration to approximate growth-rate maximization. *FEBS J* 282: 2029–2044

Bremer H & Churchward G (1977) An examination of the Cooper-Helmstetter theory of DNA replication in bacteria and its underlying assumptions. *J Theor Biol* 69: 645–654

Bremer H & Dennis PP (1996) Modulation of chemical composition and other parameters of the cell by growth rate. In *Escherichia coli and Salmonella: cellular and molecular biology* pp 1553–1569. Washington, DC: ASM Press, Washington, DC

Bremer H & Dennis PP (2008) Modulation of chemical composition and other parameters of the cell at different exponential growth rates. *EcoSal Plus* 3: 765–77

Bruggeman FJ, Planqué R, Molenaar D & Teusink B (2020) Searching for principles of microbial physiology. *FEMS Microbiol Rev* 44: 821–844

Calabrese L, Grilli J, Osella M, Kempes CP, Lagomarsino MC & Ciandrini L (2021) Role of protein degradation in growth laws. *bioRxiv*: 2021.03.25.436692

Campo N, Dias MJ, Daveran-Mingot M-L, Ritzenthaler P & Le Bourgeois P (2004) Chromosomal constraints in Gram-positive bacteria revealed by artificial inversions. *Mol Microbiol* 51: 511–522

Chan PP, Lin BY, Mak AJ & Lowe TM (2021) tRNAscan-SE 2.0: Improved detection and functional classification of transfer RNA genes. *bioRxiv*: 614032

Chang A, Jeske L, Ulbrich S, Hofmann J, Koblitz J, Schomburg I, Neumann-Schaal M, Jahn D & Schomburg D (2021) BRENDA, the ELIXIR core data resource in 2021: new developments and updates. *Nucleic Acids Res* 49: D498–D508

Chen Y & Nielsen J (2019) Energy metabolism controls phenotypes by protein efficiency and allocation. *Proc Natl Acad Sci U S A* 116: 17592–17597

Cherry JM, Hong EL, Amundsen C, Balakrishnan R, Binkley G, Chan ET, Christie KR, Costanzo MC, Dwight SS, Engel SR, *et al* (2012) Saccharomyces Genome Database: the genomics resource of budding yeast. *Nucleic Acids Res* 40: D700-5

Condon C, Philips J, Fu ZY, Squires C & Squires CL (1992) Comparison of the expression of the seven ribosomal RNA operons in *Escherichia coli*. *EMBO J* 11: 4175–85

Cooper S & Helmstetter CE (1968) Chromosome replication and the division cycle of *Escherichia coli* B/r. *J Mol Biol* 31: 519–40

Couturier E & Rocha EPC (2006) Replication-associated gene dosage effects shape the genomes of fast-growing bacteria but only for transcription and translation genes. *Mol Microbiol* 59: 1506–1518

Czyzyk J, Mesnier MP & More JJ (1998) The NEOS Server. *IEEE Comput Sci Eng* 5: 68–75

Dai X, Zhu M, Warren M, Balakrishnan R, Patsalo V, Okano H, Williamson JR, Fredrick K, Wang Y-P & Hwa T (2016) Reduction of translating ribosomes enables *Escherichia coli* to maintain elongation rates during slow growth. *Nat Microbiol* 2: 16231

Dauner M & Sauer U (2001) Stoichiometric growth model for riboflavin-producing *Bacillus subtilis*. *Biotechnol Bioeng* 76: 132–143

Davidi D, Noor E, Liebermeister W, Bar-Even A, Flamholz A, Tummler K, Barenholz U, Goldenfeld M, Shlomi T & Milo R (2016) Global characterization of in vivo enzyme catalytic rates and their correspondence to in vitro k_{cat} measurements. *Proc Natl Acad Sci U S A* 113: 3401–6

Dennis PP & Bremer H (1974) Differential rate of ribosomal protein synthesis in *Escherichia coli* B/r. *J Mol Biol* 84: 407–422

Desouki A (2016) Algorithms for improving the predictive power of flux balance analysis (Doctoral dissertation).

Diether M, Nikolaev Y, Allain FH & Sauer U (2019) Systematic mapping of protein-metabolite interactions in central metabolism of *Escherichia coli*. *Mol Syst Biol* 15: e9008

Doi RH & Igarashi RT (1964) Relation of ribonucleic acid composition to growth rate and dormancy in *Bacillus subtilis*. *Nature* 203: 1092–4

Dong H, Nilsson L & Kurland CG (1996) Co-variation of tRNA abundance and codon usage in *Escherichia coli* at different growth rates. *J Mol Biol* 260: 649–663

Dourado H & Lercher MJ (2020) An analytical theory of balanced cellular growth. *Nat Commun* 11: 1226

Dourado H, Maurino V & Lercher M (2017) Enzymes and substrates are balanced at minimal combined mass concentration in vivo. *bioRxiv*: 128009

Dourado H, Mori M, Hwa T & Lercher MJ (2021) On the optimality of the enzyme–substrate relationship in bacteria. *PLOS Biol* 19: e3001416

Ecker RE & Schaechter M (1963) Ribosome content and the rate of growth of *Salmonella* typhimurium. *Biochim Biophys Acta* 76: 275–9

Edgar RC (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5: 113

Ehrenberg M & Kurland CG (1984) Costs of accuracy determined by a maximal growth rate constraint. *Q Rev Biophys* 17: 45–82

Erickson DW, Schink SJ, Patsalo V, Williamson JR, Gerland U & Hwa T (2017) A global resource allocation strategy governs growth transition kinetics of *Escherichia coli*. *Nature* 551: 119–123

Espah Borujeni A, Channarasappa AS & Salis HM (2014) Translation rate is controlled by coupled trade-offs between site accessibility, selective RNA unfolding and sliding at upstream standby sites. *Nucleic Acids Res* 42: 2646–2659

Fang X, Lloyd CJ & Palsson BO (2020) Reconstructing organisms in silico: genome-scale models and their emerging applications. *Nat Rev Microbiol* 18: 731–743

Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, Karp PD, Broadbelt LJ, Hatzimanikatis V & Palsson BØ (2007) A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol Syst Biol* 3: 121

Felsenstein J (1985) Phylogenies and the Comparative Method. *Am Nat* 125: 1–15

Fischer E & Sauer U (2003) A novel metabolic cycle catalyzes glucose oxidation and anaplerosis in hungry *Escherichia coli*. *J Biol Chem* 278: 46446–46451

Forchhammer J & Lindahl L (1971) Growth rate of polypeptide chains as a function of the cell growth rate in a mutant of *Escherichia coli* 15. *J Mol Biol* 55: 563–8

Fuentes LG, Lara AR, Martínez LM, Ramírez OT, Martínez A, Bolívar F & Gosset G (2013) Modification of glucose import capacity in *Escherichia coli*: physiologic consequences and utility for improving DNA vaccine production. *Microb Cell Factories* 12: 42

Furano A V. (1975) Content of elongation factor Tu in *Escherichia coli*. *Proc Natl Acad Sci* 72: 4780–4784

Gao N, Lu G, Lercher MJ & Chen WH (2017) Selection for energy efficiency drives strand-biased gene distribution in prokaryotes. *Sci Rep* 7: 1–10

Gelius-Dietrich G, Desouki A, Fritzemeier C & Lercher MJ (2013) sybil – Efficient constraint-based modelling in R. *BMC Syst Biol* 7: 125

Gerosa L, Haverkorn Van Rijsewijk BRB, Christodoulou D, Kochanowski K, Schmidt TSB, Noor E & Sauer U (2015) Pseudo-transition analysis identifies the key regulators of dynamic metabolic adaptations from steady-state data. *Cell Syst* 1: 270–282

Goelzer A, Fromion V & Scorletti G (2011) Cell design in bacteria as a convex optimization problem. *Automatica* 47: 1210–1218

Goelzer A, Muntel J, Chubukov V, Jules M, Prestel E, Nölker R, Mariadassou M, Aymerich S, Hecker M, Noirot P, *et al* (2015) Quantitative prediction of genome-wide resource allocation in bacteria. *Metab Eng* 32: 232–243

Gordon J (1970) Regulation of the in vivo synthesis of the polypeptide chain elongation factors in *Escherichia coli*. *Biochemistry* 9: 912–917

Gordon J & Weissbach H (1970) Immunochemical distinction between the *Escherichia coli* polypeptide chain elongation factors Tu and Ts. *Biochemistry* 9: 4233–4236

Gourse RL, Gaal T, Bartlett MS, Appleman JA & Ross W (1996) rRNA transcription and growth rate-dependent regulation of ribosome synthesis in *Escherichia coli*. *Annu Rev Microbiol* 50: 645–677

Goyal A, Belardinelli R, Maracci C, Milón P & Rodnina MV (2015) Directional transition from initiation to elongation in bacterial translation. *Nucleic Acids Res* 43: 10700–10712

Gromadski KB, Wieden H-J & Rodnina M V (2002) Kinetic mechanism of elongation factor Ts-catalyzed nucleotide exchange in elongation factor Tu. *Biochemistry* 41: 162–9

de Groot DH, Lischke J, Muolo R, Planqué R, Bruggeman FJ & Teusink B (2020) The common message of constraint-based optimization approaches: overflow metabolism is caused by two growth-limiting constraints. *Cell Mol Life Sci* 77: 441–453

Hanegraaf PPF & Muller EB (2001) The dynamics of the macromolecular composition of biomass. *J Theor Biol* 212: 237–251

Harvey RJ & Koch AL (1980) How partially inhibitory concentrations of chloramphenicol affect the growth of *Escherichia coli*. *Antimicrob Agents Chemother* 18: 323–337

Haverkorn van Rijsewijk BRB, Nanchen A, Nallet S, Kleijn RJ & Sauer U (2011) Large-scale 13C-flux analysis reveals distinct transcriptional control of respiratory and fermentative metabolism in *Escherichia coli*. *Mol Syst Biol* 7: 477

Heckmann D, Campeau A, Lloyd CJ, Phaneuf PV, Hefner Y, Carrillo-Terrazas M, Feist AM, Gonzalez DJ & Palsson BO (2020) Kinetic profiling of metabolic specialists demonstrates stability and consistency of in vivo enzyme turnover numbers. *Proc Natl Acad Sci* 117: 23182–23190

Heckmann D, Lloyd CJ, Mih N, Ha Y, Zielinski DC, Haiman ZB, Desouki AA, Lercher MJ & Palsson BO (2018) Machine learning applied to enzyme turnover numbers reveals protein structural correlates and improves metabolic models. *Nat Commun* 9: 5252

Hill CW & Gray JA (1988) Effects of chromosomal inversion on cell fitness in *Escherichia coli* K-12. *Genetics* 119: 771–8

Hu X-P, Dourado H, Schubert P & Lercher MJ (2020) The protein translation machinery is expressed for maximal efficiency in *Escherichia coli*. *Nat Commun* 11: 5260

Hu X-P & Lercher MJ (2021) An optimal growth law for RNA composition and its partial implementation through ribosomal and tRNA gene locations in bacterial genomes. *PLOS Genet* 17: e1009939

Hui S, Silverman JM, Chen SS, Erickson DW, Basan M, Wang J, Hwa T & Williamson JR (2015) Quantitative proteomic analysis reveals a simple strategy of global resource allocation in bacteria. *Mol Syst Biol* 11: 784

Ibarra RU, Edwards JS & Palsson BO (2002) *Escherichia coli* K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. *Nature* 420: 186–189

Ikemura T (1981) Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes. *J Mol Biol* 146: 1–21

Ingraham JL, Maaløe O & Neidhardt FC (1983) Growth of the Bacterial Cell Sunderland, Massachusetts: Sinauer Associates Inc.,U.S.

Ishii N, Nakahigashi K, Baba T, Robert M, Soga T, Kanai A, Hirasawa T, Naba M, Hirai K, Hoque A, *et al* (2007) Multiple high-throughput analyses monitor the response of *E. coli* to perturbations. *Science* 316: 593–7

Jun S, Si F, Pugatch R & Scott M (2018) Fundamental principles in bacterial physiology—history, recent progress, and the future with focus on cell size control: a review. *Rep Prog Phys* 81: 056601

Kempes CP, Dutkiewicz S & Follows MJ (2012) Growth, metabolic partitioning, and the size of microorganisms. *Proc Natl Acad Sci* 109: 495–500

Kempes CP, Wang L, Amend JP, Doyle J & Hoehler T (2016) Evolutionary tradeoffs in cellular composition across diverse bacteria. *ISME J* 10: 2145–2157

Keseler IM, Mackie A, Santos-Zavaleta A, Billington R, Bonavides-Martínez C, Caspi R, Fulcher C, Gama-Castro S, Kothari A, Krummenacker M, *et al* (2017) The EcoCyc database: reflecting new knowledge about *Escherichia coli* K-12. *Nucleic Acids Res* 45: D543–D550

Kim D, Hong JS-J, Qiu Y, Nagarajan H, Seo J-H, Cho B-K, Tsai S-F & Palsson BØ (2012) Comparative analysis of regulatory elements between *Escherichia coli* and *Klebsiella pneumoniae* by genome-wide transcription start site profiling. *PLoS Genet* 8: e1002867

Kjeldgaard NO & Kurland CG (1963) The distribution of soluble and ribosomal RNA as a function of growth rate. *J Mol Biol* 6: 341–348

Klumpp S, Scott M, Pedersen S & Hwa T (2013) Molecular crowding limits translation and cell growth. *Proc Natl Acad Sci U S A* 110: 16754–9

Koch AL (1971) The adaptive responses of *Escherichia coli* to a feast and famine existence. *Adv Microb Physiol* 6: 147–217

Koch AL (1988) Why can't a cell grow infinitely fast? *Can J Microbiol* 34: 421–426

Korem Kohanim Y, Levi D, Jona G, Towbin BD, Bren A & Alon U (2018) A bacterial growth law out of steady state. *Cell Rep* 23: 2891–2900

Kothapalli S, Nair S, Alokam S, Pang T, Khakhria R, Woodward D, Johnson W, Stocker BAD, Sanderson KE & Liu SL (2005) Diversity of genome structure in *Salmonella enterica* serovar Typhi populations. *J Bacteriol* 187: 2638–2650

Kubitschek HE, Baldwin WW, Schroeter SJ & Graetzer R (1984) Independence of buoyant cell density and growth rate in *Escherichia coli*. *J Bacteriol* 158: 296–9

Kumar S, Stecher G, Li M, Knyaz C & Tamura K (2018) MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol* 35: 1547–1549

Lalanne J-B & Li G-W (2021) First-principles model of optimal translation factors stoichiometry. *eLife* 10: e69222

Link H, Kochanowski K & Sauer U (2013) Systematic identification of allosteric protein-metabolite interactions that control enzyme activity in vivo. *Nat Biotechnol* 31: 357–361

Louarn JM, Bouché JP, Legendre F, Louarn J & Patte J (1985) Characterization and properties of very large inversions of the *E. coli* chromosome along the origin-to-terminus axis. *Mol Gen Genet MGG* 201: 467–76

Louie A & Jurnak F (1985) Kinetic studies of *Escherichia coli* elongation factor Tu-guanosine 5'-triphosphate-aminoacyl-tRNA complexes. *Biochemistry* 24: 6433–9

Luo H & Gao F (2019) DoriC 10.0: an updated database of replication origins in prokaryotic genomes including chromosomes and plasmids. *Nucleic Acids Res* 47: D74–D77

Lynch M & Marinov GK (2015) The bioenergetic costs of a gene. *Proc Natl Acad Sci U S A* 112: 15690–15695

Maaløe O (1969) An analysis of bacterial growth. *Dev Biol Suppl*: 33–58

Maaløe O (1979) Regulation of the protein-synthesizing machinery—ribosomes, tRNA, factors, and so on. In *Biological Regulation and Development* pp 487–542. Boston, MA: Springer US, Boston, MA

Maaløe O & Kjeldgaard NO (1966) Control of macromolecular synthesis: a study of DNA, RNA, and protein synthesis in bacteria New York: W. A. Benjamin

Mahmoudabadi G, Milo R & Phillips R (2017) Energetic cost of building a virus. *Proc Natl Acad Sci U S A* 114: E4324–E4333

Mann N & Carr NG (1973) A constant ratio of transfer to ribosomal ribonucleic acid in *Anacystis nidulans* grown with differing mean generation times. *Biochem Soc Trans* 1: 702–704

Metzl-Raz E, Kafri M, Yaakov G, Soifer I, Gurvich Y & Barkai N (2017) Principles of cellular resource allocation revealed by condition-dependent proteome profiling. *eLife* 6: e28034

Michelsen O, Teixeira de Mattos MJ, Jensen PR & Hansen FG (2003) Precise determinations of C and D periods by flow cytometry in *Escherichia coli* K-12 and B/r. *Microbiology* 149: 1001–1010

Milo R (2013) What is the total number of protein molecules per cell volume? A call to rethink some published values. *BioEssays News Rev Mol Cell Dev Biol* 35: 1050–5

Milón P & Rodnina MV (2012) Kinetic control of translation initiation in bacteria. *Crit Rev Biochem Mol Biol* 47: 334–348

Miyajima A & Kaziro Y (1978) Coordination of levels of elongation factors Tu, Ts, and G, and ribosomal protein S1 in *Escherichia coli*. *J Biochem (Tokyo)* 83: 453–462

Molenaar D, van Berlo R, de Ridder D & Teusink B (2009) Shifts in growth strategies reflect tradeoffs in cellular economics. *Mol Syst Biol* 5: 323

Monk JM, Lloyd CJ, Brunk E, Mih N, Sastry A, King Z, Takeuchi R, Nomura W, Zhang Z, Mori H, *et al* (2017) iML1515, a knowledgebase that computes *Escherichia coli* traits. *Nat Biotechnol* 35: 904–908

Mori M, Schink S, Erickson DW, Gerland U & Hwa T (2017) Quantifying the benefit of a proteome reserve in fluctuating environments. *Nat Commun* 8: 1225

Nanchen A, Schicker A & Sauer U (2006) Nonlinear dependency of intracellular fluxes on growth rate in miniaturized continuous cultures of *Escherichia coli*. *Appl Environ Microbiol* 72: 1164–1172

Nanninga N & Woldringh C (1985) Cell growth, genome duplication and cell division in *Escherichia coli*. In *Molecular Cytology of Escherichia coli*, Nanninga N (ed) pp 259–318. London: Academic Press

Narula J, Kuchina A, Lee DD, Fujita M, Süel GM & Igoshin OA (2015) Chromosomal arrangement of phosphorelay genes couples sporulation and DNA replication. *Cell* 162: 328–337

Neidhardt FC & Magasanik B (1960) Studies on the role of ribonucleic acid in the growth of bacteria. *Biochim Biophys Acta* 42: 99–116

Nilsson A, Nielsen J & Palsson BO (2017) Metabolic models of protein allocation call for the kinetome. *Cell Syst* 5: 538–541

Noor E, Flamholz A, Bar-Even A, Davidi D, Milo R & Liebermeister W (2016) The protein cost of metabolic fluxes: prediction from enzymatic rate laws and cost minimization. *PLOS Comput Biol* 12: e1005167

O'Brien EJ, Lerman J a, Chang RL, Hyduke DR & Palsson BØ (2013) Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. *Mol Syst Biol* 9: 693

O'Brien EJ, Utrilla J & Palsson BO (2016) Quantification and classification of *E. coli* proteome utilization and unused protein costs across environments. *PLoS Comput Biol* 12: e1004998

Oldewurtel ER, Kitahara Y & Teeffelen S van (2021) Robust surface-to-mass coupling and turgor-dependent cell width determine bacterial dry-mass density. *Proc Natl Acad Sci* 118

Orth JD, Thiele I & Palsson BØ (2010) What is flux balance analysis? *Nat Biotechnol* 28: 245–8

Pál C, Papp B & Hurst LD (2001) Highly expressed genes in yeast evolve slowly. *Genetics* 158: 927–31

Pang TY & Lercher MJ (2017) Supra-operonic clusters of functionally related genes (SOCs) are a source of horizontal gene co-transfers. *Sci Rep* 7: 1–10

Panos C, Hynes LM & Cohen M (1965) On RNA composition and growth rates of A streptococcus and derived stable L-form. *Biochem Biophys Res Commun* 19: 62–66

Paradis E & Schliep K (2019) ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35: 526–528

Peebo K, Valgepea K, Maser A, Nahku R, Adamberg K & Vilu R (2015) Proteome reallocation in *Escherichia coli* with increasing specific growth rate. *Mol BioSyst* 11: 1184–1193

Pine MJ (1970) Steady-state measurement of the turnover of amino acid in the cellular proteins of growing *Escherichia coli*: existence of two kinetically distinct reactions. *J Bacteriol* 103: 207–215

Pine MJ (1973) Regulation of Intracellular proteolysis in *Escherichia coli*. *J Bacteriol* 115: 107–116

Potrykus K, Murphy H, Philippe N & Cashel M (2011) ppGpp is the major source of growth rate control in *E. coli*. *Environ Microbiol* 13: 563–575

Proshkin S, Rahmouni AR, Mironov A & Nudler E (2010) Cooperation between translating ribosomes and RNA polymerase in transcription elongation. *Science* 328: 504–508

Pruitt KD, Tatusova T, Brown GR & Maglott DR (2012) NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. *Nucleic Acids Res* 40: D130-5

R Core Team (2020) R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing

Radzikowski JL, Vedelaar S, Siegel D, Ortega ÁD, Schmidt A & Heinemann M (2016) Bacterial persistence is an active σ^S stress response to metabolic flux limitation. *Mol Syst Biol* 12: 882

Rocha EPC (2008) The Organization of the bacterial genome. *Annu Rev Genet* 42: 211–233

Rodnina MV (2018) Translation in prokaryotes. *Cold Spring Harb Perspect Biol* 10: a032664

Rosset R, Julien J & Monier R (1966) Ribonucleic acid composition of bacteria as a function of growth rate. *J Mol Biol* 18: 308–320

Rudorf S & Lipowsky R (2015) Protein synthesis in *E. coli*: Dependence of codon-specific elongation on tRNA concentration and codon usage. *PLoS ONE* 10: e0134994

Rudorf S, Thommen M, Rodnina MV & Lipowsky R (2014) Deducing the kinetics of protein synthesis in vivo from the transition rates measured in vitro. *PLoS Comput Biol* 10: e1003909

Rui B, Shen T, Zhou H, Liu J, Chen J, Pan X, Liu H, Wu J, Zheng H & Shi Y (2010) A systematic investigation of *Escherichia coli* central carbon metabolism in response to superoxide stress. *BMC Syst Biol* 4: 122

Russell JB & Cook GM (1995) Energetics of bacterial growth: balance of anabolic and catabolic reactions. *Microbiol Rev* 59: 48–62

Sauer U, Canonaco F, Heri S, Perrenoud A & Fischer E (2004) The soluble and membrane-bound transhydrogenases UdhA and PntAB have divergent functions in NADPH metabolism of *Escherichia coli*. *J Biol Chem* 279: 6613–6619

Schaechter M, Maaløe O & Kjeldgaard NO (1958) Dependency on medium and temperature of cell size and chemical composition during balanced growth of *Salmonella* typhimurium. *J Gen Microbiol* 19: 592–606

Schmidt A, Kochanowski K, Vedelaar S, Ahrne E, Volkmer B, Callipo L, Knoops K, Bauer M, Aebersold R & Heinemann M (2016) The quantitative and condition-dependent *Escherichia coli* proteome. *Nat Biotechnol* 34: 104–110

Schuster S, Boley D, Möller P, Stark H & Kaleta C (2015) Mathematical models for explaining the Warburg effect: a review focussed on ATP and biomass production. *Biochem Soc Trans* 43: 1187–1194

Scott M, Gunderson CW, Mateescu EM, Zhang Z & Hwa T (2010) Interdependence of cell growth and gene expression: origins and consequences. *Science* 330: 1099–102

Scott M, Klumpp S, Mateescu EM & Hwa T (2014) Emergence of robust growth laws from optimal regulation of ribosome synthesis. *Mol Syst Biol* 10: 747

Si F, Li D, Cox SE, Sauls JT, Azizi O, Sou C, Schwartz AB, Erickstad MJ, Jun Y, Li X, *et al* (2017) Invariance of initiation mass and predictability of cell size in *Escherichia coli*. *Curr Biol* 27: 1278–1287

Skjold AC, Juarez H & Hedgcoth C (1973) Relationships among deoxyribonucleic acid, ribonucleic acid, and specific transfer ribonucleic acids in *Escherichia coli* 15T- at various growth rates. *J Bacteriol* 115: 177–187

Slager J, Kjos M, Attaiech L & Veening J-W (2014) Antibiotic-Induced replication stress triggers bacterial competence by increasing gene dosage near the origin. *Cell* 157: 395–406

Slager J & Veening J-W (2016) Hard-wired control of bacterial processes by chromosomal gene location. *Trends Microbiol* 24: 788–800

Soler-Bistué A, Mondotte JA, Bland MJ, Val M-E, Saleh M-C & Mazel D (2015) Genomic location of the major ribosomal protein gene locus determines *Vibrio cholerae* global growth and infectivity. *PLOS Genet* 11: e1005156

Soler-Bistué A, Timmermans M & Mazel D (2017) The proximity of ribosomal protein genes to oric enhances *Vibrio cholerae* fitness in the absence of multifork replication. *mBio* 8: e00097-17

Stoebel DM, Dean AM & Dykhuizen DE (2008) The cost of expression of *Escherichia coli lac* operon proteins is in the process, not in the products. *Genetics* 178: 1653–1660

Sykes J & Young TW (1968) Studies on the ribosomes and ribonucleic acids of *Aerobacter aerogenes* grown at different rates in carbon-limited continuous culture. *BBA Sect Nucleic Acids Protein Synth* 169: 103–116

Tadmor AD & Tlusty T (2008) A coarse-grained biophysical model of *E. coli* and its application to perturbation of the rRNA operon copy number. *PLoS Comput Biol* 4: e1000038

Tawarmalani M & Sahinidis N V (2005) A polyhedral branch-and-cut approach to global optimization. *Math Program* 103: 225–249

Towbin BD, Korem Y, Bren A, Doron S, Sorek R & Alon U (2017) Optimality and sub-optimality in a bacterial growth law. *Nat Commun* 8: 14123

Tria FDK, Landan G & Dagan T (2017) Phylogenetic rooting using minimal ancestor deviation. *Nat Ecol Evol* 1: 1–7

Tuller T, Carmi A, Vestsigian K, Navon S, Dorfan Y, Zaborske J, Pan T, Dahan O, Furman I & Pilpel Y (2010) An evolutionarily conserved mechanism for controlling the efficiency of protein translation. *Cell* 141: 344–54

Valgepea K, Adamberg K, Seiman A & Vilu R (2013) *Escherichia coli* achieves faster growth by increasing catalytic and translation rates of proteins. *Mol Biosyst* 9: 2344

Vazquez A (2010) Optimal cytoplasmatic density and flux balance model under macromolecular crowding effects. *J Theor Biol* 264: 356–359

Vieira JP, Racle J & Hatzimanikatis V (2016) Analysis of translation elongation dynamics in the context of an *Escherichia coli* cell. *Biophys J* 110: 2120–2131

Vieira-Silva S & Rocha EPC (2010) The systemic imprint of growth and its uses in ecological (meta)genomics. *PLoS Genet* 6: e1000808

Volkmer B & Heinemann M (2011) Condition-dependent cell volume and concentration of *Escherichia coli* to facilitate data conversion for systems biology modeling. *PLoS ONE* 6: 1–6

Waldron C & Lacroute F (1975) Effect of growth rate on the amounts of ribosomal and transfer ribonucleic acids in yeast. *J Bacteriol* 122: 855–65

Wang M, Weiss M, Simonovic M, Haertinger G, Schrimpf SP, Hengartner MO & von Mering C (2012) PaxDb, a database of protein abundance averages across all three domains of life. *Mol Cell Proteomics MCP* 11: 492–500

Weiße AY, Oyarzún DA, Danos V & Swain PS (2015) Mechanistic links between cellular trade-offs, gene expression, and growth. *Proc Natl Acad Sci U S A* 112: E1038-1047

Wong F, Dutta A, Chowdhury D & Gunawardena J (2018) Structural conditions on complex networks for the Michaelis–Menten input–output response. *Proc Natl Acad Sci* 115: 9738–9743

Yoshikawa H, O'Sullivan A & Sueoka N (1964) Sequential replication of the *Bacillus subtilis* chromosome, III. Regulation of initiation. *Proc Natl Acad Sci* 52: 973–980

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

Zeng H & Yang A (2019) Modelling overflow metabolism in *Escherichia coli* with flux balance analysis incorporating differential proteomic efficiencies of energy pathways. *BMC Syst Biol* 13: 3

Zheng H, Bai Y, Jiang M, Tokuyasu TA, Huang X, Zhong F, Wu Y, Fu X, Kleckner N, Hwa T, *et al* (2020) General quantitative relations linking cell growth and the cell cycle in *Escherichia coli*. *Nat Microbiol* 5: 995–1001

Zhu M & Dai X (2019) Growth suppression by altered (p)ppGpp levels results from non-optimal resource allocation in *Escherichia coli*. *Nucleic Acids Res* 47: 4684–4693